World Bank Center for Development Data (C4D2)

Partnership for Capacity Development in Household Surveys for Welfare Analysis

Measuring Consumption

Instructor Manual

This draft: April 2020

This draft was prepared by Giulia Mancini and Giovanni Vecchi (University of Rome Tor Vergata). We would like to acknowledge the contribution provided by Nicola Amendola and Sédi-Anne Boukaka (University of Rome Tor Vergata) in preparing the material underlying the course. We are grateful to Gero Carletto, Michelle Jouvenal, Shelton Kanyanda and Alberto Zezza (World Bank) for helpful suggestions.

Introduction: how to use this Manual

This Manual is addressed to the Instructor of the course "Measuring Consumption", prepared by the World Bank's Center for Development Data (C4D2). It is part of a **Teaching Package** that also contains the following materials:

1. Course syllabus
2. Lecture slides (classroom version, PowerPoint, and handout version, pdf)
3. Reading package
4. Final exam and grading sheet
5. Course evaluation form

The Manual explains how best to make use of each of these items, and provides practical guidance to support the Instructor in the preparation and delivery of the course.

The rest of the Manual is organized as follows. First, the section titled **The Course** gives a general outline of the course, describes its overall learning objectives, explains how they will be reached, and lays out the Syllabus. This section also contains some important logistical information (required facilities, teaching materials, etc.), that should be assessed in preparation for the course.

The remaining sections provide information for each of the **15 lectures** included in the course. For each lecture, you will find *(i)* learning objectives; *(ii)* a description of any preparatory work that the Instructor is advised to do before giving the lecture; *(iii)* a time allocation plan, which advises Instructors on how best to allocate the available classroom time; *(iv)* a writeup of the topics covered in the lecture, designed to support the Instructor's understanding of the slides; and *(v)* keys to any exercises associated with the lecture.

The Course

Target audience and pre-requisites

The target audience is Master students enrolled in Graduate Programs in Statistics, Official Statistics, and related fields, offered by the Regional Training Centers.

This course is designed to be self-contained, therefore it has no compulsory pre-requisites. However, it is an advantage if students have a basic command of Statistics, Sampling Design, and Microeconomics.

Learning objectives of the course

By training students who will pursue careers within Statistical Offices and Institutions, this course aims at improving the **quality** and **comparability** (both over time and across countries) of household survey data for living standards measurement.

To do so, the course focuses on two main topics: first, it lays out the **conceptual framework** underlying the measurement of living standards; second, it offers practical **guidelines for survey design and data collection**, in the specific context of household consumption and expenditure modules.

The **emphasis** of the course **is on data collection, rather than data analysis**. At the end of the course, participants will have learned how to best approach the design of consumption modules in household budget surveys. The computation of poverty and inequality indicators is covered only briefly in this course, as a way of informing data producers about the uses of household consumption data.

Course syllabus

The course contains **15 lectures**. Teaching materials are optimized for each lecture to last two hours, including a short break of about 15 minutes.

A number of **readings** are associated to each lecture: all of them are provided in the form of a Reading Package, to be shared with students. Students should be encouraged to complete readings before each lecture. Then, classroom presentations provide guidance for the sometimes technical and challenging published material. A few selected readings are to be considered essential to a student's learning, and are marked **required**. All other readings are optional, but useful for those interested in knowing more.

Lectures also contain a number of **exercises**, to be assigned as homework. This Manual includes keys and evaluation guidelines for the exercises.

Plan of the lectures:

1) **Measuring living standards: a conceptual framework**
   From theory to practice: Income vs. Expenditure/Consumption vs. Wealth

2) **The Consumption Aggregate**
   Four building blocks: (i) food, (ii) non-durable non-food items, (iii) housing, and (iv) durable goods. Adjustments: for household size and needs, for within-year inflation.

3) **Understanding household surveys**
   Types of surveys; defining features of household surveys that measure consumption, as opposed to other statistical instruments.

4) **Principles of questionnaire design**
   Overview of the survey process; general principles of questionnaire design, as they apply to the measurement of consumption expenditures.

5) **Measuring food consumption: the foundations**
   Concepts (*e.g.* acquisition vs. consumption). Questionnaire design (recall vs. diary)

6) **Measuring food consumption: questionnaire design**
   Questionnaire design (list of food items, meal participation, seasonality)

7) **Food away from home and the use of non-standard units for measuring food consumption**
   Questionnaire design (non-standard measurement units, food away from home)

8) **Measuring consumption of non-durable non-food goods**
   Analytical needs; questionnaire design.

9) **Durable goods**
   Analytical needs and questionnaire design.

10) **Housing**
    Analytical needs; questionnaire design.

11) **Data validation and diagnostics**
    Key principles of statistical data editing.

12) **Outlier detection and treatment**
    Key principles of outlier detection and treatment.

13) **Measuring inequality**
    Lorenz curve and Gini index; other selected inequality measures.

14) **Measuring poverty**
    Poverty lines and measures.

15) **Describing data**
    Preparation of a "tabulation report".

Course evaluation

The Teaching Package contains a **course evaluation form** designed to collect feedback from students. The evaluation form is anonymous, and should be handed out after the last lecture. Instructors are encouraged to share results with the C4D2 team, to help improve future iterations of the course.

Final exam

The final exam is a take-home assignment. The text of the final exam is provided as part of the Teaching Package, to serve as a blueprint for the Instructor's convenience.

Logistics

The total amount of **classroom time** required for the course is 30 hours (there are 15 lectures, each lasting two hours). Training Centers are advised to schedule one or, at most, two of these lectures in one day. Because the final exam is a take-home assignment, no further classroom time is required after the end of the course.

The course calls for the use of **PowerPoint** presentations to support the Instructor's delivery: therefore, a room outfitted with a **projector** is required. The use of a **microphone** by the Instructor is strongly recommended: it greatly facilitates the engagement of participants throughout the presentations. One or more microphones should be available for participants as well, to facilitate class discussions.

Instructors can use the following **checklist** to organize the logistics of the course:

Before registration:

- ❑ Share the syllabus of the course, clarify learning objectives: make sure participants' expectations are aligned with the contents of the course
- ❑ Confirm availability of required facilities (projector, microphones)

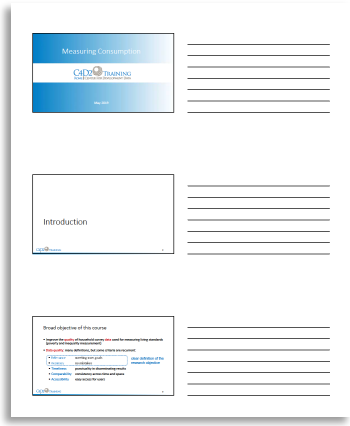
Before the first lecture:

- ❑ The Instructor gets familiar with the readings associated to each lecture
- ❑ Edit Lecture 1's "Contacts" slide, adding the Instructor's email address, office hours, etc.
- ❑ Print out slides (handout version) for each student
- ❑ Prepare to share digital copy of reading package with students

Before the final lecture:

- ❑ Print out evaluation forms for each student
- ❑ Print out (or prepare to share digital copy of) final exam

Please note: printing out the **handout version** of the course slides, so that each student can use the printed slides for notes and exercises, is strongly encouraged.



Lecture 1
Measuring living standards: a conceptual framework

Learning objectives

The goal of this lecture is to provide the theoretical basis for the choice of a measure of living standards, which analysts need in order to estimate both poverty and inequality. The material covered in the lecture justifies the focus on a consumption-based measure of living standards.

Suggested preparation

The Instructor is assumed to be familiar with the material contained in section 2.2 ("Money-metric utility") of Deaton and Zaidi (2002), as well as with references therein listed.

Time allocation

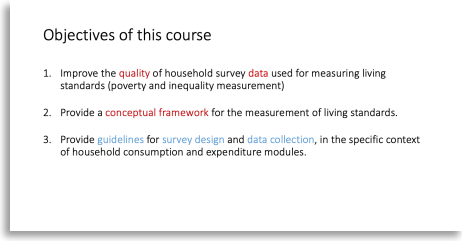
Students' introductions (name, background)	**15 min**
Introduction	15 min
What is the standard of living?	30 min
Break	**15 min**
Choosing a measure of living standards	45 min

Annotated lecture

As the first lecture begins, it is a good idea to break the ice by getting to know each student, and setting their **expectations** for the course. The Instructor can do a round of introductions and ask each student what they expect to learn. The students should be informed in advance of the objectives and contents of the course; the description of the course and the syllabus should be circulated well in advance of registration: please check the Logistics section of this manual.

If that has been done, students' expectations should not differ too much from what the Instructor is about to deliver. The course is *not* about data analysis: it is about data collection. The course is *not* about hands-on practical sessions: it is about concepts. This moment gives the Instructor the opportunity to clarify this once more.

The first slide helps to do just that, giving students a bird's-eye view of the whole course. This task does not require much time, but clarifying what the course is about, and why it is important – and hopefully interesting, too – is key. The aim is to convey **motivation** (students should feel they are about to learn important concepts and tools). This is why the first bunch of slides gives an overview of the topics that will be covered in the following classes, and, even more importantly, explains why students should take the time to study them.

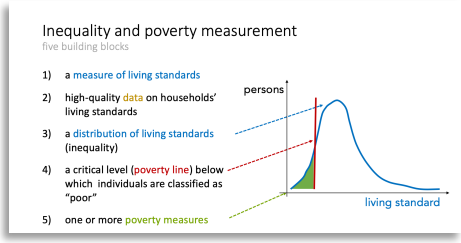


The broad, long-term objective of the course, that justifies an interest in the topics covered, is that of improving the quality of household survey data used for measuring living standards (which in turn are the basis for poverty and inequality measurement).

What do we mean by **data quality**, exactly? Before going into details, we can take this opportunity to make a general point: learning to use words and terms as precisely and technically as possible is part of the aims of this course. The attitude that we encourage in students is to ask questions such as "what do we mean by 'something', exactly?" where emphasis is on '**exactly**'. Answers are in the **literature**, which students will be encouraged to consult throughout the course, instead of relying entirely on 'common sense'. Here, we can consider engaging with students, and asking *them* "how would you define and assess 'data quality', exactly?". We allocate only a couple of minutes to discussing answers, then move on.

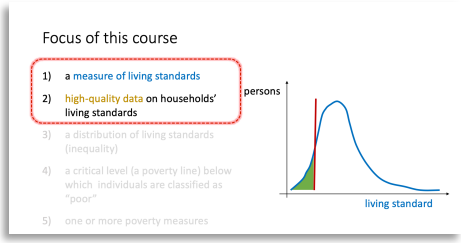
Back to our question 'What do we mean by **data quality**, exactly?'. In their manual on survey quality, Biemer and Lyberg (2003: 13-18) review different ways in which data quality has been defined by various statistical institutions: some common criteria emerge. One of the most used criteria is that, in order to be of good quality, data should be **relevant**

for the needs of the users. Therefore, it is crucial for data producers to understand and clearly define the research objective that justifies their data collection efforts (what will the data be used for?). This is why we have the next two bullets: the course will provide a conceptual framework, to clarify the goals and methods of welfare analysis, and in light of those, it will offer specific guidelines for survey design and data collection.



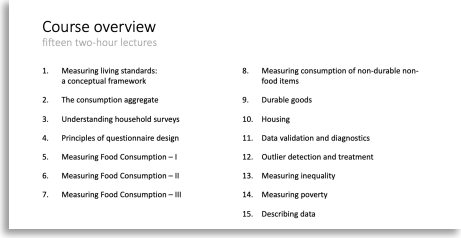
Because the course is focused on the collection of data for the purpose of inequality and poverty measurement, this slide explains the 5 building blocks that are needed to construct estimates of inequality and poverty:

1) a **measure of living standards**, that is, a quantifiable concept of what living standards are and how they can be defined in practice;
2) high-quality **data** on households' living standards, typically collected through surveys;
3) a distribution of living standards, that is, the actual data that come from the field, and that describe the variation of living standards across the population of interest. At this stage, it is possible to produce **inequality** estimates;
4) a critical level, a **poverty line**, below which individuals are classified as "poor", that is, a value, a fixed amount of whatever quantity indicates individual living standards, that gives us a threshold below which individuals are classified as poor;
5) one or more **poverty measures**, such as a simple headcount of how many poor individuals there are, or any other more complex measure of our choice, to summarize the results of the analysis.

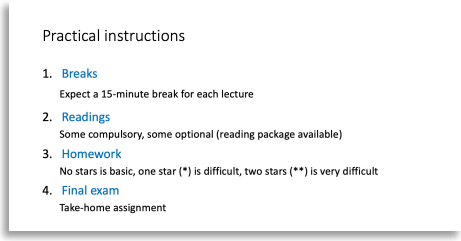


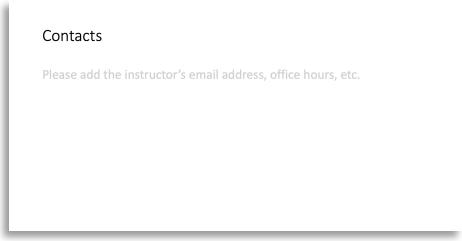
The focus of this course will be on the first 2 building blocks: a measure of living standards (again, this is the conceptual foundation of the analysis, which must be embedded in the design of the survey); and high-quality data that describe households' living standards. This is a good time to comment on the importance of '**theory**': the course will combine 'theory' (first building block) with 'practice' (second building block). There is no such a thing as good data collection in the absence of a theoretical framework. Building blocks 3) to 5)

venture into analytical territory, and are complex enough to be the subject of a different course.



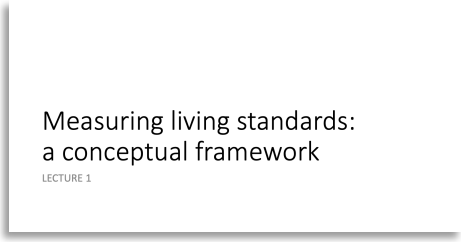
This slide gives an overview of the topics to be covered in each of the 15 lectures of the course. We should *not* read and comment item by item – this would take too long, and probably bore students to death. We simply remind students that there is a Syllabus, and encourage them to check it and use it as a roadmap throughout the duration of the course.



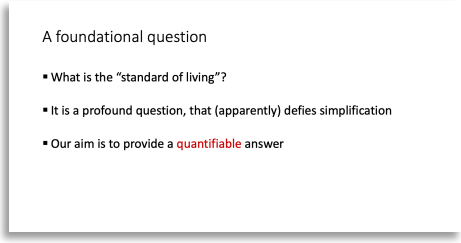


Next, some practical instructions that are of interest for students.

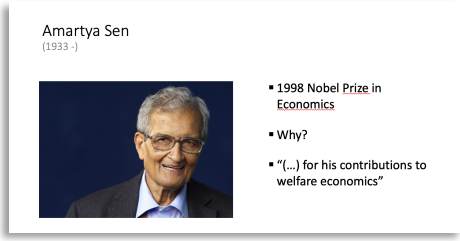
Do not forget: the **Instructor's contacts** should be personalized.

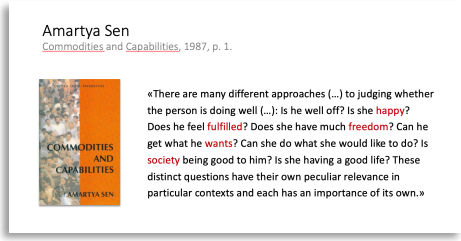


After the introduction, we enter into the actual contents of lecture 1. This lecture is about **concepts**: it provides the theoretical foundation for the need of data on household consumption. It uses a bit of economic theory and technical jargon in order to give students some common ground with data analysts. The importance of a theoretical foundation for the way we measure living standards is one of the main takeaways of the lecture: to make the point effectively, the Instructor should be familiar with the literature cited throughout.

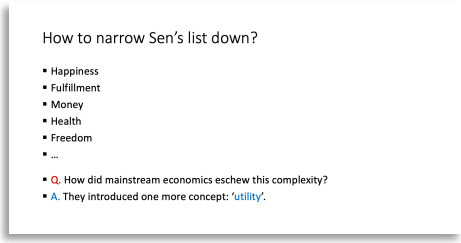


The first part of the lecture answers the question: "What is the standard of living?". Here, the Instructor can involve students, and encourage them to give their own answers to the question. The point that should emerge from this short discussion is that the answer is highly complex, highly subjective, and in no way easy to agree on. We emphasize that is good to keep this **conceptual complexity** in mind, although statisticians, economists, data analysts are interested in providing a quantifiable answer – one that allows for an estimate of how many poor people live in a country, for instance.



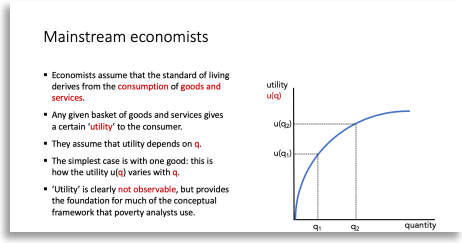


The next few slides introduce the views of a notable economist and philosopher, Amartya Sen. Sen's position is very critical of the simplification that mainstream economics adopts in order to arrive at a measurement of living standards – which will be covered in the rest of the lecture. However, the quote from Sen's book 'Commodities and Capabilities' is a good summary of the complexity of the concept of well-being, and conveys the need of delimiting the issue, if one wants to make this concept measurable.

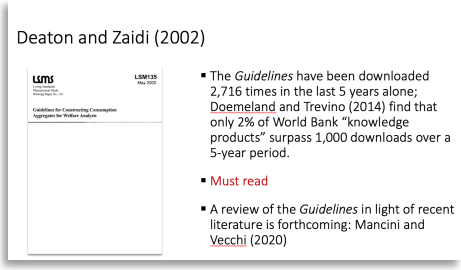


This slide brings the audience back to the fundamental question of defining the standard of living, after having established that such a definition is a complex matter, not just intuitively (see the short discussion that the class just had about well-being), but also in the literature (see Sen's thought). However, mainstream economics provides a way to **reduce** this **complexity**. It does so by introducing the concept of "utility".

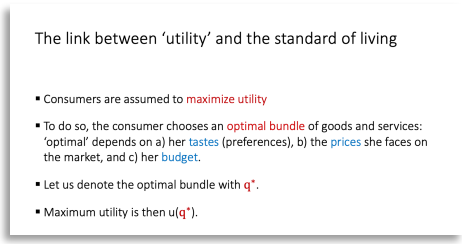
The next slides explain the basics of **utility theory**. If more details are needed, students can be encouraged to consult a manual of microeconomics (chapters on consumer choice). One of the most widely used is Varian (2010), while Deaton and Muellbauer (1980) or Varian (1992) are at a more advanced level.



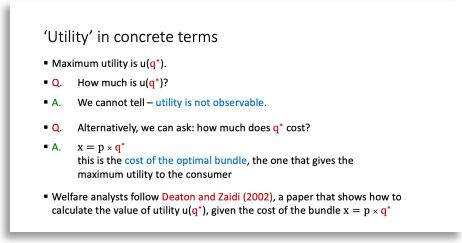
The answer that mainstream economists give to the complexities of defining well-being is based on the concept of utility, introduced in this slide. "Utility" is an abstract term, a catch-all, standing for whatever a person pursues. It is a useful tool for standard economic theory, which is concerned with individual choice.



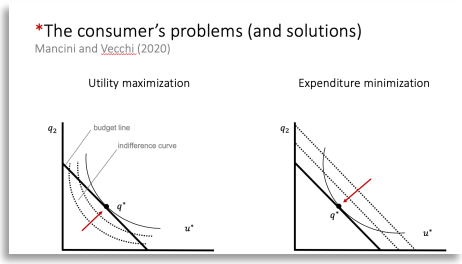
These slides introduce the work of another notable economist, Angus Deaton, who will be cited extensively in this lecture: he will help us connect utility theory with the measurement of living standards. Students should become familiar with Deaton's work, especially with Deaton and Zaidi (2002), a paper entitled *Guidelines for Constructing Consumption Aggregates for Welfare Analysis*. In this lecture, we will cover the theoretical section of the *Guidelines* (in turn based on the textbook by Deaton and Muellbauer).

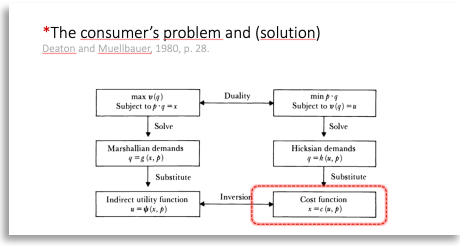


Individuals (consumers) are assumed to derive utility from the consumption of goods and services, which means that utility depends on q (a vector of goods and services). In this theoretical setting, consumers are assumed to make their choices based on utility. How? They choose the particular combination of goods and services (also called a "bundle") that gives them the highest possible utility, given the circumstances. These circumstances include the consumers' individual tastes (also called preferences), the prices that they face on the market, and their budget (the resources they are able to spend on consumption). In other words, consumers are assumed to **maximize utility**. Consider a single consumer, and let us introduce a bit of notation: the optimal bundle (the one that the consumer ends up picking) is indicated by q*, and therefore the consumer's maximum utility is u(q*).

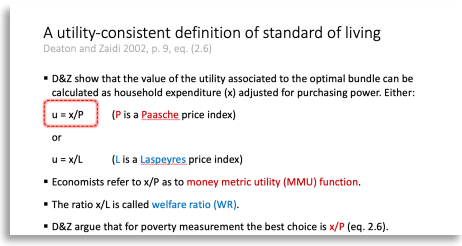


We would like to know how much is $u(q^*)$, because that would be a measure of the consumer's living standard. However, we cannot observe utility: it is just an abstract concept. What we can observe is the cost of q^*, which can be indicated by $x = q^* \times p$ (where p are the prices paid by the consumer– remember that both q^* and p are vectors, so in our notation, $q^* \times p = \sum_k p_k q_k^*$, where k indicates each of the goods consumed). Welfare analysts follow Deaton and Zaidi (2002), a paper that shows how to calculate the cost of utility given the cost of the bundle.



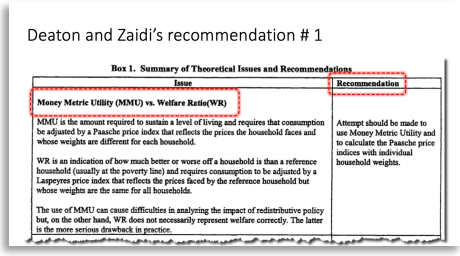


Optional slides are denoted by a 'star' (*). Depending on the circumstance, we suggest that, especially the first times the course is thaught, these slides are omitted from the presentation, unless we feel confident that it will not take to long to convey the message, and the audience is well equipped to absorb its content. These optional slides go into more detail on one key point: the fact that the cost of the optimal bundle indicates the utility achieved by the consumer when she chooses a given bundle. The justification for this is the equivalence between the utility maximization and expenditure minimization problems, known as *duality*. A detailed discussion is in Deaton and Muellbauer (1980), p. 28; also refer to Mancini and Vecchi (2020) for a non-technical version. In what follows, we assume the Instructor has *not* shown to the class and commented on the primal and dual consumer problems in the starred slide above.



Deaton and Zaidi (2002) show that the utility associated to the optimal bundle can be approximated by (calculated as) the cost of the optimal bundle, divided by a price index, which transforms the expenditure into real terms, correcting for differences in purchasing power that arise across different consumers. Price indices can be computed according to

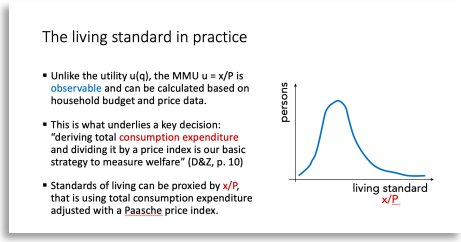
the Paasche or Laspeyres formula: these details are not important for now, and they will be covered in the following lecture. Suffice it to say that x/P is called money-metric utility (MMU), and it is the quantifiable concept of living standard that modern economic theory offers. In practice, this concept corresponds to total consumption expenditure adjusted with a Paasche price index (the index will be explained in detail in lecture 2 – so the suggestion is not to waste time going into details). This is what equation (2.6) in Deaton and Zaidi (2002), possibly the single most important result underlying the way we measure consumption in poverty and inequality analysis, establishes. We do not expect students to get down to the details and the math required to derive this equation, but it is important to explain that this recommendation is not one of many available, but the only one consistent with modern economic theory. Other options, no matter how attractive, do not have this advantage.


Deaton and Zaidi's recommendation # 1

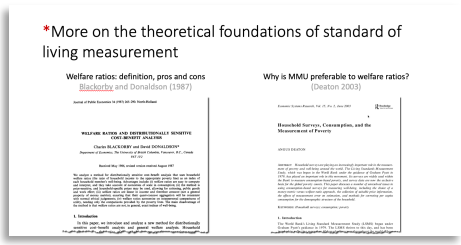
Box 1. Summary of Theoretical Issues and Recommendations

Issue	Recommendation
Money Metric Utility (MMU) vs. Welfare Ratio(WR) MMU is the amount required to sustain a level of living and requires that consumption be adjusted by a Paasche price index that reflects the prices the household faces and whose weights are different for each household. WR is an indication of how much better or worse off a household is than a reference household (usually at the poverty line) and requires consumption to be adjusted by a Laspeyres price index that reflects the prices faced by the reference household but whose weights are the same for all households. The use of MMU can cause difficulties in analyzing the impact of redistributive policy but, on the other hand, WR does not necessarily represent welfare correctly. The latter is the more serious drawback in practice.	Attempt should be made to use Money Metric Utility and to calculate the Paasche price indices with individual household weights.

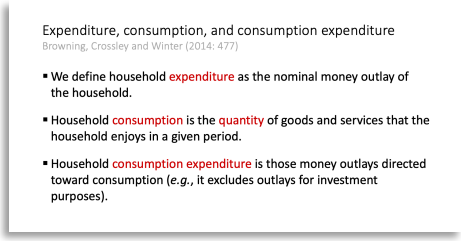
This slide is meant as a final reminder for students that the topics discussed so far, although technical, are essential. The use of money-metric utility (MMU) as a welfare indicator is no less than **number 1** in the list of recommendations made by Deaton and Zaidi in their *Guidelines* for the welfare analyst. MMU is the pillar on which everything rests! *If this concept is unclear, the importance of measuring consumption will be lost on trainees.*



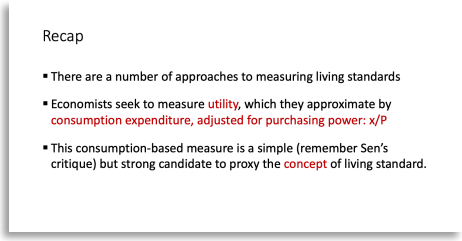
In practice, Deaton and Zaidi's recommendation – to use total consumption expenditure divided by a Paasche price index to proxy living standards – answers the question posed at the beginning of the lecture: **what is the standard of living? x/P** is no longer abstract, or unobservable, like the concept of utility. Instead, it is a concrete and achievable measure of well-being, consistent with economic theory.



This optional slide suggests further readings on the foundations of living standards measurement to interested students.

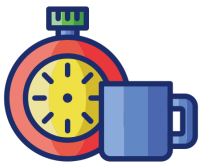


Because we are now dealing more and more with the *practice* of measuring living standards, some clarifications are in order. This slide makes some important distinctions: "expenditure", "consumption", and "consumption expenditure" are different concepts. "**Expenditure**" is the nominal money outlay of the household: simply put, it is the total amount of money spent by the household over a given reference period. "**Consumption**" refers to the quantity of goods and services that the household enjoys, or uses, during the reference period. "**Consumption expenditure**" refers only to the expenditure that is directed toward consumption: in other words, the amount of money spent to buy goods and services that are actually used (as opposed to being stored, which qualifies as an investment) over the reference period. An example may help: suppose that over a given reference period, say one month, a household spends $30 to buy 30 kg of rice. Of these, 20 kg are actually eaten by the members of the family, while the remaining 10 kg are stored, and remain available for future needs (maybe they will be consumed, maybe they will be given away, or wasted: for now, we do not know). In this example, $30 is household expenditure; 20 kg is household consumption of rice; $20 is household consumption expenditure.

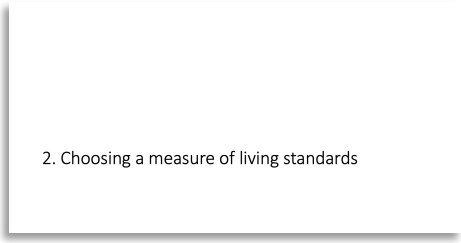


The next slides recap the take-home points from the first part of the lecture. There are a number of different approaches to measuring living standards, which were not all covered here; so far, the focus has been on one approach, which is put forward by mainstream economics. Economists seek to measure utility, which they approximate by consumption expenditure, adjusted for purchasing power. Granted, expenditure excludes potentially important contributors to utility, such as publicly provided goods, or leisure – anything that does not have a measurable price. It also measures a "simplified" concept of well-being which Amartya Sen critiques, as seen in the opening slides. However, despite these limitations, this approach is advantageous enough to be the preferred one for the World Bank's analysis of global poverty.

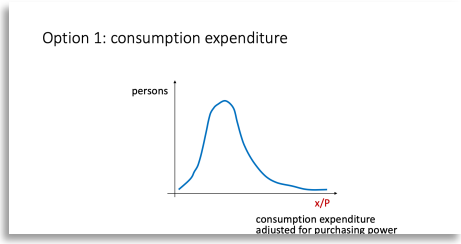
This optional slide mentions an important exception to the points just made: **multidimensional poverty measurement** is increasingly popular around the world. The cited reference is suggested to interested students.



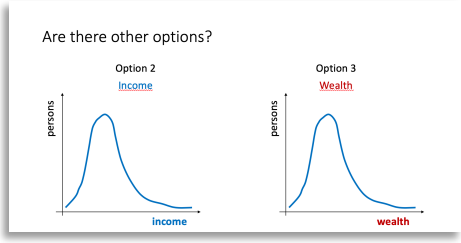
Now is a good time for a break.



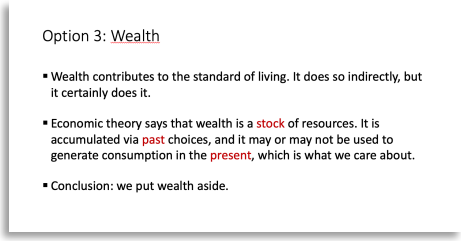
After having laid the theoretical foundations for the choice of consumption expenditure as a measure of living standards, we move to considering other potential candidates, some of which are very common in practice: the final part of the lecture deals with the pros and cons of these competing measures.



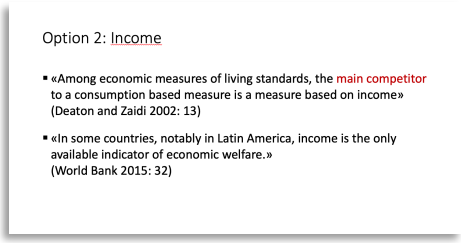
The choice of **consumption expenditure** adjusted for purchasing power follows directly from the theoretical framework of money-metric utility, which was just discussed.



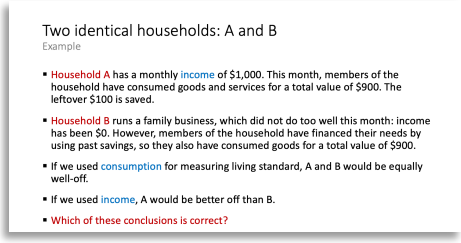
Other potential candidates are: household **income**, that is, receipts, whether monetary or in kind (goods and services) that accrue to the household or to individual members of the household over a given period of time (labor income, capital income, public and private transfers received); and household **wealth** or net worth, that is, the value of savings, investments, real estate and cash, less any debts, possessed by the household or by its members at one point in time.

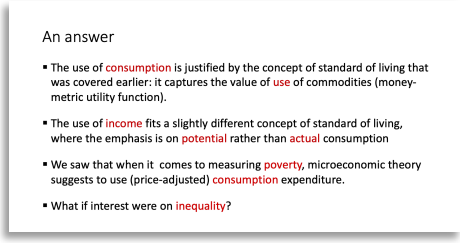


Should these other options be considered viable candidates for the purpose of measuring living standards? **Wealth** is easy to rule out, therefore it is examined first. While it seems logical to believe that wealth contributes to the standard of living, there are both conceptual and practical arguments against it. Conceptually, wealth can be defined as a stock of resources. While consumption expenditure is a flow, which means that it is defined over an interval of time (what was spent over a day, or a month, or a year), wealth is measured at one specific point in time (say, today at midnight), and may have accumulated in the past. Individuals may or may not choose to use their stock of wealth to generate consumption (and thus utility) in the present, which is what we care about. There are also practical reasons that discourage the use of wealth as an indicator of living standards: it is difficult to measure, and data on wealth are only rarely available.

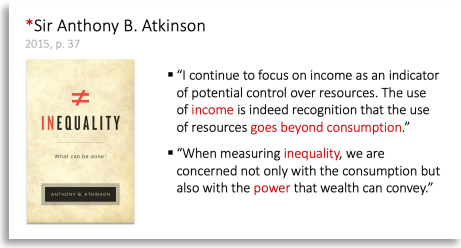


Income, on the other hand, seems like a viable alternative: in fact, many countries use it.

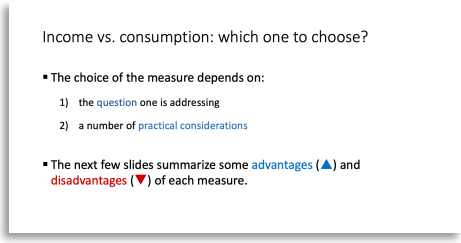




An example of the difference between income and consumption expenditure clarifies that we are talking about two very different objects. The answer to the final question, "which of these two conclusions is correct?", is, in a way, ambiguous. If one accepts the conceptualization of standard of living that was covered in the first part of the lecture, then consumption expenditure is the correct measure, because it captures the value of use of commodities (and therefore, money-metric utility). However, the choice of income can be justified by a slightly different concept of standard of living, where the emphasis is on potential rather than actual consumption. This concept turns out to be particularly fitting when the main focus is inequality, rather than poverty. This is a complex point, which will be the subject of a future course prepared by the World Bank's Center for Development Data (C4D2, focused on *Measuring Income and Wealth*.

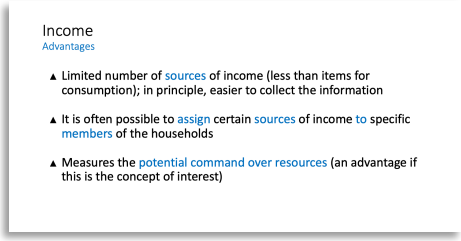


This optional slide elaborates on the point of income vs. consumption. This slide shows quotes from a book by Anthony B. Atkinson, a well-known scholar and expert in the field of inequality, where this alternative conceptual framework is briefly explained. This is just a quick hint, as the main frame of reference for poverty measurement, and for this course, remains money-metric utility; interested students can go to Atkinson's book for more information.

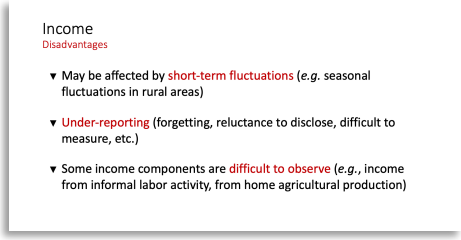


The next slides list some of the **advantages and disadvantages** of income and consumption expenditure as measures of living standards. We suggest to comment on each bullet rapidly, inviting students to think about how convincing each argument is given the

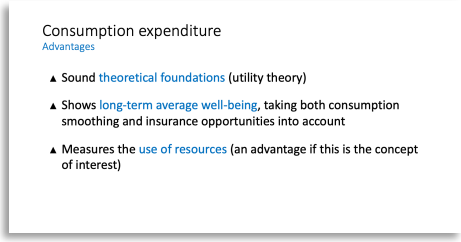
context – we emphasize that context matters, that there is an on-going discussion among experts on the relevance of these arguments, and no easy way exist to reach an agreement. What is important is awareness of these issues, and ability to assess each of them in a specific context.



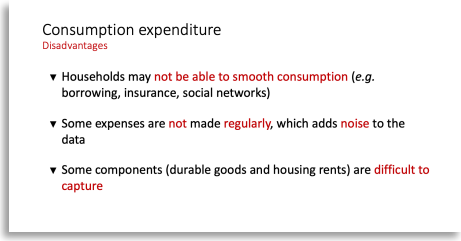
The **advantages of income** are the following: *(i)* in practice, household-level totals of both consumption expenditure and income are often computed as a sum of sub-components (expenditures in various items, various sources of income), and sources of income are fewer in number than consumption items: this makes it easier to gather information on income, at least in this regard; *(ii)* some sources of income are directly connected to the individuals earning them, which allows for analyses of individual living standards in a way that consumption expenditure does not; *(iii)* the final point reiterates what was covered earlier: income is consistent with a concept of well-being as potential command over resources, and this makes income a good candidate when analysts are interested in such an approach.



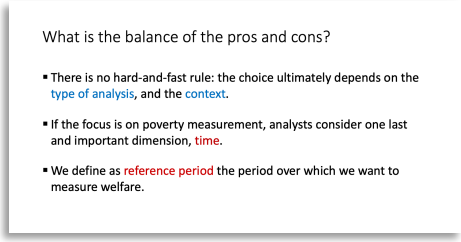
The **disadvantages of income** are the following: *(i)* income is frequently affected by short-term fluctuations, especially in rural areas: we are usually interested in a measure of living standards that is representative of a longer reference period, say a year (more on this later), and it is challenging to adjust for these fluctuations at the data collection stage, and at the analysis stage; *(ii)* under-reporting is a serious issue when it comes to collecting data on income, and exposes the measure to significant measurement error; *(iii)* some income components are difficult to observe: two cases that are especially relevant in poorer countries are that of revenues from informal labor, and that of valuing incomes from agricultural home production (which are in-kind).



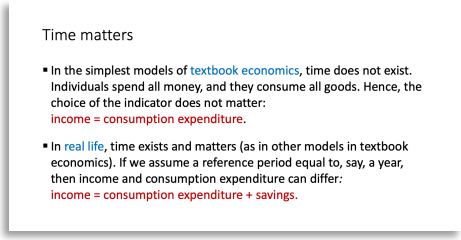
The **advantages of consumption expenditure** are the following: *(i)* the choice of this indicator is grounded in utility theory (the framework that was developed at the beginning of the lecture), and this is a desirable feature for empirical analysis; *(ii)* consumption is typically easier to recall than income, which decreases measurement error; *(iii)* again, the final point reiterates the consistency of consumption expenditure with a concept of well-being that coincides with the actual fruition of resources, which makes it a good candidate if that is the target of analysis.



The **disadvantages of consumption expenditure** are the following: *(i)* certain expenditures are directed toward items that do not enhance well-being (some economists call these items "bads", in contrast with the term "goods" that is used generally): fines are an example; *(ii)* families may make extraordinary expenditures, that are not representative of their usual living standards: weddings and other social functions are an example; *(iii)* some components of consumption expenditure, like the use-value of durable goods and housing, are difficult to measure and require estimation procedures (more on this in a coming lecture).

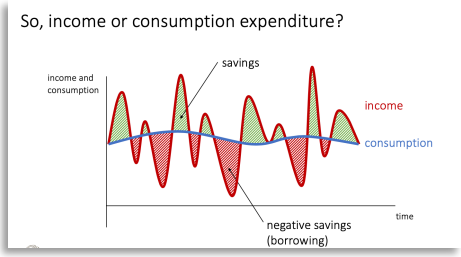


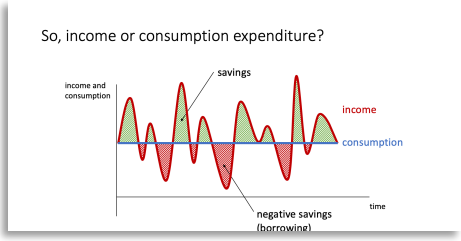
So, what is the balance of the pros and cons that have just been reviewed? As previously mentioned, the choice depends on the goals of the analysis and on the context. However, a final argument can be made regarding pros and cons of the two competing measures, and that concerns the reference period, which measured living standards are supposed to be representative of.



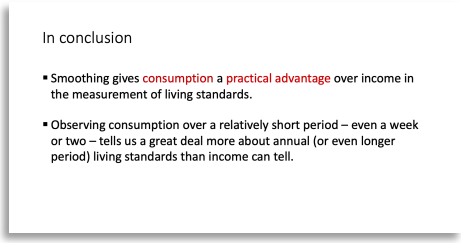
According to the simplest models of consumer choice presented in textbooks of economics, **time** does not exist. Individuals make their consumption choices just once, selecting the

optimal bundle, and spending all of their available budget. In this extreme case, income and consumption expenditure are the same. However, in real life, time matters. In general, income and consumption expenditure can differ*:* income = consumption expenditure + savings (where savings may be negative, which is called dissaving, and occurs when there is borrowing). We are interested in the living standards of households over a given period of time, typically **a year**, and we want our estimates to be representative of that period.





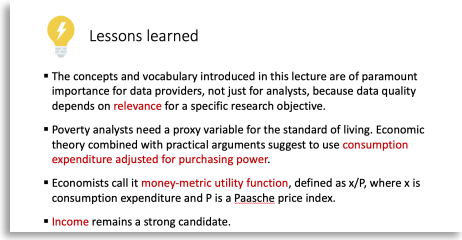
The implication of this discussion for our choice of income versus consumption expenditure is explained by this figure. If the reference period we are interested in is a year, then we must evaluate income and consumption expenditure (or even wealth, or a combination of all three) over that period of time. Empirically, consumption expenditure fluctuates less (in other words, it is **smoother** over time) in the short-run than income does – refer to Deaton and Zaidi (2002: 14) for a discussion of this topic. The graph helps intuitively grasp what "smoothing" is.



This slide wraps up the discussion on **income vs. consumption**: among the advantages of using consumption expenditure as an indicator of living standards, we must count its (comparative) stationarity over time: because of smoothing, observing consumption over a relatively short period – even a week or two – tells us a great deal more about annual (or even longer period) living standards than income can.

At this point of the presentation the audience is most likely tired… so the way we conclude is by showing them something concrete, and hopefully interesting. This bunch of slides provides some examples of the choices made by countries around the world when it comes

to measuring poverty and inequality, to provide a sense of what the common practice is, and how it varies. This overview is meant to give students some practical examples, and to ground the concepts covered in the lecture into practice.



The final slide is a recap of the whole lecture: it is a good idea to spend a few minutes on it, to make sure that the main bring-home conclusions are clear. The first point is especially important, and justifies the time and effort spent on the theoretical issues that were the focus of the lecture: the concepts and vocabulary just introduced are of paramount importance for data providers, not just for analysts, because data quality depends on **relevance** for a specific research objective. If data providers and analysts do not share the same conceptual tools, data will never match user needs and expectations, no matter the amount of money and effort spent in collecting them. The remaining points summarize the key messages of the lecture.

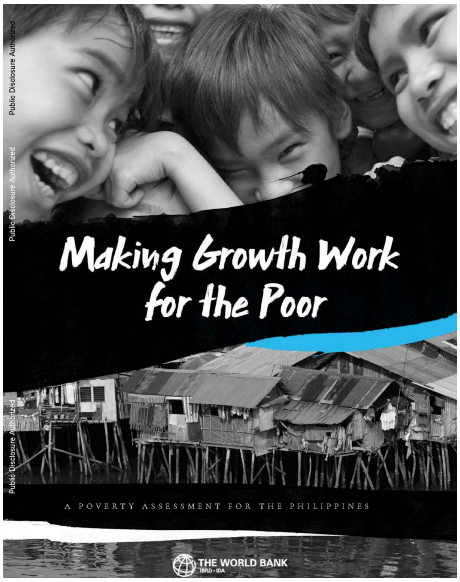
Homework

Exercise 1 – Engaging with the literature

This exercise is conceived as an invitation to get involved with relevant pieces of literature that are related to the topics covered in the lecture. The emphasis here is not on the essay itself, which should be brief and synthetic, but on the student's understanding of the key conclusions of the papers, and how they relate to the lecture. Obviously, there is not a specific and unique solution for this exercise: the Instructor should evaluate each short essay by keeping in mind the intent that was just described. The same holds for all "Engaging with the literature" exercises in the course.

Exercise 2 – Income or consumption?

This exercise gives students an opportunity to see how the theoretical principles covered in the lecture operate in practice, by exploring actual Poverty Assessment reports and the analytical choices they make. Students may find the cross-country variation in the choice of income or consumption interesting, and may also face instances in which the indicator of choice is not entirely clear – which is a valuable lesson in transparency and accountability. An example of what students may find is from the latest for the Philippines (2018), "Making Growth Work for the Poor: A Poverty Assessment for the Philippines":



Box 1.1. Poverty estimates using national and international poverty lines

National official poverty estimates in the Philippines are produced by the Philippine Statistics Authority (PSA). These are derived using income welfare aggregates evaluated against per capita poverty lines that are set broadly following the cost-of-basic-needs (CBN) approach. Using a national reference food bundle based on expert opinion of what constitutes a nutritionally adequate bundle, province-specific bundles are set separately for urban and rural areas and reflect locally consumed commodities. These locally

Exercise 3 – Multidimensionality of well-being

The goal of this exercise is to show that drawing conclusions on a country's living standards based on multiple indicators of well-being, while attractive (and useful in some contexts), is challenging, as different indicators do not normally agree with each other. Once again, the emphasis here is not on the essay itself, but on students' ability to engage critically with some of the concepts covered in the lecture.

Lecture 2
The consumption aggregate

Learning objectives

The goal of this lecture is to complete the conceptual framework underlying the construction of a measure of living standards. Building on the conclusion reached by Lecture 1 – preference should be given to a consumption-based measure of living standards – this lecture provides a working definition of consumption expenditure, and discusses the adjustments that analysts should apply to it, in order to construct a *living standard indicator*.

Suggested preparation

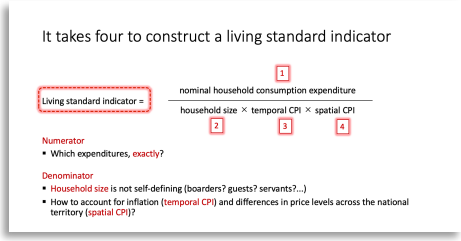
The Instructor is assumed to be familiar with the material contained in sections 4 and 5 of Deaton and Zaidi (2002).

Time allocation

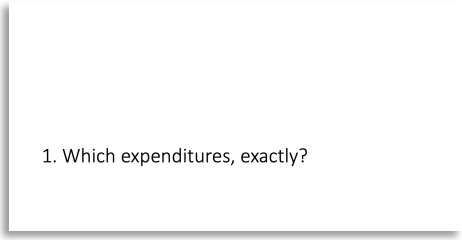
Roadmap and 'which expenditures?'	**15 min**
Adjusting for household size and composition	**30 min**
Adjusting for purchasing power	
Terminology	10 min
Break	**15 min**
Inflation example, price indices	45 min
Lessons learned	**5 min**

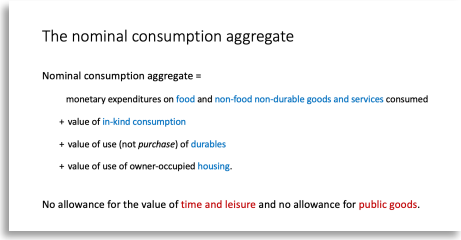
Annotated lecture

The lecture opens with the agenda for the day.

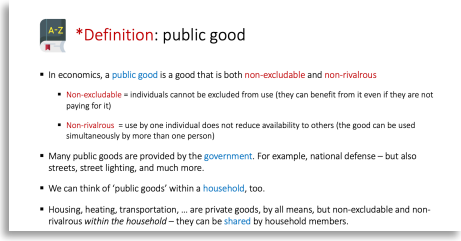


This slide provides the roadmap for the whole lecture. Our objective, constructing a living standard indicator on the basis of survey data, consistently with the theoretical framework covered in Lecture 1, is going to be reached by combining four elements, as shown here. This lecture will be discussing each of the four elements.

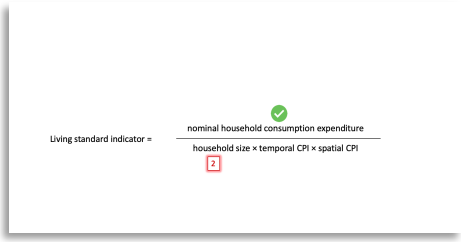




The next few slides discuss the numerator of the living standard indicator: nominal household consumption expenditure. The takeaway for this section is that this step does not simply imply adding up all household expenditures, precisely because what matters is *consumption*, rather than *expenditure* per se. The discussion of this topic is intentionally kept short, as it relates more to analysis than to data collection. More space will be given to adjustments (the denominator of the living standard indicator), which have more bearing on survey design choices.



A quick note on the slide above: slides with the "encyclopedia" icon provide **definitions** for key concepts. They may be technical concepts, or, as in this case, ideas that are quite well-known, but for which it is useful to give a *precise* definition, so that everybody in the room is on the same page. Here we use the concept of 'public good' in a general fashion, to indicate all those goods that are owned by the household, but are shared between household members (*i.e.* they are on-exclusive and non-rivalrous within the household). For a precise definition see Stiglitz (2015), *Economics of the Public Sector*, ch. 5.



This slide signals that the numerator of the living standard indicator has been dealt with: it is time to move on to the next component, that is, the adjustment for household size and composition.

2. Adjusting for household size and composition

...

Adjusting for household size and economies of scale

- Larger households consume more, because there are "more mouths to feed"
- One possibility is to consider per capita consumption
- There is a subtler issue: housing, heating, transportation etc. are shared between members: an analogy can be drawn with the idea of public goods
- Example: housing. Consumption by one member of the household does not necessarily reduce the amount available for consumption by another person within the same household. Economists say that there are significant economies of scale for housing.
- By failing to adjust for economies of scale, one might underestimate the wellbeing of large households (and overestimate that of small households).

Some introductory slides explain why we should concern ourselves with differences in household size (including the possibility of accounting for economies of scale) and household composition.

Economies of scale: adjustment

- A popular strategy is to rescale household consumption expenditure as follows:

$$\tilde{x}_i = \frac{x_h}{(n_h)^{\alpha}} \qquad \alpha \in [0,1]$$

- $\alpha = 1$ means we assume that no goods consumed are public in the household, in which case consumption is equally divided among household members. No adjustment for economies of scale is made.
- $\alpha = 0$ means we assume that all goods consumed are public in the household. This is a purely hypothetical situation in which each individual is assumed to consume the total consumption of the household.
- In practice, α assumes conventional values. E.g. $\alpha = 0.5$ implies that a household of four persons needs twice as much as a single-person household.

Economies of scale: to adjust or not to adjust?

- When a high percentage of budget is devoted to public goods (that is, if price and quantity of housing, utilities, and durable goods are high), economies of scale are likely to be significant.
- Analysts look at the shares in the data, and decide whether to adjust.
- Rule of thumb:
  large share = adjust
  small share = do not adjust

The following slides explain the most common approach used in practice to adjust for economies of scale (using the coefficient 'alpha' to modify household size), and shares a general principle to help decide whether or not such an adjustment is necessary.

Adjusting for household composition

- It is usually assumed that children and the elderly need less than working-age adults.
- Similarly, it is thought that women need less consumption than men.
- If that is the case, our standard of living indicator should account for differences in household composition.

Equivalence scales: adjustment

- If adjustments are to be made, we use equivalence scales.
- An equivalence scale calculates the number of equivalent adults in the household.
- For example, an equivalence scale may look like this:

$$ES = (n_{males\ 15+} \times 1) + (n_{females\ 15+} \times 0.8) + (n_{kids\ 0-14} \times 0.5)$$

where ES denote the equivalent household size, that is, the number of equivalent adults.

- Different categories have different 'weights': adult males may count for 1, adult females for 0.8, ...
- Once the equivalent household size has been calculated, we use it to rescale household consumption expenditure as follows:

$$\tilde{x}_i = \frac{x_h}{ES}$$

Next, we introduce the concept of adjusting for differences in household composition. This is relatively intuitive: a household made up by one adult and two kids may need less consumption than a household made up of three adults, therefore the two households may enjoy a different level of welfare, even though they may spend the same for consumption. Differences in average human energy requirements (minimum calories needed) are an obvious demonstration of the fact that needs vary with demographic characteristics.

To account for these differences in needs, analysts use equivalence scales. An example illustrates: the basic idea is that to each relevant demographic category, we associate a weight, that describes how much lower the needs of that category are with respect to a reference (usually adult males).

Equivalence scales: OECD and Eurostat

- The OECD equivalence scale (OECD-I) is defined as follows:

$$ES_{OECD-I} = 0.3 + 0.7 \times A + 0.5 \times K$$

- where ES_{OECD-I} is the number of 'adult equivalents', that is, the household equivalent size, as measured by the OECD type-I scale.
- The first adult (A) is given a weight of 1. Other adults are given a weight of 0.7, to reflect economies of scale. Children (K) are given a weight of 0.5 to reflect their lower needs.
- In the late 1990s Eurostat adopted the so-called OECD-modified scale (OECD-II):

$$ES_{OECD-II} = 0.5 + 0.5 \times A + 0.3 \times K$$

Equivalence scales: USA

- Originally suggested by Cutler and Katz (1992), and subsequently adopted by the US National Research Council (1995):

$$ES_{NRS} = (A + \alpha \times K)^{\theta}$$

where ES_{NRS} denotes the number of adult equivalents that is, the household equivalent size, as measured by the NRS scale;

- As before, A = number of adults, K = number of children;
- α in [0,1] = cost of a child relative to that of an adult;
- θ in [0,1], $(1 - \theta)$ measures the extent of economies of scale.

The choice of a particular equivalence scale amounts to the choice of the weights. The most commonly used scales are listed here: the slides show the general formulas for the OECD-I and OECD-II scales, and for the so-called NRC scale. The latter accounts for both economies of scale (note the theta coefficient) and household composition, in a single formula.

Equivalence scales in practice

Household composition	Equivalent household size				
	Per capita	OECD-I scale	OECD-II scale	Square root scale (α = 1, θ = ½)	Per household
1 adult	1	1	1	1	1
2 adults	2	1.7	1.5	1.4	1
2 adults, 1 child	3	2.2	1.8	1.7	1
2 adults, 2 children	4	2.7	2.1	2.0	1
2 adults, 3 children	5	3.7	2.4	2.2	1

The table compares common adjustments for household size. The first column, per capita, is the baseline: it simply shows the size of each of the household types listed. The other categories show the equivalent size that results from using OECD-I and II, and the 'square-root scale' (a simple adjustment for economies of scale, that does not account for differences in composition). Finally, the last column shows what happens if we do not adjust at all, and simply use total household consumption.

Equivalence scales: to adjust or not to adjust?

- If children/elderly are as "expensive" as adults despite their lower nutritional requirement (*e.g.* because of very high costs for education or health), less need for adjustment.
- Rule of thumb:
  large differences in the "cost" of different household members = adjust
  small differences = do not adjust.

To wrap up, we share some advice on how to decide whether or not we should adjust for differences in household composition.

The international practice

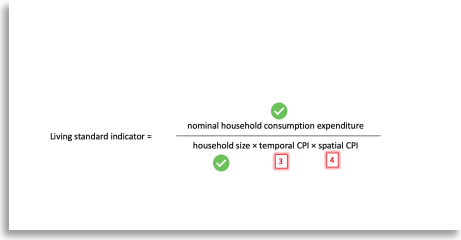
Where do countries around the world fall when adjusting for household size and composition?

...

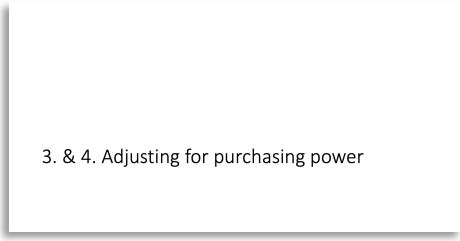
Sub-Saharan Africa

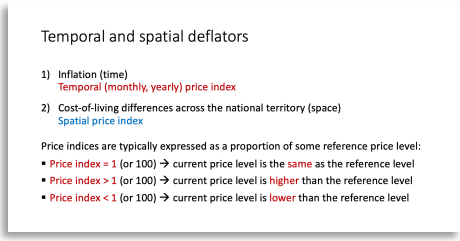
Surveys	Consumption vs Income	Household Size
Côte d'Ivoire 2015	Consumption	Per Capita
Kenya 2015	Consumption	Per Adult Equivalent
Malawi 2016	Consumption	Per Capita
Mali 2018	Consumption	Per Capita
Mozambique 2014	Consumption	Per Capita
Nigeria 2010	Consumption	Per Adult Equivalent
South Africa 2014	Consumption	Per Capita
Tanzania 2018	Consumption	Per Adult Equivalent
Uganda 2016	Consumption	Per Adult Equivalent
Zambia 2015	Consumption	Per Adult Equivalent
Zimbabwe 2017	Consumption	Per Capita

Then, some evidence on what countries around the world actually implement is presented. Different regions of the world adopt different strategies, but it is shown that expressing the living standard indicator *at least* on a per capita basis is routine.

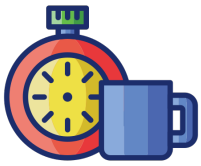


Again, this slide signal that the lecture is now moving on to the last topic: the adjustment for **differences in purchasing power**, which encompasses elements 3 and 4.

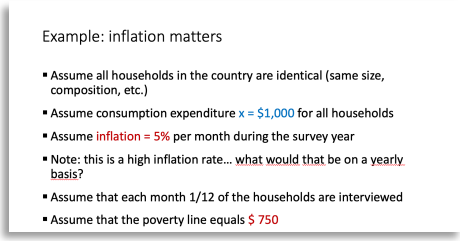


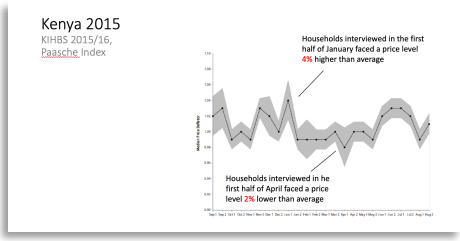


These slides provide some basic terminology: nominal vs. real, consumer price indices (CPI), temporal and spatial deflation. The way CPIs should be interpreted is also clarified.

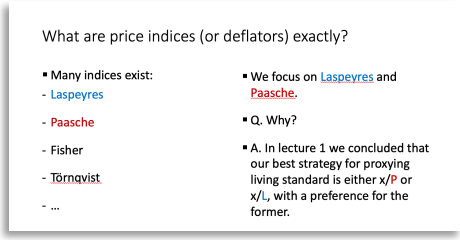


Now is a good time for a break.

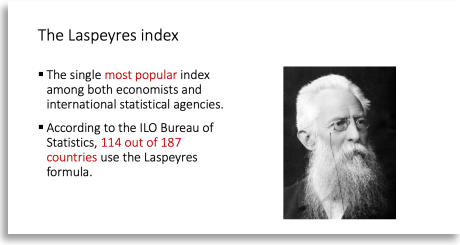


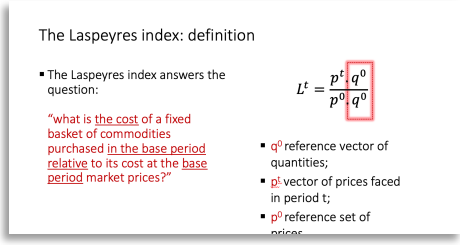


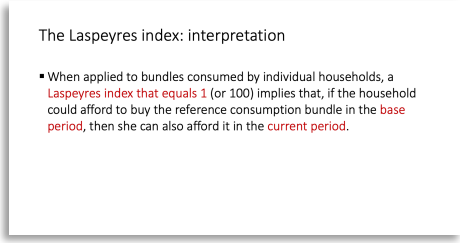
These slides present a hypothetical example, that demonstrates why adjusting for differences in purchasing power is important. The example focuses on temporal price differences, i.e. inflation, but the same reasoning holds for spatial price differences. The upshot is that using nominal or real expenditure produces different poverty estimates, and that the second measure best represents actual living standards. A real example from Kenya concludes the discussion, showing the relevance of this issue in practice.



The final batch of slides discusses different price indices, namely Laspeyres and Paasche. Section 4 of Deaton and Zaidi (2002) is the main reference for this topic. The goal is for students to understand that there are several ways to adjust for price differences, and that they are not equivalent.







The presentation of both indices follows a simple structure: first, we introduce the index using its formula. Then, we give the interpretation of the index 'in words'. Finally, we comment on its advantages and disadvantages.

Regarding the formula, the only real difficulty here is understanding the notation: keeping in mind the example of temporal differences in prices (but everything applies to spatial differences, too) we can think of the apex '0' as the reference or base period, and of the apex 't' as the current period. The letter q indicates a set of quantities of goods consumed by households; p indicates the corresponding prices. Note that the 'dot-product' here indicates the sum of each quantity times its price (that is, the total cost of the consumed bundle).

Once this is clear, the expression for the Laspeyres index has no secrets: it is the ratio of the costs of a reference bundle of goods, evaluated at the base period and at the current period.

The Laspeyres index: comment

- A key feature of the Laspeyres formula is that it tends to overstate the rise in the cost of living by not allowing any substitution between goods to occur (Diewert, 2001).
- To the extent to which price and demanded quantity are negatively correlated, the Laspeyres index provides an upper bound to the "true cost of living" faced by a household.

To comment on the Laspeyres index, we share that one of its key features is that it tends to *overstate* cost of living differences, by not allowing any substitution between goods to occur when prices vary. This is a consequence of holding quantities q^0 fixed at the base (reference) level: in reality, when faced with a price change, households tend to reallocate their resources away from relatively pricier and towards relatively cheaper goods. Consequently, the Laspeyres index provides an upper bound to the cost of living faced by a household.

The Paasche index

- The Paasche index is the one that most welfare analysts opt for.
- Deaton and Zaidi (2002) explain why.

The Paasche index: definition

- The Paasche index:

"what is the cost of a fixed basket of commodities purchased in period t relative to its cost at the base period market prices?"

$$P^t = \frac{p^t \boxed{q^t}}{p^0 \boxed{q^t}}$$

- q^t vector of quantity purchased in period t;
- p^t vector of prices faced in period t;
- p^0 reference set of prices

The Paasche index: interpretation

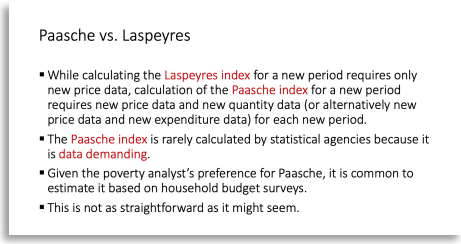
- When applied to bundles consumed by individual households, a Paasche index that equals 1 (or 100) states that, in the base period, a household could have consumed the same bundle as she is consuming in the current period.

Using the same logic, we define and interpret Paasche. The key difference between the two indices lies in how prices are *weighted* by the quantities consumed: while both indices account for *price relatives*, that is, the prices in the current period relative to the reference prices (p^t/p^0), the Paasche index also accounts for the specific consumption pattern of each period t (q^t), something that is not true of a Laspeyres index, which instead uses a set of reference quantities, q^0, that is the same when the index is computed in all subsequent periods. The importance of this distinction is clearer in the spatial context: Laspeyres uses weights (quantities) that are common to all households, while Paasche uses household-specific weights.

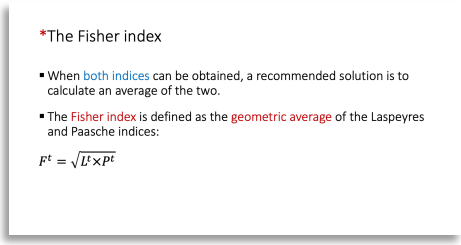
The Paasche index: comment

- The Paasche formula does not allow for the substitution of products or services at the base period prices.
- To the extent to which price and demanded quantity are negatively correlated, it provides a lower bound to the "true cost of living" faced by the household.

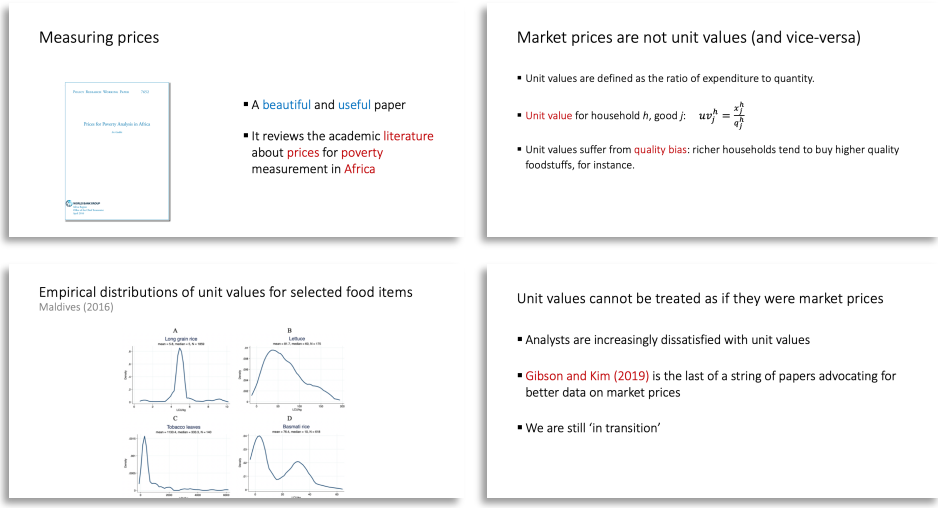
The Paasche formula suffers from the opposite drawback as the Laspeyres formula, because the weights are set at the level that is optimal at prices – in a way, after consumers have completed all substitutions of goods and services in response to a change in prices. To the extent that price and demanded quantity are negatively correlated, the Paasche index provides a lower bound to the cost of living faced by the household.



In our context, however, other considerations are more important. When the focus is welfare measurement, the preferred price index is Paasche, for the reasons explained in Lecture 1. The challenge for NSOs is to collect the data necessary to its estimation.

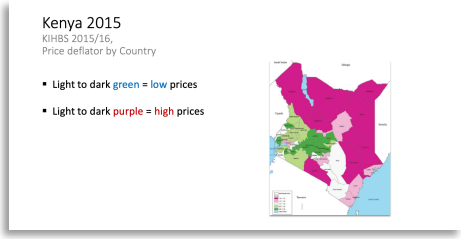


This optional slide touches on another index that is widely used in practice, the Fisher index.

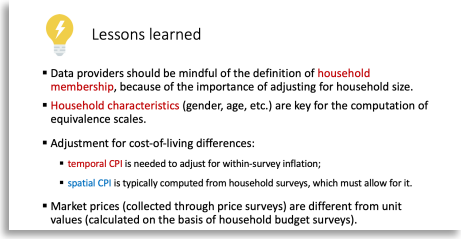


A common source of information on prices comes from household surveys themselves: *unit values*, that is, the ratio of expenditures to quantities, are often easy to compute based on the food modules of most expenditure surveys. Historically, they have been used as

proxies for market prices. However, unit values have become increasingly embattled in the recent literature on price indices, because of the many empirical problems they have. Goods of very different qualities may be grouped under the same category in household surveys, and the resulting unit values may be averages of a miscellaneous bunch. The graphs from Maldives illustrate this point: the top-left distribution looks reassuring, because it indicates that reported unit values for rice are concentrated around a single value, but the other examples suggest the presence of underlying quality differences. However, the literature on this topic has not yet converged on a concrete alternative: the transition away from unit values is not yet complete.



Again, the discussion wraps up by showing a real example of spatial differences in prices.



The final slides summarize the topics discussed in the lecture, with a focus on their implications for survey data collection.

Homework

Exercises 1 and 2 – Engaging with the literature

See exercise 1, Lecture 1.

Lecture 3
Understanding household surveys

Learning objectives

The goal of this lecture is to provide an overview of the many different survey instruments developed and implemented by statistical institutions around the world. The material covered in the lecture puts household consumption and expenditure data – the main focus of the course – into a broader context.

Suggested preparation

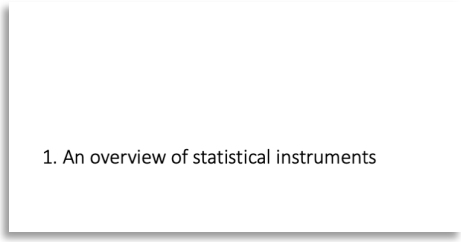
The paper by Grosh and Glewwe (1998) provides background information on the Living Standard Measurement Study and the surveys developed under its umbrella. Familiarity with it is assumed.

Time allocation

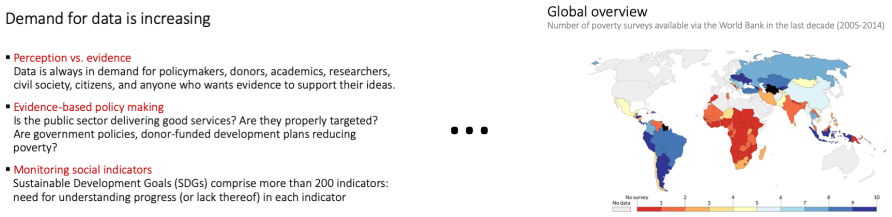
Overview of statistical instruments	**20 min**
Overview of household surveys	
Introduction and terminology	30 min
Break	**15 min**
Taxonomy: common surveys, quick HCES, large HCES	50 min
Lessons learned	**5 min**

Annotated lecture

The set opens with a roadmap: after concepts, which were covered in the previous lectures, it is now time to move to data. There are many different ways for statistical institutions to collect data, and this lecture provides broad classifications and terminology for different data collection efforts.



The first part of the lecture provides an overview of statistical instruments *in general*.

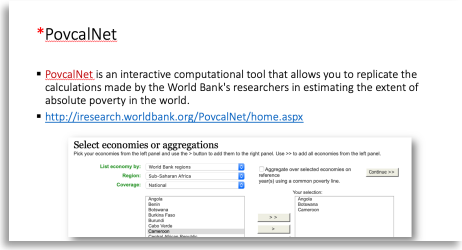


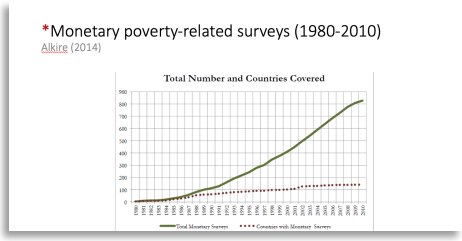
By way of motivation, the first few slides make the point that the demand for data (by policymakers, researchers, society in general) is increasing; the **supply of data** is increasing too, but there is still much work to do. The maps presented here come from the website Our World in Data (https://ourworldindata.org/extreme-poverty), and represent the number of recent household surveys that are available around the world as a basis for the World Bank's estimates of global poverty.

The countries in grey are those where not a single survey is available to the World Bank in the last decade. Many of these countries are rich countries, in which extreme poverty is very low. But there is also missing data for some poorer countries, in which surely a considerable share of the population is living in extreme poverty.

The countries in red and orange are those with very few available surveys. This is the case for many African countries, where there is only one survey available in the last decade.

The last map gives a general overview of 'data deprivation' in the field of poverty measurement. The key point is that efforts toward better data collection are important.

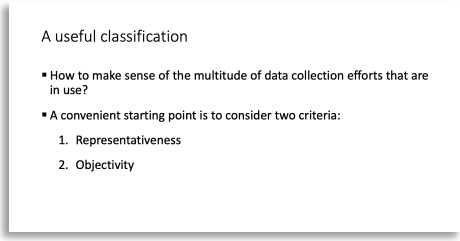


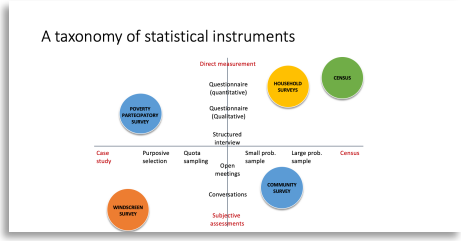


These two optional slides provide two more examples of data availability in the field of poverty measurement. **PovcalNet**, in particular, can be an interesting tool for students to explore: it allows to replicate the World Bank's official poverty estimates, and can be used to compute poverty measures on imported data as well.[1]

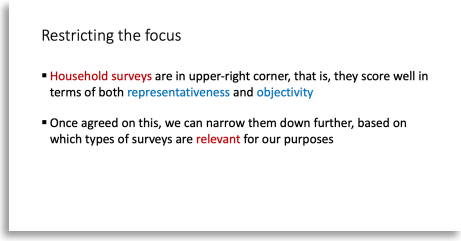
The evidence of an increasing demand for data is used to justify the multitude of different statistical instruments currently in use.

[1] For an introduction to PovcalNet, see http://iresearch.worldbank.org/PovcalNet/introduction.aspx.

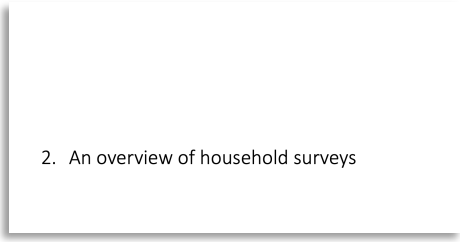


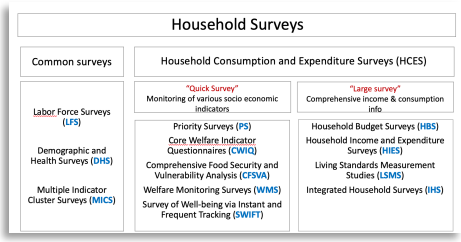


A classification based on two dimensions – **representativeness and objectivity** – can be a useful tool to evaluate these different instruments. Notable examples of information gathering are placed on a diagram defined by these two dimensions. Note that a *windscreen survey* is not a survey at all, but rather an informal, subjective assessment of one's surroundings, done from one's seat in a car; and a *poverty participatory survey* is similar to a focus group, ran with the aim of understanding poverty determinants by incorporating the perspectives of the poor themselves.

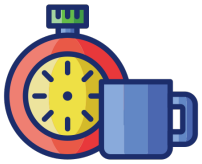


Representativeness and objectivity are desirable properties, and they help restrict our focus to statistical instruments that possess these qualities: **household surveys**.

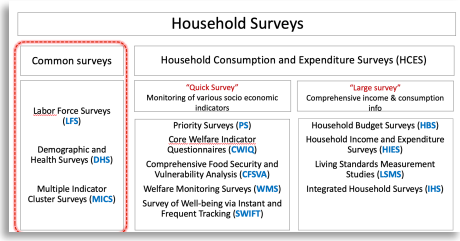


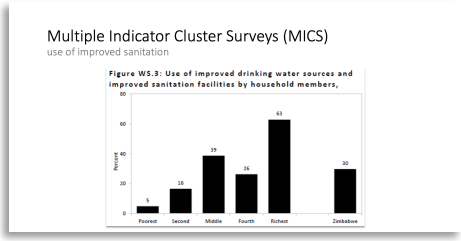


The second part of the lecture focuses precisely on household surveys. The topic is introduced with some historical background and some terminology (for '**household**' – the case of polygamy, which is common in some countries, is used to engage students and give an example of a local context where defining a household may be challenging – and '**probabilistic sample**'). Once the basics are covered, a slide introduces a classification of the main types of household surveys. On the one hand, we have **'common' surveys** – common in the sense that they are virtually ubiquitous, frequently mentioned, essentially a staple for most statistical offices around the world. On the other hand, we have Household Consumption and Expenditure Surveys (**HCES**): not necessarily less 'common' than other surveys, what characterizes them is their focus on measuring consumption and expenditure; they can be divided into **'quick' and 'large' surveys**. The following slides go into each of these groupings.

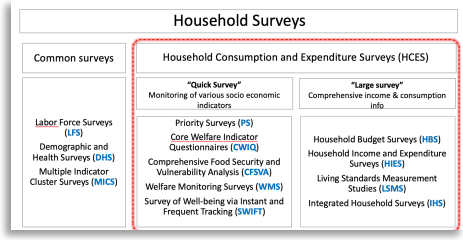


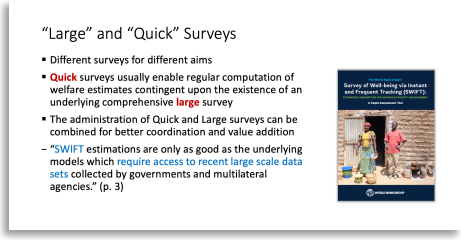
Now is a good time for a break.





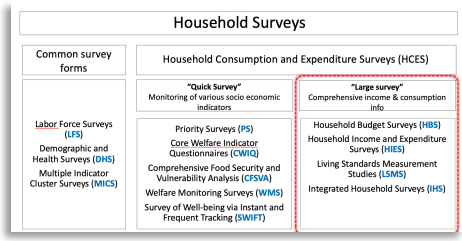
The group of slides on **common surveys** is structured as follows: each survey type (LFS, DHS, MICS) is introduced with a general description; a recent example from a country is presented, to show what the survey might look like in practice; a final slide shows an indicator or statistic that is usually computed from that type of survey. This section can be covered rather quickly, as students may already be familiar with some of the material.

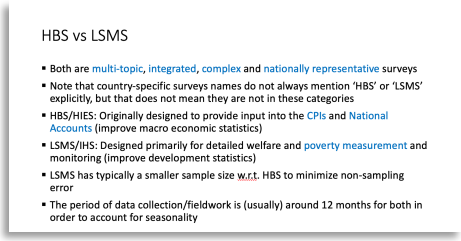




The discussion then moves to **HCES**. **'Quick' surveys** are described synthetically, as they are complementary to 'large' surveys. Quick surveys are especially relevant in specific context, like countries impacted by conflict and violence, where efficiency and cost-effectiveness are paramount. A full discussion of this rather specialized instrument goes beyond the scope of the course: interested students are directed to the suggested references for more details.

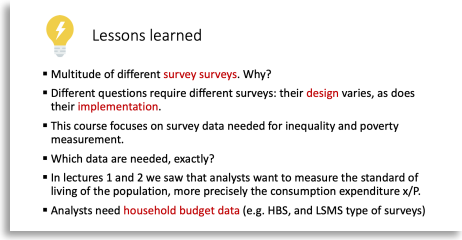
Optional (starred) slides discuss an example from Indonesia, and the Nigeria CWIQ 2006.





Finally, the lecture zooms into **'large' HCES**. This part should be emphasized over the others, as it is more important for what will follow in the rest of the course. In particular,

HBS and **LSMS** are two main approaches to conducting HCES, they are both routinely used for measuring living standards, poverty and inequality, and the differences and similarities among them should be adequately discussed.



The bring-home conclusion is that welfare analysts need household budget data (e.g. HBS, and LSMS type of surveys). By the end of the lecture, the characteristics of this type of surveys, and the general context in which they are situated, should be clear to students.

Homework

Exercise 1 – Engaging with the literature

See exercise 1, Lecture 1.

Exercise 2 – Household?

This exercise gives students an opportunity to see how the theoretical principles covered in the lecture operate in practice, by exploring actual survey documentation material and the definitions underlying the design of questionnaires. An example of what students may find is from the enumerator manual of Malawi's Integrated Household Panel Survey 2010-2013:

"A household may be either a person living alone or a group of people, either **related or unrelated**, who **live together** as a single unit in the sense that they **have common housekeeping arrangements** (that is, share or are supported by a common budget). A standard definition of a household is: "**a group of people who live together, pool their money, and eat at least one meal together each day**". It is possible that individuals who are not members of the household may be residing with the household at the time of the survey. In most cases, but not all, someone who does not live with the household during the survey period is not a current member of the household."

The students' findings can be used as a basis for discussion of the differences, or similarities, of such a fundamental definition across countries.

Exercise 3 - Household surveys

This exercise encourages students to identify the main features that characterize a survey, and pushes them to read and get familiar with technical documentation. Again, there is no single right answer for this exercise: the students' findings can be used to start a discussion on the topic of harmonization. A good solution might look like this:

country	survey	year	Sample size	Sampling method
Tanzania	DHS	2015-2016	13,376 hh	Stratified sample selected in two stages. Probability proportional-to-size selection at the first stage of sampling
Tanzania	NPS	2014-2015	3,265 hh	multi-stage cluster sample design.
Ethiopia	DHS	2016	18,008 hh	Stratified sample selected in two stages. Probability proportional-to-size selection at the first stage of sampling
Ethiopia	ESS	2015-2016	5,469 hh	Two stage sampling. Probability proportional-to-size selection (of EAs) at the first stage of sampling
Zambia	LCMS	2015	12,260 hh	two-stage stratified cluster sample design. Probability Proportional to Estimated Size (PPES) at the first stage of sampling
Zambia	LFS	2014	11,520 hh	disproportionate allocation to strata

Exercise 4 - Sherlock Holmes

The goal of this exercise is to strengthen the students' awareness that any aggregated indicator originates from survey data, and that different surveys are used for different purposes (typically, labor force and poverty statistics are computed on the basis of Labor Force Surveys and Household Budget Surveys, respectively).

Exercise 5 - The Gemini Project

The goal of this exercise is to acquaint students with the Gemini Project, an important initiative launched to help redesign the US Consumer Expenditure Survey, addressing issues of measurement error and respondent burden. The Project has spurred considerable research effort, and many of the papers published under this umbrella contain findings that are relevant for other countries.

Exercise 6 – Historical Household Budget Surveys

This exercise is similar to exercise 3, but this time the focus is on considering the issue of harmonization of survey methods over time.

Lecture 4
Principles of questionnaire design

Learning objectives

The goal of the lecture is to review some general principles of questionnaire design, as they apply to the design of surveys to measure consumption. Principles include both general planning recommendations, and specific tips on formulating questions, applicable to any module of the survey.

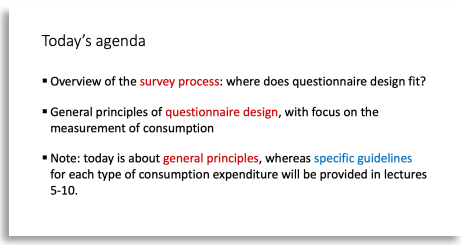
Suggested preparation

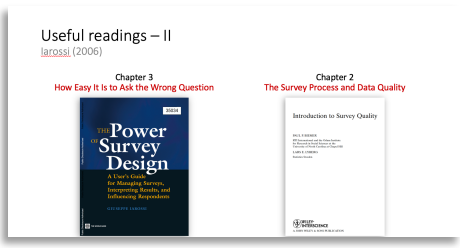
Chapters 2, 3 and 5 of Grosh and Glewwe (2000) are a classic reference for the topics reviewed in this lecture, and are especially useful given their focus on LSMS surveys. The more recent Glewwe (2005) revisits some of the same material, in a more synthetic fashion. Familiarity with both works is assumed. Reviewing chapter 2 of Biemer and Lyberg (2003), and chapter 3 of Iarossi (2006) is also advised.

Time allocation

The survey process	**20 min**
Principles of questionnaire design	
Choice of topics and respondents	40 min
Break	**15 min**
Formulation of questions, order of questions, field testing	40 min
Lessons learned	5 min

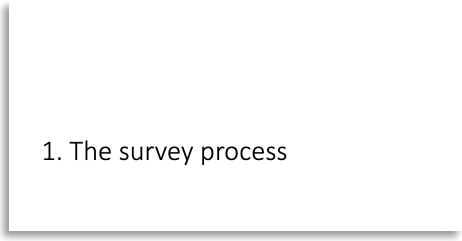
Annotated lecture

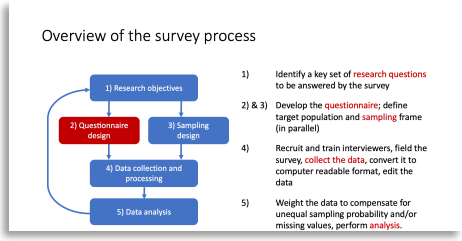




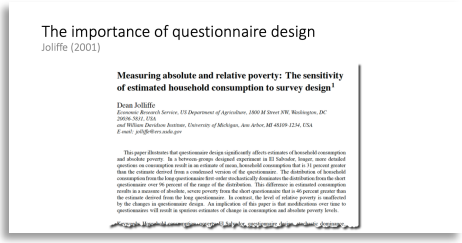
The set opens with the agenda for the day. We remind students that the lecture will focus on **general principles** underlying the questionnaire design stage, and that guidelines that apply to the design of individual modules will be discussed in upcoming lectures. We also

invite students to engage with the recommended readings, which provide the material covered in this lecture.

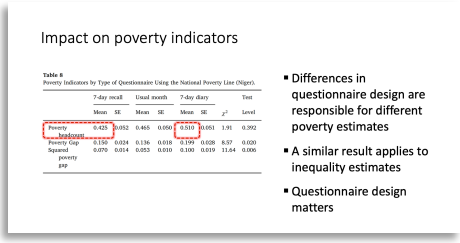




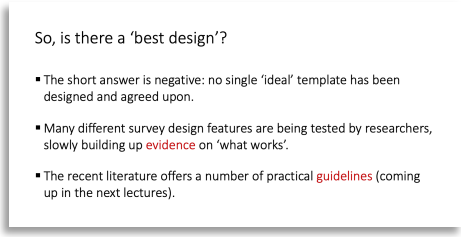
First, it is useful to put questionnaire design into **context**, among the different steps that make up the survey process. The design of the questionnaire is informed by research objectives, which in turn are the goal of data analysis: the idea that the last phase of the whole process – the actual use of the data – defines priorities at the very beginning of it is represented by the arrow that circles back from step 5 to step 1 in the figure.



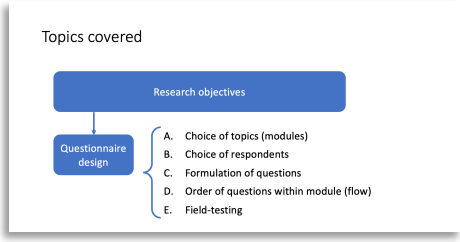
...



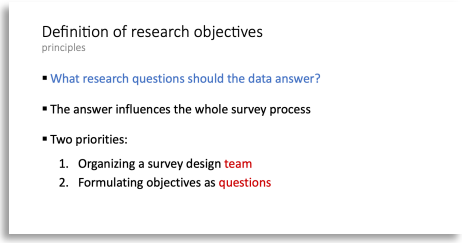
The papers cited in these slides, Joliffe (2001) and Backiny-Yetna et al. (2017), are here as an example of the fact that questionnaire design makes a great deal of difference to final estimates. Both papers will be discussed further in the next lecture, so there is no need to spend much time on this part. Suffice it to say that these are two examples of research comparing questionnaires that differ only in seemingly small and irrelevant features, and conclude that these features are responsible for large differences in the estimates of interest. The takeaway is: **details matter**.



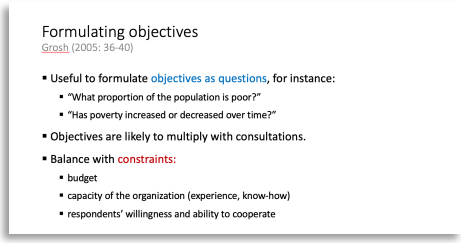
At this point, seeing that there is no definite wrap-up of the findings of this literature, students may ask the question: which one among those tested is the 'best' questionnaire design? This slide gives the Instructor a chance to reiterate that the idea of a single, definitive 'best' design is a fallacy. Empirical evidence helps to define best practices for deciding among alternatives. Many of those best practices and recommendations will be reviewed in the next lectures.



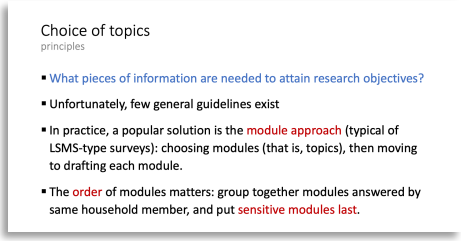
This diagram shows the steps that a questionnaire design team follows when constructing a questionnaire. We will discuss them one by one.



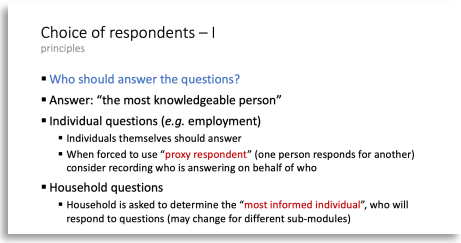
• • •

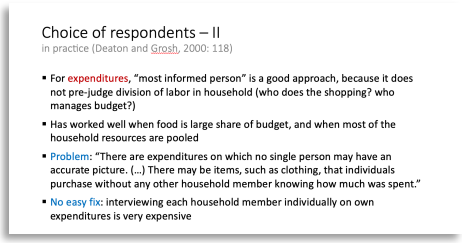


First, a prerequisite for the whole process: the **definition of research objectives**. The Instructor should review the references cited in grey for more details.



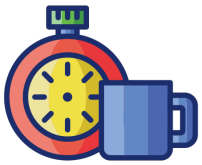
After having defined the research objectives addressed by the survey, it is good practice to proceed by choosing broad **topics** to be included in the questionnaire. These two slides deliver general advice – adopting a **module approach** – and cite the LSMS survey template as an example.



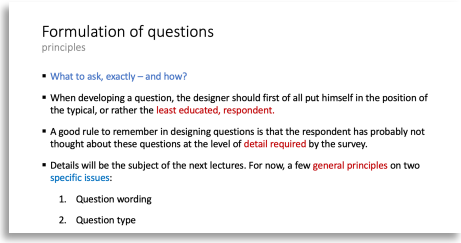


Next, who in the household should be addressed as a **respondent**? It is recommended to select whomever is **most knowledgeable** on the question or module at hand. Deaton and Grosh (2000) point out that for some consumption items, it is unrealistic to expect a single person to have an accurate picture of each household member's own expenditures. In some cases, individual modules are indeed preferred (some examples will be cited in the next

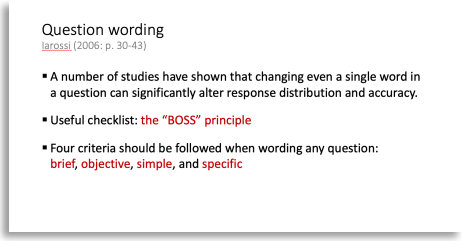
lectures: one is food away from home) but the benefits of implementing such an approach should be weighted against the increased costs with respect to a 'proxy respondent' module.



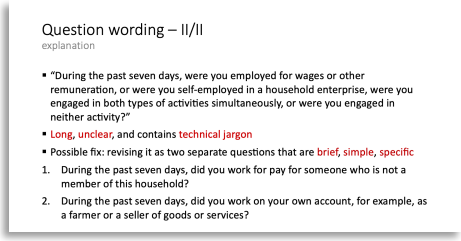
Now is a good time for a break.



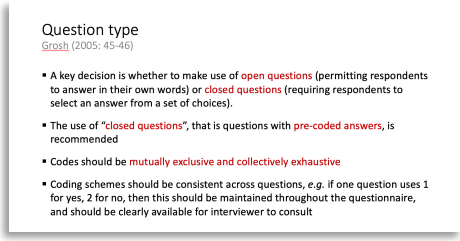
What to ask, and how? The next lectures will address the 'what'; for now, a few words of advice on the 'how' (formulation of questions).

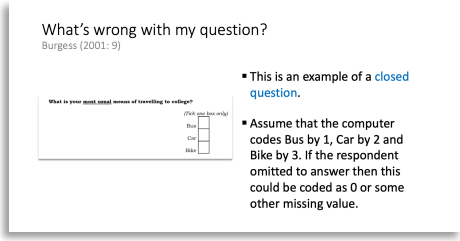


• • •



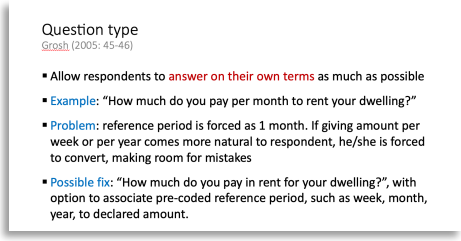
In terms of **question wording**, a useful rule to remember is the so-called 'BOSS principle': questions should be brief, objective, simple, and specific. A few slides explain each criterion with an example: students can be encouraged to try and give their own answers before we discuss 'possible fixes'.



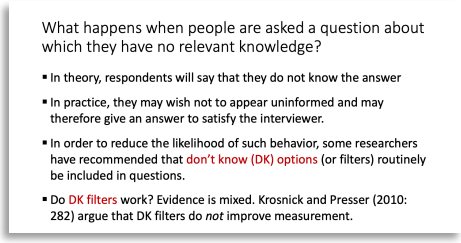


The choice between an open or closed question is about **question type**. Closed questions with pre-coded answers are most common in expenditure surveys: there are a few basic recommendations that help to design them correctly. The example reproduced in the 'bus, car, bike' question highlights the importance of well-designed response codes. The problem with this questions is that, although codes are mutually exclusive, they are not

exhaustive: there is no option for ‘none’, which leads to confusion between ‘none’ and missing.

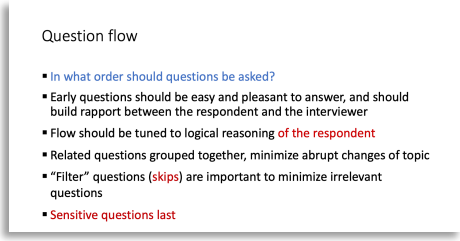


Another recommendation on pre-coded question is to allow enough flexibility to accommodate the heterogeneity of respondents and reduce the cognitive burden of answering an unfamiliar question.

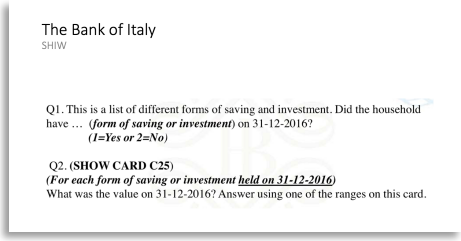


This slide touches on ‘DK filters’. The Instructor is directed to the reference provided for more details.

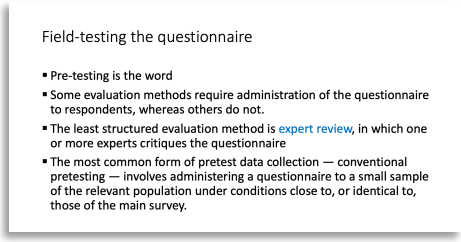
We remind students how much ground was covered, and where we are in the survey process.

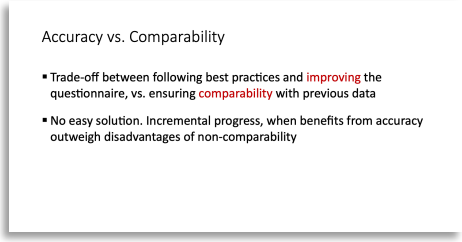


• • •



The order in which questions are asked (the question ‘flow’) is also important. One of the main topics to have been discussed by the literature is that of sensitive questions, and how to ask them: the paper by Barton (1958) is a funny way to convey the issue. A common solution in modern surveys is the **unfolding brackets** approach, which is often used for questions on income.





Finally, a few tips about pre-testing the questionnaire. The fact that no questionnaire is ever perfect, and that arriving to a good design is likely to be an incremental process, leads to some considerations about the benefits and pitfalls of accuracy (continually improving the design of the questionnaire) and those of comparability (adopting the same design as previous waves. These considerations are echoed in the final slide (lessons learned).

Homework

Exercise 1 – Engaging with the literature

See exercise 1, Lecture 1.

Exercise 2 – Question wording

Questions 1 and 3 are too subjective: they will not necessarily yield the present value of the house (according to the market), but how much it is worth according to the owner. Question 2 is worded in a much too technical way, with excessive jargon. Question 4 is just right: it is easy to understand, but at the same time grasps the right concept (what people would be *willing* to pay for the house). In fact, this question is often found in the housing section of many HCES. Question 5 is inaccurate: it does not ask for the net present value of the house, but for its original purchase price.

Lecture 5
Measuring food consumption: the foundations

Learning objectives

This lecture, together with lectures 6 and 7, has the goal of providing detailed recommendations on the collection of food consumption and expenditure data through household surveys. Lecture 5 focuses on the concepts of *acquisition* and *consumption*, and on determining the mode of data collection (*recall* vs. *diary*).

Suggested preparation

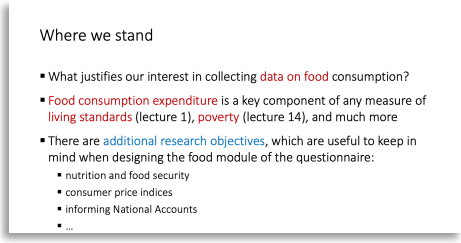
Much of the material covered in lectures 5, 6, and 7 is adapted from two references: FAO and The World Bank (2018), and Smith, Dupriez and Troubat (2014). The instructor is assumed to be familiar with both. This lecture also summarizes the evidence from a few important studies (Beegle et al. 2012, de Weerdt et al. 2016, Backiny-Yetna et al. 2017), which the instructor is encouraged to review.

Time allocation

Introductory slides	**5 min**
Acquisition vs. consumption	
Concepts	20 min
Examples from questionnaires	15 min
Recommendations	10 min
Break	**15 min**
Recall vs. diary and length of reference period	
Concepts	20 min
Evidence from Tanzania and Niger	20 min
Recommendations	10 min
Lessons learned	**5 min**

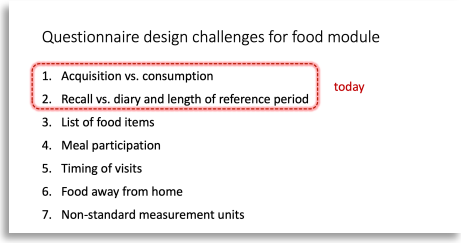
Annotated lecture

The set opens with a brief introduction, which has the goal of re-acquainting the students to the main topic of the course – measuring consumption – after the brief detour taken during Lecture 4 on general principles of questionnaire design.



We begin by saying that today's lecture is the first of a group of three lectures (5, 6, and 7) focused on **measuring food consumption**. What justifies such a strong focus on this topic? First and foremost, food consumption expenditure is a key component of any measure of living standards (lectures 1 and 2), and therefore, it is needed for poverty measurement. But there are several other important research topics that justify an interest in measuring food consumption with household surveys. We may 'wake up' the audience by soliciting answers (2 mins discussion might help). Either way, it should be clear that accurately measuring food consumption is important, for different types of data users and researchers.

This and the following lectures go into the specifics of questionnaire design, and in particular, of the design of the food module. The main references for this group of lectures are the FAO/World Bank Guidelines on food data collection (left), and a recent paper by Smith, Dupriez and Troubat (right). The former illustrates best practices, while the latter is more focused on what countries actually do.



This slide is an outline of what is to come: the topics covered are the main 'challenges' that a questionnaire design team will have to face when implementing the food module of a household consumption and expenditure survey. The red rectangle indicates that Lecture 5

will deal with the first two topics (acquisition vs. consumption, and recall vs. diary), while the remaining issues will be tackled in Lectures 6 and 7.

1. Acquisition vs. consumption

Acquisition vs. consumption is our first topic.

Definitions

- Acquisition
  coming into possession, taking control of goods
- Consumption
  utilizing goods (*i.e.* eating, in the case of food)
- Mode of acquisition:
  - purchase
  - own-production
  - in-kind receipt

Acquisition vs. consumption
what to do with the chicken?

- All goods that are consumed have been acquired in some way
- However, acquisition and consumption do not necessarily take place during the same reference period
- During a given period, say previous week, three possibilities:
  - a chicken is acquired and eaten (A = C)
  - a chicken is acquired, but not eaten (A > C)
  - a chicken is eaten, but has been acquired earlier (A < C)

The distinction between acquisition and consumption is a conceptual one, regarding *what* exactly is captured by the food module. The first of the slides pictured above gives rigorous *definitions* of acquisition and consumption (and adds some terminology on *modes* of acquisition); the second slide, "What to do with the chicken?", clarifies these concepts by means of an *example*. For more details, you can check Section 2 (p. 44) of Conforti, Grünberger and Troubat (2017), which elaborates further.

Why definitions matter

- Acquisition and consumption are measured for different purposes:
  1) Interest in consumption is justified by interest in estimating a number of things: standard of living, calorie intake, etc.
  2) Interest in acquisition is justified by interest in food security (availability)
  3) Interest in acquisition from purchases (*i.e.* food expenditure) is justified by CPI weighting, and informing national accounts
- Based on survey objectives, concept(s) of interest must be clear, and the questionnaire must be unambiguous

Does the distinction between acquisition and consumption really matter? Acquisition and consumption, as we have just seen, are different. Neither is intrinsically superior: each may be useful for different analytical purposes (the slide lists only a few examples). What is important is that the choice of what to measure be consistent with these objectives, and that the measure of interest be clear and unambiguous for both interviewers and respondents.

Current practice
Smith et al. (2014)

- Smith et al. (2014) review 100 surveys from developing countries
- They find that both consumption and acquisition are commonly collected, but poor questionnaire design is common
- About 25% of surveys were found to include poorly worded questions, ambiguity

Approaches to data collection
Conforti et al. (2017: 44)

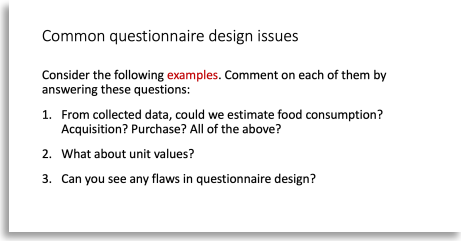
Typically, data on food are collected in one of three ways:

A. Acquisition
Households report on food they acquired through purchases, own production and in-kind transfers. Actual consumption of the same food is not reported.

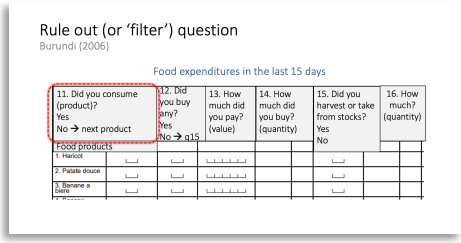
B. Combination of acquisition and consumption
Households report on food they acquired through purchases, without specifying the amount of food consumed. Food consumption derived from own-production or received from transfers is reported.

C. Consumption
Households report on food actually consumed, and on whether that same food was purchased, own-produced or received as a transfer.

We now transition **from concepts to practice**. What do household surveys around the world record: acquisition or consumption? The assessment by Smith et al. (2014) shows that both concepts are routinely measured in practice. However, poor questionnaire design is common: it will soon be clear what this means, specifically (slides on 'questionnaire design issues' are coming up). Conforti, Grünberger and Troubat (2017) summarize the main approaches to data collection found in practice (neither of which, again, is intrinsically superior): check page 44 of their paper for more details.

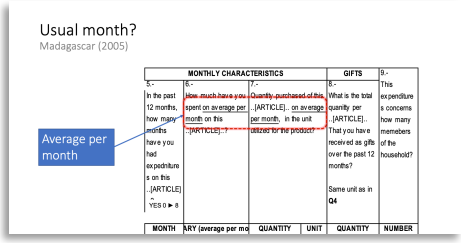


The idea of flaws in questionnaire design is introduced by some examples that students are encouraged to interact with. We suggest to ask students to comment on a few questionnaires, by answering three questions. Questions 1 and 2 (see slide above) are purely descriptive: would the data collected through each questionnaire be suitable to measure food consumption, acquisition, purchases, or unit values? These questions check whether the concepts covered in the previous slides are clear to students. Question 3 encourages students to think about possible flaws in the way the questionnaires are designed. More of these examples are used as homework for this lecture.



This page is from **Burundi** QUIBB 2006 (translated from French). With this questionnaire, it is impossible to fully capture either consumption or acquisition. If the objective is to measure *consumption*, after the filter question (question 11) data are collected on food acquired (purchased, harvested) rather than food consumed. When this occurs, quantities and expenditures on food acquired include those entering into the households' stocks – not the household pantry for immediate consumption – and are systematic overestimates of food consumed from home production. If the objective is to measure *acquisition*, then question in M11 rules out any following question for food that were acquired but not consumed in the reference period. These are flaws in questionnaire design, because they lead to incomplete enumeration of both acquisition and consumption.

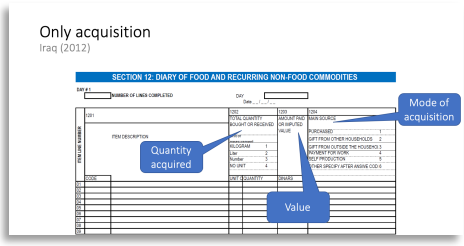
On the other hand, *unit values* can be computed based on this module: it would be sufficient to divide question 13 by question 14 (provided local units can be converted into standard units).



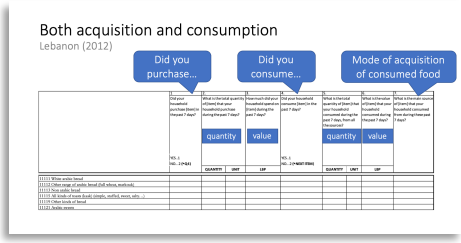
This example is from the **Madagascar** Enquete Permanent Menages (EPM) 2005. The questionnaire captures *acquisition*, rather than consumption, although in this page we cannot see any question on own-produced foods, which would therefore be missing from the total amount and value of acquisitions.

Unit values can be obtained as question 6 divided by question 7.

A problem with the design of this questionnaire, besides the apparent lack of questions on own-production, is that it is not clear if the "average per month" in questions 6 and 7 refers to the average of the past 12 months, or the average over those months in which consumption actually took place. A better approach would have been to clarify the ambiguity, or even to get rid of the "usual month" approach (more on this later).

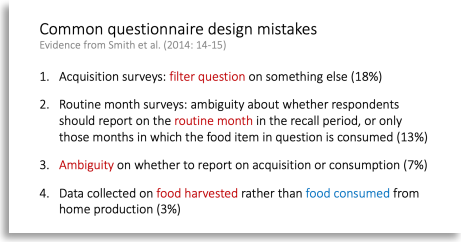


This page is from the **Iraq** Household Socio Economic Survey 2012. The questionnaire only records food *acquisition* and not consumption. However, all modes of acquisition are recorded, and there are no apparent flaws in the design of the questionnaire. *Unit values* could be computed as question 1203 divided by 1202 (provided that non-standard units are convertible into kg).

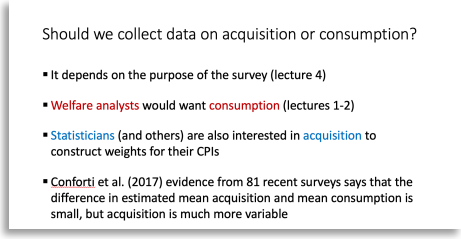


This page is from the **Lebanon** Household Budget Survey. This questionnaire partially records food *acquisition*: it focuses on *purchases* only, and does not record food acquired through other sources. On the other hand, the questionnaire captures food *consumption* in full: question 7 allows to record all sources of food consumed. There are no flagrant mistakes in questionnaire design, assuming that the goals of the survey did not include

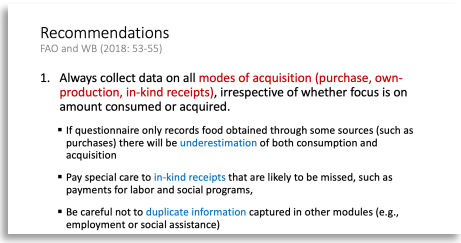
measurement of total food acquisition. Again, *unit values* may be computed both from purchases and consumption, by computing the ratio between reported value and quantity.



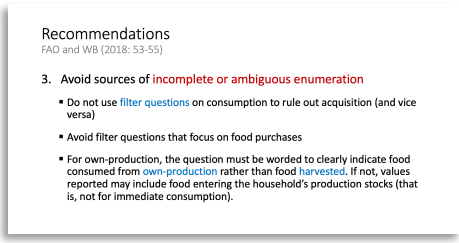
This slide summarizes the most common issues found in household survey questionnaires in developing countries, as listed by Smith et al. (2014). Some of these were exemplified by the questionnaires discussed with students (point 1 can be seen in the Burundi example, point 2 in Madagascar, point 4 in Burundi again). These should be seen as a catalogue of the most common mistakes to be avoided when designing the food module.



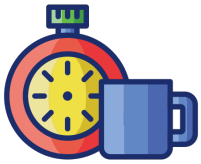
Provided that mistakes and incomplete enumeration are avoided, one question remains: **is it best to record acquisition, or consumption?** The answer depends on the goals of the survey. Based on the conceptual framework covered in Lectures 1 and 2, we could argue that welfare analysts prefer food *consumption*. However, recording food *acquisition* is important for other purposes, the most common of which is probably the computation of consumer price index (CPI) weights. The evidence presented by Conforti et al. (2017) is reassuring: the two measures are not too distant, so that one may be used in place of the other, in case of necessity. A common approach in practice, as we have seen, is to record both.



• • •



A set of recommendations closes this topic. Recommendations are adapted from section 3.3 of the FAO/World Bank Guidelines (2018): further details are found there.



Now is a good time for a break.

2. Recall vs. diary
   and length of reference period

Recall vs. diary, and the determination of the length of the reference period is our next topic for the lecture.

Definitions: recall and diary

Data on household food consumption (or acquisition) commonly collected in one of two ways:

1. Respondents are interviewed and asked to recall consumption during a specified period (past week, past month…).
2. Households are asked to keep a diary over a reference period (days, weeks…) and record consumption at the moment it takes place.

…

Example of diary

Kenya Integrated household budget survey 2015, seven-day reference period

CD01	CD02	CD03	CD04	CD05	CD06
Date	ITEM CODE	Description of item	Mode of acquisition of consumption item	Quantity consumed	Unit of consumption
	FOR OFFICIAL USE ONLY		PURCHASED ..... 01 OWN PRODUCE ..... 02 OWN STOCK ..... 03 GIFT ..... 04 RELIEF ..... 05 PAYMENT IN KIND ..... 06 GATHERING/HUNTING/FISHING ..... 07 BARTER ..... 08 OTHER (SPECIFY) 09		
				Code	
2015-07-13		maize flour	Own Produce	02	3	14
2015-07-13		cabbage	Purchased	01	1	15
2015-07-13		milk	Purchased	04	500	03

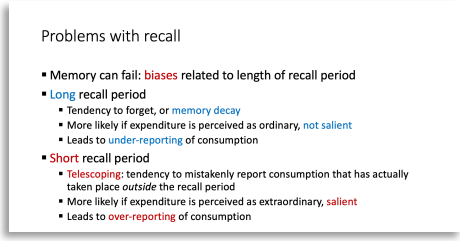
GRAMS	01
KILOGRAMS	02
MILLITRES	03
LITRES	04
NO / PIECE	15

First, the slides give definitions for diary and recall, recall period and reference period. Two examples are used to substantiate the difference between diary and recall.

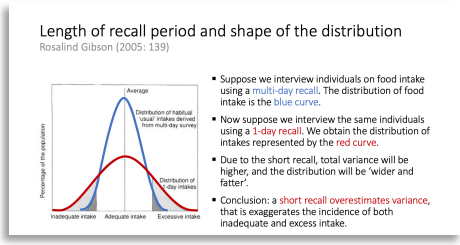
Diary or recall?

- Which approach is better, in terms of the quality of collected data?
- Both methods have pros and cons
- In particular, they both have the potential to generate measurement error, for different reasons
- Let us see how to evaluate the risks associated to either choice

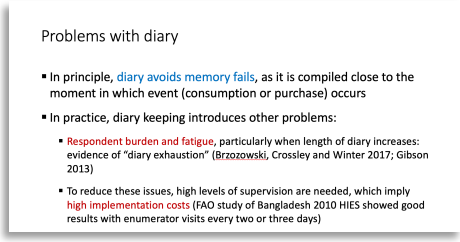
Which of the two approaches guarantees the collection of quality data? The answer is not easy: both diary and recall surveys have the potential to generate **measurement error**. Empirical evidence helps to evaluate the risks implied by each of the two methods. The next slides illustrate.



The most problematic aspect of using **recall** methods is related to memory and its failures: respondents' inability to remember correctly leads to biases in data collected through recall surveys. Depending on the length of the recall period, different types of bias may arise. *Memory decay* and *telescoping* are related to long and short recall periods, respectively. For a detailed explanation of these biases, see the discussion at pages 110 and 111 of Grosh and Glewwe (2000) (the 'Consumption' chapter by Deaton and Grosh).

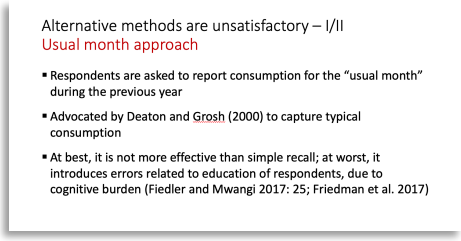


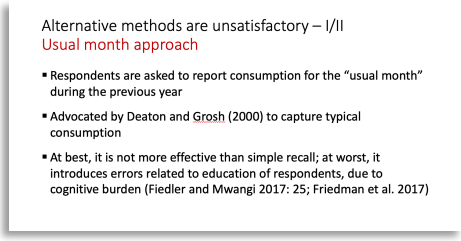
Another issue that emerges when using recall surveys, and that is related to the **length of the recall period**, is that of the variability of reported consumption. Shorter recall periods record 'episodic' consumption (consumption that takes place on a given day or week) which may be randomly higher or lower than usual. These anomalies are smoothed out by longer recall periods. As a result, data recorded with short recall periods are *more variable* than those recorded with longer ones. This slide illustrates the point by displaying the distribution of reported food intake from a short recall (red curve) and a long recall (blue curve). The two curves have the same mean (on average, both estimates are correct) but the red curve has higher variance. This is a problem if data users are interested in the 'tails' of the distribution of consumption – households with very low calorie intakes, for instance – because the likelihood of extremely high and extremely low will be overestimated by surveys using short recall periods. Crossley, Fiedler and Mwangi (2016) discuss this issue in rather technical terms on page 19 of their IFPRI Discussion Paper.



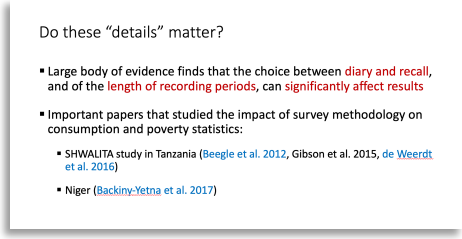
Using **diary** surveys to collect data on food consumption may seem like a solution to the memory biases that characterize recall surveys: with diaries, respondents note down their consumption as soon as it happens. In practice, however, the diary approach introduces

other problems, mainly linked to respondent burden and fatigue, and to the high implementation costs needed to minimize them.





Some authors have suggested **alternative methods**, a 'third way' besides diary and recall. However, empirical evidence has failed to support these alternatives.



Now that the trade-offs of the diary and recall approaches have been covered, one may ask: **do these details really matter?** The short answer is yes. The following slides summarize results from a few notable experimental studies that tested the effects of different questionnaire designs on estimated food expenditure and other outcomes.

...

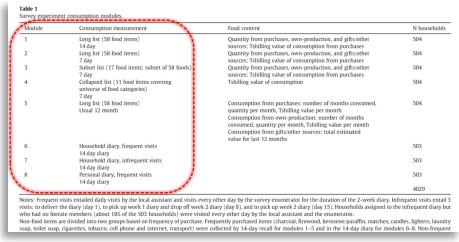
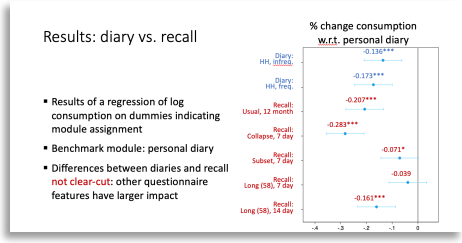


Table 1
Survey experiment consumption modules.

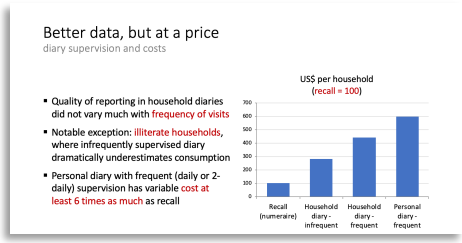
Module	Consumption measurement	Food content	N households
1	Long list (58 food items) 14 day	Quantity from purchases, own-production, and gifts/other sources; Tshilling value of consumption from purchases	504
2	Long list (58 food items) 7 day	Quantity from purchases, own-production, and gifts/other sources; Tshilling value of consumption from purchases	504
3	Subset list (17 food items; subset of 58 foods) 7 day	Quantity from purchases, own-production, and gifts/other sources; Tshilling value of consumption from purchases	504
4	Collapsed list (11 food items covering universe of food categories) 7 day	Tshilling value of consumption	504
5	Long list (58 food items) Usual 12 month	Consumption from purchases: number of months consumed, quantity per month, Tshilling value per month Consumption from own-production: number of months consumed, quantity per month, Tshilling value per month Consumption from gifts/other sources: total estimated value for last 12 months	504
6	Household diary, frequent visits 14 day diary		503
7	Household diary, infrequent visits 14 day diary		503
8	Personal diary, frequent visits 14 day diary		503
			4029

Notes: Frequent visits entailed daily visits by the local assistant and visits every other day by the survey enumerator for the duration of the 2-week diary. Infrequent visits entail 3 visits: to deliver the diary (day 1), to pick up week 1 diary and drop off week 2 diary (day 8), and to pick up week 2 diary (day 15). Households assigned to the infrequent diary but who had no literate members (about 18% of the 503 households) were visited every other day by the local assistant and the enumerator.
Non-food items are divided into two groups based on frequency of purchase. Frequently purchased items (charcoal, firewood, kerosene/paraffin, matches, candles, lighters, laundry soap, toilet soap, cigarettes, tobacco, cell phone and internet, transport) were collected by 14-day recall for modules 1–5 and in the 14-day diary for modules 6–8. Non-frequent

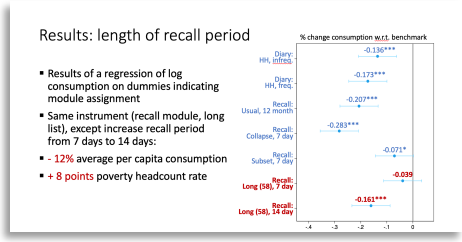
First is **Beegle et al. (2012)** with an experimental study conducted in Tanzania. The instructor is encouraged to read the paper: the slides offer a quick summary of the main findings, as they relate to the topics of Lecture 5. The study implemented eight different questionnaire designs, assigned them randomly to 4,000 households, and then compared average food consumption expenditure resulting from each of the questionnaires. One design – a personal diary with frequent visits – is considered the most accurate way of collecting data on food consumption, and is therefore used as a benchmark, against which all other questionnaire designs are stacked up.



This slide shows the **comparison of diary and recall** questionnaires, by using a type of chart that is going to appear several more times in the lecture. *The chart requires some explanation in order to make its interpretation clear, but it is an important one, and should be given as much time as needed.* It makes a crucial point: we can find answers to questions on 'best' questionnaire designs, if we look to the evidence. Allowing students to become comfortable with this literature, and its findings, is the real added value of the lecture. Each design is compared to the benchmark (personal diary, frequent visits). The chart plots the coefficients of a regression of log consumption on dummies indicating assignment to questionnaire designs. Asterisks indicate statistical significance: *** significant at 1%, ** at 5% level. No asterisks indicate no statistical significance (at traditional significance levels). Interpretation of coefficients is as follows (take the first line as an example): average food consumption expenditure obtained with a household diary with infrequent visits is 13.6% lower than the value obtained with a personal diary with frequent visits (the benchmark design). Now, if we compare the coefficients for the two different *diary* designs (in blue), to those for the five different *recall* designs (in red), we note that results are mixed. All numbers are negative, meaning that all alternative designs return a lower average food expenditure than the benchmark; but *how much* lower seems to depend on characteristics of the design *other than* the choice of diary vs. recall.

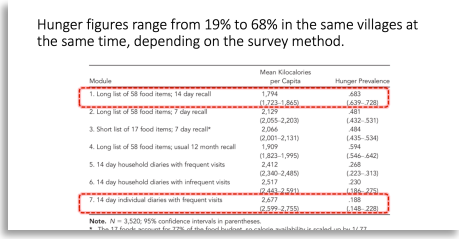


Other selected results of the study concern implementation costs for different designs: diaries give better results when closely supervised, at least for less-educated households; however, they are also much more expensive to implement than recall modules, as shown by the bar chart).

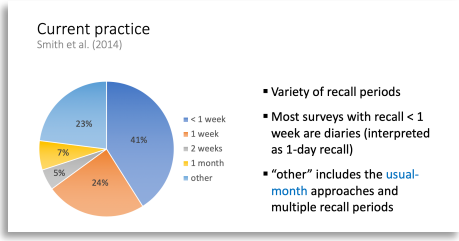


Another important finding concerns the length of the recall period: comparing the final two designs in the chart (in red), that only differ in recall length, shows that the 7-day recall comes much closer to the benchmark than the 14-day recall. In other words, if we increase the recall period from 7 to 14 days, all else being equal, the average reported consumption falls by 12% (we obtain this as the difference between the two coefficients corresponding to the 7 and 14-day recall, 0.161-0.039). Elsewhere in the paper (page 14) the author reports that this also corresponds to an increase of 8 percentage points of the poverty headcount rate.

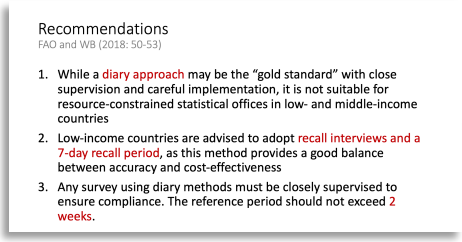
Backiny-Yetna et al. (2017) is another experimental study, conducted in **Niger**. In this case, the comparison between diary and recall gives somewhat unexpected results: a 7-day recall module returns higher estimated food expenditure than a 7-day diary. This is reflected by poverty estimates (the poverty headcount is higher according to the diary module). Respondent burden is likely to be responsible for the low results obtained with the diary, which suggests that supervision matters.

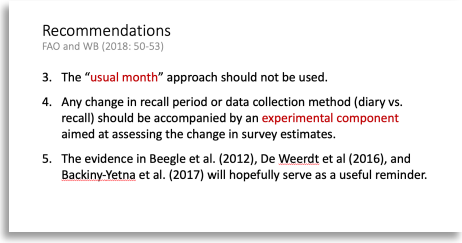


Finally, De Weerdt et al. (2016) use the same data as Beegle et al. (2012), but compare results on nutrition. In this case, recall questionnaires yield lower estimated calorie intakes than diary designs.

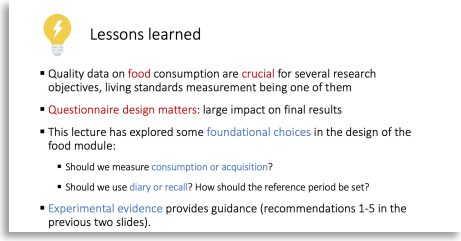


To close the discussion of empirical evidence, we once again turn to the review of the current practice by Smith et al. (2014), to see which approaches are actually implemented in practice. In this graph, diaries are interpreted as surveys using a recall period of 1 day. Note that using different recall periods for different foods ('other' in the graph) is common.





Recommendations are adapted from Section 3.1 of FAO and World Bank (2018).



The final slide summarizes the main takeaways of Lecture 5: it is important to take some time for this final recap, as it provides an opportunity to cement the students' understanding of topics covered, and take any final questions.

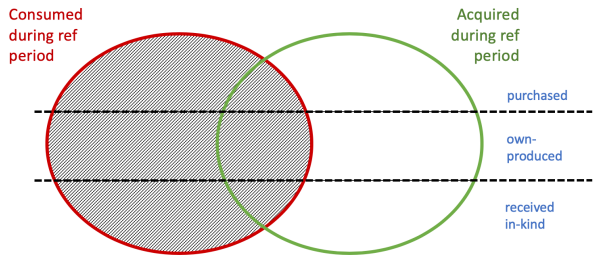
Homework

Exercise 1 – Engaging with the literature

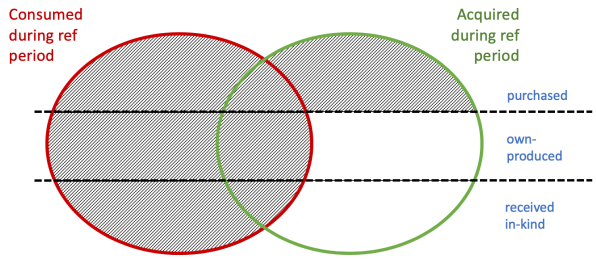
See exercise 1, Lecture 1.

Exercise 2 – Acquisition vs. consumption

Example 1 allows to estimate the total value of consumption. All three modes of acquisition are accounted for. The corresponding diagram is pictured below.



Example 2 also allows to estimate the total value of consumption. However, question C02 asks for the *total* value of purchases during the last 7 days (not just those going towards consumption): this makes it possible to estimate at least part of the value of acquisitions, precisely the portion acquired through purchases. The corresponding diagram is pictured below.



Lecture 6
Measuring food consumption: questionnaire design

Learning objectives

This lecture, together with lectures 5 and 7, has the goal of providing detailed recommendations on the collection of food consumption and expenditure data through household surveys. Lecture 6 focuses on the definition of the list of food items for which information is collected, on why and how to record meal participation, and on the timing of visits to interviewed households.

Suggested preparation

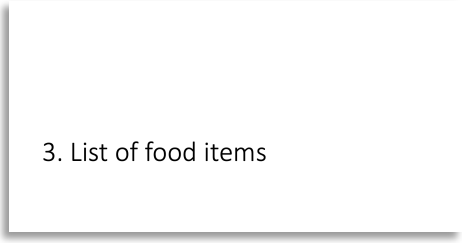
Much of the material covered in lectures 5, 6, and 7 is adapted from two references: FAO and The World Bank (2018), and Smith, Dupriez and Troubat (2014). The Instructor is assumed to be familiar with both. This lecture also summarizes the evidence from several studies (Jolliffe 2001, Beegle et al. 2012, Pradhan 2009, Gibson and Rozelle 2002, Troubat and Grunberger 2017), which the Instructor is encouraged to review.

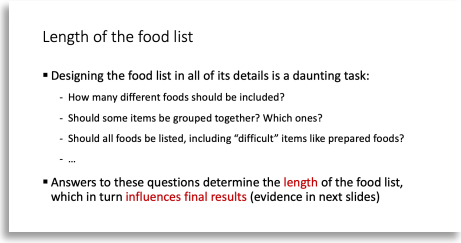
Time allocation

Introductory slides	**5 min**
List of food items	**40 min**
Meal participation	
Concepts: what partaking is and why it matters	10 min
Break	**15 min**
Evidence and examples, recommendations	20 min
Timing of visits	**25 min**
Lessons learned	**5 min**

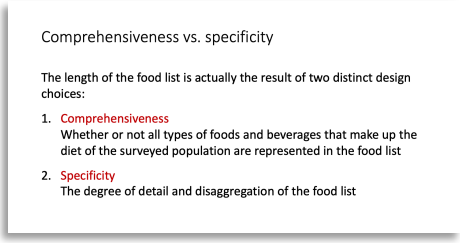
Annotated lecture

After a brief introduction, aimed at reminding students the general framework in which the current lecture is inscribed and the outline for the day, we dive into the first topic: determining the list of items for the food module.

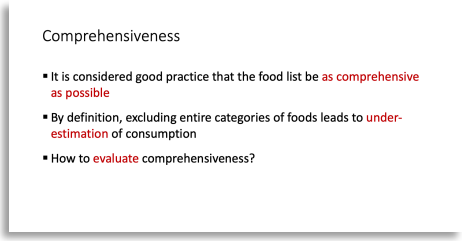


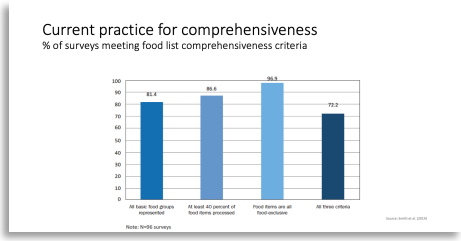


The goal of these first slides is to define what is meant by '**food list**', and to convey the fact that defining it from scratch is a difficult, non-trivial, and important task. The decisions that go into the definition of the food list are sometimes just summarized as "determining the length of the list".

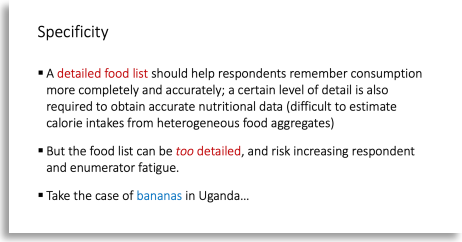


It helps to think about the 'length' of the food list as resulting from two separate choices: the level of **comprehensiveness** of the list, and its **specificity**.





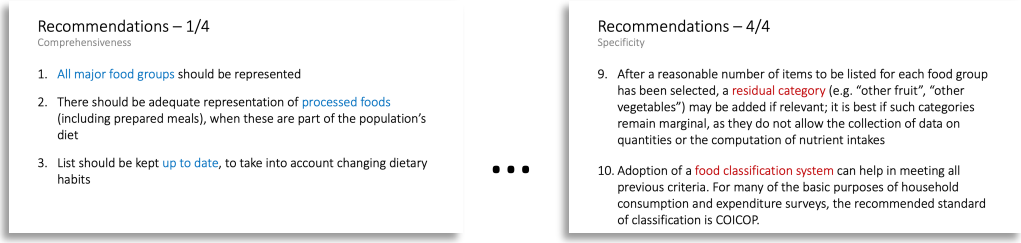
Comprehensiveness is, in a way, uncontroversial: the more comprehensive the list, the better. For this reason, the point is discussed quickly. Smith et al. (2014) set useful **criteria** that help to evaluate the level of comprehensiveness of the food list, and their assessment shows that most questionnaires perform well in this dimension.



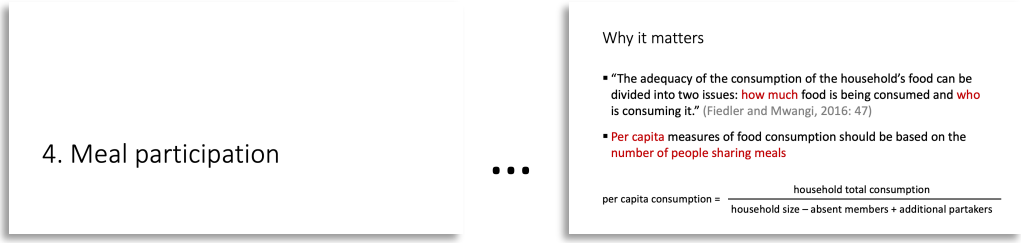
Unlike comprehensiveness, **specificity** – the level of detail, or disaggregation, of the food list – has pros and cons. According to the literature, both too specific and too aggregated food lists can lead to measurement error. The case of Uganda, and its many varieties of bananas, is cited as an example: would it be a good idea to ask about consumption of all 95

varieties, specifically, in the food module? The Instructor can engage students for a couple of minutes on this, just to liven up the discussion.

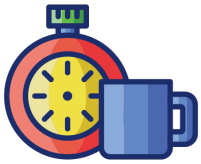
At this point, we once again bring in the empirical evidence. The studies cited make the point that specificity matters, in that it greatly affects final estimates. The example from El Salvador (Jolliffe 2001) shows that there was a 27% increase in median reported consumption when the specificity of the food list was increased. The SHWALITA study of Tanzania is cited again, as specificity was among the many design features tested. The chart with the blue and red coefficients has been introduced and explained in lecture 5, and can be interpreted in the same way (the Instructor is encouraged to revisit it). In this case, a more specific food list caused a 23% increase in average reported consumption (this is obtained as 0.283-0.039).



The final **recommendations** give general guidelines to construct the food list. Section 3.6 of the FAO/World Bank Guidelines can be checked for further details.



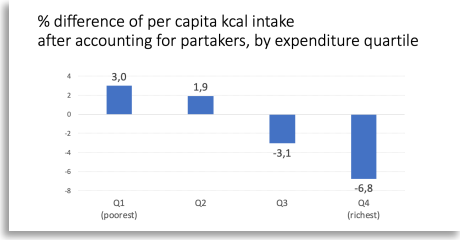
Next, the lecture tackles the issue of adjusting for **meal partakers**. The goal of the first slides is to give a definition of partakers, and to motivate our focus on the issue, by conveying the message that knowing the exact *denominator* of per capita consumption (*how many* people are sharing the total?) is fundamental to obtaining an accurate estimate of individual welfare (be it consumption, expenditure, or nutrition).



Now is a good time for a break.

Evidence on the impact of partakers

- Accounting for partakers reduces inequality of consumption
- Bouis, Haddad, and Kennedy (1992) and Bouis (1994) show that the difference between mean calorie intakes of the poorest and richest quartiles is much lower when partakers are accounted for (Kenya and the Philippines)
- Gibson and Rozelle (2002) finds similar evidence (Papua New Guinea)

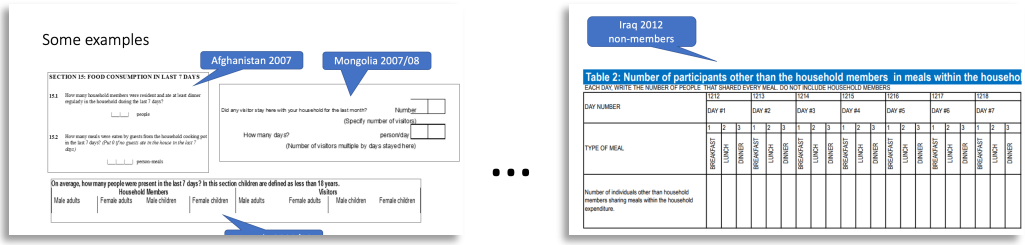


Empirical evidence is brought in to show that **adjusting for partakers matters** in practice – that is, dividing total household consumption, or total household calorie intake, by the number of meal partakers instad of simply the number of household members, significantly changes the distribution of per capita intakes. With the adjustment, estimated per capita consumption decreases among the rich, and increases among the poor. The graph taken from Gibson and Rozelle (2002) shows the percentage difference between unadjusted (i.e. per household member) and adjusted (i.e. per partaker) per capita calorie intake, for each quartile of per capita expenditure: the difference is positive for the poorest households (the adjustment has increased their per capita intake) and negative for the richest (the adjustment has decreased their intake). This is a key reason why the adjustment is needed, given that estimating the per capita *distribution* of consumption (not just the average) is one of the main uses for household consumption and expenditure survey.

Current practice

- Assessment of 81 recent surveys by Fiedler and Mwangi (2016)
- Most commonly, surveys do not collect information on meal partakers
- When they do, approaches are heterogeneous
- Lack of research to tell us what works

In contrast to the importance of collecting information on partakers, Fiedler and Mwangi (2018) show that most countries fail to collect data on this aspect, and those that do have very heterogeneous approaches.



These slides show some miscellaneous **examples**, in order of increasing complexity of the partakers module. For the time being, there is no need to comment on which approach is expected to work better: the examples are a way to give more substance to the idea of collecting data on partakers, and also to show how questionnaires can vary widely in practice.

Heterogeneity of approaches
Smith et al. (2014: 32)

Table 7: Collection of data on food given to non-household members (Percent of surveys)	
Data are collected on the presence and/or household meal consumption of non-household members during the recall period	15
Data collected on the number of visitors in the household	11
Data collected on visitors' length of stay	5
Data collected on the number of meals consumed by visitors/guests	10
Data collected by type of meal (breakfast, lunch, dinner)	7
Data collected on the age of visitors/guests	7
Data collected on the sex of visitors/guests	6

Note: N=100 surveys.

A typology of approaches
FAO and WB (2018: 55)

A. Food consumer: count the number of people usually partaking to household's meals, and divide total household consumption by this number.
Limitation: Counting heads of partakers is not precise. The method has difficulties to account for situations in which people do participate only at some meals per day, e.g. employees.

B. Meal partakers: requires an exact accounting of the number of meals taken by household members and non-household members over the same reference period as that for which food data is collected.
Limitation: difficult to implement.

These slides give more structure to the **heterogeneity of approaches** exemplified by the questionnaires shown. Smith et al. (2014) summarize the different pieces of information collected by questionnaires in developing countries; the FAO/World Bank guidelines reduce these different approaches to two main ways of adjusting for partakers – 'food consumer' and 'meal partakers'.

Recommendation # 1
FAO and WB (2018: 55-56)

Information should be collected on the number of meals and the number of individuals (household and non-household members) who participated in each meal.

The addition of an individual household member-based meal module should be considered for all surveys that do not yet have it.

...

Recommendation # 3
FAO and WB (2018: 55-56)

If the entire individual household member-based meal module cannot be added, survey design teams should consider adding questions to a proxy respondent.

The aim is to capture the number of meals were taken at home by household members and others, during the reference period.

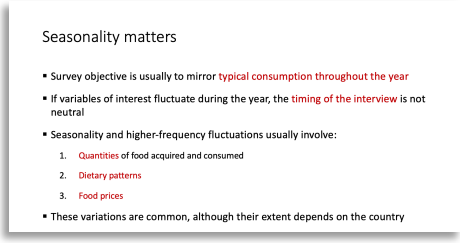
The discussion on partakers ends with some recommendation from the FAO/World Bank manual. More details can be found there, if the Instructor wishes to elaborate.

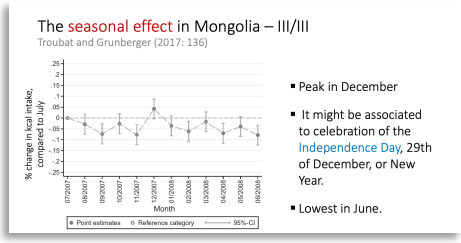
5. Timing of visits

Temporal fluctuations

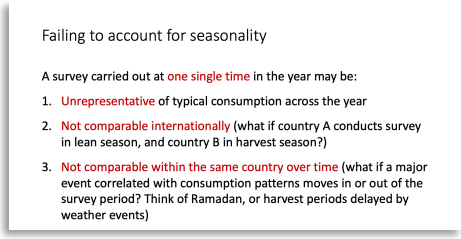
- Fluctuations in consumption and expenditure within the year are common
- Variation between months, also called seasonality:
  - Agricultural season(s), cyclical food production cycles, festivals and holidays
- But there is also cyclical variation within months and weeks:
  - Payday for wage workers, market day, transfer-day' for households receiving cash transfers, Friday, Saturday, Sunday (depending on culture) consumption may differ from 'usual'

Next is the issue of determining the **timing of visits** to interviewed households. The choice matters because of fluctuations of consumption over time (seasonality, but also within-month or within-week variation).

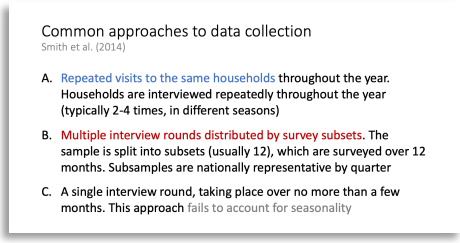


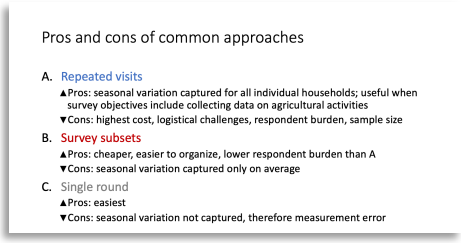


Some evidence is presented on the **magnitude** of seasonal (and higher-frequency) variation in consumption. In Afghanistan, the poverty headcount computed by quarter was found to oscillate widely, due to seasonal variation in both prices and consumption. In Mongolia, Troubat and Grunberger (2017) found that per capita calorie intake varied significantly by day of the week, by day of the month, and by month of the year: the graphs show the percentage change in per capita intakes with respect to a reference day or month (vertical bars indicate confidence intervals for each point estimate). The main point to be made through both examples is that, because consumption is found to fluctuate significantly within the survey year in most countries, the timing of interviews matters for final survey estimates.

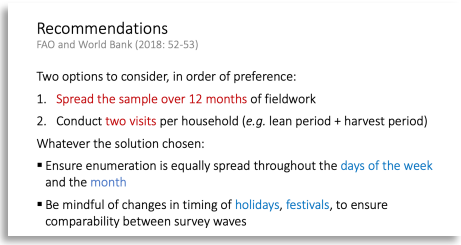


This point is further stressed by this summary of the negative consequences of failing to adjust for seasonality.

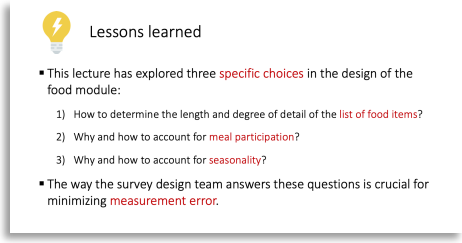




The next slides discuss the **best ways to account for seasonality**. The two most common approaches are introduced, and their frequency in practice is evaluated – together with the frequency of surveys that *do not* account for seasonality at all.



Recommendations prioritize the 'survey subsets' approach, although 'repeated visits' are also acceptable.



The final slide is used to remind students of the three topics discussed during the lecture. We do not repeat all recommendations – students can refer back to each topic to refresh their memory – but we stress that the ultimate goal of any 'best practice' is to minimize measurement error.

Homework

Exercise 1 – Engaging with the literature

See exercise 1, Lecture 1.

Exercise 2 – Food module, international comparisons

This exercise gives students an opportunity to see a practical application of the concepts discussed in the lecture, and become familiar with questionnaires used in different countries. The goal of the exercise is to assess the students' grasp of the guidelines on the length of the food list, and see whether they can critique questionnaires based on the recommendations learned during the lecture.

Exercise 3 – Meal participation

Example 1 (Somaliland, 2012) collects information on individuals *usually* taking part in the household's meals. Data collected through this questionnaire would only allow to compute per capita intakes using the 'food consumer' approach (estimating the number of people usually partaking during the reference period).

Example 2 (Namibia, 2016) collects daily information on who takes part in each meal (breakfast, lunch, dinner). Data collected through this questionnaire would allow to compute per capita intakes using the 'meal partakers' approach (accounting precisely for the number of meals taken by household and non-household members on each day).

Lecture 7
Food away from home (FAFH) and the use of non-standard units for measuring food consumption

Learning objectives

This lecture, together with lectures 5 and 6, has the goal of providing detailed recommendations on the collection of food consumption and expenditure data through household surveys. Lecture 7 focuses on measuring the consumption of food prepared away from home, and on using non-standard measurement units.

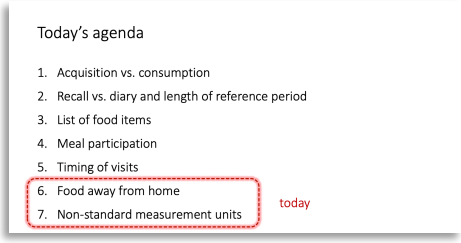
Suggested preparation

Much of the material covered in lectures 5, 6, and 7 is adapted from two references: FAO and The World Bank (2018), and Smith, Dupriez and Troubat (2014). The Instructor is assumed to be familiar with both. In addition, Lecture 7 discusses experimental evidence from Farfàn, McGee, Perng, and Vakis (2019) in some detail, and summarizes guidelines from Oseni, Durazo, and McGee (2017). The Instructor is encouraged to review these additional references.

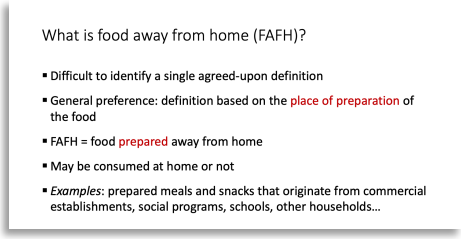
Time allocation

Food away from home	
Concepts	10 min
Empirical evidence	30 min
Recommendations	10 min
Break	**15 min**
Non-standard measurement units	
Concepts and checklist for implementation	40 min
Recommendations	10 min
Lessons learned	**5 min**

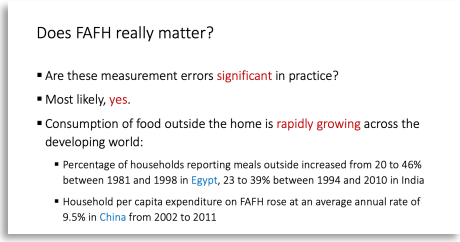
Annotated lecture

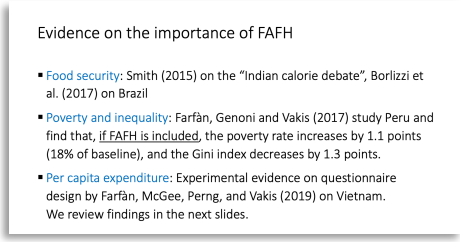


The last lecture on measuring food consumption tackles two very specific topics: **food away from home**, and **non-standard measurement units**.

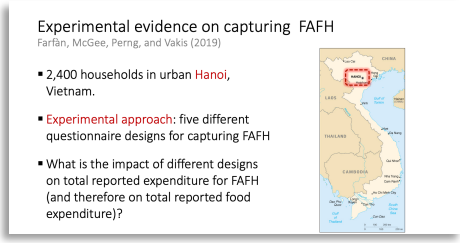


The first few slides define **what exactly is meant by food away from home**: the focus is on food that is *prepared* outside of the home, and therefore does not enter the budget of the household, unless it is explicitly accounted for.

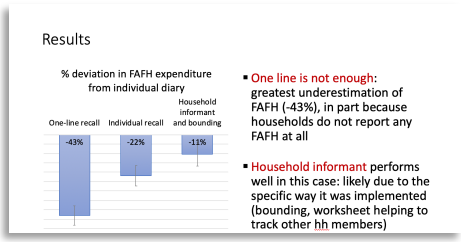




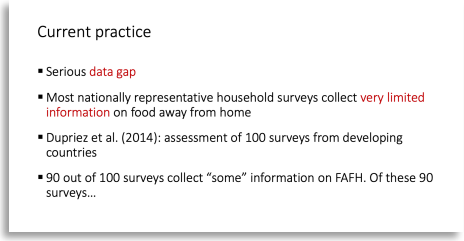
Some evidence is cited to motivate the emphasis on correctly measuring food away from home. Different studies document that the share of FAFH in total food consumption is increasing in many countries, therefore measuring it correctly is becoming more and more important; several papers estimate the impact of including the expenditure for FAFH in measures of total consumption on food security, poverty and inequality measures. We focus our attention on the findings from a recent experimental study conducted in Hanoi, Vietnam.

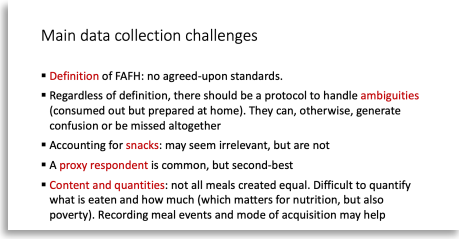


…

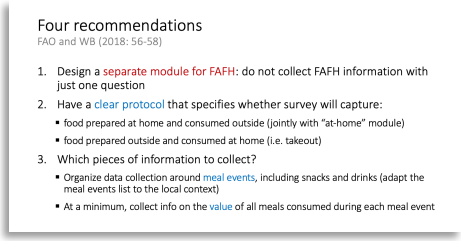


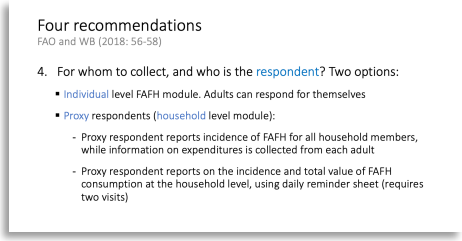
Evidence from an experimental study by Farfan et al. (2019) is used to showcase different approaches to the measurement of FAFH, and to establish which method performs best. The Instructor is encouraged to review the paper in advance, in order to be able to deliver these slides effectively. Similarly to the studies conducted in Tanzania, and mentioned in previous lectures, this experiment also tests different questionnaire designs, and compares them to a 'benchmark' (an individual diary with close supervision, or the "gold standard"). The result that should be most stressed is that **a single question is not enough to adequately capture FAFH**. The issue, then, is how to choose the best among the remaining approaches.



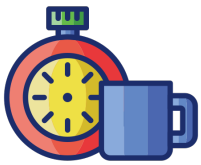


A review of the **current practice** shows that in many cases, the measurement of FAFH is lacking. Almost one in four surveys capture FAFH using just one question, which, as documented in the literature, is likely to severely underestimate actual consumption; the majority of surveys do not account for snacks; less than one in five surveys collect information on FAFH at the individual level.





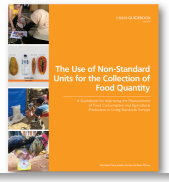
Recommendations adapted from Section 3.5 of the FAO/World Bank guidelines show how to meet the challenges related to the collection of data on FAFH.



Now is a good time for a break.

7. Non-standard measurement units

Main reference for this topic



Oseni, Durazo, and McGee (2017)
Non-standard units LSMS Guidebook

The second part of the lecture, on the use of non-standard measurement units, is modelled closely after the **Guidebook** by Oseni et al. (2017). The Instructor is encouraged to consult sections 3 and 4 of the Guidebook for a detailed discussion of this topic.

What are non-standard units (NSU)?

- Standard units (SUs): standardized across all locations, items. Typically, metric units: grams, kilograms, liters…
- Nonstandard units (NSUs): weights can vary by item, location, customs, type of food preparation…
  Pieces, bushels, heaps, boxes…

What are non-standard units (NSU)?
Malawi, Third Integrated Household Survey (IHS3) 2010/11

IHS3 Item Code	Item Name and Type	Unit Code	Unit Description	Regional Mean in KG: North Region	Central Region	South Region
411	Okra / Therere	9	Piece	0.08	0.08	0.08
		9A	Piece (small)	0.04	0.04	0.04
		9B	Piece (medium)	0.08	0.08	0.08
		10	Heap	0.11	0.21	0.20
		10A	Heap (small)	0.09	0.14	0.13
		10B	Heap (medium)	0.11	0.21	0.20
		10C	Heap (large)	0.26	0.25	0.32
		16	Cup	0.38	0.38	0.38

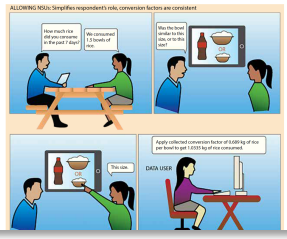
First, we **define** what is meant by non-standard measurement units, by contrasting them with standard (metric) units. An example from Malawi allows students to form an idea of the kind of NSUs that are used in practice.

Advantages of NSUs

1. May be more familiar to respondents
   - SUs may not be used in markets
   - Respondents may not encounter SUs in their daily activities
   - Items may traditionally be consumed in NSUs
   - Evidence: When given the choice, respondents prefer reporting in NSUs. In Malawi IHPS 2013, 73% of consumed item entries are in NSUs
2. Simplified recall: reduces on-the-spot calculations, cognitive burden.

…

In practice
non-standard units



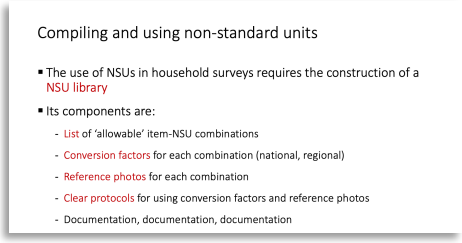
The use of NSUs in a household consumption and expenditure survey has some clear **advantages**: NSUs may be more familiar to respondents, and allowing them to report their consumption in terms of these more everyday units reduces their cognitive burden (they are not forced to make the conversion into standard units in their heads) – a vignette illustrates this point.

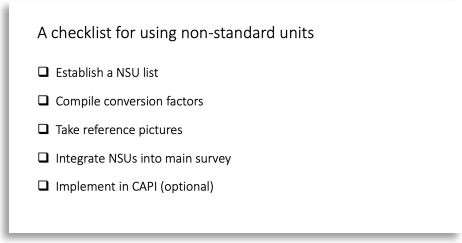
Disadvantages of NSUs

1. Complex to collect:
   may vary by item, location…
2. Not always clearly defined:
   - A 'piece' of bread?
   - A container can be filled level or heaped
   - A common unit in Ethiopia: chinet (donkey load)
3. Need to be converted into SUs for analysis

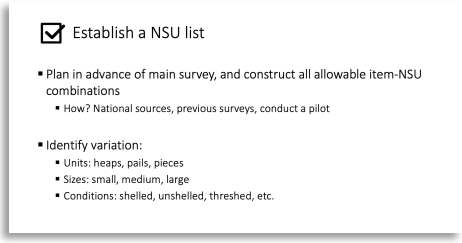
Dengu (basket) in Malawi

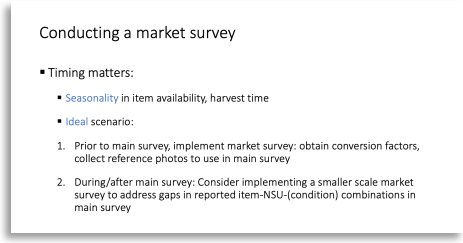
However, the use of NSUs has some **disadvantages**, too, mainly linked to the amount of preliminary work that the Statistical Office is required to do in order to make them work properly. The next slides identify a strategy to help meet these challenges.





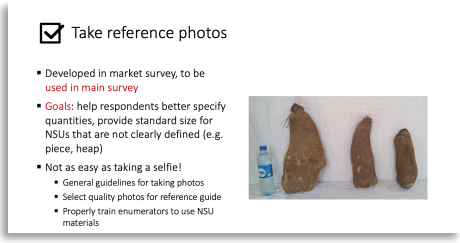
Making NSUs work requires to build a **NSU library**. To do that, one must go through several steps, summarized here in the form of a **checklist** – this is a condensed version of the guidelines offered by Oseni et al. (2017).

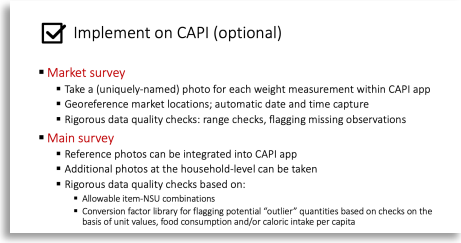




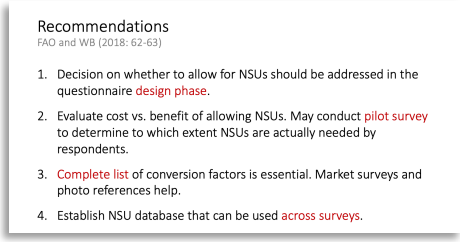
First, a **list of NSUs** must be established, in advance of the main survey. For each of the NSU that will be used in the main survey, there must be a **conversion factor**, that allows the analyst to recover metric quantities ex-post. Without conversion factors, data collected using NSUs will be useless: data users will not be able to compute calorie intakes, unit values, and other variables that are crucial for poverty measurement. In practice, the instrument used to collect conversion factors is a **market survey**, to be conducted before the main household expenditure survey. This is a survey collecting data about the nature and weight in grams of NSUs (as well as, in some cases, market prices) from a number of markets where food items are traded.

Both the **coverage and timing** of the market survey need to be planned carefully: in particular, survey specialists must be mindful of the seasonal availability of food items, to make sure all relevant foods are covered. There are several advantages to conducting the market survey in advance of the main survey: reference photos can be taken and then used during fieldwork for the main survey; NSUs that may be missing from any already-existing list can be identified on time and added. However, there are some drawbacks to this choice, too: what if respondents are inclined to report consumption using units that were not covered by the market survey? That would determine conversion factor gaps, which is to be avoided. For this reason, "(…) the ideal plan is to conduct two market surveys—one before and one after the main survey. Both market surveys need not be equally rigorous; one will likely be more comprehensive than the other. For example, the ex-ante survey could be limited, aiming to collect reference photos and weights (for conversion-factor calculations) for the most common NSUs, while the ex-post survey could comprehensively collect weights for all additional NSUs reported during the main survey." (Oseni et al. 2017: 11).

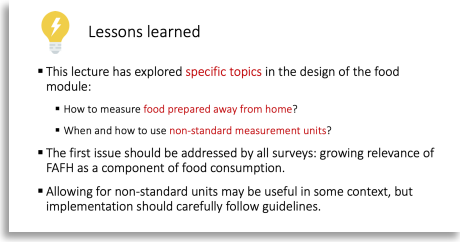




The next slides go into details on the other aspects of compiling a NSU list – from taking reference photos to implementing the use of NSUs on CAPI – the Instructor is encouraged to consult the corresponding sections of Oseni et al. (2017) for any clarifications.



To wrap up the discussion on NSUs, we share a summary of recommendations, taken from the FAO and World Bank guidelines.



The last slide recaps the topics of the lecture for students. In just two bullets, we summarize the general takeaways for both FAFH and NSUs.

Homework

Exercise 1 – Engaging with the literature

See exercise 1, Lecture 1.

Exercise 2 – FAFH modules

An in-depth analysis of the two questionnaires may raise some or all of the following points.

Both questionnaires capture the data on food consumed away from home (FAFH) with a recall period of 7 days. Tanzania has a distinct module/section dedicated to FAFH. By

contrast the data is incorporated in the overall food and beverage consumption in Uganda. The major differences between the two questionnaires are twofold:

1) FAFH in Uganda is collected for the entire food list, whereas Tanzania concentrates on 7 (somewhat ambiguous) composite items (full meals, barbecued meat chips and snacks, local brews, tea coffee samosa cake and snacks, alcoholic and non-alcoholic beverages).
2) Tanzania accounts for FAFH received as gift and differentiate them from other food received for free (collected in another module). In Uganda FAFH is collected for food purchases only (as the phrasing of the question seems to indicate).

The minimum reliability criteria from Smith et (2014: 23) are met for Tanzania only. In both questionnaires the sources of FAFH received in kind are missing.

Exercise 3 – Non-standard measurement units

Picture 1 lacks a familiar reference, like a pen or a bottle, that would allow to gauge the size of the fruits. Picture 2 contains a reference, but it is not very visible, and perspective is skewed: the largest heap is closest to the camera, thus appearing even larger than its actual size. A better way to picture the units would be to frame them straight on against the backdrop. A similar problem occurs with picture 3: this time the bottle is clearly visible, but the baskets are pictured from above, and the largest appears closer to the camera, again skewing perspective and making it difficult to estimate the size of the containers.

Lecture 8
Measuring consumption of non-food non-durable goods

Learning objectives

The goal of this lecture is to provide detailed recommendations on the collection of data on expenditures on non-food, non-durable goods through household surveys. First, the lecture gives a broad description of the analytical procedures usually adopted when analysts construct a proxy of non-food consumption; this knowledge is then used to inform data collection recommendations.

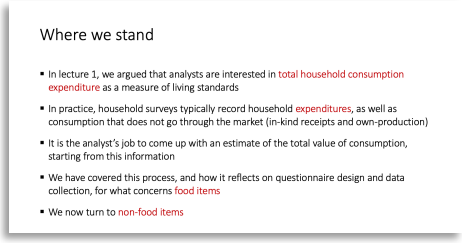
Suggested preparation

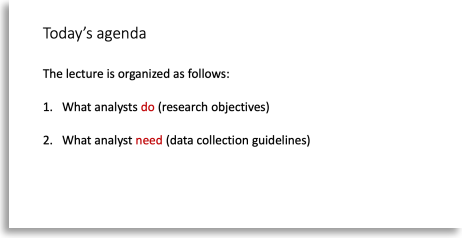
The Instructor is assumed to be familiar with the material contained in section 3 ("Constructing the Household Consumption Aggregate") of Deaton and Zaidi (2002). Guidelines on data collection of non-food non-durable expenditures are from Deaton and Grosh (2000) (general guidelines), Oseni et al. (2018) (education), Gertler, Rose, and Glewwe (2000) (health): this material should be read in advance. The lecture also cites evidence from Heijink et al. (2011) and Lu et al. (2009).

Time allocation

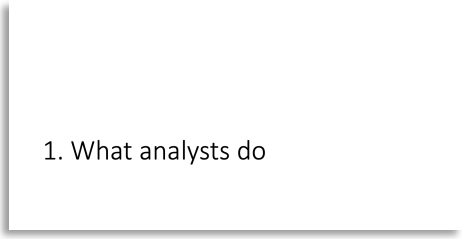
What analysts do, 'in or out?'	**45 min**
Break	**15 min**
What analysts need	
General guidelines	15 min
Education	20 min
Health	20 min
Lessons learned	**5 min**

Annotated lecture

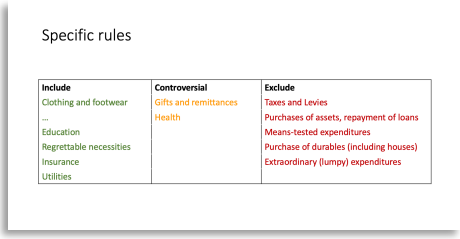




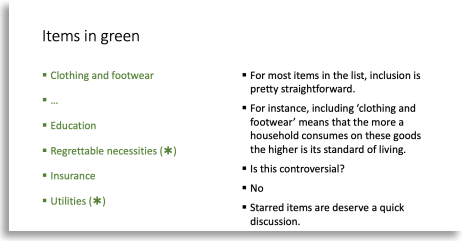
The first slide of this set is meant to help students find their bearings again, after the three lectures on measuring food consumption: it is now time to move on to **non-food consumption**. The outline clarifies the 'grassroots' approach of the lecture: the discussion starts from research objectives (what do welfare analysts do when constructing the non-food component of the consumption aggregate?), and then proceeds to data collection guidelines (how can the questionnaire meet the research objectives?).



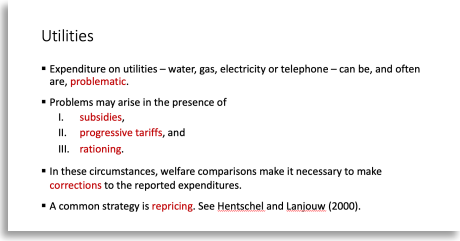
• • •



The takeaway from this first part of the lecture is that **not all expenditures** should be indiscriminately considered to be part of the living standard indicator. The *consumption* aggregate is constructed by analysts based on *expenditure* data, which requires some fine-tuning: some expenditures should not be included in the aggregate, because they do not represent *typical*, *welfare-enhancing consumption*.

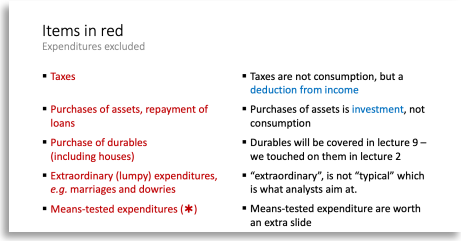


• • •

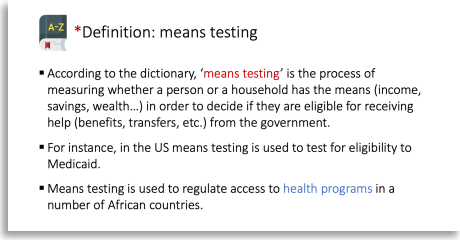


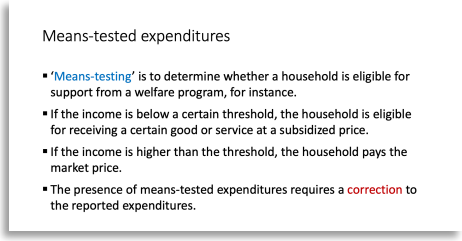
Next, the lecture discusses **inclusion and exclusion** 'rules' in further detail. Items to include are generally uncontroversial. Two categories deserve a comment. First, regrettable necessities: these are goods and services that yield no welfare in their own right, but that have to be purchased, for example, in order to earn income, such as work clothes or transport to work. Although one might argue not to include these expenditures in the consumption aggregate (they are not welfare-enhancing), the practical difficulties of telling them apart from ordinary consumption are insurmountable: "Transport to work is a regrettable necessity for someone who has little choice of where to work or where to live, but is consumption for someone who chooses to live in a pleasant suburb. Out-of-pocket

medical expenses are a necessity for some, but a choice for others, as in curative versus cosmetic medicine. It is hard to see how guidelines could be constructed that would allow one and not the other." (Deaton and Zaidi, p. 20). Ultimately, the recommendation is to include the expenditures that may be considered regrettable, to avoid any ambiguity. Another caveat concerns utilities, which can be problematic when, for different reasons, they are not purchased at market price: in those cases, the price paid by households is not a correct representation of the true market value of the service, and a market price must be estimated by the analyst (Deaton and Zaidi touch on this on p. 34).

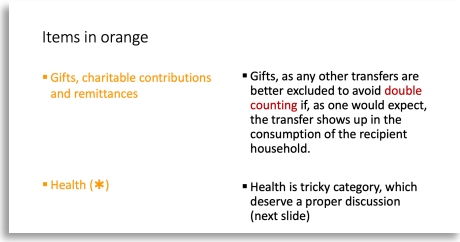


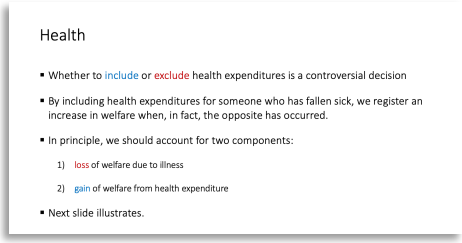
Expenditures to be excluded are only briefly discussed: two upcoming lectures will focus specifically on the two most important categories, durable goods and housing.



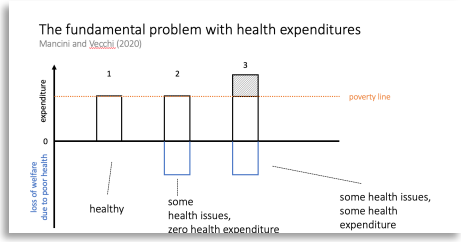


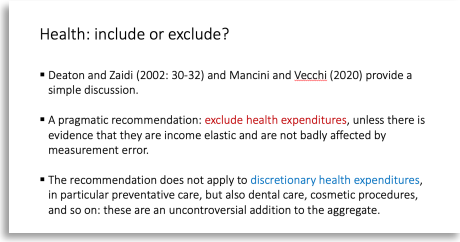
A short discussion is dedicated to means-tested expenditures. First, the Instructor has the option to show the starred slide with the definition of means testing, in case the audience requires it. In terms of construction of the consumption aggregate, the problems related to means-tested expenditures are similar to those arising in the case of subsidized items: these expenditures must be excluded, unless their price can be corrected and equalized across households.





Finally, the slides touch on mention the controversial topic of **health expenditures**. Why controversial? The fundamental problem is that we can observe and measure the increment in welfare due to receiving (consuming) healthcare, but not the loss in welfare determined by the need for care, arising from a deterioration of one's health status.





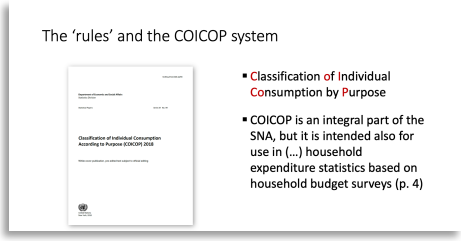
A **graph** taken from Mancini and Vecchi (2020) illustrates the point. Observed welfare, measured by consumption expenditure, is represented by the black bars above the horizontal axis. The taller the bar, the higher the expenditure, the better off the household. Below the horizontal axis is what we cannot observe and measure: the loss of welfare due to a compromised health status (in blue). Pictured are three households: household 1 is healthy (there is no bar below the horizontal axis, indicating a loss of zero), and its consumption aggregate puts it right at the poverty line (the bar above the horizontal axis is the same height as the poverty line). Household 2 experiences some health problems (this is represented by the negative blue bar below the horizontal axis), and is also at the poverty line in terms of total expenditure: for the sake of our example, let us assume that household 2 cannot afford any healthcare, and that its expenditure is exactly equal to that of household 1 (bars above the axis for households 1 and 2 are equally sized). Household 3 experiences the same health issues as household 2, but is able to spend some of its resources in healthcare (the shaded portion of the bar), which puts its total expenditure above the poverty line. Comparisons between these households clarify the conundrum related to health expenditures.

Compare households 2 and 3. Their health status is equally compromised, but household 3 is getting healthcare, unlike household 2. If health expenditures were included in the NCA, we would see that 3 is better off than 2, thanks to its ability to consume more of the goods and services it desires. If we were to exclude health expenditures from the NCA, we would lose this distinction: the comprehensiveness of our measure of welfare would deteriorate. Overall, in this situation, inclusion appears to be a better strategy than exclusion.

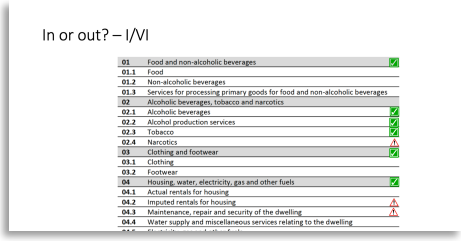
Now turn to households 1 and 3. With health expenditure included in the NCA, 3 would appear to be better off than 1, despite the fact that it only consumes more goods and services because of a health crisis. As Deaton and Zaidi (2002: 32) put it, “by including health expenditures for someone who has fallen sick, we register an increase in welfare when, in fact, the opposite has occurred”. If we excluded health expenditures, 1 and 3 would appear equally as well off. This is also problematic, because the household whose members are sick is likely to be worse off than the healthy one – but it is also unavoidable, because we cannot measure the loss of welfare due to poor health.

The bottom line is that each choice has its drawbacks: if we exclude health expenditures from the NCA, we miss the welfare-enhancing value of healthcare: keeping health status fixed, we would ideally like to capture the increased welfare of households that enjoy better care. If we include it, we mistakenly attribute higher living standards to households that are actually struggling: health expenditures have the peculiar characteristic of being associated with a reduction in welfare rather than an increase, as is the case with other expenditures.

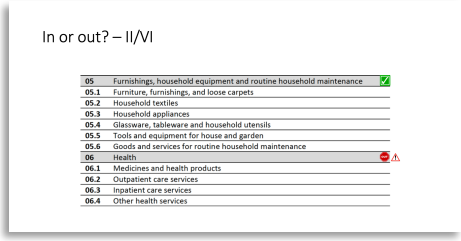
Deaton and Zaidi (p. 30-32) argue that the first mistake – a loss of coverage of the NCA – is less serious than the second one – a mis-ranking of households.



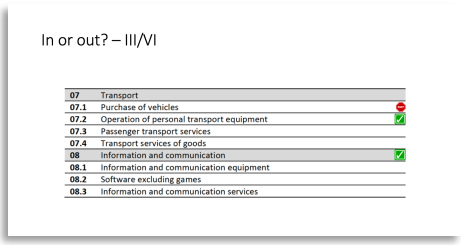
Students may already be aware of the **COICOP classification of expenditures**; if not, the availability of this system is an important takeaway of this lecture. The United Nations COICOP classification provides a basis that each country can extend at will, and it is often used in household consumption and expenditure survey questionnaires. The COICOP classification is used to put the 'rules' discussed in the previous slides to the test. Students should be invited to attempt to explain why certain items should or should not be included in the measure of living standards.

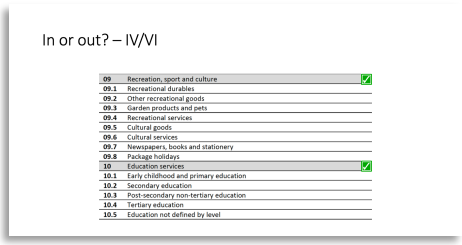


Food, beverages, clothing, housing and utilities should all be included. The danger sign on narcotics indicates the potential criticism of narcotics not being welfare-enhancing (but Deaton and Zaidi recommend their inclusion, nonetheless).



Expenditures in these categories are to be included, but students should be reminded of the controversial nature of health expenditures.





Vehicles are classified as durable goods. The remaining items are to be included.

In or out? – V/VI

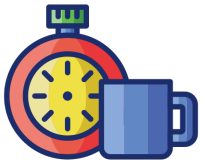
11	Restaurants and accommodation services
11.1	Food and beverage serving services
11.2	Accommodation services
12	Insurance and financial services
12.1	Insurance
12.2	Financial services
13	Personal care, social protection and miscellaneous goods and services
13.1	Personal care
13.2	Other personal effects
13.3	Social protection
13.9	Other services

If the label 'financial services' indicates savings or investment, they should be excluded. Note that insurance services are, in principle, consumption (consumers purchase a service, which is the insurance company's promise to protect them from a potential bad outcome), but in some countries, insurance is used as a form of investment. In that case, it should also be excluded.

In or out? – VI/VI

14	Individual consumption expenditure of non-profit institutions serving households (NPISHS)
14.1	Housing
14.2	Health
14.3	Recreation and culture
14.4	Education
14.5	Social protection
14.6	Other services
15	Individual consumption expenditure of general government
15.1	Housing
15.2	Health
15.3	Recreation and culture
15.4	Education
15.5	Social protection

Finally, these categories are used for national accounting, and are typically not needed when designing household surveys.



Now is a good time for a break.

2. What analysts need

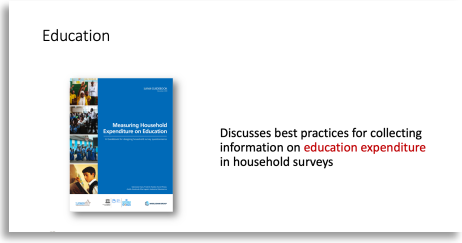
...

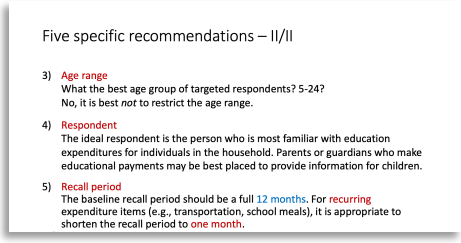
Organization of topics: General advice – II/II

Deaton and Grosh (2000: 122)

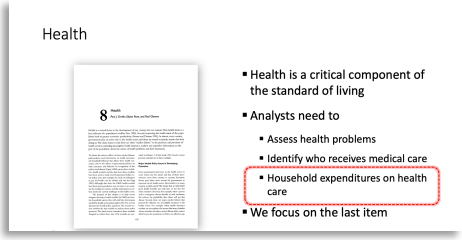
1. Some expenditures are best collected in a dedicated module, together with related non-expenditure information
2. Housing, durables, health, education usually have their own module; employment module is best place to gather information on household consumption of non-food items provided in kind
3. Create a cross-check to avoid double-counting in these instances; make sure questions are worded carefully so analysts are able to exclude duplicate measures of the same item, if needed.

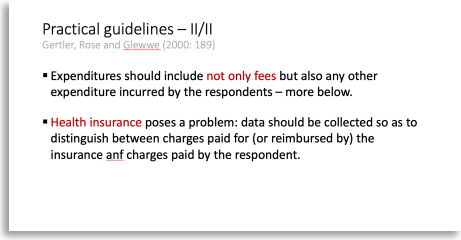
Next, the lecture moves on to consider **data collection** recommendations, on the basis of the analytical priorities discussed so far. It should be acknowledged that guidelines on the collection of non-food, non-durable expenditure data are rather broad, and for many choices that questionnaire design teams will face, there is no single universal answer. This portion of the lecture is based on Deaton and Grosh (2000), which describes the general approach adopted by LSMS surveys.



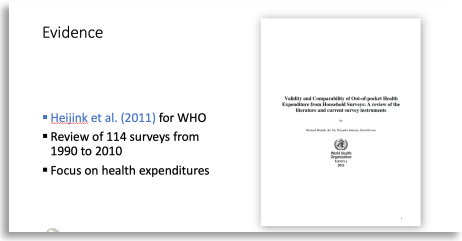


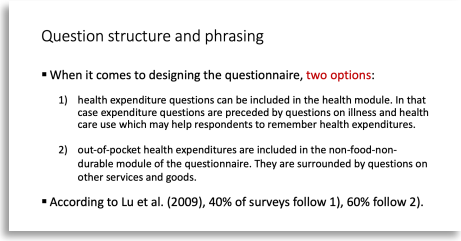
Oseni et al. (2018) provide more specific guidelines on education. The Instructor is referred to the Guidebook for details.





A useful reference for specific guidelines on the health module is Gertler et al. (2000), which draws once again from the LSMS experience.





The lecture is wrapped up by a discussion of the **current practice**. The point to be made is that there is some leeway in designing many of the important features of the health module (recall period, disaggregation, structure).

Homework

Exercise 1 – Engaging with the literature

See exercise 1, Lecture 1.

Exercise 2 – 'In or out?'

Students must read Section 3 of Deaton and Zaidi (2000) in order to engage with this exercise. However, an exact solution is not the goal here. Not all items will be uncontroversially in or out: the exercise can be used to spur a discussion on analytical choices, and their bearing on data collection.

Exercise 3 – The treatment of health expenditures

The goal of the exercise is to let students evaluate the practical repercussions of the concepts discussed during the lecture. Both countries exclude health expenditures from the consumption aggregate; the motivation given by both is in accord with the arguments presented by Deaton and Zaidi (2000), except for the inclusion of purchased medicine in the case of Bhutan, which is essentially left unmotivated.

Exercise 4 – The education module

This exercise gives students an opportunity to see a practical application of the concepts discussed in the lecture, and become familiar with questionnaires used in different countries. The goal of the exercise is to assess the students' grasp of the guidelines on the education module, and see whether they can critique questionnaires based on the recommendations learned during the lecture.

Lecture 9
Durable goods

Learning objectives

The goal of this lecture is to provide detailed recommendations on the collection of data on durable goods. The lecture explains how analysts treat this category of expenditures when constructing a measure of living standards; these analytical requirements guide the design of a survey module on durable goods.

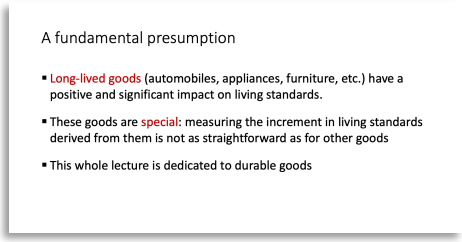
Suggested preparation

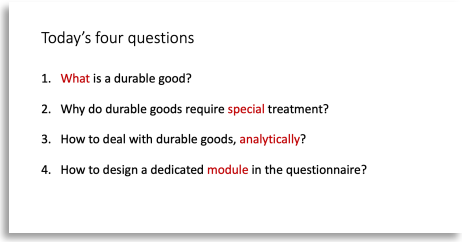
The Instructor is assumed to be familiar with Section 3.4 of Deaton and Zaidi (2000) (Consumer Durables), and with Amendola and Vecchi (2014).

Time allocation

What are durables, and why do they require special treatment?	**20 min**
How to deal with durable goods	
Notation. Acquisition and rental equivalence approaches	40 min
Break	**15 min**
User cost approach, data requirements	25 min
How to design a dedicated module in the questionnaire?	**15 min**
Lessons learned	**5 min**

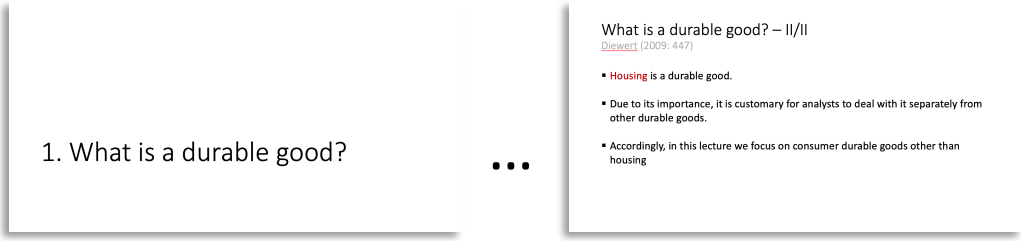
Annotated lecture



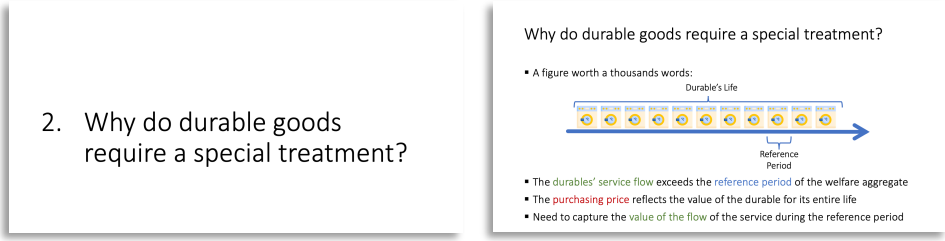


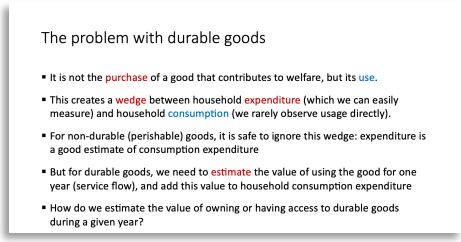
The first few slides of this set make an important point: **durable goods** are not like other consumption goods that contribute to living standards, and they **require special treatment** when we construct a consumption-based measure of welfare. *Why* this is the case and *how*

to treat these goods, both in terms of analysis and if questionnaire design, will be the subject in this lecture.

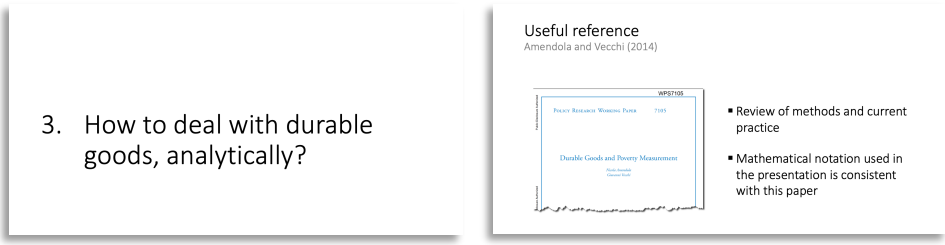


First, we give an exact **definition** of durable goods, using Diewert (2009). Typical examples of durable goods are housing (which requires separate treatment, and will be discussed in the next lecture), cars and other vehicles, household appliances, and so on.

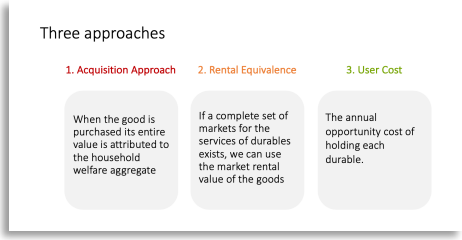


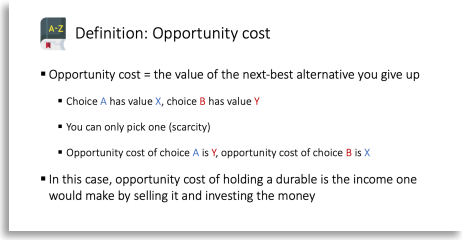


Next, why is it that durable goods require a special treatment? The key concept is that welfare analysts want to measure *consumption* during the reference period, and not expenditure. For most goods, which are purchased and consumed right away, expenditure is a good proxy for consumption. But the use of a durable good exceeds, by definition, the reference period, and its purchase price reflects this long-lasting flow of consumption that the consumer is able to enjoy from the good: if you buy a washing machine today, you are paying for the consumption of it throughout its duration, which is much longer than a year. Instead, welfare analysts want to **measure only the portion of consumption that is enjoyed during the reference period**.

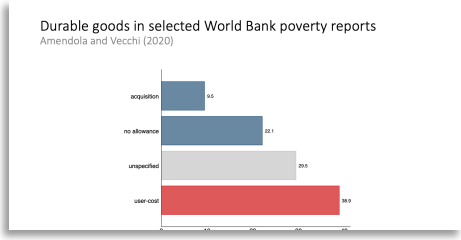


The next question is, how do welfare analysts treat durable goods? Our discussion in the rest of this lecture is a condensed version of the reference cited here: this is required reading for anyone who wishes to understand the details of what is about to follow.



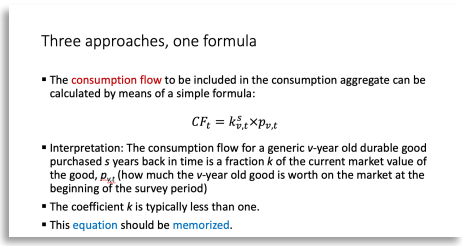


There are three main analytical approaches to the treatment of durable goods, and each of them requires different pieces of information – which in turn reflects on how the durable goods module of the questionnaire should be designed. The first approach (**acquisition**) is cited as a baseline: it consists in ignoring the peculiar nature of durable goods and the 'special treatment' due to them, and simply using the purchase value as a representation of the value of consuming them. This is obviously incorrect, given the previous discussion. The **rental equivalence** and **user cost** approaches are, instead, two ways of computing the use-value of the durable good during the reference period. One of our 'definition' slides with the encyclopedia icon reminds students of the precise definition of the economic concept of opportunity cost, which is crucial for the user-cost approach (the one that, as we will see, is recommended).



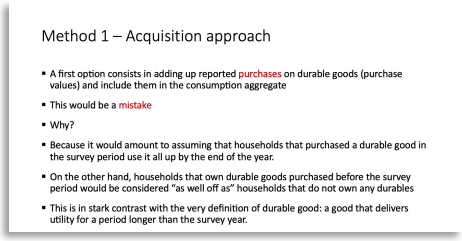
The bar chart shows the relative frequency of the approaches in recent poverty assessments around the world: the user-cost approach is most frequent, but the use of the (incorrect) acquisition approach is not rare, and as many as 22% of poverty assessments do not include the value of durable goods in the welfare indicator. Note that the lack of suitable data may be behind such a phenomenon.

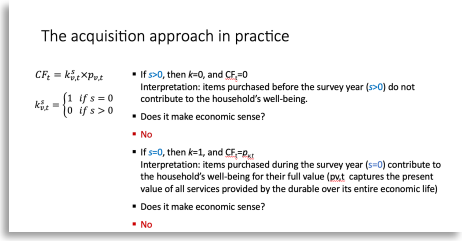
...



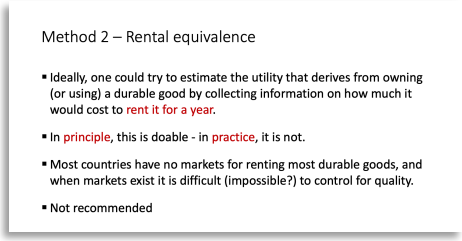
How do the different approaches work in practice? We introduce some conventional notation and a basic formula that helps us think more clearly about the problem at hand: in

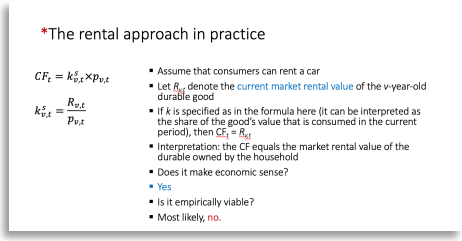
a nutshell, the **consumption flow** from the durable good during the reference period, CF_t, is a fraction k of its current market price.



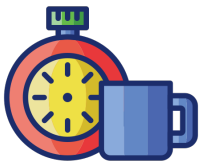


First, we discuss the acquisition approach – basically, *not* treating durable goods any different than any other good. If a household purchases a durable good, say a car, during the survey period, then they are assumed to have consumed its entire value that same year; if a household did not purchase any cars during the durvey period, they are assumed to not enjoy any returns from it, *whether or not they own a car*. This is not correct from a conceptual standpoint: again, it is *use*, not *purchase* that counts for well-being.

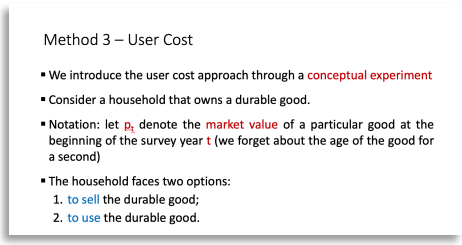


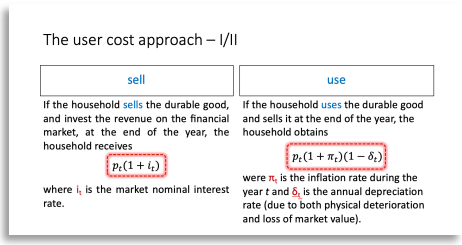


Then, the **rental equivalence** approach. In theory, because rent as a representation of the use-value of a good for a particular time period, it is very sensible to try to measure CF_t as a rental value. The problem is, a **rental market does not exist** for most durable goods, so that this approach is generally unfeasible in practice. We will see that, unsurprisingly, this is not the case for housing. The optional (starred) slide presents the rental equivalence approach in terms of the basic equation introduced at the top of this section.

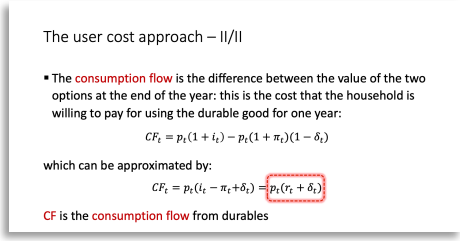


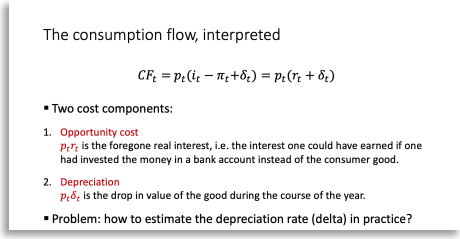
Now is a good time for a break.

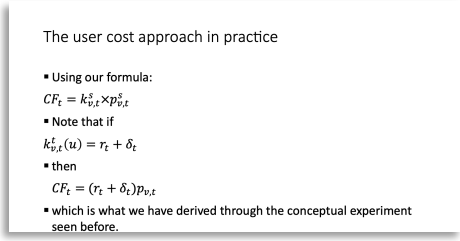




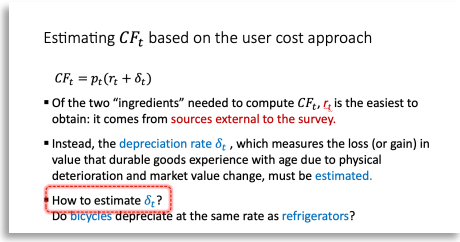
Finally, we discuss the **user cost method**, which requires a bit more time, because it is the most used in practice. The first step is to recognize that a household that owns a durable good can be seen as facing **two options** at the beginning of the survey period: to sell it right away, or to use it for a year, and then sell it.

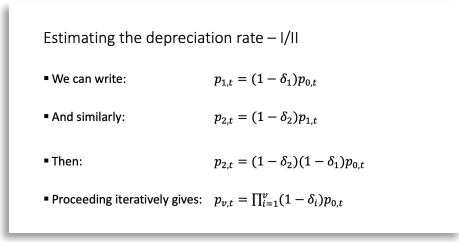




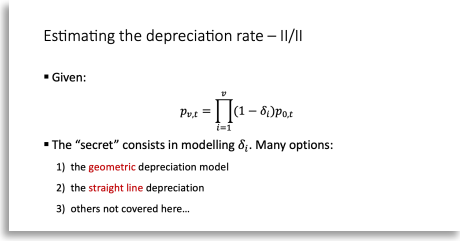


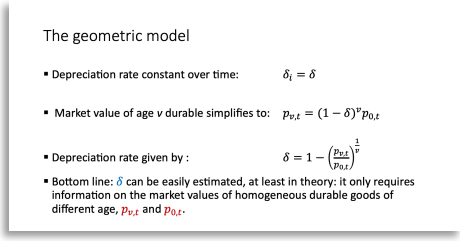
The difference between the values of the two options is an estimate of CF_t. If a household does not sell the good, but uses it, it is because the good's utility compensates for that difference. Another way to see the final equation is that, by holding the good for a year, the household is 'paying' the *opportunity cost* of selling it, plus the economic depreciation of the good. This can be brought back again to the general formula seen at the start.





We now have a formula for CF_t, but we still do not know what the depreciation rate **delta** is. If we think about delta as relating the market value of a new good at time t ($p_{0,t}$), with the market value of a 1-year-old good of the same type, again at time t ($p_{1,t}$), then we can proceed iteratively until we reach a general expression. This expression is one step closer to knowing delta (assuming that we know both $p_{v,t}$ and $p_{0,t}$); but still, the expression says that we have a different delta for each year of the good's life, which is highly complex.





To overcome this complexity, we need to make assumptions, and to **model delta**. The most common assumption is that of a **geometric depreciation model**: this simply means that we assume delta to be the same every year. This delivers a simple way to estimate delta – at least in theory.

*The straight line model

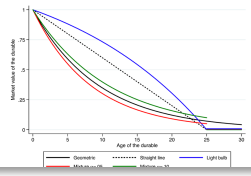
- Finite economic life. After T years CF falls down to zero. Linear pattern

$$\frac{p_{v,t}}{p_{0,t}} = \begin{cases} \frac{T-v}{T} & if\ v \leq T \\ 0 & otherwise \end{cases}$$

- The depreciation rate increases over time

$$\delta_i = \begin{cases} \frac{1}{T-i} & if\ i < T \\ 1 & otherwise \end{cases}$$

*Depreciation models compared



The optional slides mention some alternative models for estimating delta.

Recap

- User cost is the more appropriate concept to evaluate the consumption flow from durables
- In terms of data requirements, the geometric depreciation model is a good compromise
- We need to estimate:

1) Current market value of the durable: $p_{v,t}$
2) Current real interest rate: $r_t = i_t - \pi_t$
3) Depreciation rate: δ

This slide gives an opportunity to recap the user cost approach, and the 'ingredients' that are needed to implement it: this is where data comes in. **Questionnaire design must match the analysts' needs**, and gather enough information to allow for the user cost approach to be feasible in practice.

Data requirements: first best

- Current market value of item of vintage v: $p_{v,t}$
- Current market value of a new item: $p_{0,t}$
- Age v of the durable
- Current nominal interest rate: i_t
- Current yearly inflation rate: π_t

$$CF = (i_t - \pi_t + \delta)p_{v,t} \qquad \delta = 1 - \left(\frac{p_{v,t}}{p_{0,t}}\right)^{\frac{1}{v}}$$

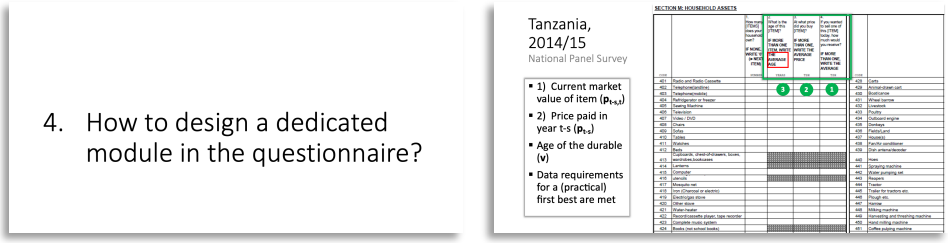
Ideally, all of the 'ingredients' of the formulas reproduced at the bottom of the slide (the user cost method for estimating the CF from durable goods) would need to be gathered by the questionnaire, except for the nominal interest rate and inflation rate, which are typically available from external sources. If this is the case, the analyst would simply plug survey data into the formula: this is what is meant by **first best**.

Some practical considerations

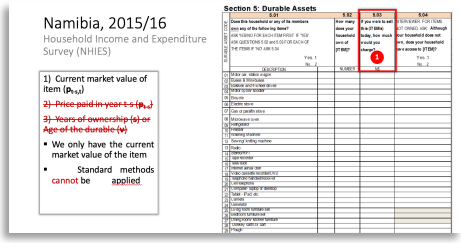
$$\delta = 1 - \left(\frac{p_{v,t}}{p_{0,t}}\right)^{\frac{1}{v}}$$

- In place of v (age of the durable): years of ownership can be a more practical approximation
- In place of $p_{0,t}$ (current market value of a new item): more practical to ask for the price originally paid for the good when it was purchased, $p_{v,t-v}$
- The analyst will use an inflation rate π in order to approximate what is needed for estimating delta: $p_{0,t} \approx (1+\pi)^v p_{v,t-v}$

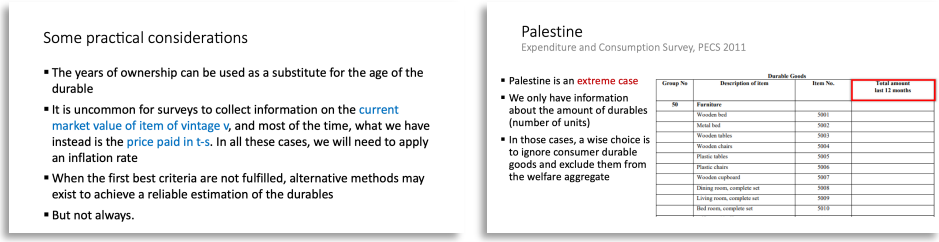
Realistically, obtaining enough reliable information on all of the ‘first best’ variables is not easy. In practice, a more realistic ‘**second best**’ approach is to ask households about the purchase time, purchase value, and current market value of items they own. The years of ownership of the good can be used as a proxy for the ‘vintage’ or age of the durable, *v*, if needed. This still allows to apply the user cost method, provided that some changes are made to the formulas: the new expression is reproduced below.



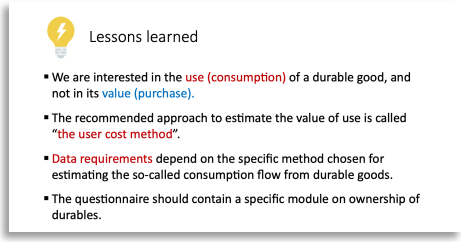
This is an expample of a survey that correctly gathers information needed to implement the second best approach (here called a ‘practical first best’). In this case, we know *v*, rather than *s*, which is ok from an analytical standpoint.



This, on the other hand, is an example of a survey that does not gather enough information to implement even the second best approach.



In practice, even if the ‘second best’ approach is not feasible, we can still apply alternative methods, and estimate the necessary variables in an indirect way. In a case like Palestine this is really impossible: the available data is too limited.



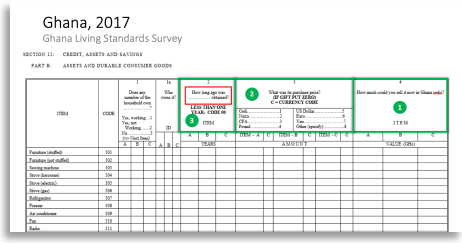
Lessons learned deserve a few minutes of our time. The takeaways are that durable goods require special treatment when constructing a measure of welfare, because welfare analysts want to proxy consumption; to do that, the user cost method is recommended. The questionnaire should definitely include a module on durable goods, and gather as much information as needed to apply the user cost method.

Homework

Exercise 1 – Engaging with the literature

See exercise 1, Lecture 1.

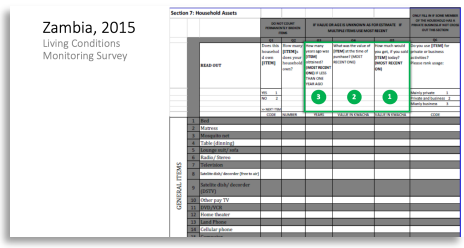
Exercise 2 – The durable goods module



Data requirements for the second best are met even in the absence of the age v of the durable:

1) Current market value of item ($\mathbf{p_{t-s,t}}$)
2) Price paid in year t-s ($\mathbf{p_{t-s}}$)
3) Years of ownership (**s**)

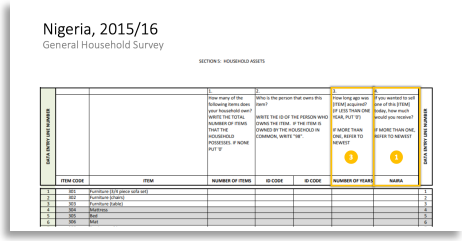
Note a potential problem with the question: "How long ago was […] <u>obtained</u>". If many durables were obtained as gifts, then the reported purchase value of goods will often be zero, and analysts may need to impute these values.



Similarly to Ghana, the data requirements for a second best are met even in the absence of the age *v* of the durable:

1) Current market value of item ($\mathbf{p_{t-s,t}}$)
2) Price paid in year t-s ($\mathbf{p_{t-s}}$)
3) Years of ownership (**s**)

Here, information is collected on the "most recent" durable. How does this affect the resulting estimate (think about cases in which the household owns more than one durable good for each type)?


Nigeria, 2015/16
General Household Survey

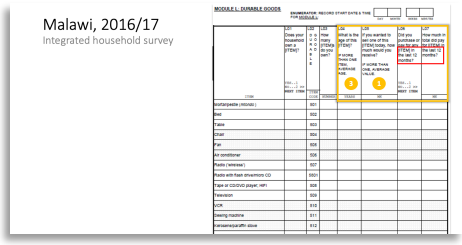
SECTION 5: HOUSEHOLD ASSETS

DATA ENTRY CODE MANAGER			1. How many of the following items does your household own? WRITE THE TOTAL NUMBER OF ITEMS THAT THE HOUSEHOLD POSSESSES. IF NONE PUT '0'	2. Who is the person that owns this item? WRITE THE ID OF THE PERSON WHO OWNS THE ITEM. IF THE ITEM IS OWNED BY THE HOUSEHOLD IN COMMON, WRITE "HH".		3. How long ago was [ITEM] acquired? (IF LESS THAN ONE YEAR, PUT '0'). IF MORE THAN ONE, REFER TO NEWEST 3	4. If you wanted to sell one of this [ITEM] today, how much would you receive? IF MORE THAN ONE, REFER TO NEWEST 1	DATA ENTRY CODE MANAGER
	ITEM CODE	ITEM	NUMBER OF ITEMS	ID CODE	ID CODE	NUMBER OF YEARS	NARA	
1	305	Furniture (5/9 piece sofa set)						1
2	302	Furniture (chairs)						2
3	303	Furniture (table)						3
4	304	Mattress						4
5	305	Bed						5
6	306	Mat						6

In this case, standard methods cannot be applied.

1) Current market value of item ($\mathbf{p_{t-s,t}}$)
~~2) Price paid in year t-s ($\mathbf{p_{t-s}}$)~~
3) Years of ownership (**s**)

Alternative methods exist when we only have the current market value of the durable, and an estimation of the **maximum economic life of the durable.**


Malawi, 2016/17
Integrated household survey

MODULE L: DURABLE GOODS

ENUMERATOR: RECORD START DATE & TIME FOR MODULE L

ITEM	L01 Does your household own a [ITEM]? CODE: 1 = Yes, 2 = No MORE ITEMS?	L02 How many [ITEM]s do you own?	L03 What is the age of this [ITEM]? (IF LESS THAN ONE YEAR, PUT '0'). IF MORE THAN ONE, REFER TO NEWEST 3	L04 If you wanted to sell one of this [ITEM] today, how much would you receive? IF MORE THAN ONE, REFER TO NEWEST 1	L05 Did you purchase or [illegible] the [ITEM] in the last 12 months? CODE: 1 = Yes, 2 = No MORE ITEMS?	L06 How much in total did you [illegible] the [ITEM]?
		CODE	YEARS	MK		MK
Mortarpestle (Mtondo)		501				
Bed		502				
Table		503				
Chair		504				
Fan		505				
Air conditioner		506				
Radio (wireless)		507				
Radio with flash drive/micro CD		5501				
Tape or CD/DVD player, MP3		508				
Television		509				
VCR		510				
Sewing machine		511				
Kerosene/paraffin stove		512				

The questionnaire only collects information about the price paid if the durable was acquired *in the last 12 months* (in year t!).

1) Current market value of item ($\mathbf{p_{t-s,t}}$)
2) Price paid in year t ($\mathbf{p_t}$)
3) Age of the durable (**v**)

No information is collected on the price paid in year t-s; however, note that, in this case, $\mathbf{p_t}$ and $\mathbf{p_{t-s,t}}$ will probably be very similar value. This case is similar to **Nigeria,** but the fact that there is information on the age of the durable puts the analyst in a better position to use alternative methods for estimating the CF.

Lecture 10
Housing

Learning objectives

The goal of this lecture is to provide detailed recommendations on the collection of data on housing expenditures. The lecture explains how analysts treat this category of expenditures when constructing a measure of living standards; these analytical requirements guide the design of the survey module on housing.

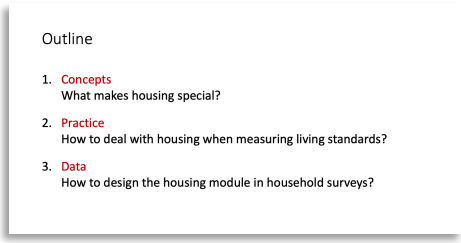
Suggested preparation

The main reading for the lecture is Malpezzi (2002), which focuses on data collection. For a thorough understanding of the analytical requirements underlying the design of the questionnaire, it is also important to be familiar with Deaton and Zaidi (2002), Section 3.5, and Balcázar et al. (2017).

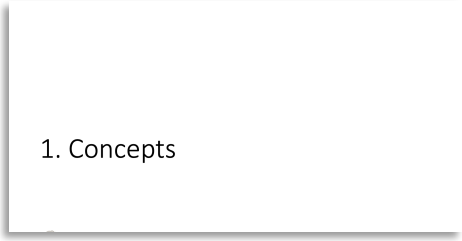
Time allocation

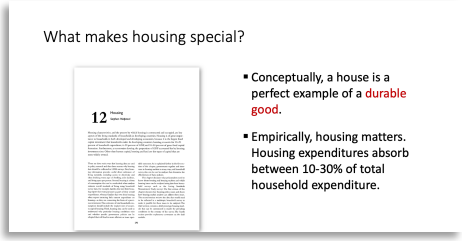
Concepts	30 min
Practice	30 min
Break	**15 min**
Data	40 min
Lessons learned	5 min

Annotated lecture

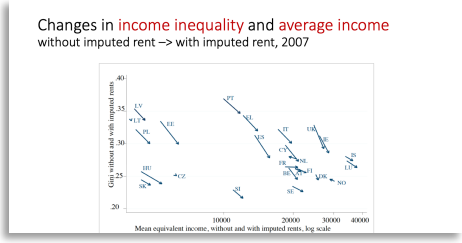


The set opens with a roadmap of the lecture, which is organized around three topics: concepts, practice, and data.

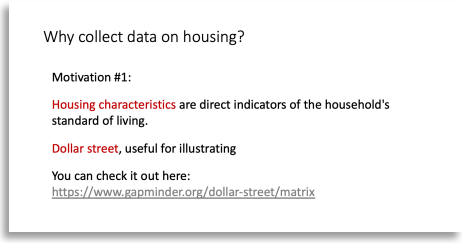


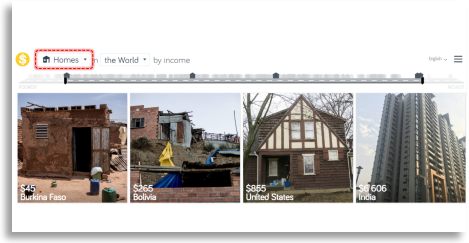


Conceptually, **housing is a durable good**, and presents the same issues as those discussed in the previous lecture.

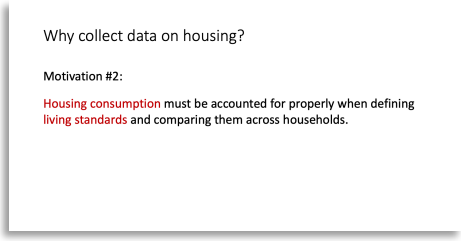


These slides present evidence of the **empirical relevance of housing expenditures**, in the context of welfare measurement: inequality and poverty estimates are significantly affected by housing. The graph from Törmälehto and Sauli (2010) shows, for a selection of European countries, the trajectory of two variables, income inequality (measured by the Gini index, vertical axis) and average income (horizontal axis), before and after including imputed rent in the welfare aggregate: most countries move towards south-east, that is, including imputed rent decreases estimated inequality and increases average incomes.

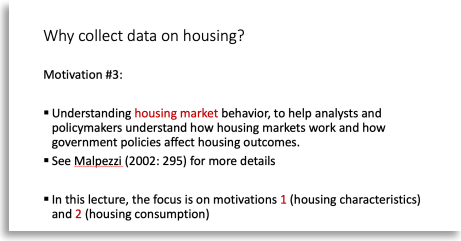




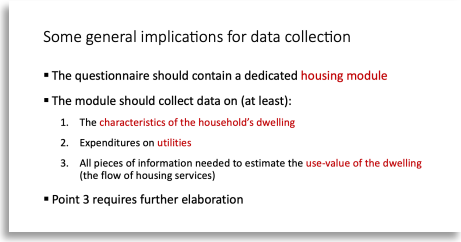
Data on housing are interesting for several reasons. One is that housing characteristics are themselves an **indicator of living standard**: Dollar Street, a project developed by the Gapminder Foundation, is used to illustrate how dwelling characteristics go hand in hand with well-being. The example of "flying toilets" is a vivid way of communicating the importance of some of these housing features for the standard of living of families.



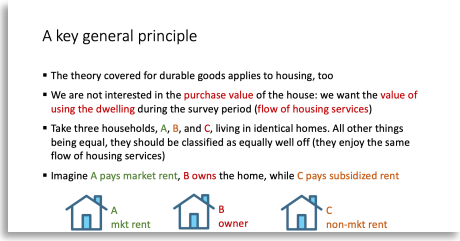
Another reason for collecting housing data is that housing is a part of total household consumption, and therefore we need to **account for it in the consumption aggregate**.

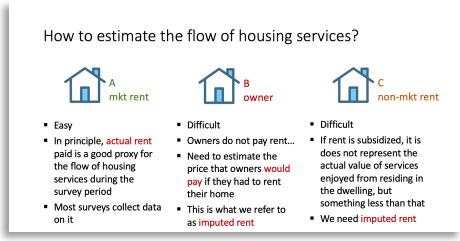


Finally, data on housing expenditures are of interest for analysts that study the **behavior of housing markets**. Of course, our focus is on welfare analysis, therefore motivations 1 and 2 will drive the rest of the discussion.

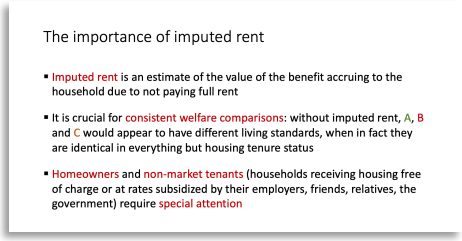


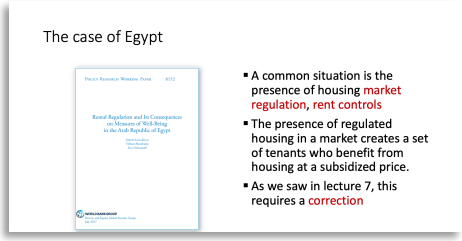
Motivations 1 and 2 imply that the housing module must contain at least the pieces of information here listed. But what is the **use-value of the dwelling**? The next few slides explain.



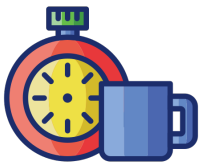


Theoretical considerations here are very similar to those that apply to durable goods. Welfare analysts are interested in the **value of consumption** of durable goods, including the dwelling, not in their purchase value. What differentiates a house from other durables is the fact that **rent** is usually available, and as mentioned in the previous lecture, rent is an adequate representation of the use-value of a good. For households that rent their dwelling, the case is closed. However, households that own the dwelling do not pay any rent, of course. A similar issue arises when a household pays a rent that is not representative of the true value of their dwelling, but is artificially lowered (by subsidies, for instance). In both of these cases, what we need is the rent that owners or non-market tenants *would* pay, if they were to rent their home on the housing market. This is called **imputed rent**.

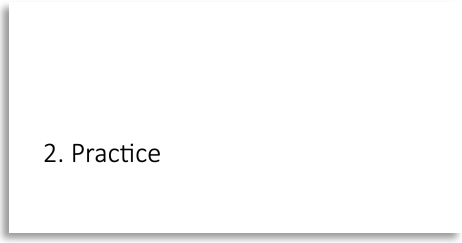




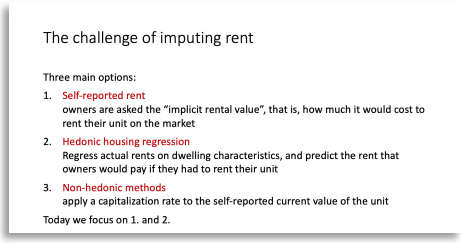
These slides reinforce the concept of imputed rent and its importance for making correct welfare comparisons. Egypt is a real-life case in which rent controls (legally enforced lowered rents for certain tenants) pose this kind of problem to analysts.



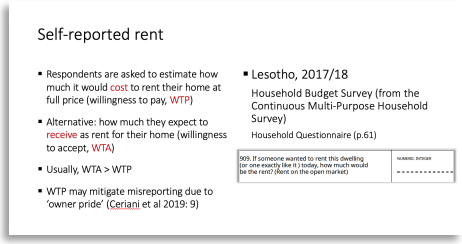
Now is a good time for a break.

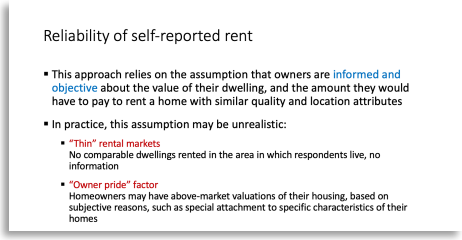


...



In practice, computing imputed rent is **not trivial**. Different methods are available. Students are directed to the article by Balcazàr et al. for a complete review: this is currently the most comprehensive and updated source on the topic, and it is a fundamental reading for this lecture. In fact, it will be difficult to comprehend the rest of the discussion on housing without having gained at least some familiarity with it. There are three main options for estimating imputed rent, but we will focus on the first two (the last one is less often used in practice).

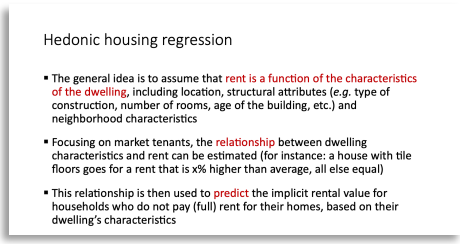


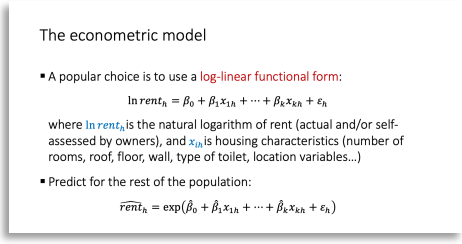


The first approach, **self-reported rent**, relies entirely on the collection of data directly from respondents, by asking homeowners a question of the type: "How much would you pay if you were to rent the dwelling you are currently living in?". This question records the respondent's willingness to pay (WTP) for hypothetically renting her own dwelling;

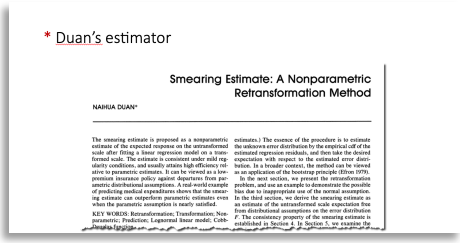
another approach is to record the respondent's willingness to accept (WTA) payment in exchange of her dwelling, with a question like "How much would you receive as a rent if you were to lend your apartment?". The two questions are very similar, and both are used in questionnaires around the world. Ceriani et al (2019) cites some of the literature comparing WTP and WTA approaches, and mentions that usually, WTA questions yield higher reported amounts than the equivalent WTP questions, "due to the fact that a person's perceived value of an item is often higher when the person owns the item (i.e., the role of the landlord), then when she does not own it (i.e. the role of the tenant). This would suggest that the second version of the question (WTP), may mitigate the possibility of misreporting due to sentimental attachment to the property." (p. 9).

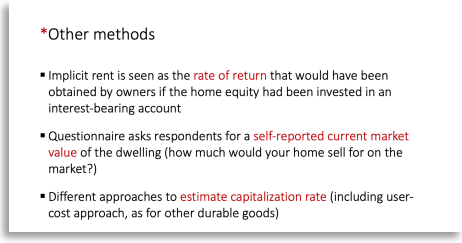
The self-reported rent question should always be included in the housing module; however, there may be challenges to the quality of the data collected in this way, when respondents are not informed (the housing market in the area does not exist, or is very small, so respondents have no information to base their response on) or not objective (the perceived value of one's own dwelling is unrealistic).





An alternative is to estimate rent based on other information, namely the characteristics of the dwelling (a nocer house would go for more on the rental market than a dilapidated one: if the analyst is able to select the characteristics that influence market rent, she can "predict" it). This approach is called **hedonic regression**, usually implemented as a log-linear specification.





These optional (starred) slides mention two extensions of the topic that was just discussed. **Duan's estimator** provides more accurate predictions of rents estimated via hedonic regression, for reasons having to do with the logarithmic transformation of the dependent variable. The Instructor is invited to read the paper for more details. The second slide hints at another avaulable method besides hedonic regression, which may be used if all else fails.

Main takeaways

- Estimating an "implicit" rent for all those households who do not pay actual market rent is one of the main challenges facing welfare analysts
- Several estimation approaches, based on different assumptions: choice depends on context
- Methodology aside, success of estimation rests on the availability of the necessary information from surveys (data availability) and its accuracy (data quality)
- Self-reported rent and dwelling characteristics emerge as crucial data requirements

Now is the time to recap the main messages of this section: the last bullet is especially important, and leads to the final part of the lecture.

3. Data

Examples of definitions for the housing module

Malpezzi (2002: 307)

Definitions are context-specific: this is just a potential starting point

- Structure
  "A structure is a physically separate entity such as a house, an apartment building, or a tent. It may contain one or more dwelling units."
- Dwelling unit
  "A dwelling is an accommodation unit that contains one or more households. There may be several dwellings in a structure."
- Room
  "Whole rooms used for living purposes, such as living rooms, dining rooms, bedrooms (…). Not included are bathrooms (…). If a room is used by occupants of more than one unit, the room is included with the unit from which it is most easily reached."

These slides offer a few practical recommendations for the **design of the housing module**, mostly from Malpezzi (2002). First, it should be clear that concepts must be defined, however familiar they may seem.

Components of the housing module

Malpezzi (2002: 305-310)

Given our focus on living standards measurement, we summarize recommendations for the following components of the housing module:

1. Dwelling characteristics and housing services
2. Dwelling expenditures

Dwelling characteristics and housing services

- Characteristics of the structure or dwelling
  - Building materials (roof, walls, floor…)
  - Age and size of the structure
  - Number of rooms/bedrooms/bathrooms
- Characteristics of the neighborhood, availability of services
  - Location of the dwelling
  - Availability and distance from services (water, sanitation…)

Next, what should the housing module should ask for, in practice? **Dwelling characteristics** are crucial.

Dwelling expenditures

a. Questions about expenditures are closely linked with questions about housing tenure
b. Main expenditure to be recorded: rent (actual and self-reported)
c. Other housing expenditures: utilities, maintenance and repairs

Housing tenure

- Property rights and tenure vary considerably depending on the country's context: these questions must be customized
- At a minimum, questionnaire should differentiate owning vs. renting
- Length of tenure is important because it often impacts rent paid

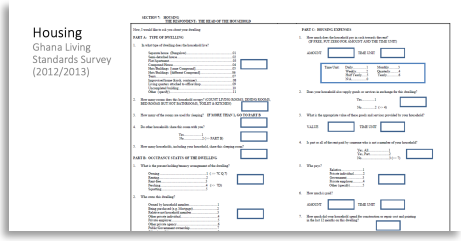
Rent

- Actual rent (for renters)
  - "How much does the household pay towards rent?"
    - Data must be collected on "arms-length transactions", i.e. transactions between two counterparts who have no special relationship that would suggest that rent paid differs from market prices
  - Crucial to differentiate between households that pay market rent, and households under rent controls or subsidy, related to the landlord, etc.
- Self-reported rent (for non-renters)
  - "How much would you charge if you were to rent out this dwelling?"

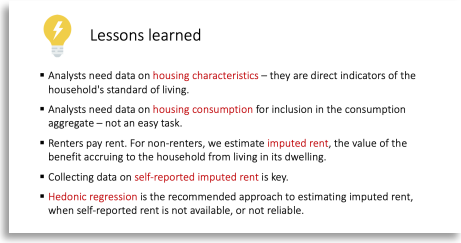
Utilities, maintenance and repairs

- Distinction between housing expenditures inclusive of utilities or not is crucial
- *E.g.* some renters pay for utilities separately, but others pay a monthly rent that includes utility charges
- Questionnaire must be designed to distinguish between these cases
- Expenditure for utilities may be collected elsewhere in the questionnaire (together with other expenditures with same recall period)

In terms of **dwelling expenditures**, the crucial pieces of information are tenure status, rent, and utilities/maintenance/repairs.



The final slides show **examples** of housing modules in recent HCES. Students are invited to examine them and point out which questions collect the different pieces of information mentioned during the lecture.



As usual, the Instructor is encouraged to take a few minutes to recap the main messages of the lecture. In terms of data collection, the main targets for those designing questionnaires are self-reported rent (it is fundamental for welfare analysts) and dwelling characteristics (both because of their use as direct indicators of living standards, and as regressors in hedonic models).

Homework

Exercise 1 - Engaging with the literature

See exercise 1, Lecture 1.

Exercise 2 – Secondary residences

This exercise gives students an opportunity to explore the housing module in a number of different questionnaires, and pushes them to examine their structure in depth, by asking a question that was not covered during the lecture. Each survey will of course collect data on secondary residences in a slightly different way: rather than the answer, what is important here is the students' hands-on engagement with the questionnaire.

Exercise 3 – Housing in theory and practice

Similarly to Exercise 2, this question asks students to comment on questionnaires, this time pushing them to make a connection with the theory. Ultimately, students should be able to conclude (i) whether or not the design of the housing module allows for the implementation of at least one of the recommended methods for the estimation of the use-value of the dwelling, and (ii) what was actually done when constructing the welfare indicator.

Lecture 11
Data validation and diagnostics

Learning objectives

This lecture deals with data validation techniques and data diagnostics. The main goal of the lecture is to provide a gentle introduction to a variety of problems that arise when survey data are collected in the field and processed by the NSI prior to dissemination with final users. The lecture is closely linked to Lecture 12, which focuses on outliers, as the latter are often a specific form of (gross) errors, one of the many covered in this lecture.

Suggested preparation

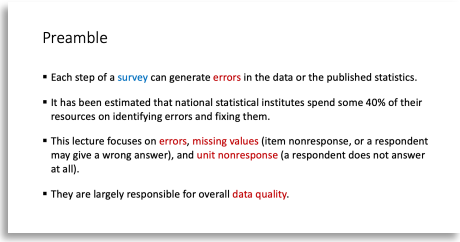
The single most important reference for this lecture is the manual by de Waal et al. After becoming familiar with the sections indicated in the syllabus, Instructors are encouraged to customize the references with country specific material, as appropriate.

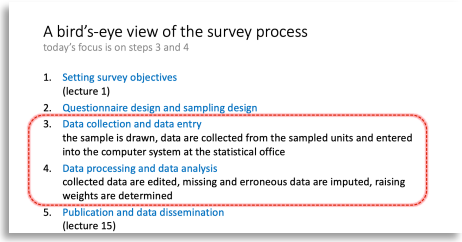
The number of slides in this lecture is intimidating. Many are examples, which only require a few seconds of discussion. Nevertheless, the Instructor is invited to check whether cuts are needed *before* delivering the lecture.

Time allocation

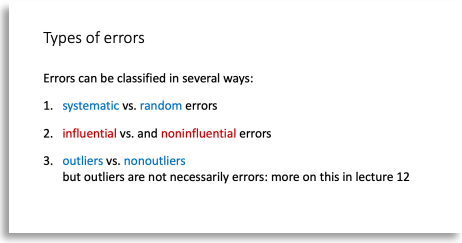
Types of errors	20 min
Data editing	15 min
Missing data	25 min
Break	**15 min**
Data validation and diagnostics	40 min
Lessons learned	5 min

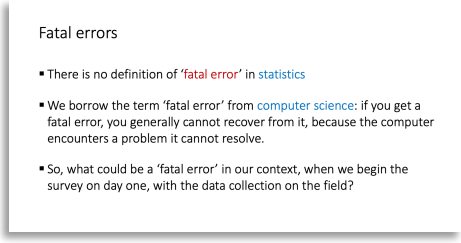
Annotated lecture



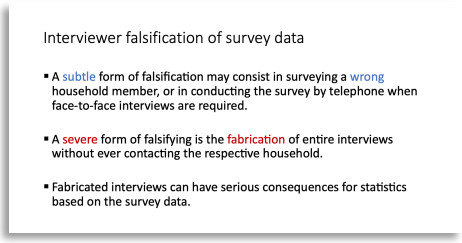


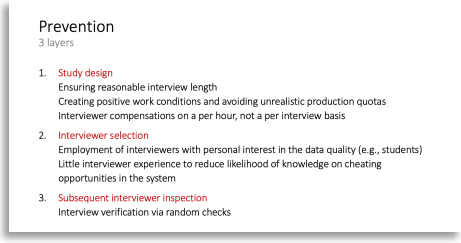
We begin by ackowledging that errors are ‘part of the game’. After showing the first bullet of the first slide we can open the floor and engage with a discussion with students, asking them what kind of errors they can think of. We do not reply, we simply ‘chair’ the discussion. Then, we remind students that this lecture focuses on errors that arise in steps 3 and 4 of the survey process.



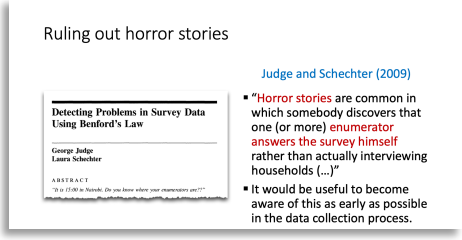


Next, we provide students with a quick overview of different types of errors. The goal is twofold: i) to introduce a few key *concepts* on the different types of error and ii) to build up the students’ statistical vocabulary.





To liven up this ‘catalogue of errors’, which is useful but potentially a bit dull, we touch on the issue of **interviewer falsification** of survey data. This is a potentially serious threat to data quality.



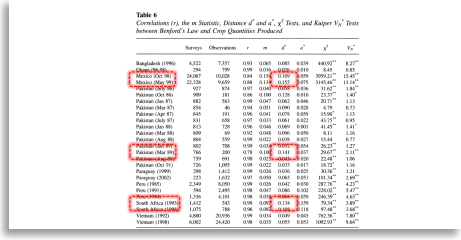
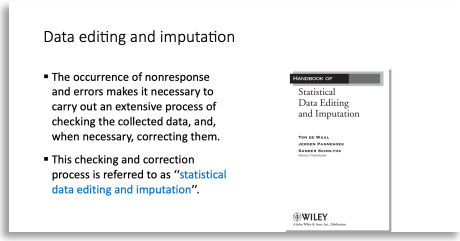
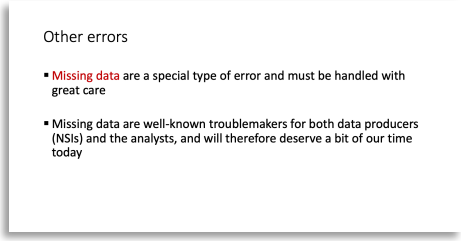


Table 6
Correlations (r), the m Statistic, Distance d^* and a^*, χ^2 Tests, and Kuiper V_N^* Tests between Benford's Law and Crop Quantities Produced

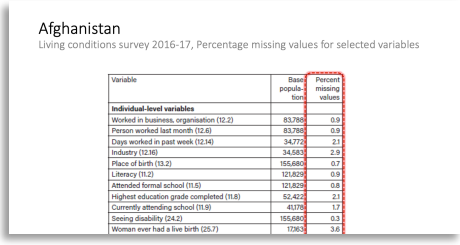
	Surveys	Observations	r	m	d^*	a^*	χ^2	V_N^*
Bangladesh (1996)	4,522	7,357	0.93	0.065	0.085	0.038	440.92**	8.27**
China (Dec 86)	284	799	0.99	0.019	0.026	0.010	8.69	0.85
Mexico (Oct 91)	24,087	10,026	0.84	0.154	0.169	0.059	3959.21**	15.45**
Mexico (May 94)	22,328	9,659	0.88	0.147	0.152	0.075	3145.46**	11.94**
Mexico (Oct 94)	927	874	0.97	0.048	0.054	0.016	31.62**	1.86**
Pakistan (Oct 86)	989	181	0.86	0.100	0.128	0.016	23.37**	1.46*
Pakistan (Jan 87)	882	563	0.99	0.047	0.062	0.046	20.71**	1.13
Pakistan (Mar 87)	854	46	0.94	0.051	0.094	0.028	4.79	0.73
Pakistan (Apr 87)	845	191	0.98	0.041	0.079	0.059	15.98*	1.13
Pakistan (Oct 87)	811	658	0.97	0.033	0.061	0.022	43.75**	0.95
Pakistan (Jan 88)	833	728	0.96	0.046	0.089	0.001	41.45**	1.41*
Pakistan (Mar 88)	833	728	0.96	0.046	0.089	0.001	41.45**	1.41*
Pakistan (Aug 88)	804	559	0.99	0.022	0.036	0.027	15.44	0.73
Pakistan (Oct 89)	802	758	0.99	0.018	0.025	0.054	26.23**	1.27*
Pakistan (Mar 90)	796	200	0.78	0.108	0.141	0.037	29.67**	2.17**
Pakistan (Aug 90)	759	691	0.98	0.025	0.033	0.020	22.48**	1.06
Pakistan (Oct 91)	726	1,095	0.99	0.022	0.035	0.017	18.72*	1.16
Paraguay (1999)	298	1,412	0.99	0.026	0.036	0.025	30.36**	1.21
Paraguay (2002)	223	1,632	0.97	0.050	0.063	0.051	101.34**	2.09**
Peru (1985)	2,349	8,050	0.99	0.026	0.042	0.030	287.78**	4.25**
Peru (1991)	594	2,495	0.98	0.047	0.086	0.102	220.02**	5.47**
Russia (1994)	1,336	4,101	0.98	0.038	0.064	0.059	246.39**	4.85**
Russia (1995)	1,336	4,101	0.98	0.038	0.064	0.059	246.39**	4.85**
South Africa (1993)	1,075	780	0.96	0.091	0.128	0.118	97.68**	3.68**
Vietnam (1992)	6,880	20,936	0.99	0.034	0.049	0.045	782.36**	1.80**
Vietnam (1998)	6,002	24,420	0.98	0.035	0.055	0.053	1082.93**	8.66**

We provide the Instructor with the option to take a *short* digression on a specific tool, **Benford’s Law**, that can help motivate the audience. Benford’s Law *per se* is a topic that should not take more than a couple of slides in the context of this course – we prepared 6. The reason is that this is an engaging story, it keeps the audience’s attention up and helps to ease into the rest of the discussion. In short, the 6 slides are a rethoric device, that should be used only by Instructors familiar with the references provided in the syllabus, and confident that the opportunity cost of this longer-than-strictly-needed presentation is not too high.

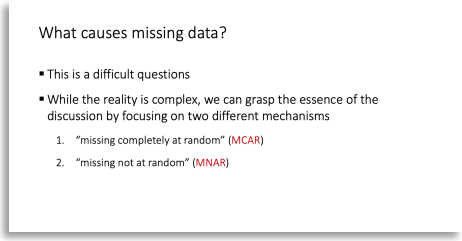


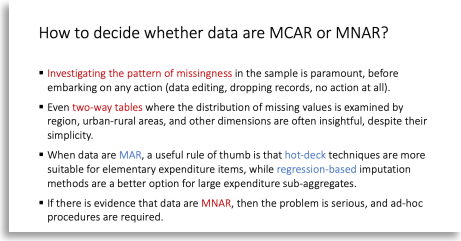


The next bunch of slides discusses the concept of 'data editing'. Slides are self-explanatory, we believe and can be improved by adding specific examples, if the Instructor wishes so. Our advice is emphasise and elaborate on *over-editing* – it is a common bad practice and the discussion in De Waal et al. deserves some attention.



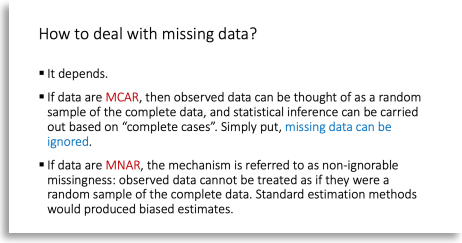
The discussion of **missing data** is kept within the space of about a dozen slides. Given the relevance and complexity of the topic, the choice of contents is clearly a challenge. We lead with a fundamental idea: the single most important priority when tackling missing data is to understand the *mechanism* driving the missingness itself, that is, why are data missing in the datasets? Different answers imply different coping strategies. The example of the Afghanistan Living Conditions Survey 2016 helps to demonstrate the fact that missing data are ubiquitous, and shows a case of transparency and disclosure of this issue in the official documentation. The reference highlighted here – the manual by Little and Rubin – is a classic, and students are highly encouraged to make use of it.

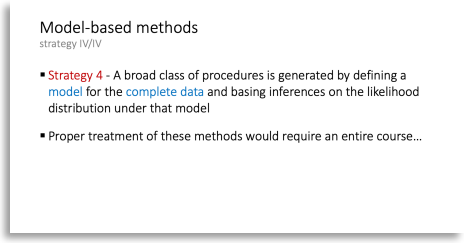




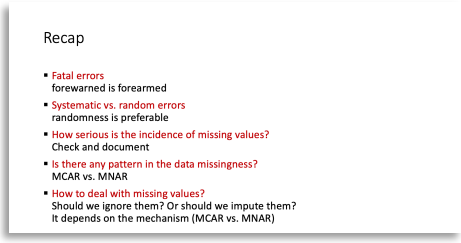
Next, we illustrate and discuss the two extremes of the possible missing data **mechanisms**: data could be MCAR, that is, missing completely at random, or MNAR, missing not at random. There is a third case, that of MAR – data missing at random, or data that are MCAR *conditionally* on some observed characteristic – that is not explicitly recalled here, but the Instructor might decide to delve deeper if needed, with the help of the references associated with this lecture. In general, determining which mechanism is behind the observed pattern of missingness is challenging. In practical situations, even simple two-way tables can help shed light on the issue. Making up one's mind on the mechanism at

work is important, because the optimal course of action to treat missing data differs under MCAR or MNAR.

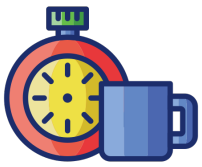




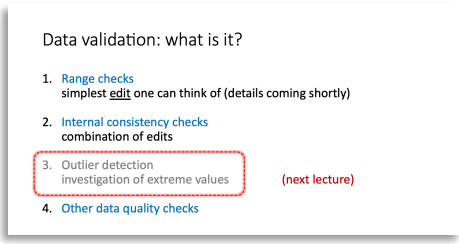
The remaining slides deal with the **treatment** of missing data. To impute or not to impute? In general, when data are MCAR, one can effectively ignore the missing observations and perform all analyses on the subset of complete cases: estimates will remain unbiased, albeit less efficient. On the other hand, when data are MNAR, more complex strategies must be deployed. If the analyst were to proceed as if under MCAR, estimates would be biased. We briefly mention the main strategies available to analysts – but the Instructor can move relatively quickly through these slides, as they are not designed to offer a complete presentation of such a complex topic. Again, interested students can be directed to the references. In particular, regarding weighting procedures, a clear and effective introduction can be found in Deaton (1997: section 1.4).

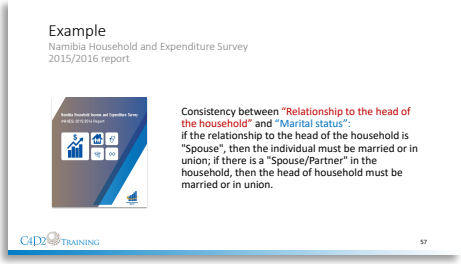


A recap of the discussion so far leads us to the break.

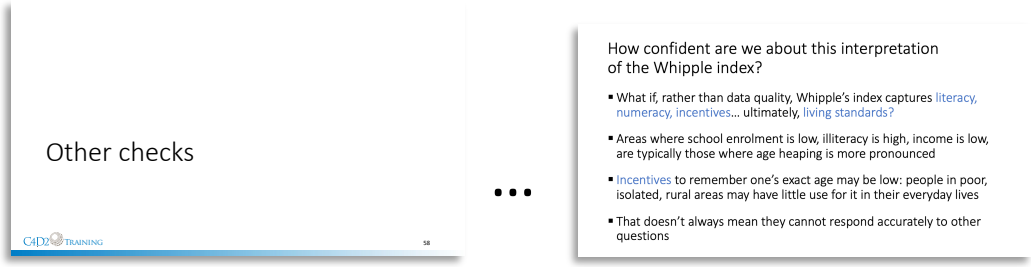


Now is a good time for a break.

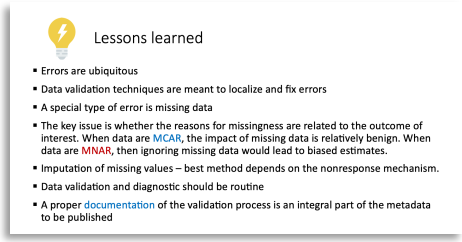




After the break, we introduce the last topic of the lecture: **data validation and diagnostics**. We start by defining the meaning of these terms, and we provide a selection of examples that can be discussed with students interactively. “Range checks” are usually simple to cover (here a reminder of the discussion on data capture methodolgies might be in order), and so are the examples of “consistency checks”. Emphasis here is not much on specific examples, but rather oin the fact the there is a need to introduce complex checks involving multiple variables, possibily collected in different sections of the questionnaire.



The final part of the lecture introduces two tools that are often used as a check of **overall data quality**. The first tool simply consists of a visual inspection of the population pyramid, as estimated by the survey. Does it show suspicious dents or spikes, or any other anomaly? If so … how to interpret it? If students are not too tired, this is a nice topic for discussion. The second tool focuses on the age variable, and calculates, the so-called Whipple index, a measure of age heaping (the tendency to report one’s age as an approximate, ‘round’ number, namely a multiple of 5). Despite its popularity, the interpretation of this index is controversial. If age heaping is common in the data, this may say less about data quality, and more about the respondents’ background, education, habits – ultimately, living standards. The Instructor might wish to discuss pros and cons with students, and end the lecture leaving it as an open issue.



The lecture closes with a plain recap of the main points covered: a) multifaceted and pervasive errors (not just probabilistic, as we as seen, but a long catalogue), b) among errors, missing data deserve special attention. As many problems as there are solutions: this is how we used the second part of the lecture, discussing data validation and imputation methodologies.

We conclude with an important message: the need to produce (and/or ask for) proper and detailed **documentation** of all choices made.

Homework

Exercises 1, 2 and 3 - Engaging with the literature

See exercise 1, Lecture 1.

Exercise 4 – Data imputation

Instructor are encouraged to design and implement this exercise as they see it appropriate. One way is to ask students to write a short essay discussing the pros and cons of the imputation procedures described in the reports in the slide (but others can be chosen).

Lecture 12
Outlier detection and treatment

Learning objectives

This lecture is an extension of the topics discussed in Lecture 11 on data validation and diagnostics. The goal of this lecture is to provide a conceptual framework and some techniques to detect and treat outliers (*i.e.* extreme values), which are commonly found in most datasets. We will focus, in particular, on consumption expenditure.

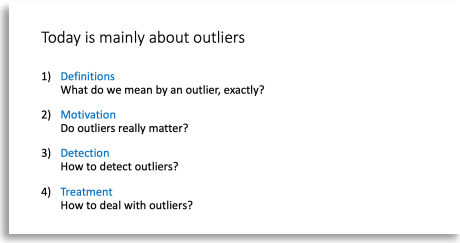
Suggested preparation

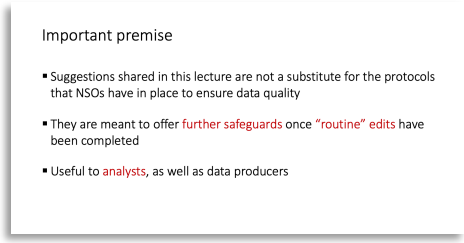
Chapters 1 and 2 of Barnett and Lewis (1994) provide background for the topics of the lecture, and the Instructor is encouraged to review this material.

Time allocation

Definitions	**10 min**
Do outliers matter?	**25 min**
How to detect outliers	
Visual inspection	25 min
Break	**15 min**
Statistical methods	35 min
How to deal with outliers	**5 min**
Lessons learned	**5 min**

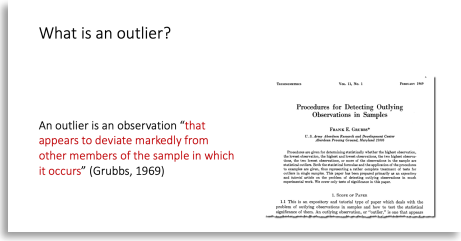
Annotated lecture

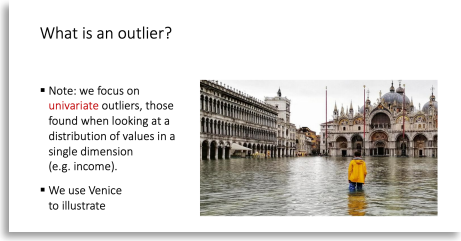




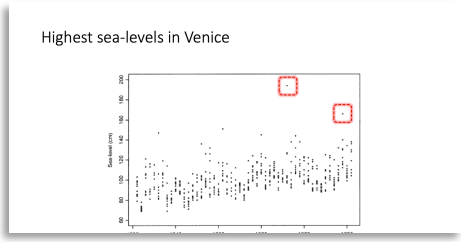
The lecture opens with a roadmap: four questions – what are outliers, how influential are they for our statistics of interest, how do we detect them, and what do we do with them – will guide the discussion for the day. Before going into them, though, the Instructor should

share an important disclaimer: none of the recommendations offered during this lecture is intended as a substitute for the routine edits and checks that NSOs have in place to ensure data quality, and to minimize, among other things, the occurrence of extreme and implausible values. Our recommendations are meant to offer further safeguards once "routine" edits and checks – by and large, those discussed in the second part of lecture 11 – have been completed.

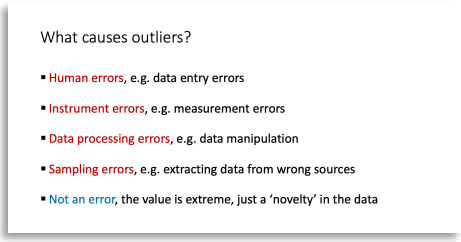


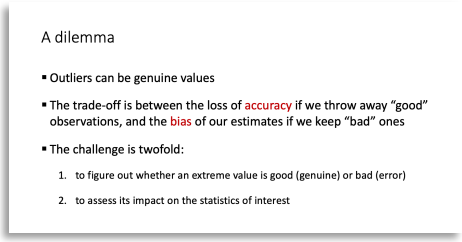


First, the basics: what is an **outlier**? Many definitions exist (a recent study, Aguinis et al. 2013, has documented the use of 14 different definitions, based on a literature review of 28 papers), but the one by Grubbs (1969) is intuitive, and has become a 'classic' definition. The note on *univariate* outliers simply means that our focus for the lecture (and the course) is on outliers within the distribution of a *single* variable; a more complex case would be that of *multivariate* outliers, which are defined based on a combination of different variables – but we will not go into that.

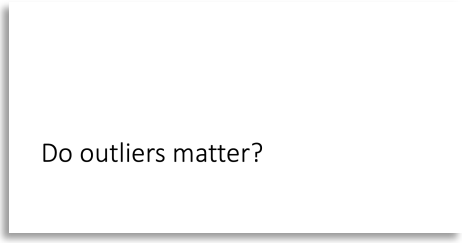


As will become apparent in the discussion that follows, **visualizing data** helps with spotting outliers. The distributions shown in the graph (sea levels measured every year from 1930 to 1985) are concentrated around a certain range, and what immediately can be perceived as 'extreme' values become visible. Here the Instructor can engage students for a minute on which data points they may want to classify as outliers, exactly, and why. The Venice example should convey the idea that, although the idea of outlier is quite natural and intuitive, there is no hard-and-fast rule to determine what counts as an outlier, and what does not.

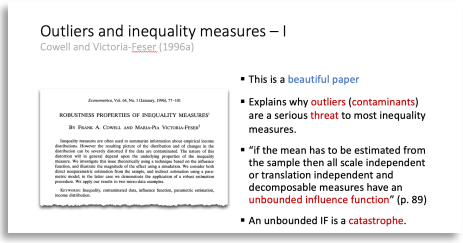




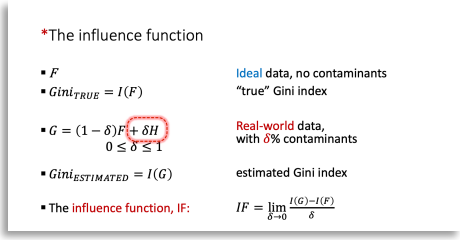
Outliers may be the result of various types of error, but they could also be **genuine** values. This is an important message to deliver: outlier is *not* synonym with 'mistake'. It is not unheard of, for variables like, say, consumption or income, to have extreme values: very rich and very poor people do exist, and they may be the very focus of our analysis. Therefore, data producers and data analysts face a dilemma: assessing whether outliers are genuine values, and, if they are not, understanding their impact on the statistics of interest.

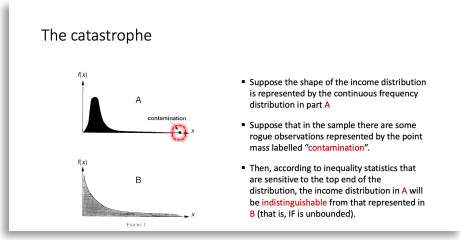


...

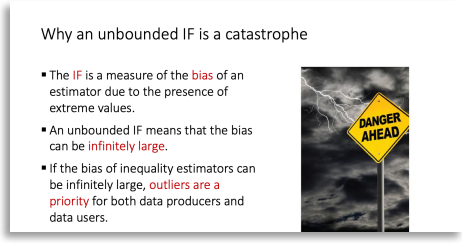


Why should one care about outliers? The next batch of slides presents some evidence demostrating that outliers do matter in practice, in terms of their **impact on outcomes** of interest for data users. We summarize the results from at least two, at most three papers (Cowell and Flachaire 2007 may be skipped, as it is quite advanced). First, Cowell and Victoria-Feser (1996): the paper is technical in nature, but its main conclusions are easy to grasp intuitively, and important to understand. The quote from page 82 of the paper is in a language that is inaccessible to most readers, but its meaning will soon be clear.

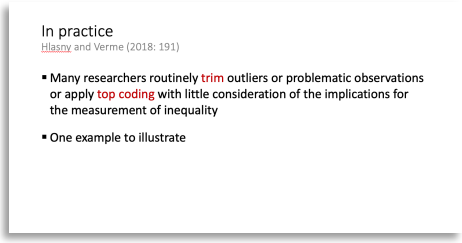


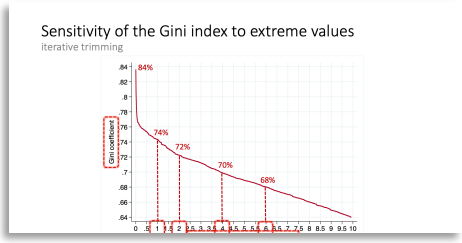


These optional slides elaborate on the concept of **influence function** (IF). The advice to the Instructor is to get into these slides only if she feels comfortable with the contents of the paper. If that is the case, the Instructor can explain that the IF is a measure of the difference between the 'true' statistic of interest – in this case, the Gini coefficient – and the same statistic, computed using contaminated data (*i.e.* data containing observations that should not be part of the datasets, but that are present by mistake – they are called 'contaminants' in statistics). A **contamination**, that is, one or more 'wrong' observations that are far away from most other observations in the distribution, will make the distribution appear much more skewed than it actually is, according to summary statistics and inequality indicators. An unbounded IF means that no matter how few the contaminants (or, in our case, outliers), their impact on the Gini can be infinitely large – that is, the difference between the 'true' Gini (what we would observe in the absence of outliers) and the estimated Gini (what we estimate in the presence of outliers) can be very large, infinitely large. This is why an unbounded IF is a problem.

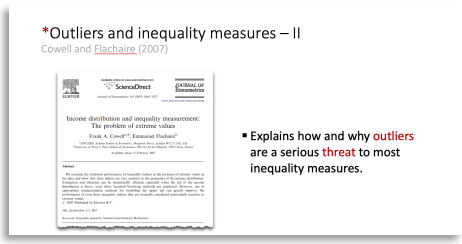


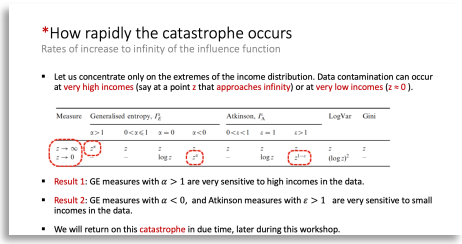
Even if the starred slides have *not* been shown, the implications of Cowell and Victoria-Feser's findings can still be made intuitively clear, by way of this non-technical slide.



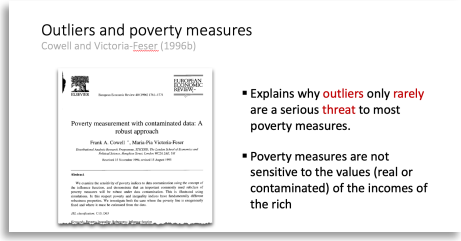


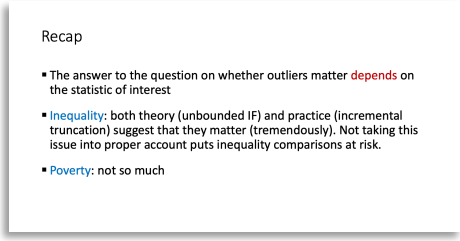
A practical **example** can help solidify the idea of the sensitivity of inequality measures to outliers. 'Trimming' extreme values in a distribution is common practice: this just means that a researcher may drop a few of the largest observations in the sample – the top 1% households with the highest expenditure, for instance. This is similar to 'contamination': the distribution is a certain shape before trimming, and changes after. The graph shows how the estimated Gini coefficient varies when the largest observation in the distribution of per capita consumption is dropped; then the second largest; then the third largest; and so on. The horizontal axis keeps track of how many observations are dropped, as a percentage of the total sample size. As we can see, dropping just a few observations has a huge impact on Gini. There is no limit, in fact, to how large that initial drop may be, regardless of how few observations we are modifying.



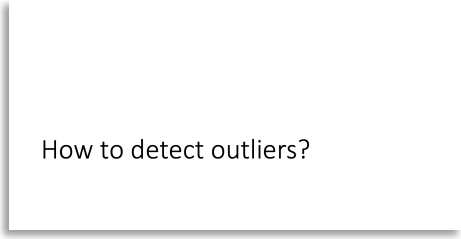


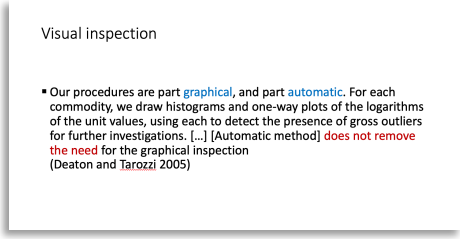
These optional slides elaborate on the **impact of outliers on inequality measures**. The Instructor should get familiar with the paper, in case this part is covered in class. The main takeaway is that different statistics of interest (in this case, inequality indicators) show different reactions to outliers.



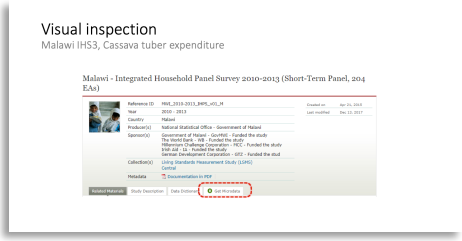


To conclude this discussion, we mention another paper by Cowell and Victoria-Feser, which investigates outliers and poverty measures, and reaches a rather opposite conclusion with respect to the previous one: **poverty** measures are not sensitive to extreme values.

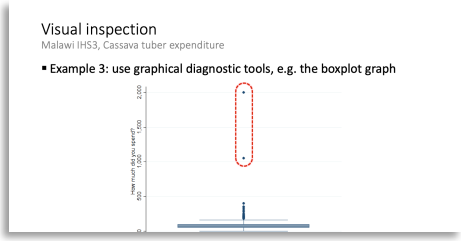




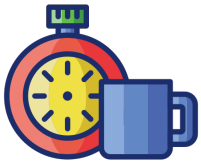
Assuming now that we do care about outliers, as they gravely threaten some of the statistics of interest to the welfare analyst, how should we **detect** them? The question is not trivial, given that there is no single definition of outlier that is valid in every context. Graphical inspection or authomatic methods usually work in tandem to detect outliers.



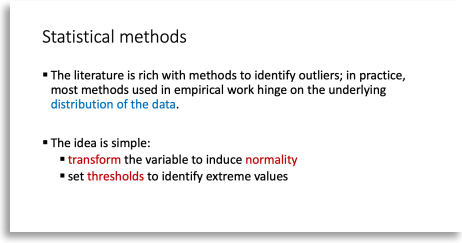
...



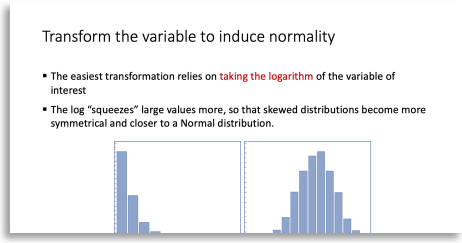
A practical example is used to illustrate **visual inspection** of the distribution of expenditure. We use publicly available microdata from Malawi. Descriptive statistics, graphs of the density of the variable, and box-plot graphs are all tools that help detect any extreme values. In this case, there are two very large, clearly anomalous observations in the expenditure distribution under examination. These methods are effective in pinpointing them – but there are other methods available.



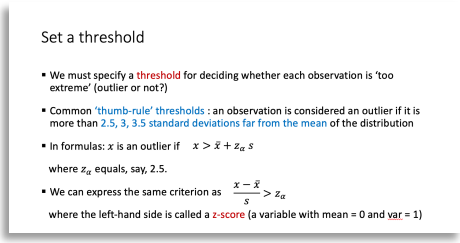
Now is a good time for a break.

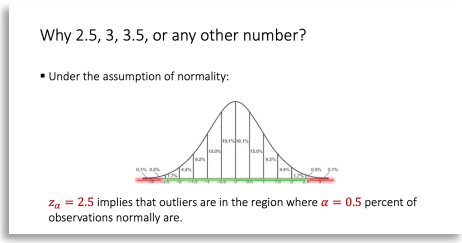


These are **statistical (automatic) outlier detection techniques**. The idea behind a number of popular outlier detection methods is a simple two-step strategy: (i) transform the consumption aggregate variable to induce *normality*, and (ii) set *thresholds* to identify extreme values. The main benefit of transforming target distributions into approximately normal (Gaussian) ones (step 1), is the ability to work with a well-known, simple distribution, that is computationally straightforward and familiar to virtually all analysts. More specifically, the exact value of the probabilities underlying a normal distribution at different intervals is known. An important consequence of this is that, if we are able to successfully transform the unknown, misshapen, mis-behaved distribution of our variable of interest – whose outliers we want to detect – into a well-known and well-behaved normal, the task of setting specific thresholds beyond which observations are considered 'extreme' (step 2) becomes simple, and, perhaps more importantly, easily replicable, whatever the initial distribution. Given that areas underlying the tails of a Normal distribution are known, analysts have a simple decision to make: where in the 'tails' of the Normal is an observation become extreme (i.e. unlikely) enough, that we decide to flag it as an outlier? The next slides elaborate.

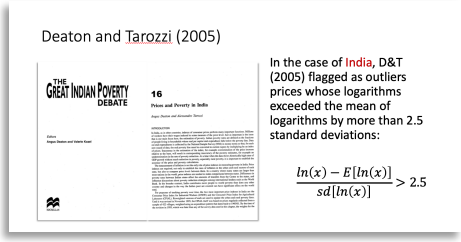


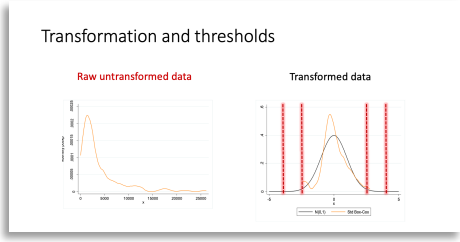
The easiest way to transform the variable of interest into something close to a normal distribution is to take the **logarithm** of the variable, which, as illustrated, changes the shape of a skewed distribution, making it closer to a Normal (in most cases): the logarithm "squeezes" large values more than small ones, so that the logarithm of a skewed distribution becomes more symmetrical.



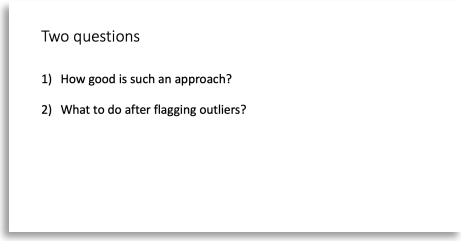


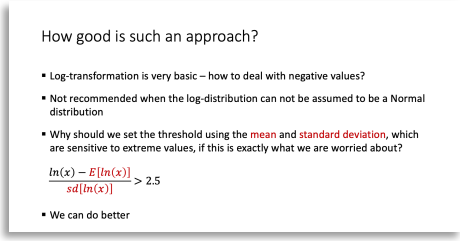
Next, it is convenient to introduce **z-scores**. By taking a variable and subtracting its mean, we center (translate/shift) the distribution of the variable at zero, and by dividing by the standard deviation, we rescale the variable so that its variance equals one. This amounts to using a standard Normal instead of a Normal, of course. The final step is to set a threshold, after which observations are to be considered outliers. The general rule is: observation x is an outlier if $x > \bar{x} + z_\alpha\, s$, where $\bar{x}$ is the sample mean of the variable, s is its standard deviation, and z_α is a conventional value, say 2.5. The conventional values of z_α are well known in the context of the Normal distribution: because the Normal is well known, we know that each of these values corresponds to a certain probability left in the tails.



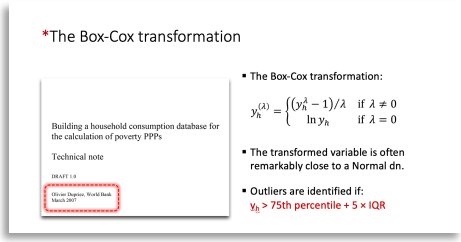


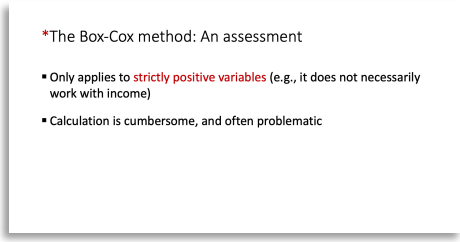
A notable **application** of this simple procedure – normalizing the variable of interest by taking the log, calculating a z-score, and applying a conventional threshold to define an 'outlier detection region – is in Deaton and Tarozzi (2005). The graph with the orange curves provides a graphical representation of the method. This is simply an example, that should help students consolidate what they have just seen in formulas.



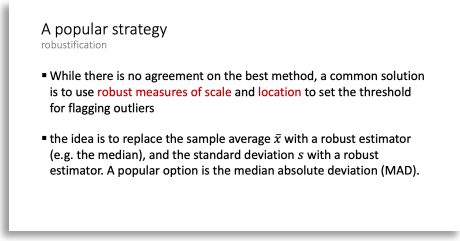


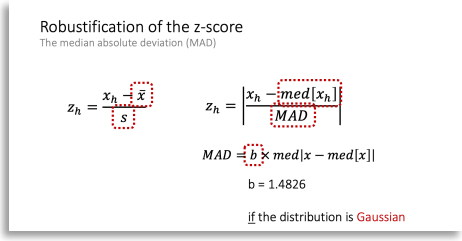
The approach can, however, be critiqued: it cannot deal with negative values (because the logarithm only exists for positive numbers), it does not apply to variables that are not normal after the transformation, and – perhaps most importantly – the thresholds that should help to identify outliers are *themselves* sensitive to outliers. As is well known, the sample mean is very much affected by extreme values, and so is the standard deviation. All in all, **we can do better** than the simple 'take the log and run' method illustrated so far. How so?



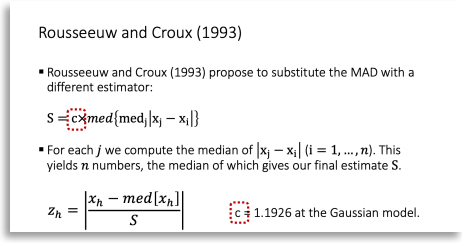


These optional slides hint at one of these alternatives: it relies on a different 'normalization' of the target variable, more complex and more flexible than the logarithm, called the **Box-Cox transformation**. The definition of the threshold is also different, as it relies on the **interquartile range (IQR)** instead of the standard deviation, and on a given percentile, for instance, the 75th, instead of the mean. These statistics are more robust to the presence of extreme values. The main downside of the method is that it is, ultimately, hard to implement, as the Box-Cox transformation cannot always be calculated. Our suggestion is to skip these two slides, unless there are specific questions on methods better than the log to induce normality in the distribution of a variable.





One solution is to **'robustify' the z-score**. Instead of using a z-score based on the sample mean and the standard deviation, we replace them by plugging in robust estimates of scale and location. Instead of the mean, we use the **median**; instead of the standard deviation, we use the **median absolute deviation (MAD)**. Note that the MAD is defined as the median of all distances between each value of x and the median, times the parameter b (this is just a constant that is required for MAD to be a consistent estimator of the standard deviation under the hypothesis that the distribution is Normal).



But even this approach can be improved. Rousseeuw and Croux (1993) discuss the pros and cons of the MAD at some length, and end up suggesting that more efficient alternatives exist, especially in the presence of skewed distributions, which is exactly the case of interest here (this is a technical paper, and it is not essential for Instructors or students to master it: the fundamental intuition is enough for the purposes of the lecture). Following this research, we suggest the following, alternative robustification of the z-score: at the numerator of the score, continue using the median, in place of the mean, but at the denominator, switch to the **'S-estimator'** instead of MAD. To calculate the S estimator, we simply need to compute, for each household i in the sample, the expression $\{med_j|x_i - x_j|\}$ for $j=1, \ldots, n$. This gives n numbers, the median of which gives the estimate S (the number 1.1926 in the formula is required for making S a consistent estimator of the standard deviation under the assumption of normality). These are details that are *not important* for students to be confident with – it is enough to explain why those parameter are there. Statistical softwares take care of these details.

Recap

- "take the log and run" is not a recommended practice
- taking the log and robustifying the z-score is a better practice
- Belotti and Vecchi (2019) provide `outdetect.ado`

Take the log and run vs. robust z-scores

Countries	Year	Outliers (%)					
cutoff = 3		log-transformation			best normalization		
		overall	left	right	overall	left	right
		(1)	(2)	(3)	(4)	(5)	(6)
Malawi	2017	0.75	0.14	0.61	0.30	0.22	0.08
Nigeria	2012	1.35	0.11	1.24	0.72	0.32	0.40
India	2012	1.39	0.03	1.36	0.62	0.13	0.49
Pakistan	2014	1.58	0.02	1.56	0.39	0.21	0.18
Guatemala	2014	1.14	0.06	1.08	0.61	0.15	0.46
Peru	2015	0.36	0.09	0.27	0.28	0.16	0.12
Armenia	2013	0.91	0.08	0.83	0.68	0.17	0.51
Georgia	2015	0.75	0.25	0.50	0.73	0.32	0.41

To conclude this topic, we can say that the approach of simply taking the logarithm of the target variable and then using z-scores to detect outliers is not the best available method. Using other statistics to robustify the z-score gives better results in practice. Data from selected countries are used to show the output of the **outdetect.ado** Stata command (Belotti et al. 2020), which performs the Rousseeuw and Croux (1993) robustification.

How to deal with outliers?

(in one slide)

Treatment of outliers

Three main methods of dealing with outliers, apart from removing them from the dataset:

1) reducing the weights of outliers (trimming weight)
2) changing the values of outliers (Winsorisation, trimming, imputation)
3) using robust estimation techniques (M-estimation).

- Documentation, transparency & reproducibility

The final topic for the lecture is **outlier treatment**, which is discussed only briefly. It bears repeating that checking for "gross mistakes" and other context-specific sources of measurement error is a necessary first step when dealing with extreme values – indeed with any data points that data producers deem "implausible". Given the impact of outliers on inequality estimates, however, taking *some* action, once these routine checks are completed and if outliers are still present, is recommended. To explain, the Instructor can recall a prescription from Huber's classical manual on Robust Statistics (Huber 1981: 4): "Any reasonable, formal or informal, procedure for rejecting outliers will prevent the worst.". Provided choices are motivated and documented by the analyst, such procedures may fall into the broad categories recalled here.

Lessons learned

- Outliers can be genuine observations… be gentle to the data and document each and every step of the data processing
- As far as inequality is concerned, outliers are the worst enemy (unbounded IF)
- Outlier detection:
  - go beyond the "take the log and run" strategy. It works well only if you can describe the data with a Gaussian distribution. Typically, however, distributions are skewed.
  - Use a "take the log, robustify the z-score and run", strategy.
- Outlier treatment: it depends. Quantile regression is a good candidate.

As usual, it is good practice to take some time to end the lecture, and summarize the main takeaways. The first point should emerge strongly: documentation of each of the steps discussed during the lecture is key for any type of data cleaning or analysis.

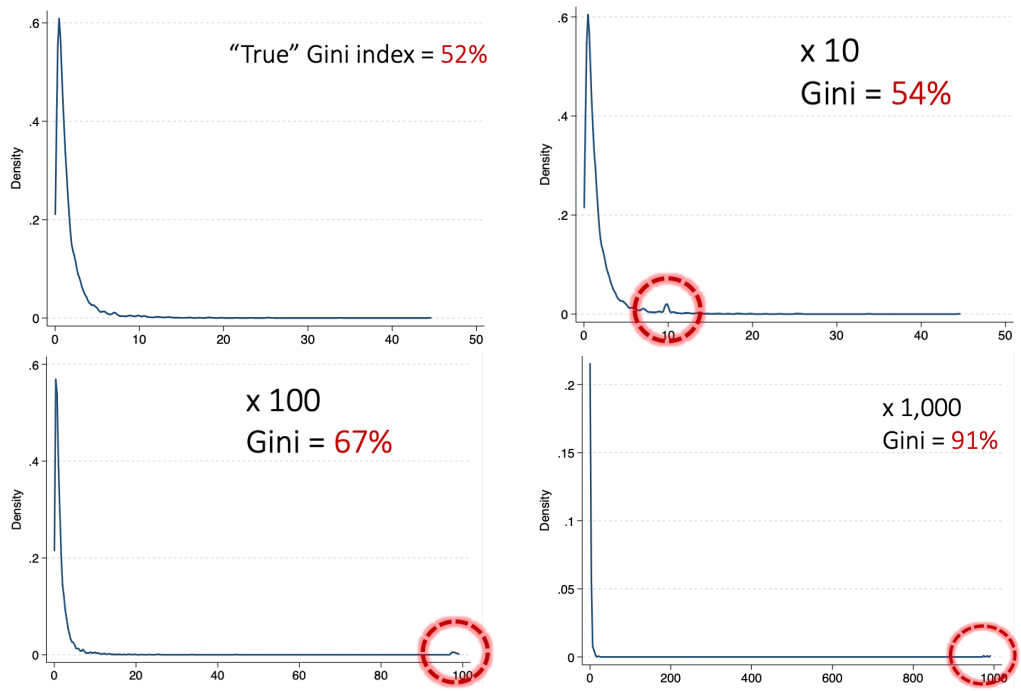
Homework

Exercise 1 - Engaging with the literature

See exercise 1, Lecture 1.

Exercise 2 – Do-it-yourself…

This exercise is meant to give students an opportunity to do a hands-on experiment on the impact of outliers on a statistic of interest (in this case, the Gini index), and to get a sense of how these extreme values may arise in practice (wrong placement of a decimal separator is a common data entry mistake). The code pictured in the slide is written for Stata, but the same operations can be performed with any statistical software. Results should look similar (not necessarily identical) to those pictured below:



Exercise 3 – Inequality measures

The table referenced by the exercise is an example of how strong the impact of the method chosen to treat outliers can be on final estimates. It should be stressed that none of the columns in the table report the 'correct' estimates, necessarily: they are simply the results of different ways to deal with extreme values. The importance of documenting outlier detection methods should also be emphasized.

Lecture 13
Measuring inequality

Learning objectives

The goal of this lecture is to provide an overview of both concepts and methods underlying inequality measurement. The task of estimating inequality is greatly facilitated by the availability of suitable data – thus the lectures provides the theoretical framework required for data producers and users to share a common view of what is meant by 'inequality' and how to measure it.

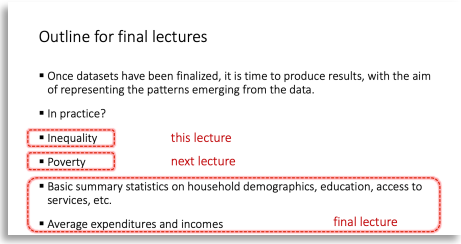
Suggested preparation

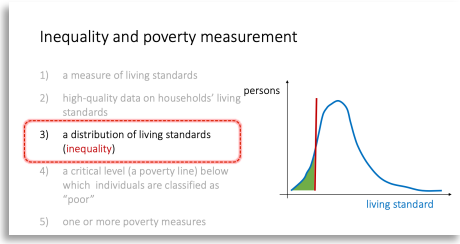
The lecture follows chapters 1 and 2 of Cowell (2011). The Instructor is assumed to be familiar with this material.

Time allocation

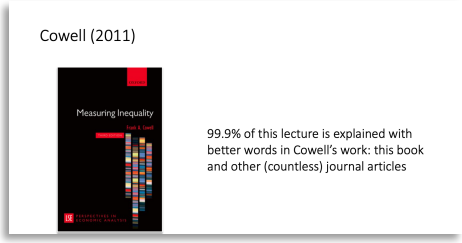
Tables and graphs	45 min
Break	**15 min**
Indicators	55 min
Lessons learned	5 min

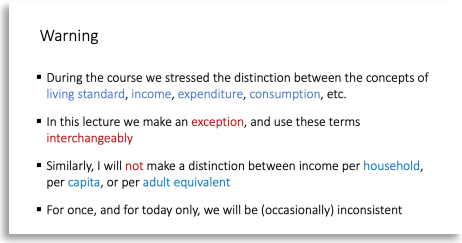
Annotated lecture



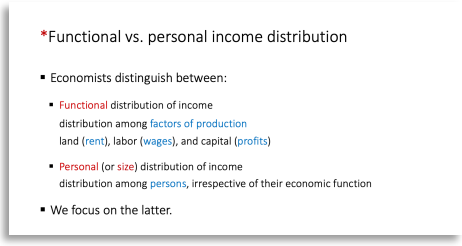


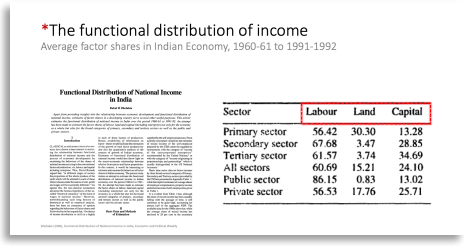
The lecture opens with two introductory slides, where we provide a short recap of where we stand, and discuss the next steps. The main message is that today's class is the first of a twin set of *theoretical* lectures on **inequality** (today) and **poverty** (tomorrow) **measurement**.





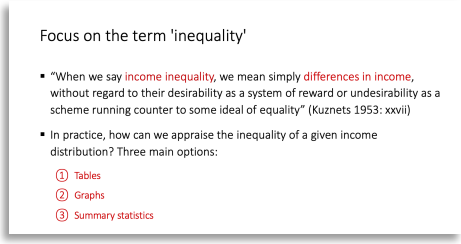
This lecture is modelled after the reference cited: students interested in inequality measurement should certainly aim to get familiar with Cowell's book. We also share a disclaimer about **terminology**: in contrast to the rest of the course, this lecture will move beyond the construction of a welfare indicator, into analysis of its distribution. For this reason, we will not be as specific or precise about what the welfare indicator actually *is*: our attention here will be on inequality *measures*, not on data, per se. Data, however, still continue to be important, of course: learning about analysis will help students be aware of how data are ultimately put to use.



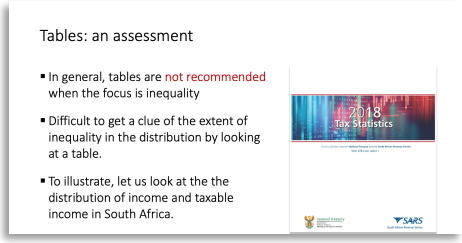


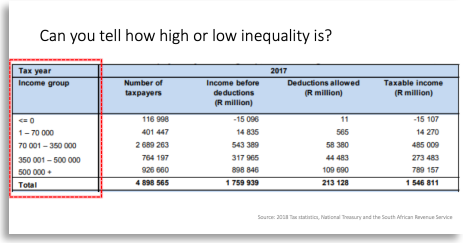
Sector	Labour	Land	Capital
Primary sector	56.42	30.30	13.28
Secondary sector	67.68	3.47	28.85
Tertiary sector	61.57	3.74	34.69
All sectors	60.69	15.21	24.10
Public sector	86.15	0.83	13.02
Private sector	56.53	17.76	25.71

These slides are optional, and the Instructor is encouraged to skip them, unless the audience demonstrates a specific interest on the difference between **functional and personal inequality**. Our focus in this class is on the latter: functional inequality is typically more of interest to macroeconomists than to welfare economists.



Next, we spend a few words on **what we mean by 'inequality'**. The point here is to clarify that in this lecture we aim at defining and measuring inequality from a purely statistical standpoint, that is, we do *not* discuss a number of interesting issues/questions related to fairness and equity – just 'differences in income' as in the Kuznets' quote. The bottom part of the slide is the outline for the rest of the lecture. The plan is to answer the question: what is the **best way of describing (measuring) inequality**? Using tables? Graphs? Indicators? Each option will be examined and evaluated, so that ultimately, we will be able to motivate the use of well-known inequality measures. The spoiler – the Instructor may share it in advance – is that tables are *not* a good idea, the Lorenz curve 'wins' among graphical tools, and Gini and GEI 'win' among inequality measures.

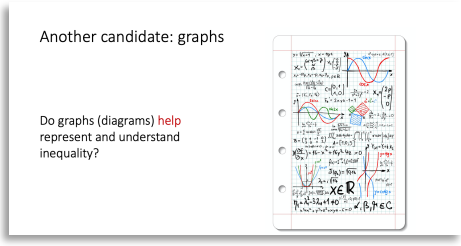




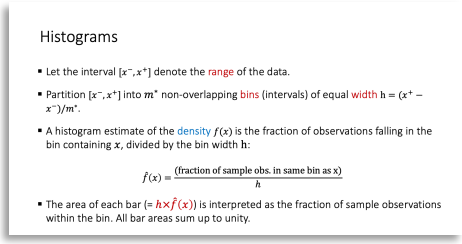
Tax year	2017			
Income group	Number of taxpayers	Income before deductions (R million)	Deductions allowed (R million)	Taxable income (R million)
<= 0	116 998	-15 096	11	-15 107
1 – 70 000	401 447	14 835	565	14 270
70 001 – 350 000	2 689 263	543 389	58 380	485 009
350 001 – 500 000	764 197	317 965	44 483	273 483
500 000 +	926 660	898 846	109 690	789 157
Total	4 898 565	1 759 939	213 128	1 546 811

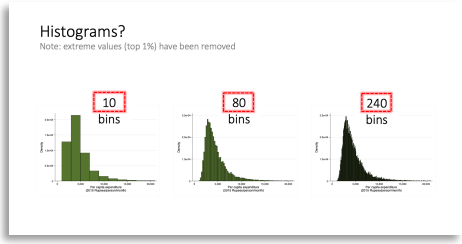
Source: 2018 Tax statistics, National Treasury and the South African Revenue Service

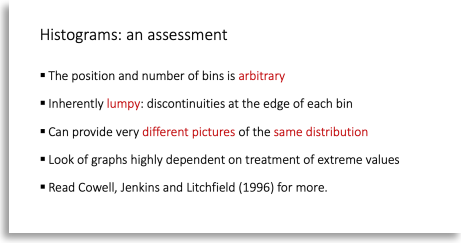
Let us quickly evaluate our first candidate: **tables**. The Instructor may show the example from South Africa to the audience, and ask the question: is this table effective in conveying the extent of inequality? We *briefly* engage with students (a couple of minutes are enough) and listen to their opinions. After weighing pros and cons, we argue that tables are *not* the best way to summarize inequality.



If tables are not good enough to summarize inequality, what about **graphs**? Again, we briefly engage with students. Which graphs would they use? Why? We listen, we do not reply and/or comment on each question. Then we take students through the next slides, where we assess a selection of popular graphs. We start with the **histogram**.

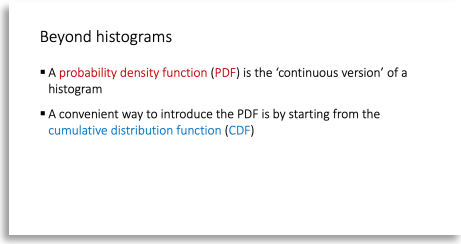




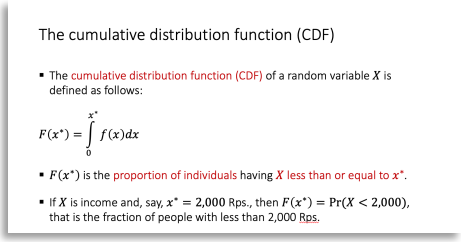


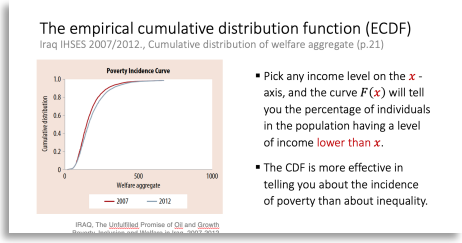
The aim of first slide of this bunch is to check with the students whether they understand histograms as we think they should. Would they be able to explain how to construct a histogram from scratch? Faced with a histogram, would they be able to answer technical question (e.g. how to interpret the height of a bar, how many bins to choose, etc.)? The message: take this opportunity to refresh your knowledge of these basic tools. Only by understanding *the details* can they answer the question of which is the 'best' histogram out

of the three shown in the slides. The point that the Instructor should make here is that the number of bins, which greatly influences the appearance of the graph, and therefore our perception of the distribution and of inequality, is ultimately arbitrary. In addition to that, the note says that "the top 1% observations have been removed": this is common when drawing graphs of highly skewed distributions, given that plotting the top values often makes the graph impossible to read. But in our context, this robs us of crucial information: top values are precisely the interesting ones – how can we gauge inequality without them? After discussing pros and cons of histograms we conclude that, despite their popularity, they are certainly not the optimal way of summarizing inequality.

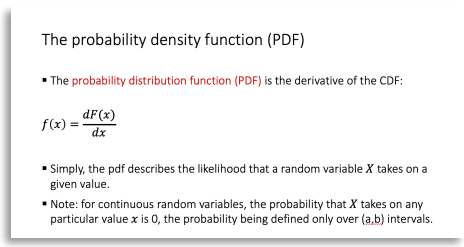


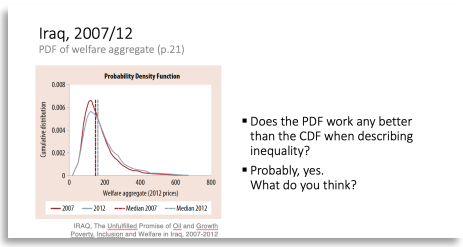
The next candidates are the **probability density function (PDF)**, and the **cumulative distribution function (CDF)**. If these statistical tools are already well-known to students, as they likely are, there is no need to cover them in great detail; if not, the following slides can be used to refresh their understanding of some of the most fundamental tools in statistics. Familiarity with these concepts is key when working with inequality measures,



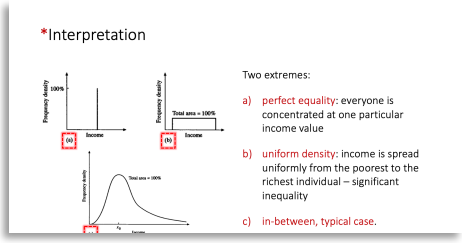


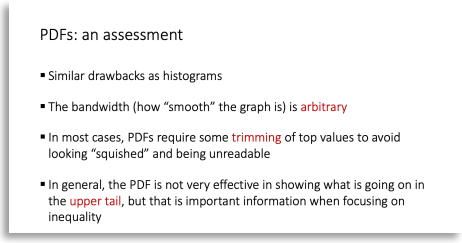
We follow the same logic as we did for histograms, and define the **CDF** first; then, we show an example from a recent report, and wonder whether the graph provides useful information about inequality. The CDF returns the share of the population with incomes lower than any given value: a very handy piece of information indeed, when you are concerned about poverty. Plug the poverty line into the CDF, and you get the poverty headcount rate back. This is not really what we need when we care about inequality. After arguing this, we move on to the PDF.



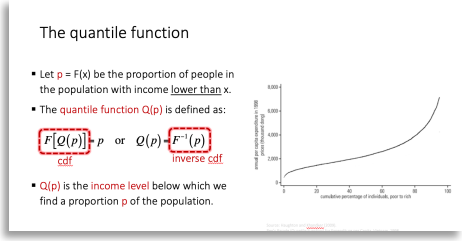


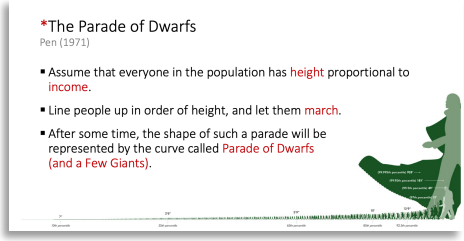
We introduce the **PDF** (two hidden slides give the Instructor the options to delve into the details, but should not be needed with students who have a background in statistics).



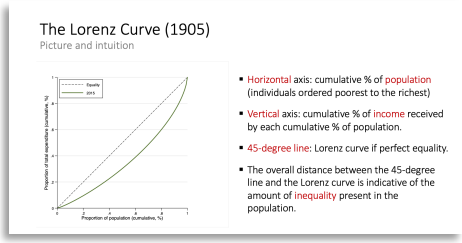


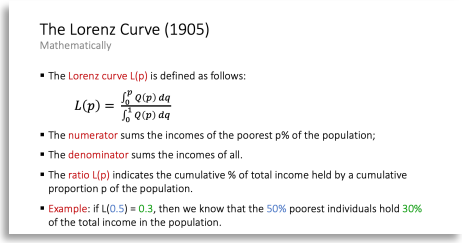
The starred slide titled "Interpretation" may help to explain the way one might approach the PDF as a graphical tool to evaluate inequality. The wo extreme scenarios (a and b) are clear enough; the problem is that in any practical situation, we find ourselves in c, and visually discriminating between more or less unequal PDFs is not easy. We conclude that the PDF ultimately suffers from the same limitations as the histogram.



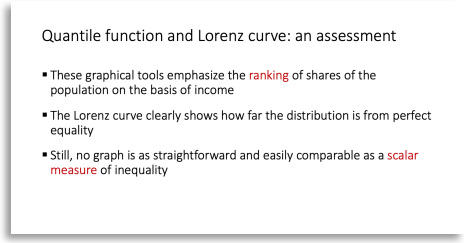


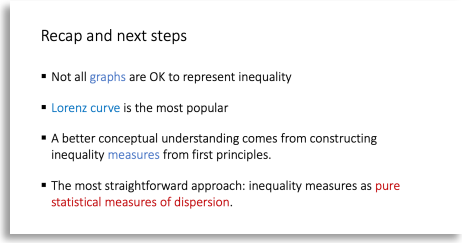
Next, we introduce the **quantile function**, the inverse function of the CDF. We concede that perhaps it is not totally ineffective at summarizing inequality, and if time permits, we explain why by means of the 'parade of dwarfs' (see the excellent explanation in Cowell 2011: the famous methaphor describes the quantile function as a parade of people, that are as tall as their incomes – the parade is made up by many dwarfs and a few giants). We conclude that ultimately, we hope to find better graphical tools. It is finally time to introduce *the* graph that – so we argue – succeeds in representing inequality: the Lorenz curve.



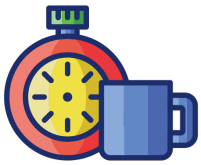


We discuss the **Lorenz curve** at some length, and conclude that this is a useful tool, one that helps welfare analysts tremendously. This must become part of every student's analytical toolbox.



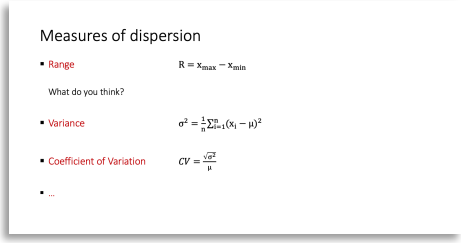


With the Lorenz Curve, we conclude the review of graphical tools for analysing inequality, and this is when we are likely to need a break.

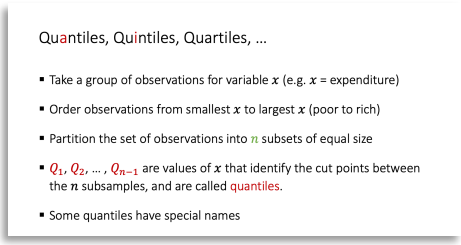


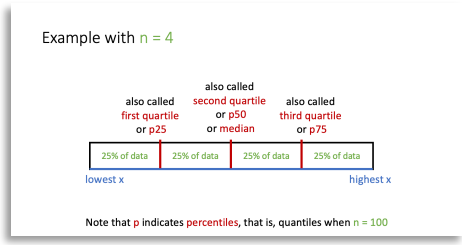
Now is a good time for a break.

The second part of the lecture is devoted to **inequality indicators**: scalar measures of inequality, that summarize our phenomenon of interest with a single number. Again, we follow the same logic: we evaluate and dismiss candidates, until we arrive at our 'winner'.

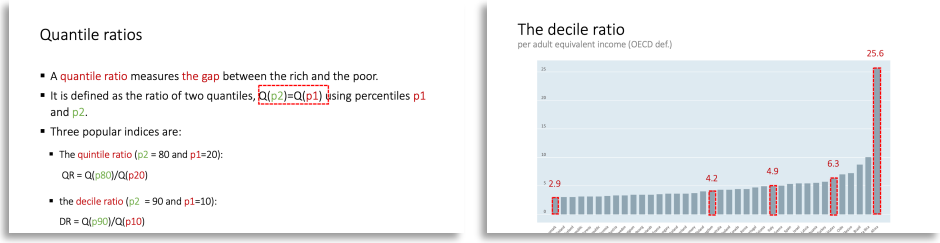


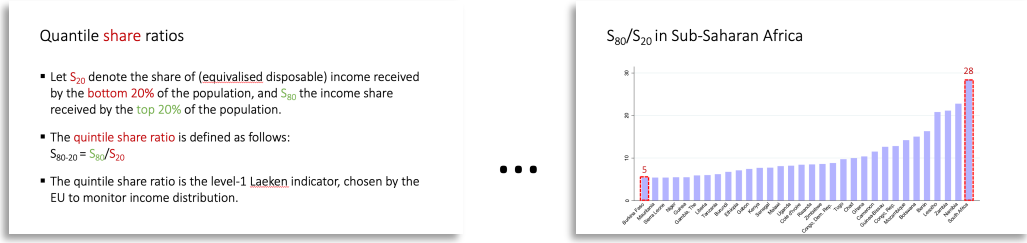
There is an obvious solution to the problem of measuring inequality using a single number: what about familiar **measures of dispersion**, like the variance? No need to spend much time discussing this: each measure has some advantages, and some (many more) disadvantages – the point is that we have many candidates, and we don't really know how to choose. We would like *one* measure that we can trust (or even a few of them), not a multitude of different ones that we are not very confident in. Ultimately, we can do much better.



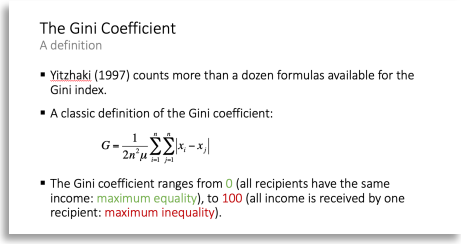


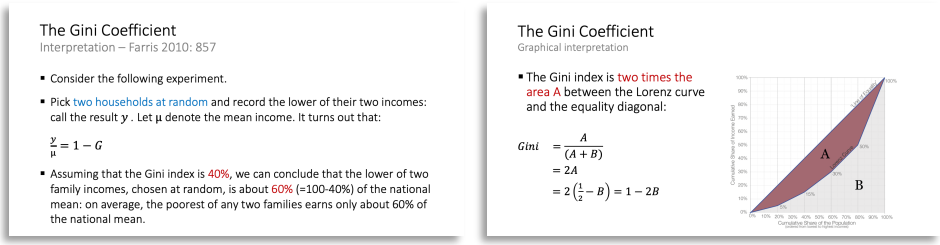
These slides review the definition of **quantiles**: again, this is likely to be well known to students, so it may be covered quickly. Quantiles are fundamental building blocks for some widely used inequality indicators.



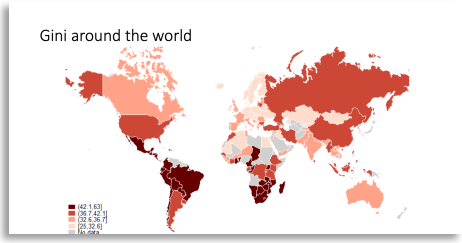


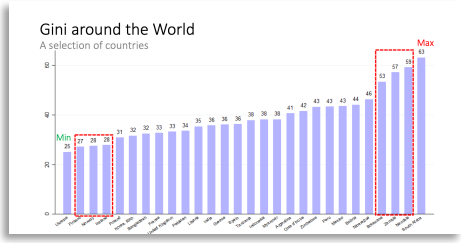
These are **quantile ratios** and **quantile share ratios**, which are defined and 'shown in action' with a few examples. The graphs help students form an idea of what is the expected range of each of these indicators. Both measures are commonly encountered in practice, but we move on to an even more useful one.



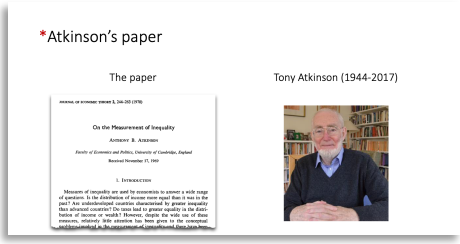


We are talking, of course, about the **Gini coefficient**, arguably the single most used inequality indicator. We give a mathematical definition, and then a few interpretations, among the many available (the one by Farris, about average income differences, and a graphical one based on the Lorenz curve).

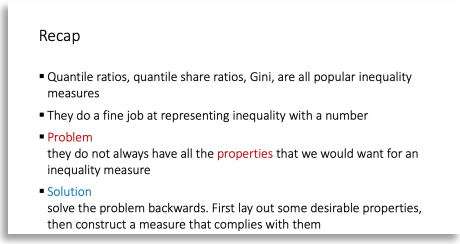




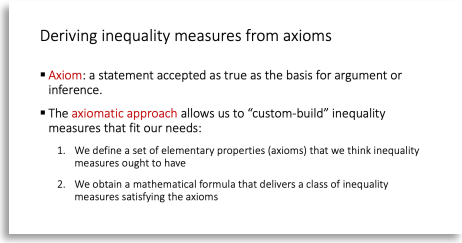
A few more graphs help once again with grasping the usual order of magnitude of the Gini index across countries in the world. We invite students to get as familiar with the Gini coefficient as much as possible, due to its wide practical application. We emphasize that, after the Lorenz curve, this is the second analytical tool that everybody should have in her toolbox.

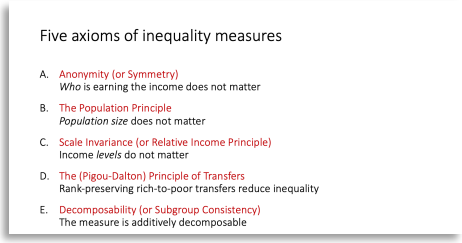


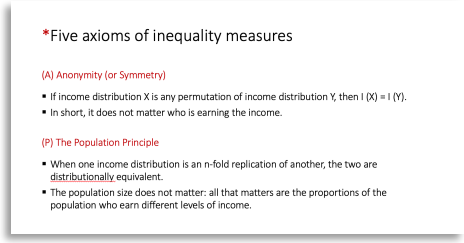
This optional slide delves deeper on the issue of **inequality comparisons**. The approach is to simply direct students to one of the foundational papers on the topic.

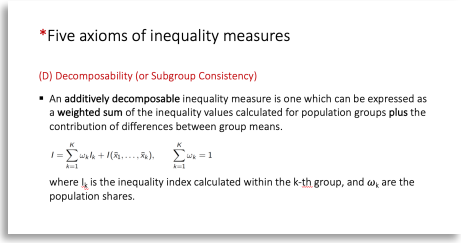


A short recap of the main takeaways leads us to the next and final topic of the lecture. We have told students that some measures of inequality are better or more popular than others: but why is that, exactly? One way to think about the choice of the 'best' indicator more rigorously is in terms of desirable properties of an inequality measure. As we might do when we buy a car or a phone, it seems logical to first identify some characteristics we care about (good reception, nice design, a certain amount of memory) and then see which phones possess these characteristics, and select them based on that. The same can be done with inequality measures: this logic is called the **axiomatic approach**.

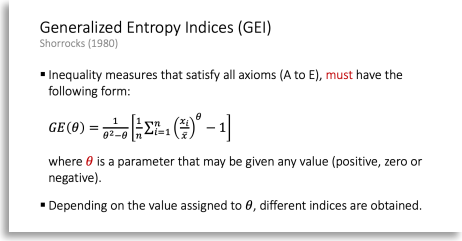


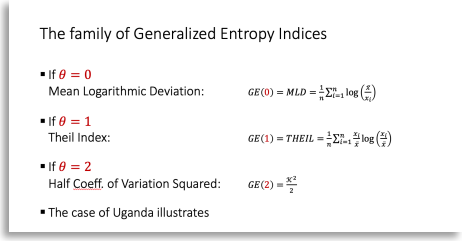


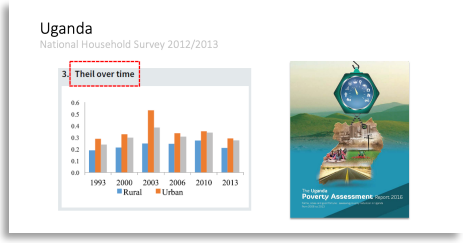




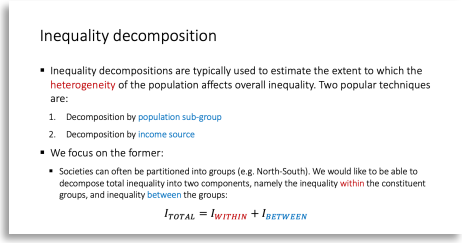
This approach has revolutionized the way inequality is measured, by laying out five axioms that a good inequality measure must comply with. The axioms are simply stated in one slide; a series of five optional slides goes into more detail as to what each axiom exactly implies. The instructor is advised to cover them only if time permits.

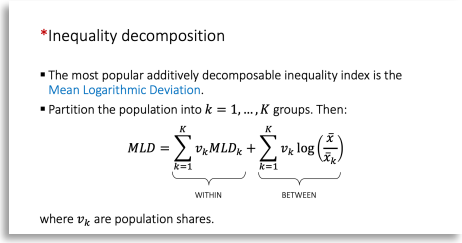


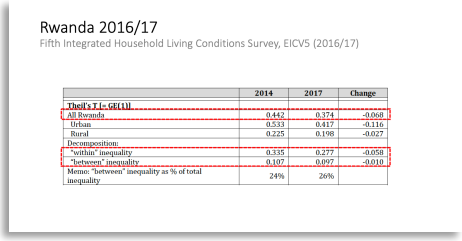




The axiomatic approach leads to the **Generalized Entropy class of Indices (GEI)**, which comply with all axioms. Gini and the GEI indices are the most commonly used by scholars the world over (note that Gini does not comply with all axioms, but it is so popular that welfare analysts are, in some way, 'stuck' with it: it is still a useful indicator, regardless). The instructor is advised to read Cowell's highly accessible account of the axiomatic approach, which underlies the slides. The take-away for students is: GEIs are important, and in particular the Mean Log Deviation (MLD).

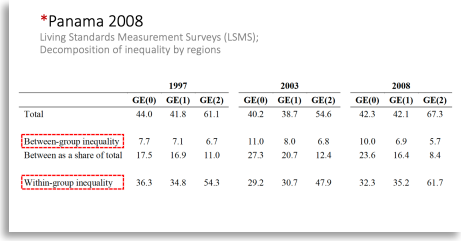






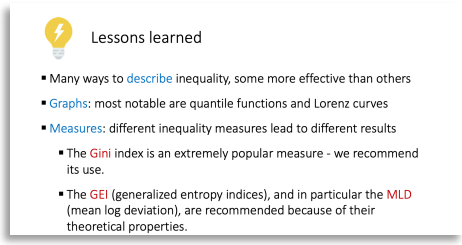
	2014	2017	Change
Theil's T [= GE(1)]			
All Rwanda	0.442	0.374	-0.068
Urban	0.533	0.417	-0.116
Rural	0.225	0.198	-0.027
Decomposition:			
"within" inequality	0.335	0.277	-0.058
"between" inequality	0.107	0.097	-0.010
Memo: "between" inequality as % of total inequality	24%	26%	

...



	1997			2003			2008		
	GE(0)	GE(1)	GE(2)	GE(0)	GE(1)	GE(2)	GE(0)	GE(1)	GE(2)
Total	44.0	41.8	61.1	40.2	38.7	54.6	42.3	42.1	67.3
Between-group inequality	7.7	7.1	6.7	11.0	8.0	6.8	10.0	6.9	5.7
Between as a share of total	17.5	16.9	11.0	27.3	20.7	12.4	23.6	16.4	8.4
Within-group inequality	36.3	34.8	54.3	29.2	30.7	47.9	32.3	35.2	61.7

If time permits, the Instructor can mention one final topic that may be of interest: the last inequality axiom has to do with decomposability of the measure, and these slides explain the benefits of such a property. Once inequality is estimated at the national level, if we care about population subgroups (such as urban and rural areas, or different regions) a decomposable measure can be additively decomposed into a within and a between component (this is explained non-technically in one slide, and a hidden, optional slide may be covered if the Instructor believes showing the mathematical formula for the decomposition can help). One or more example of inequality decomposition using an index of the GEI family can be shown (note that the Gini index is not additively decomposable).



Lessons learned can be customized, as required. Here we summarize a few key messages, namely that there are many ways to describe inequality, but only few of them 'work', that is only few have sound theoretical foundations and a wide practical application. In essence, our recommendation is to focus on the Lorenz curve (and possibly the quantile function), the Gini coefficient, and the Mean Log Deviation.

Homework

Exercise 1 - Engaging with the literature

See exercise 1, Lecture 1.

Exercise 2 – Inequality in South Asia

The main point here is that it is not a good idea to mix expenditure- and income-based inequality estimates.

Exercise 3 – Functional vs. Personal Income distribution

The exercise is for students interest in this specific topic (functionl vs. personal distribution), and as a way to check whether their understanding of the report has benefitted from attending the lecture and read Cowell's reference.

Lecture 14
Measuring poverty

Learning objectives

The goal of this lecture is to provide a conceptual framework for measuring poverty. Given the dimension and the technical contents of the literature, we limit the discussion to a short review of how analysts define a poverty line, and how they measure the incidence, depth and severity of poverty.

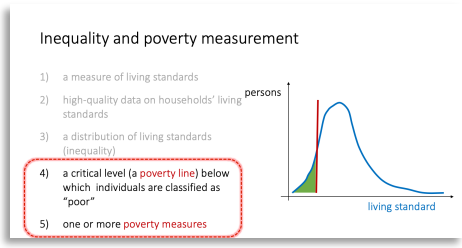
Suggested preparation

The lecture follows the sections indicated in the syllabus from Ravallion (2016). The Instructor might find it useful to become familiar with the material in Chapters 2-4 in Haughton and Khandker (2009).

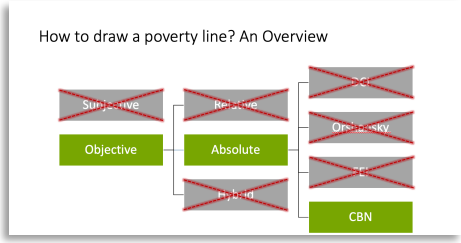
Time allocation

Poverty lines	55 min
Break	**15 min**
Poverty measures	45 min
Lessons learned	5 min

Annotated lecture

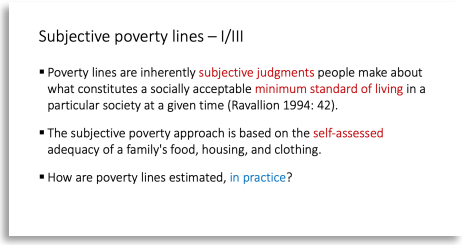


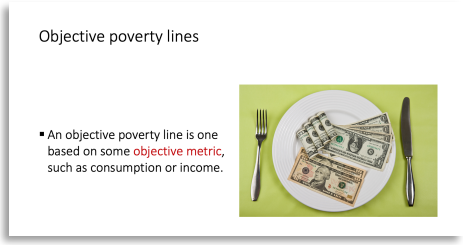
The lecture begins, as usual, with the outline of the topics that will be covered. Today is a simple two-part structure, where we explain poverty lines in part 1, and poverty measures in part 2.



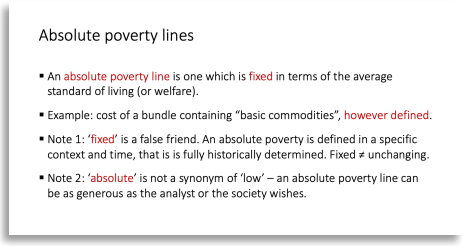
The opening slide argues that there is a **variety of approaches** that have been proposed to 'draw the line': we have chosen one, out of the many possible, which is called the Cost-of-basic-needs (CBN) method.

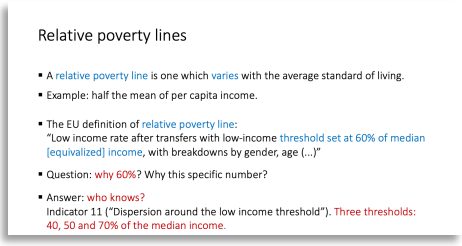
The illustration of the CBN method will take place after dismissing all other methods. We share our strategy with the audience. We announce from the onset, for instance, that we will define and explain what is meant by a 'subjective' poverty line and an 'objective' one, and explain that we will criticize and dismiss the former and focus on the latter. Next, we define and explain what is meant by 'relative', 'absolute' and 'hybrid' poverty lines, and explain that we after discussing pros and cons associated to each option we will focus on absolute poverty lines. And so on, until we reach the 'CBN'.





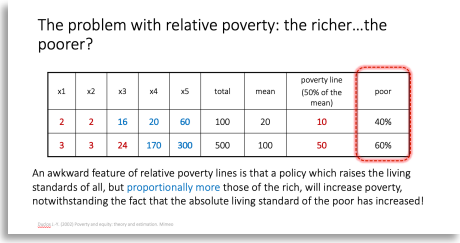
The first set of slides is on the **subjective vs. objective** poverty debate. The discussion is only supposed the share the *ideas* underlying these methods, to conclude that there is revival in the interest of the subjective approach, but currently, it is not the one implemented in national studies of poverty in most countries in the world. We leave it to interested students to read more on the topic.





The discussion on **absolute vs. relative poverty** is an important one. First, we define both concepts. The meaning of 'absolute' in this context should be made very clear: it does not mean that it is fixed and immovable. On the contrary, an absolute poverty line may change over time, if the country decides to re-calculate it in order to make it consistent with the consumption patterns that are currently prevalent in the country. 'Absolute' also does not mean 'extreme': on the contrary, it might indicate a level of consumption that is well above

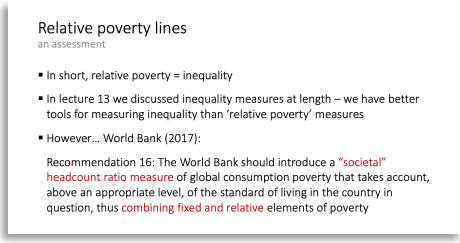
starving. The deeper meaning of these caveats will be clearer when the absolute approach is contrasted with the relative one.



The problem with relative poverty: the richer…the poorer?

x1	x2	x3	x4	x5	total	mean	poverty line (50% of the mean)	poor
2	2	16	20	60	100	20	10	40%
3	3	24	170	300	500	100	50	60%

An awkward feature of relative poverty lines is that a policy which raises the living standards of all, but proportionally more those of the rich, will increase poverty, notwithstanding the fact that the absolute living standard of the poor has increased!



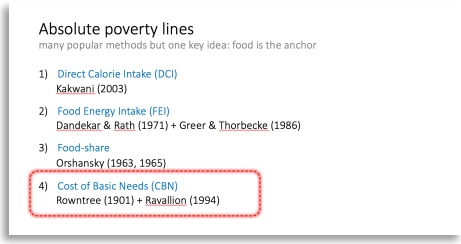
Relative poverty lines

an assessment

- In short, relative poverty = inequality
- In lecture 13 we discussed inequality measures at length – we have better tools for measuring inequality than 'relative poverty' measures
- However… World Bank (2017):

Recommendation 16: The World Bank should introduce a "societal" headcount ratio measure of global consumption poverty that takes account, above an appropriate level, of the standard of living in the country in question, thus combining fixed and relative elements of poverty

The reason why we ultimately dismiss the relative poverty approach is explained by means of a numerical 'toy' example, which leads us to a paradoxical conclusion regarding the level of poverty in two countries, or two moments in time (the two rows of the table). This should illuminate the fundamental defects of relative poverty. The key message in these slides is that relative poverty is – de facto – a measure of inequality. This message can be softened and nuanced, but this is the first bring-home conclusion from this set of slides.

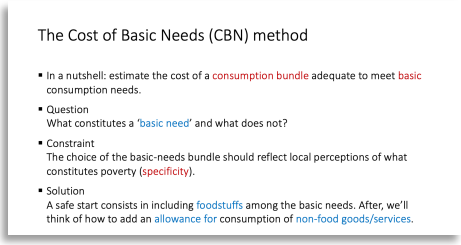


Absolute poverty lines

many popular methods but one key idea: food is the anchor

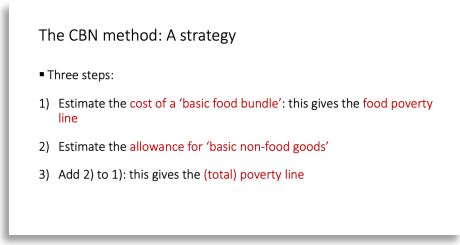
1) Direct Calorie Intake (DCI)
   Kakwani (2003)
2) Food Energy Intake (FEI)
   Dandekar & Rath (1971) + Greer & Thorbecke (1986)
3) Food-share
   Orshansky (1963, 1965)
4) Cost of Basic Needs (CBN)
   Rowntree (1901) + Ravallion (1994)

Now that our focus on objective, absolute poverty has been motivated, we focus on the **CBN approach** to setting a poverty line. Our account of the CBN method is masterly explained in Martin Ravallion's work, from his 1994 book (check the references) all the way to his recent 2016 book. There is no need to repeat all details here. The instructor is encouraged to customize these slides the way she/he thinks best for the audience. In our experience, the most effective way is first to explain the idea in words, next to use the support of one or more graphs, and finally to double check the understanding of the audience by illustrating the formulae (we provide the non-parametric version of the method, but the Instructor might wish to replace it with its parametric, regression-based counterpart. This is what students find in the Appendix of Ravallion 1994). This is the structure that the Instructor will find in the following slides. In the next pages of this Manual, we give a quick description of the main steps of our argument.



The Cost of Basic Needs (CBN) method

- In a nutshell: estimate the cost of a consumption bundle adequate to meet basic consumption needs.
- Question
  What constitutes a 'basic need' and what does not?
- Constraint
  The choice of the basic-needs bundle should reflect local perceptions of what constitutes poverty (specificity).
- Solution
  A safe start consists in including foodstuffs among the basic needs. After, we'll think of how to add an allowance for consumption of non-food goods/services.



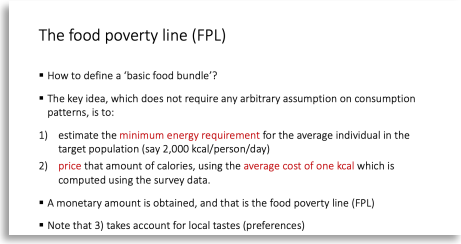
The CBN method: A strategy

- Three steps:

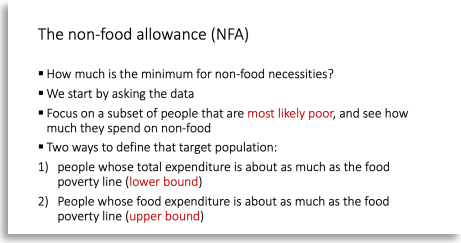
1) Estimate the cost of a 'basic food bundle': this gives the food poverty line
2) Estimate the allowance for 'basic non-food goods'
3) Add 2) to 1): this gives the (total) poverty line

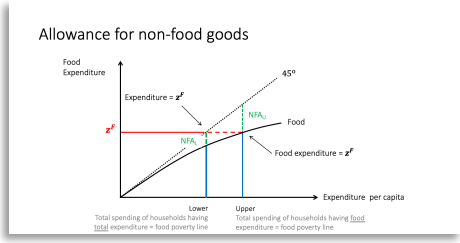
The basic premise of the CBN method is rather intuitive. The goal is to estimate the **cost of a basic consumption bundle**: this will give us a threshold, beyond which individuals

are considered to be poor, i.e. not able to satisfy basic needs. We start from the cost of a basic food bundle, and then we add the cost of any non-food items that are considered 'basic' on top of it. The whole essence of the CBN method is in how these two problems are solved.

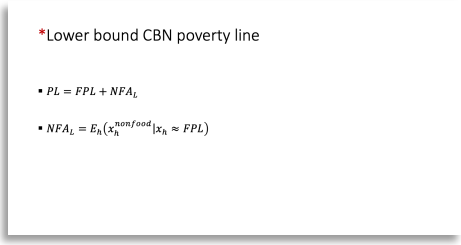


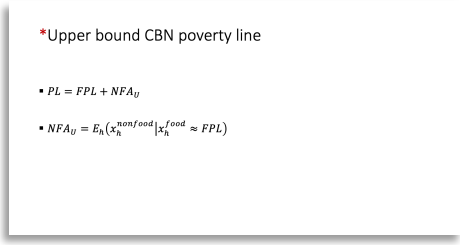
Determining the **food poverty line** – the cost of a basic food bundle – is perhaps the easiest of the two steps. The most 'neutral' way of doing it (that is, a way that does not require any assumptions on which food items are essential and which are not) relies on calories. First, we estimate the minimum caloric requirement for the average individual in the population (to do that, we rely on parameters that are disseminated by medical authorities, and on the demographic structure of the population). Then, we estimate the average cost of one calorie in our sample of household-level data (usually, by restricting the sample to the bottom part of the expenditure distribution, so that we target the right group of people). Multiplying the minimum energy requirement by the average cost of a calorie gives us a monetary amount: the FPL.





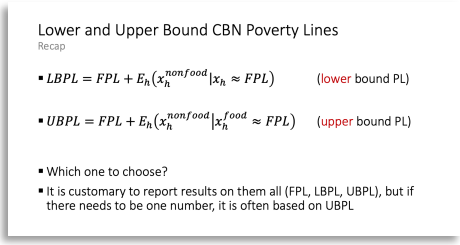
Next is the determination of a **non-food allowance**. Again, the approach is to try and assume as little as possible about what the poor need: our preference always goes to objective estimates ('asking the data'). The key idea is to define a target group of people in the population, who are most likely poor, and to calculate how much they spend on non-food items on average. According to the way we define this target, we can obtain two different non-food allowances. The method is explained by way of a graph first.



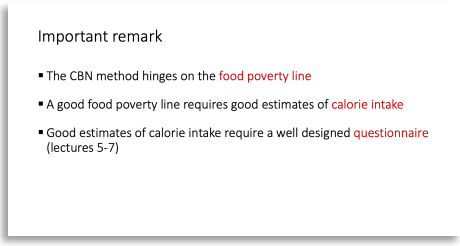


Optionally, and if deemed useful, the Instructor can use formulas, as well. The gist is that the lower bound non-food allowance (NFA) is the average non-food expenditure among

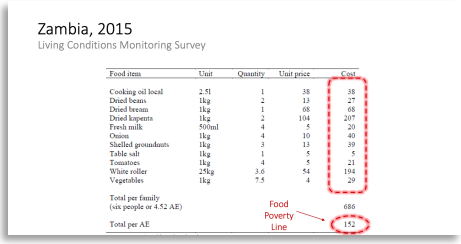
those households whose *total* expenditure is close to the food poverty line (they are extremely poor, cannot even afford the FPL, therefore whatever they spend on non-food must be essential for them); the upper-bound NFA is the average non-food expenditure among those households whose *food* expenditure is close to the food poverty line (they are still poor, although not quite as much, and obviously consume non-food items as well as food items).

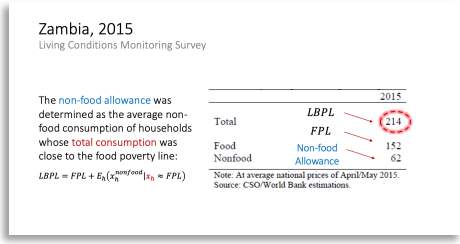


Finally, the FPL and NFAs are added together, to obtain **two poverty lines, lower and upper**. Conventionally, both poverty lines are used, although some analysts may prefer to focus on the UBPL.

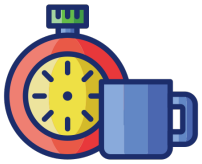


The Instructor is encouraged to stress this point: data producers should focus on the collection of data that allow for the estimation of **calorie intakes** because, among other uses, calories are the foundation of the CBN method for estimating poverty lines.

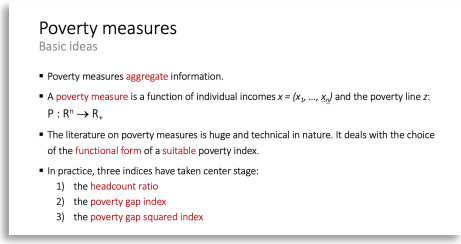




One example concludes the presentation on poverty lines.

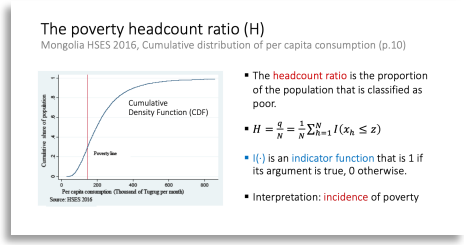


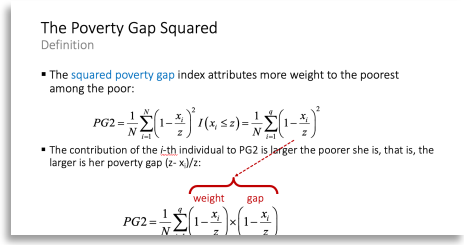
Now is a good time for a break.



After the break, the lecture resumes by the question: after estimating a poverty line, how do we summarize the extent of poverty in the society? We engage a short discussion with students. For instance, we can claim that 'the best way for measuring poverty is to count the number of households (or individuals?) whose expenditure (or income?) falls short of the poverty line. The we ask: can anyone think of a 'better' way to measure poverty? What's wrong with counting the poor? We keep the discussion short by not replying to suggestions – the point of this discussion is to deliver a first message: there are many **measures** available, different indicators deliver different results, not clear how to identity the best indicator. Hopefully, students should be motivated to attend the second part of the lecture. For our purposes, the complexity of the choice can be reduced, in practice, to the discussion of only three indices, which we denote here by **H, PG and PG2**.

The rest of the lecture is devoted to introducing and assessing the three indices.

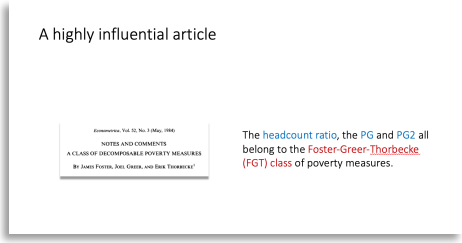


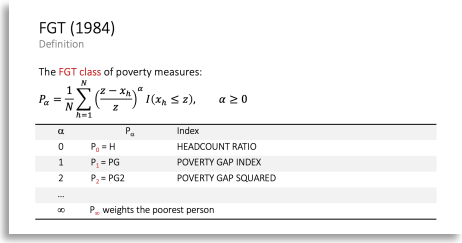


For each index, the strategy proposed in the slide is as follows. First, we provide a definition in words ("The headcount poverty rate is defined as…"). Next, we give the same definition using a formula. The notation is consistent for the three indices and the use of the **indicator function** (defined in the first slide) is introduced and described as a useful device to filter out the 'rich', that is, individuals whose expenditure is greater than the poverty line. We help students to read the math, exactly as if they would read plain English – this is a challenge for the Instructor, sometimes, but a high-return investment. In our experience students will appreciate mastering the formulas that we have selected. Third, we provide a visual illustration of each index (with the exception of the poverty gap squared index, for which none is available). So, the headcount ratio is illustrated by means of the CDF, while

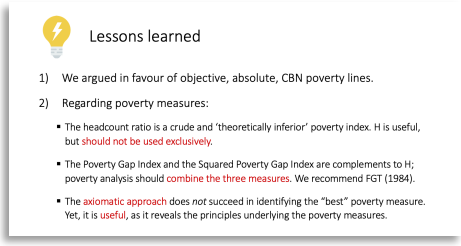
the PG by means of the quantile function. Finally, we provide a short assessment of each index, a phase that lends to engaging with students (time permitting).

The material covered in these slides is standard, and the Instructor should feel free to amend it by adding numerical examples, for instance.





The last two slides introduce the **FGT class of poverty measures**. It is important that students familiarize with this formula. The paper is a famous and classical one, but it is *not* a recommended reading for our students. Too technical and not worth the effort, for the intended audience of this course. It is important however that students understand that i) H, PG and PG2 belong to the same family of indices (the so called 'FGT class of poverty indices'), ii) different indices assign different weights to different people (this is what we explain in the slide for PG2), 3) the FGT class of poverty indices has become the most popular among welfare analysts thanks to its theoretical axiom-derived properties: the logic is similar to that introduced in the lecture on inequality.



The last slide summarizes the main points discussed during the lecture. We have been very parsimonious in summarizing the discussion on poverty lines, for instance. The Instructor can consider adding a second slide with additional conclusions/considerations, should she wish so.

Homework

Exercise 1 - Engaging with the literature

See exercise 1, Lecture 1.

Exercises 2 and 3 – DASP and ADePT

Students who have basic command of Stata can take advantage of explroing DASP and ADePT, two popular tools in use among welfare analysts. The Instructor can devise and assign specific exercises after preparing suitable small-size datasets.

Lecture 15
Describing data

Learning objectives

This final lecture closes the course by discussing the final stage of the survey process: the dissemination of final results. It presents general principles that guide the presentation of findings from a typical household survey, with emphasis on how to design effective tables and graphs, and how to present inequality and poverty estimates.

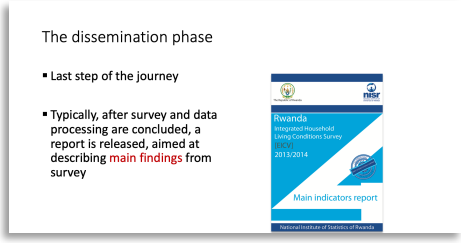
Suggested preparation

Glewwe and Levin (2005) is a useful reference on the presentation of descriptive statistics from household survey data, and the Instructor is assumed to be familiar with it. Schwabish (2014) is focused on the visualization of economic data, and is the source for some of the material presented in the lecture.

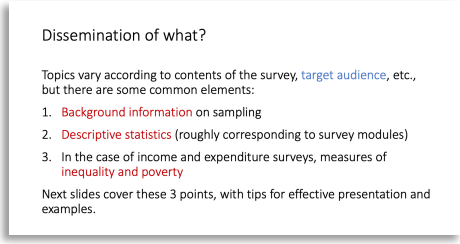
Time allocation

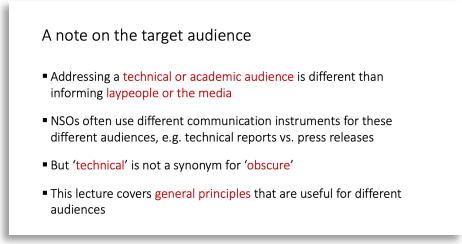
Background information	20 min
Descriptive statistics	40 min
Break	**15 min**
Poverty and inequality measures	45 min
Lessons learned	5 min

Annotated lecture

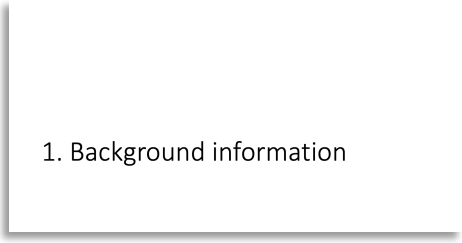


The subject of this lecture is the construction of the **final report** that is typically released in the final stages of the survey process, to disseminate the main findings from the survey, and potentially accompany the dissemination of microdata files for public use. An example of this kind of report is reproduced here (cover of the Rwanda 'Main indicators report').

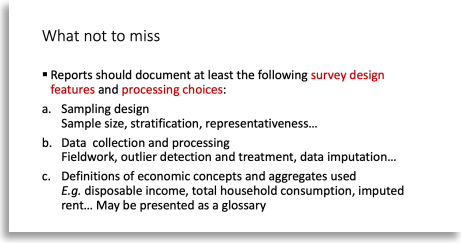




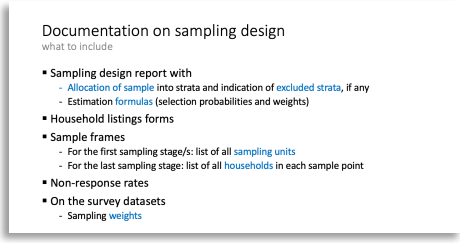
The rest of the lecture will present some guidelines that can be applied to the construction of the final report. Context matters, of course, but it is possible to make some general points. We focus on three main components of the report listed here: background information on sampling, descriptive statistics, and, more to the point of this course, the presentation of statistics on poverty and inequality. One important point that the Instructor should stress is that, of course, the **target audience** matters: the best strategy for effective communication changes depending on whether the NSO wants to reach 'technical' people or the general public. However, one of the takeaways of the lecture is that a 'technical' report does not need to be obscure: certain general rules for clarity and rigor apply to all types of dissemination.

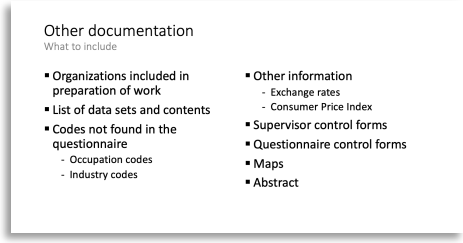


...

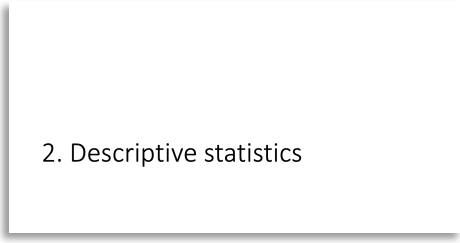


Some **background information** on the survey process should be included in any report. These slides present a 'checklist' for the pieces of information that should never be missed.

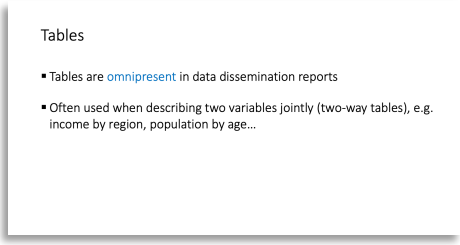


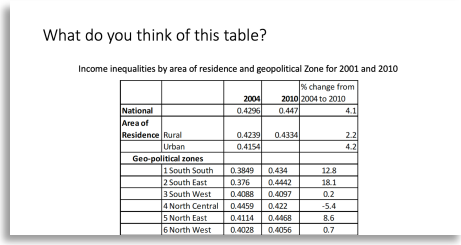


We now go into further detail, by providing some concrete examples. These slides illustrate the type of information on **sampling design, data collection, and the definition of economic concepts and aggregates** that may appear in a typical report: there is no need to dwell on each detail, usually experts will be tasked with drafting these technical sections. However, it is important for students to understand what is meant by 'documentation', in practice. Examples from recent reports – Kenya, Uganda, and South Africa – have the same illustrative purpose.



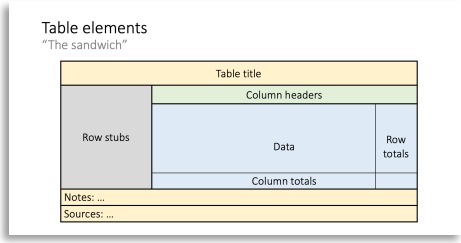
Next, we delve into the topic of presenting **descriptive statistics**. Text, tables, and graphs are our main tools when it comes to describing data; text, that is, writing about data, is a topic in and of itself, and is not going to be discussed here. Interested students are encouraged to check the cited reference.



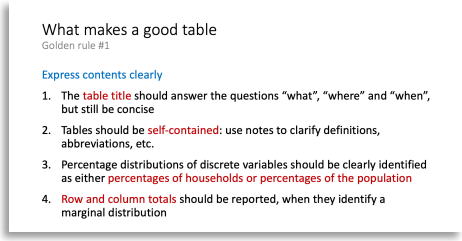


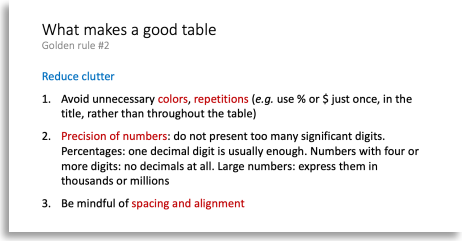
		2004	2010	% change from 2004 to 2010
National		0.4296	0.447	4.1
Area of Residence	Rural	0.4239	0.4334	2.2
	Urban	0.4154		4.2
Geo-political zones				
	1 South South	0.3849	0.434	12.8
	2 South East	0.376	0.4442	18.1
	3 South West	0.4088	0.4097	0.2
	4 North Central	0.4459	0.422	-5.4
	5 North East	0.4114	0.4468	8.6
	6 North West	0.4028	0.4056	0.7

We quickly move on to **tables**. The example shown here is simply a device to be used – if the Instructor feels like it may be useful – to engage the audience. No long debates are needed: the discussion should simply highlight a few points that students may deem important (using clear definition, consistent formatting, alignment, etc.) which the Instructor will soon organize in a few clear 'rules'.

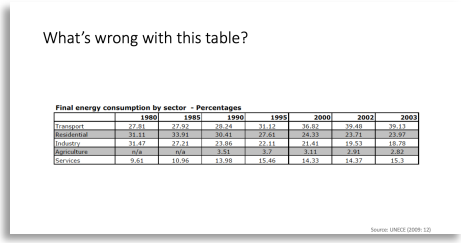


This slide is meant to give students some **terminology**, and to highlight the importance of tables being self-contained, thanks to notes and sources.

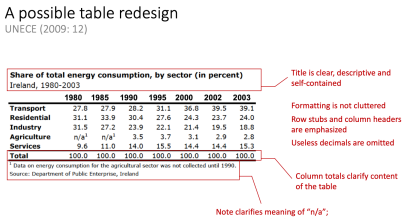




We offer a few '**golden rules**' on how to present effective tables. These boil down to two simple but crucial principles: explaining contents clearly, and reducing clutter. We elaborate on each 'rule' by offering specific tips.



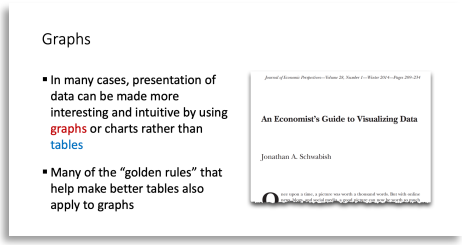
	1980	1985	1990	1995	2000	2002	2003
Transport	27.81	27.92	28.24	31.12	36.82	39.48	39.13
Residential	31.11	33.93	30.41	27.61	24.33	23.71	23.97
Industry	31.47	27.21	23.86	22.11	21.41	19.53	18.78
Agriculture	n/a	n/a	3.51	3.7	3.11	2.91	2.82
Services	9.63	10.96	13.98	15.46	14.33	14.37	15.3

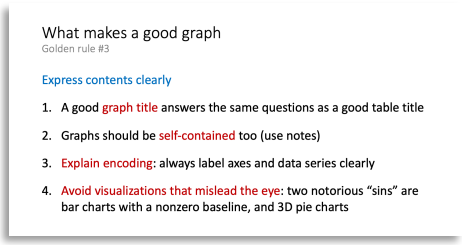


	1980	1985	1990	1995	2000	2002	2003
Transport	27.8	27.9	28.2	31.1	36.8	39.5	39.1
Residential	31.1	33.9	30.4	27.6	24.3	23.7	24.0
Industry	31.5	27.2	23.9	22.1	21.4	19.5	18.8
Agriculture	n/a[1]	n/a[1]	3.5	3.7	3.1	2.9	2.8
Services	9.6	11.0	14.0	15.5	14.4	14.4	15.3
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0

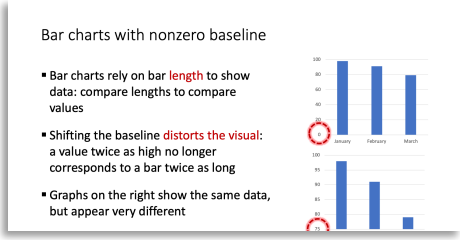
[1] Data on energy consumption for the agricultural sector was not collected until 1990.
Source: Department of Public Enterprise, Ireland

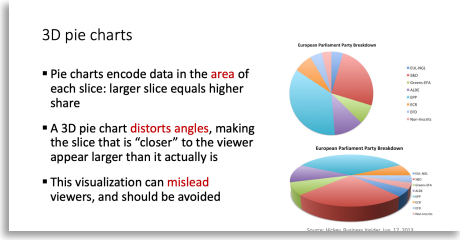
An example clarifies even further. Students can be engaged directly in commenting the first table. The 'redesigned' table is an example, and is by no means the only way to put together an effective table, but it incorporates a number of improvements following the 'golden rules': a more informative title, alignment and formatting that facilitate legibility, only one decimal digit, column totals to clarify the interpretation of cells, notes and sources.



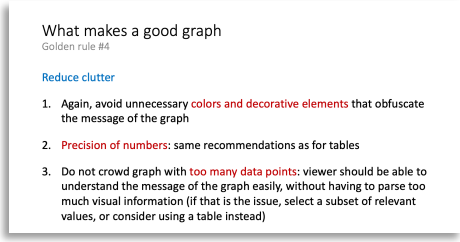


We adopt a similar approach for the case of **graphs**. The two 'golden rules' are the same in spirit, but they are obviously adapted to the context of data visualization.

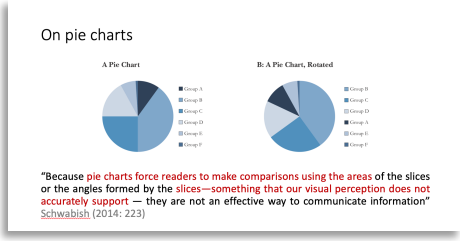


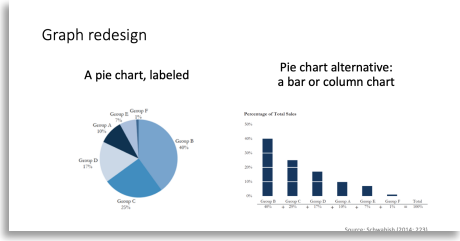


Here are two common examples of graphics that do not express their message in the clearest way. Bar charts with nonzero baseline and 3D pie charts are misleading, and should be avoided.

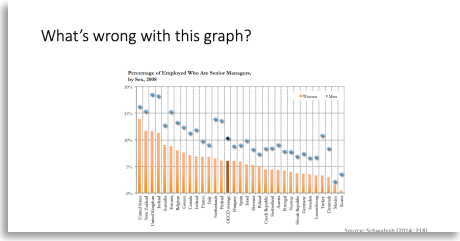


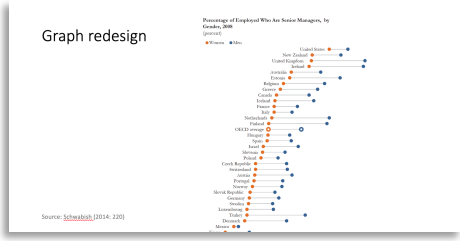
The golden rule of 'reducing clutter' applies to graphs as well.

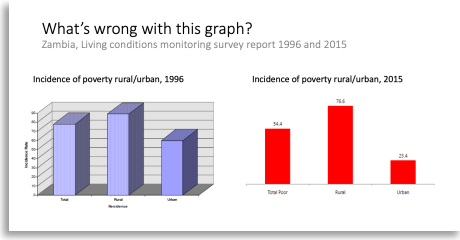




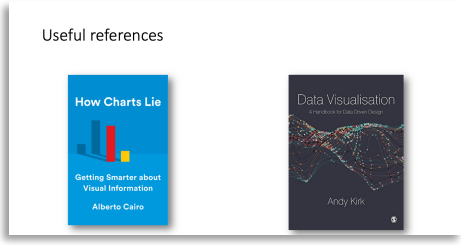
Pie charts, even when not 3D, can be a less-than-optimal way of presenting information, as mentioned by Schwabish (2014).



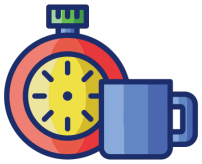




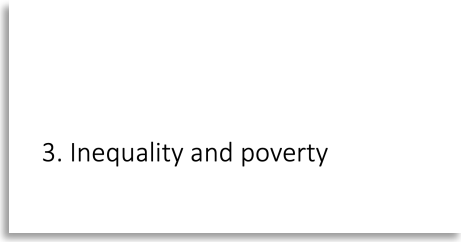
These slides show some examples of cluttered visualizations, and of some ways in which they can be modified to bring the data back into the spotlight, eliminating distractions. The 'redesigned' graphs are not supposed to be the only good solution, just an example of visualization that takes advantage of the 'golden rules'.



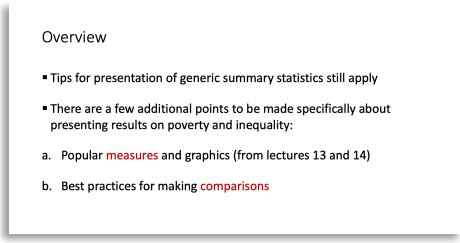
Interested students can take advantage of the references cited here: these are accessible resources on data visualization that do not focus specifically on poverty or inequality, but can certainly prove useful to Statistical Officers.



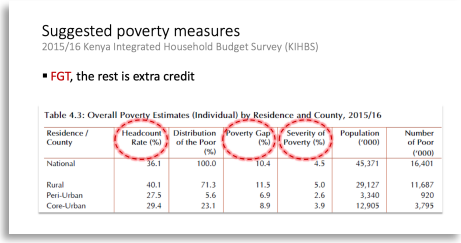
Now is a good time for a break.



Finally, some guidelines on the **presentation of results on inequality and poverty**, one of the main outcomes of consumption and expenditure surveys.

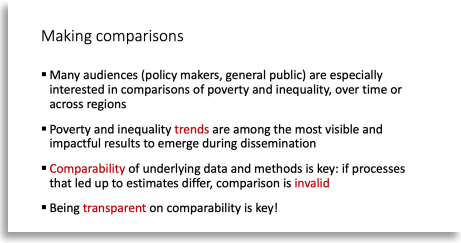


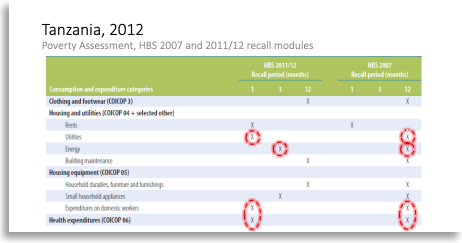
...



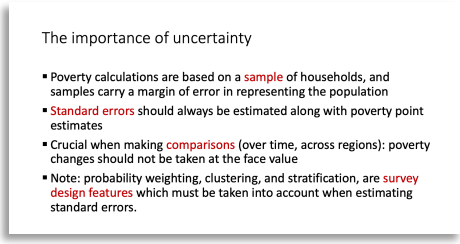
Poverty and inequality measurement have been the subject of previous lectures. At this stage – the dissemination of results – it is useful to know which of the available measures have come to be expected by the international community of readers, practitioners, and academics, who are likely in the target audience of the report. The **Gini index and FGT**

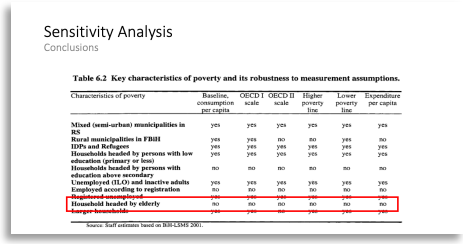
poverty measures are essential; more measures are a welcome addition. This group of slides includes examples from recent reports.





Another important point is that of **comparability**. Methodological changes are a threat to consistent comparisons over time and across countries, and it is crucial that they are thoroughly documented when results are released, to allow for a critical assessment of trends.





A final point concerns **uncertainty**. One of the recurring topics of the course has been the impact of survey design features – as well as analytical choices – on final estimates. These slides emphasize the importance of being transparent in conveying the unavoidable uncertainty of estimates, and recommends two tools, which are too rarely used in standard reports: standard errors, and sensitivity analysis. Examples from recent publications illustrate.

Homework

Exercise 1 - Engaging with the literature

See exercise 1, Lecture 1.

Exercise 2 - Standard Errors

This exercise asks students to elaborate on the meaning and importance of reporting standard errors. One way to discuss the tables is the following. The point estimate plus or minus 1.96 times its standard error gives the upper and lower bounds a 95% confidence interval. For instance, the confidence interval for the poverty headcount rate in urban areas is (27.5, 39.3), obtained as 33.4 +/- 1.96*3. If one were to repeat this calculation for all reported point estimates, one would discover that at least some of the confidence intervals for strata overlap, which would imply that, for instance, poverty rankings of geographical areas may need to be taken with caution.

Exercise 3 - Sensitivity analysis

This exercise is aimed at conveying the meaning of sensitivity in a practical setting. Changing the minimum calorie requirement from 2100 to 2200 kcal per person per day causes a large change in the poverty line (green to blue line), and, consequently, in estimated poverty. Adding just 23 more calories to the calorie norm generates a jump in the line that is almost as large (blue to grey line), while increasing the norm to 2300 only moves the line minimally. The poverty line is much more sensitive to the calorie norm at around 2200, while it becomes more robust over 2238.

References

A'Hearn, B., Amendola, N., and Vecchi, G. (2016). On historical household budgets. Rivista di Storia Economica, 32(2), 137-176.

Alkire, S. (2013). Towards frequent and accurate poverty data.

Alkire, S., and Samman, E. (2014). Mobilising the household data required to progress toward the SDGs.

Alvarez, E., Garcıa-Fernández, R. M., Blanco-Encomienda, F. J., and Munoz, J. F. (2014). The effect of outliers on the economic and social survey on income and living conditions. World Acad. Sci., Eng. Technol., Int. J. Soc., Behav., Educ., Econ., Bus. Ind. Eng, 8, 3276-3280.

Amendola, N. and G. Vecchi (2014), Durable goods and poverty measurement, World Bank Policy Research Working Paper no. 7105.

Atkinson, A. B. (1970). On the measurement of inequality. Journal of economic theory, 2(3), 244-263.

Atkinson, A. B. (2015). Inequality: What Can Be Done? Harvard University Press.

Atkinson A. B., WBG (2017). Monitoring Global Poverty, Report of the Commission on Global Poverty.

Backiny-Yetna, P., Steele, D., & Dijma, I. (2017). The impact of household food consumption data collection methods on poverty and inequality measures in Niger. Food Policy, 72, 7-19.

Balcázar, C. F., Ceriani, L. , Olivieri, S. and Ranzani, M. (2017), Rent-Imputation for Welfare Measurement: A Review of Methodologies and Empirical Findings. Review of Income and Wealth, 63: 881-898.

Barnett, V., and Lewis T. (1994). Outliers in Statistical Data. 3rd edition. J. Wiley and Sons.

Barton, A. J. (1958). Asking the embarrassing question. Public Opinion Quarterly, 22, 67-68.

Beegle, K., De Weerdt, J., Friedman, J., and Gibson, J. (2012). Methods of household consumption measurement through surveys: Experimental results from Tanzania. Journal of Development Economics, 98, 3–18.

Bouis, H., Haddad, L., and Kennedy, E. (1992). Does it matter how we survey demand for food?: Evidence from Kenya and the Philippines. Food Policy, 17(5), 349-360.

Borlizzi, A., Delgrossi, M. E., and Cafiero, C. (2017). National food security assessment through the analysis of food consumption data from Household Consumption and Expenditure Surveys: The case of Brazil's Pesquisa de Orçamento Familiares 2008/09. Food policy, 72, 20-26.

Brzozowski, M., Crossley, T. F., and Winter, J. K. (2017). A comparison of recall and diary food expenditure data. Food Policy, 72, 53-61.

Burgess, T. F. (2001). A general introduction to the design of questionnaires for survey research. Leeds: University of Leeds.

Chen, S., and Ravallion, M. (1996). Data in transition: Assessing rural living standards in southern China. China economic review, 7(1), 23-56.

Cowell, F. (2011). Measuring inequality. Oxford University Press.

Cowell, F., and Flachaire, E. (2007). Income distribution and inequality measurement: The problem of extreme values. Journal of Econometrics, 141(2), 1044-1072.

Cowell, F.A., Jenkins, S.P., Litchfield, J. (1996): The Changing Shape of the U.K. Income Distribution: Kernel Density Estimates. In: Hills, J. (Ed.): New Inequalities. The Changing Distribution of Income and Wealth in the United Kingdom. Cambridge University Press, Cambridge.

Cowell, F., and Victoria-Feser, M. (1996). Robustness Properties of Inequality Measures. Econometrica, 64(1), 77-101.

Cowell, F., and Victoria-Feser, M. (1996). Poverty measurement with contaminated data: A robust approach. European Economic Review, 40(9), 1761-1771.

Dang, H. A., Jolliffe, D., and Carletto, C. (2018). Data Gaps, Data Incomparability, and Data Imputation. Ecineq WP, 456.

De Waal, T., Pannekoek, J., and Scholtus, S. (2011). Handbook of Statistical Data Editing and Imputation. New York: John Wiley and Sons.

De Weerdt, J., Beegle, K., Friedman, J., and Gibson, J. (2016). The challenge of measuring hunger through survey. Economic Development and Cultural Change, 64(4), 727-758.

Deaton, A. (1997). The Analysis of Household Surveys: A Microeconomic Approach to Development Policy. Washington, D.C.: World Bank.

Deaton, A. , and Grosh., M. (2000), "Consumption." In M. Grosh, and P. Glewwe eds., Designing Household Survey Questionnaires for Developing Countries: Lessons from Ten Years of LSMS Experience. Washington, DC: World Bank.

Deaton, A., and Muellbauer, J. (1980). Economics and consumer behavior. Cambridge University Press.

Deaton, A., and Tarozzi, A. (2000). Prices and poverty in India. Princeton, July.

Deaton, A. and Zaidi, S. (2002). Guidelines for Constructing Consumption Aggregates for Welfare Analysis. LSMS Working Paper No. 135. Washington, DC: The World Bank.

Diewert, W. E. (2004), "Durables and User Costs" in ILO, Consumer Price Index Manual: Theory and Practice, chapter 23, ILO / IMF / OECD / UNECE / Eurostat / World Bank.

Diewert, W. E. (2009), "Durables and Owner-Occupied Housing in a Consumer Price Index" in W. E. Diewert , J.S. Greenlees and C.R. Hulten (eds.), Price Index Concepts and Measurements, University of Chicago Press.

Dupriez, O. (2007). Building a household consumption database for the calculation of poverty PPPs. Technical note. Available at: http://go. worldbank. org/4YG7I5RGT0.

FAO and The World Bank. 2018. Food data collection in Household Consumption and Expenditure Surveys. Guidelines for low- and middle-income countries. Rome.

Farfan, G., McGee, K. R., Perng, J., and Vakis, R. (2019). Poverty Measurement in the Era of Food Away from Home: Testing Alternative Approaches in Vietnam. Policy Research Working Paper Series 8692, The World Bank.

Farris, F. A. (2010). The Gini index and measures of inequality. The American Mathematical Monthly, 117(10), 851-864.

Fellegi, I. P., and Holt, D. (1976). A systematic approach to automatic edit and imputation. Journal of the American Statistical association, 71(353), 17-35.

Fiedler, J. L. and Mwangi, D. M. (2016). Improving household consumption and expenditure surveys' food consumption metrics: developing a strategic approach to the unfinished agenda. IFPRI

Friedman, J., Beegle, K., De Weerdt, J. and Gibson, J. (2017). Decomposing response error in food consumption measurement: implications for survey design from a randomized survey experiment in Tanzania. Food Policy, 72: 94–11.

Finn, A. and Ranchhod, V. (2017). Genuine Fakes : The Prevalence and Implications of Data Fabrication in a Large South African Survey. Published by Oxford University Press on behalf of the World Bank.

Foster, J., J. Greer, and E. Thorbecke (1984). A Class of Decomposable Poverty Measures. Econometrica, 52, 3: 761–65.

Gaddis, I. (2016). Prices for poverty analysis in Africa. The World Bank.

Gertler, Paul J., Elaina Rose, and Paul Glewwe. (2000), "Health." In M. Grosh, and P. Glewwe eds., Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study. Washington, D.C.: World Bank.

Gibson, J. (2007). A guide to using prices in poverty analysis. World Bank, Washington, DC.

Gibson, J., Beegle, K., De Weerdt, J., and Friedman, J. (2013). What does variation in survey design reveal about the nature of measurement errors in household consumption?. The World Bank.

Gibson, J., and Rozelle, S. (2002). How elastic is calorie demand? Parametric, nonparametric, and semiparametric results for urban Papua New Guinea. Journal of Development Studies, 38(6), 23-46.

Gibson, R. S. (2005). Principles of nutritional assessment. Oxford university press, USA.

Glewwe, P. (2005). Chapter III: Overview of questionnaire design for household surveys in developing countries. In United Nations Statistical Division, United Nations Department of Economic and Social Affairs (Eds.), Household surveys in developing and transition countries. New York, NY: United Nations.

Glewwe, P., and Levin, M. (2005). Presenting simple descriptive statistics from household survey data. In UN, Household Sample Surveys in Developing and Transition Countries. Studies in Methods Series F No. 96.

Grosh, M., and Glewwe, P. (1998). Data Watch: The World Bank's Living Standards Measurement Study Household Surveys. The Journal of Economic Perspectives, 12(1), 187-196.

Grosh, M. and Glewwe, P. (2000). Designing Household Questionnaires for Developing Countries, Lessons from 15 years of Living Standards Measurement Study, Volume One: World Bank.

Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. Technometrics, 11(1), 1-21.

Harrison, D. E., and Krauss, S. I. (2002). Interviewer cheating: Implications for research on entrepreneurship in Africa. Journal of Developmental Entrepreneurship, 7(3), 319.

Haughton, J. and Khandker, S. R. (2009). Handbook on poverty and inequality. Washington, DC: World Bank.

Heijink, R., Xu, K., Saksena, P., and Evans, D. (2011), Validity and comparability of out-of-pocket health expenditure from household surveys: a review of the literature and current survey instruments. Geneva: World Health Organization, 28.

Hentschel and Lanjouw (2000), "Household welfare measurement and the pricing of basic services", Journal of International Development, 12: 13-27.

Heston, A. and A.O. Nakamura (2009), Questions about the equivalence of market rents and user costs for owner occupied housing, Journal of Housing Economics, 18, 273—279.

Hlasny, V., and Verme, P. (2018). Top Incomes and Inequality Measurement: A Comparative Analysis of Correction Methods Using the EU SILC Data. Econometrics, 6(2), 30.

Iarossi, G. (2006). The power of survey design - a user's guide for managing surveys, interpreting results, and influencing respondents. Washington, DC: World Bank.

Ibarra, G. L., Mendiratta, V., and Vishwanath, T. (2017). Rental regulation and its consequences on measures of well-being in the Arab Republic of Egypt. The World Bank.

Jolliffe, D. (2001). Measuring absolute and relative poverty: the sensitivity of estimated household consumption to survey design. Journal of Economic and Social Measurement, 27(1, 2), 1-23.

Judge, G., and Schechter, L. (2009). Detecting problems in survey data using Benford's Law. Journal of Human Resources, 44(1), 1-24.

Krosnick, J. A., Et Presser, S. (2010). Question and questionnaire design. In J. D. Wright Et P. V. Marsden (Eds.), Handbook of survey research (second edition) (pp. 263-313). Bingley, UK: Emerald Group.

Lanjouw, P. (2012), "Consumption-Based Measures in Developing Nations. Lessons from Brazil", in Besharov and Couch (eds.), Counting the Poor. New York: Oxford University Press. Ch. 13.

Lipton, Michael and Ravallion, Martin (1995). "Poverty and policy," Handbook of Development Economics,in: Hollis Chenery and T.N. Srinivasan (ed.), Handbook of Development Economics, edition 1, volume 3, chapter 41, pages 2551-2657 Elsevier.

Little, R. J., and Rubin, D. B. (2019). Statistical analysis with missing data (Vol. 793). Wiley.

Lohr, S. L. (2009). Sampling: design and analysis. Nelson Education.

Lu C, Chin B, Li G, and Murray CJ. (2009) Limitations of methods for measuring out-of-pocket and catastrophic private health expenditures. Bull World Health Organ;87(3):238-44, 244A-244D.

MacDonald, L., Macpherson, D. A., Sirmans, G. S., and Zietz, E. N. (2006). The value of housing characteristics: a meta analysis. The Journal of Real Estate Finance and Economics, 33(3), 215-240.

Malpezzi, S. (2002). "Housing". In Grosh, M. and Glewwe, P. (eds.). Designing Household Questionnaires for Developing Countries, Lessons from 15 years of Living Standards Measurement Study, Volume One: World Bank.

Mancini, G. and Vecchi, G. (2019), On the Construction of a Welfare Indicator for Inequality and Poverty Analysis, mimeo.

Meyer, B., and Sullivan, J. (2003) "Measuring the Well-Being of the Poor Using Income and Consumption", The Journal of Human Resources, 38.

Meyer, B. D., and Sullivan, J. (2009). "Five decades of consumption and income poverty". National Bureau of Economic Research.

Meyer, B. D., and Sullivan, J. (2011). "Further results on measuring the well-being of the poor using income and consumption." Canadian Journal of Economics, 44(1), 52-87.

OECD (2013). OECD Guidelines for Micro Statistics on Household Wealth.

Oseni, G., Durazo, J., and McGee, K. (2017). The Use of Non-Standard Units for the Collection of Food Quantity. LSMS guidebook.

Oseni, G., Huebler, F., McGee, K., Amankwah, A., Legault, E., and Rakotonarivo, A. (2018), Measuring Household Expenditure on Education: a new guidebook on measurement. LSMS guidebook.

Pape, Utz Johann and Mistiaen, Johan A.. (2018). Household expenditure and poverty measures in 60 minutes : a new approach with results from Mogadishu.

Pradhan, M. (2009). Welfare analysis with a proxy consumption measure: evidence from a repeated experiment in Indonesia. Fiscal Studies, 30(3-4), 391-417.

Pyatt, G. (1976). On the interpretation and disaggregation of Gini coefficients. The Economic Journal, 86(342), 243-255.

Ravallion M. (1994). Poverty Comparisons.

Ravallion M. (2008) Poverty Lines. In: Durlauf S.N., Blume L.E. (eds) The New Palgrave Dictionary of Economics. Palgrave Macmillan, London.

Ravallion, M. (2016). The Economics of Poverty History, Measurement, and Policy. Oxford: Oxford University Press.

Ravallion, M. and B. Bidani (1994). How Robust is a Poverty Profile?, World Bank Economic Review, 8: 75-102.

Rousseeuw, P. J., and Croux, C. (1993). Alternatives to the median absolute deviation. Journal of the American Statistical Association, 88(424), 1273-1283.

Schwabish, J. A. (2014). An economist's guide to visualizing data. Journal of Economic Perspectives, 28(1), 209-34.

Sen, A. (1976). Poverty: An Ordinal Approach to Measurement, Econometrica, 44(2): 219-31.

Sen, A. (1987). The Standard of Living. Cambridge: Cambridge University Press.

Sen, A. (1987). Commodities and Capabilities. New Delhi: Oxford University Press.

Shorrocks, A. F. (1980). The class of additively decomposable inequality measures. Econometrica: Journal of the Econometric Society, 613-625.

Smith, L. C. (2015). The great Indian calorie debate: Explaining rising undernourishment during India's rapid economic growth. Food Policy, 50, 53-67.

Smith, L. C., Dupriez, O., and Troubat, N. (2014). Assessment of the reliability and relevance of the food data collected in national household consumption and expenditure surveys. International Household Survey Network.

Stigler, G. J. (1954). The early history of empirical studies of consumer behavior. Journal of Political Economy, 62(2), 95-113.

Stiglitz, J. E., Sen, A., and Fitoussi, J. P. (2009). Measurement of economic performance and social progress.

Tourangeau, R., Groves, R. M., & Redline, C. D. (2010). Sensitive topics and reluctant respondents: Demonstrating a link between nonresponse bias and measurement error. Public Opinion Quarterly, 74,413–432.

Troubat, N. and Grünberger, K. (2017). Impact of survey design in the estimation of habitual food consumption. The case of the 2007/08 Socio Economic Survey of Mongolia applied to urban households. Food Policy, 72(C): 132–145.

UNECE (2009). Making Data Meaningful, Part 1: A Guide to writing stories about numbers. United Nations, Geneva.

UNECE (2009). Making Data Meaningful, Part 2: A Guide to presenting statistics. United Nations, Geneva.

World Bank Group. (2015). A Measured Approach to Ending Poverty and Boosting Shared Prosperity: Concepts, Data, and the Twin Goals. Policy Research Report. Washington, DC: World Bank.

World Bank. (2017). Monitoring Global Poverty: Report of the Commission on Global Poverty. Washington, DC: World Bank.

Xu K, Ravndal F, Evans DB, and Carrin G. (2009), Assessing the reliability of household expenditure data: results of the World Health Survey. Health Policy;91(3):297-305.

Zheng, B. (1997). Aggregate Poverty Measures, Journal of Economic Surveys, 11(2): 123-62.