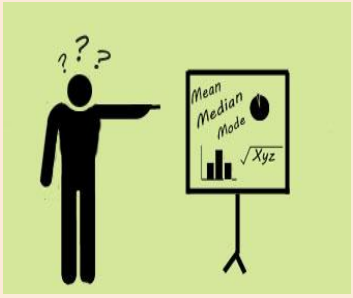
The 10 ultimate concepts you must know to survive to data requests from students

Lucia Tramonte, PhD
Department of Sociology
UNBF



2

The 10 concepts

1. Population and sample
2. Index or scale?
3. Validity
4. Reliability
5. Standard deviation and standard error
6. Significance (statistical)
7. Correlation
8. Causation
9. Power
10. Multivariate or multilevel?

3

Populations and samples

Population: The total set of individual objects/persons of interest

All identical twins reared apart in the 20th century

Sample: The subset of the population for which we actually have data

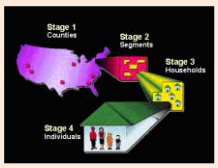
53 sets of identical twins reared apart in London from 1920-1950

We want to distinguish between two types of samples

Random samples | **Non random samples**

4

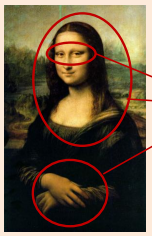
Simple random sample: extracting with close eyes from a jar of jelly beans!



Complex sample: extracting unequal groups of individuals from strata in the population

5

The idea



Elements that make you recognize "the essence" of the portrait

In a stratified sample, we divide the picture in strata, or places/areas, and extract more cases from the key places and less from the not-so-important places.

6

How does this affect modelling or analyzing data?

- The sample is no longer simply random: we have oversampled some aspects-individuals and undersampled others.
- We may need to rebalance this picture: we purposefully biased the sample to gain efficiencies to meet other goals, but we can correct this bias by applying the design weights.

7

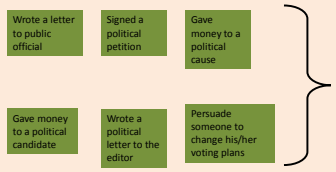
How to interpret sampling and go back to the picture

- If you looked only at the parts we sampled, you wouldn't get an accurate picture.
- All the parts would be there but not in the right proportions.
- The design weights compensate for the known distortions.
- The final weights include estimated distortions.

9

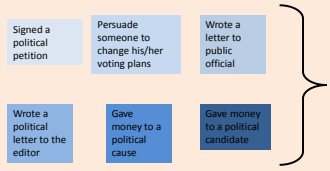
INDEXES AND SCALES

10



These items represent expressions of political activism.

To create an index of political activism we might give people 1 point for each of the actions that they have taken. Then we sum the values.



What if the items are located on a continuum of political activism, from mild to strong?

To create a scale of political activism, we might give people a higher score if they responded to the items that are stronger expression of political activism.

11

Why bothering for indexes and scales?

- In large datasets there are scores for individual items as well as composite measures of theoretical and practical constructs.
- For each construct, the data documentation offers insights on the constituents as well as on the properties of the aggregate measure.
- Capable students may want to manipulate the items; less capable students should be advised to use the available aggregate measures.

12

…the last slide brings us into defining two crucial properties of any measure

VALIDITY AND RELIABILITY

13

Validity

In science and statistics, validity generally refers to the extent to which a concept, conclusion or measurement is well-founded and corresponds accurately to the real world.

The validity of a measurement tool (for example, a test in education) is considered to be the degree to which the tool measures what it claims to measure.

In the area of scientific research design and experimentation, validity refers to whether a study is able to scientifically answer the questions it is intended to answer.

14

Why bother with validity?

If your purpose is to open a tuna can, and you decide to use a chain saw to accomplish your task, chances are that your final job won't be what you expected.

The idea of validity is per se quite simple: *choose the right instrument to do the job that you intended to do.*

15

Therefore

When you are asked about advising on a specific dataset to conduct a research and gather information on a certain topic, you are ultimately asked to do something that will impact the validity of that study.

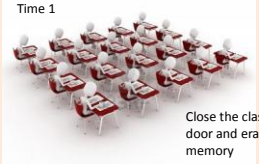
A valid study is a pre-requisite to be able to make safe use of its results.

16

Reliability

Or to be able to trust in a measure because it works consistently….

Time 1



Administrate a math test or ask questions about self esteem.
Each person will have a score in math or self esteem.

Close the classroom door and erase people's memory

Time 2

Administrate the same math test or ask the same questions about self esteem.
Then correlate the results from time 1 with results from time 2.

17

Therefore

When you are asked about advising on a specific measure available in a dataset, suggest students to consult the official documentation in order to learn about the reliability of a measure/test/construct.

High reliability is a pre-requisite to be able to conduct good statistical analyses.

18

Important

A measure/test/construct can be highly reliable but completely invalid.

A measure could be valid but function quite unreliably.

Often people look at reliability and think that they have solved issues of validity. It is not the case.

Said differently: the chain saw may work really well in cutting wood logs, but it remains an inappropriate instrument to open tuna cans.

19

STANDARD DEVIATION AND STANDARD ERROR

20

Standard deviation

Standard deviation is a measure of the amount of variability in a characteristic observed in a sample or population.

The essence of statistics and the social sciences is variability. There is nothing interesting to do unless we observe variability.

It is more practical to discuss variability in units of standard deviation than units of variance.

21

Why bothering with it?

Students should focus on choosing characteristics and then variables that display variability.

You can advise them to consult the documentation in order to gather information on frequency distributions of the variables of interests.

By the way, such distributions also help spotting potential loss of cases, aka MISSING VALUES.

22

Standard Error

It is a measure of the precision of an estimate.

You don't know the average height in Canada;

you extract a sample of Canadian adults and compute the mean.

You will use that mean as the best estimate of the population average height.

However, people may want to know how precise you have been in estimating the average height. You add the standard error as a measure of how much, on average, you missed the target.

23

24

The standard error is smaller…

…if the standard deviation in your sample is small
and if your sample size is big.

This means that you have more precise estimates from homogeneous and big samples.

Make sure that you remind students about the importance of sample size.

Minimum sample size: 30 cases!

25

THE STATISTICAL MEANING OF "SIGNIFICANT"

26

Significant, like….

Common jargon

- Important
- Noteworthy
- Surprising
- Relevant
- Large
- Strong
- Relatively different

Statistical jargon

- An effect or relationship that is systematically different from zero
- An effect or relationship that is very unlikely (p value smaller than .05) to happen under the assumption that the actual effect is zero
- An effect or relationship that is exceptional, and occurs only in 5% (or 2 or 1) of the possible samples that we can draw from a population

27

Logic of statistical hypothesis testing

In the population, one, and only one, of these statements is true—but how can we tell which one on the basis of a single sample?

$\beta_1 = 0$
There really is **no relationship** between X and Y in the population

Null Hypothesis:
H_0

$\beta_1 \neq 0$
There really is a **relationship** between X and Y in the population

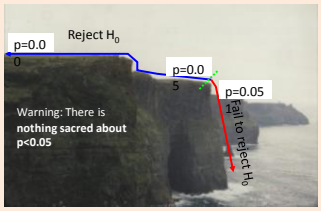
Alternative Hypothesis:
H_1 (or H_a)

Basic idea of statistical hypothesis testing:
How unusual is our sample?

Use sample data to make an educated guess by asking: assuming the null hypothesis was true, how likely is it that we would have gotten the sample result we did?

28

Evaluating the size of p-values:
Uses the 0.05 level of significance



and misuses of the term…

The vastness of this country, the high mobility rate of many of its inhabitants and its statistically significant immigrant population all contribute to the need for an efficient postal service.
-- *The New York Times*

The numbers involved in this comparison were considered too small to be statistically significant.
-- *The New York Times*

It is difficult to test our results … against observations because no statistically significant global record of temperature back to 1600 has been constructed.
-- *Science*

29

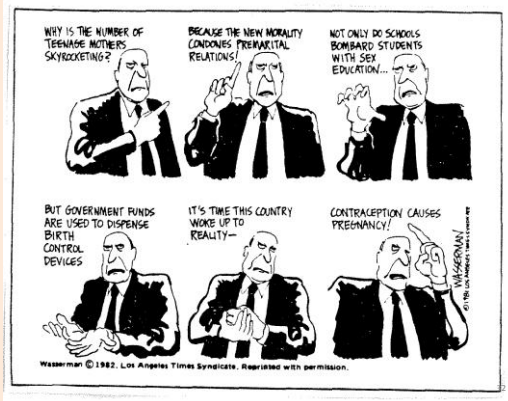
Statistical significance depends on

- Presence or absence of an actual effect/relationship
- Sample size
- Effect size
- Standard error
- Reliability
- Standard deviation

30

CORRELATION AND CAUSATION

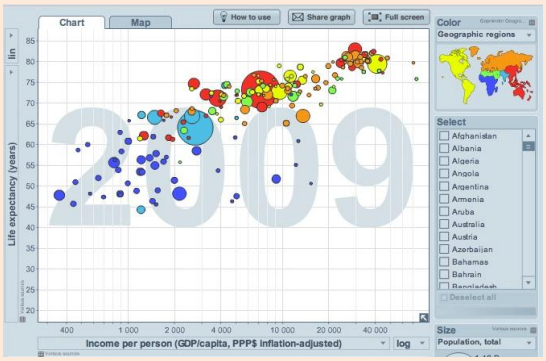
31



Association or correlation?

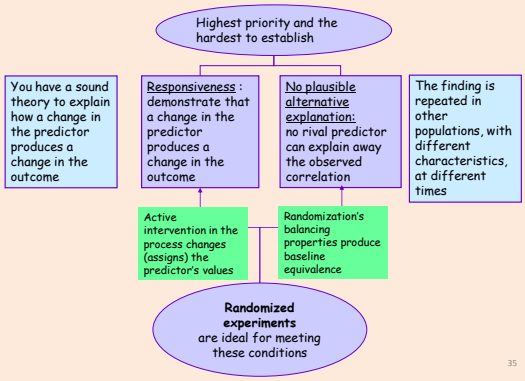
- Association→ there is something going on between X and Y
- Association=undefined relationship between X and Y.
- Association=safe term…very generic
- Correlation=linear relationship or linear association of X and Y
- Correlation→ the relationship between X and Y can be represented by a straight line with a slope different from zero (a straight line that is not flat horizontal)

33



34

Four conditions for establishing causality



35

POWER OR SUPERPOWERS…

36

The power of a statistical test is the probability that the test will do a good job, that is to reject the null hypothesis when the null hypothesis is false. Therefore power is also known as the *sensitivity*.

Power nearly always depends on the following three factors:

- the statistical significance criterion used in the test
- the magnitude of the effect of interest in the population
- the sample size used to detect the effect

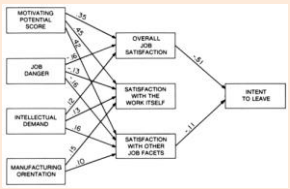
37

MULTIVARIATE OR MULTILEVEL ANALYSES

38

Multivariate analysis

A data analysis that involves simultaneously several independent variables that concur and compete to explain a dependent variable



39

Multilevel analysis

A data analysis that involves simultaneously one or more independent variables at different level of aggregation to explain a dependent variable.