

Methodological skills

rMA linguistics, week 4

Tamás Biró

ACLCLC

University of Amsterdam

`t.s.biro@uva.nl`

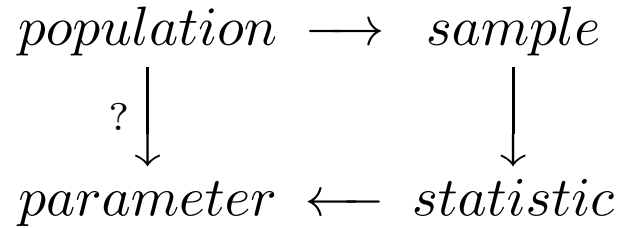
Toward inferential statistics:

Sampling distributions

Toward inferential statistics:

Sampling distributions

Remember:



Inferential statistics

Procedure of inference:

- Population \rightarrow sample \rightarrow statistic \rightarrow parameter estimated.
- What happens if we repeat this procedure many times?
- Probability theory: field of mathematics describing the behavior of random processes, such as sampling. (Intro: M&M ch. 4.)
- We have done it only once: so what can we infer?

Sampling distribution

- **Sampling distribution of a statistic:** the distribution of the statistic, if taken from all possible samples of size n .

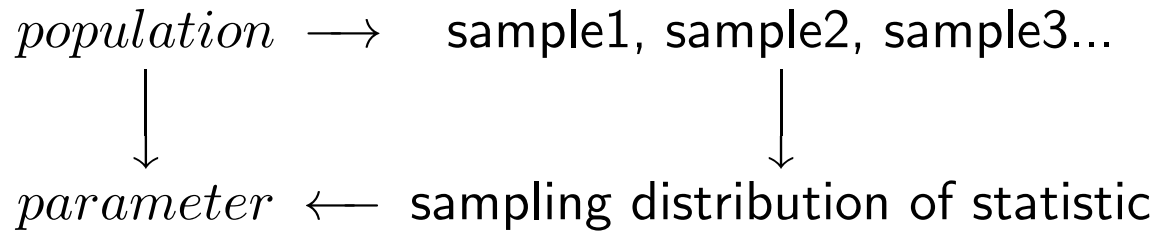
Sampling distribution

Example: throw many dice.

1 1	2 1	3 1	4 1	5 1	6 1
1 2	2 2	3 2	4 2	5 2	6 2
1 3	2 3	3 3	4 3	5 3	6 3
1 4	2 4	3 4	4 4	5 4	6 4
1 5	2 5	3 5	4 5	5 5	6 5
1 6	2 6	3 6	4 6	5 6	6 6

Statistic measured: $\text{sum} = \sum_{i=1}^n x_i$.

Sampling distribution



But usually money for one sample only!

- Will it give the correct answer? 'Reliable' and 'valid'?

Reliability and validity

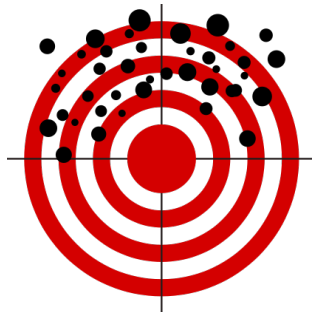
- **Reliability** of the procedure:

Do we get different results if we repeat the experiment?

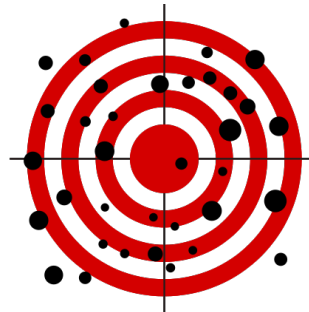
- **Validity** of the procedure:

Is the outcome of the experiment answering our question?

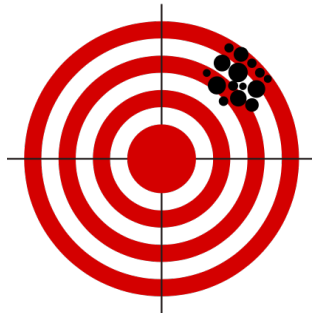
Reliability and validity



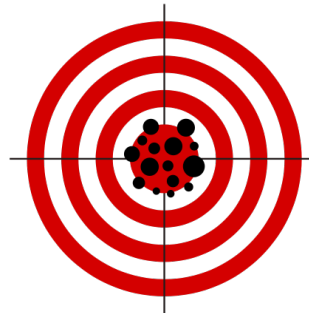
Unreliable & Invalid



Unreliable, But Valid



Reliable, Not Valid



Both Reliable & Valid

http://en.wikipedia.org/wiki/File:Reliability_and_validity.svg

Reliability and validity

- **Reliability** of the procedure:

Variability of statistic: spread of sampling distribution, given our procedure.

- **Validity** of the procedure:

Bias = center of sampling distribution.

Unbiased statistic:

true value of the parameter = mean of sampling distribution, given our procedure.

GOOD NEWS!

Sampling distribution

- To reduce bias, achieve validity:
use random sampling!
- To reduce variability, achieve reliability :
use larger sample!
- Population size N (if much larger than sample size n)
does not matter.

Towards inferential statistics

Procedure of inference:

- Population \rightarrow sample \rightarrow statistic \rightarrow parameter estimated.

To be continued next week with *Central Limit Theorem!*

Paper presentation by students

Research questions

Is there a significant relationship between the grammatical gender system in one's native language and one's perception of an object's natural gender?

Is there a significant relationship between one's own gender and one's perception of an object's natural gender?

Is the linguistic landscape around metro stations Kraaiennest and Ganzenhoef a reflection of the ethnic composition of the populations of K & G-buurt?

Is it true for Dutch as well that verbs of the type 'give' (geven) occur more often in a double object construction and verbs of the type 'sell' occur more often in a ditransitive construction?

To what extent are international students at the University of Amsterdam interested in learning Dutch?

Are the hesitation pauses of women in general longer than those of men?

A corpus study on concrete and metaphorical use of sensory verbs in academic Italian

Does the age of babysitters at Oppascentrale
Kriterion have an effect on how often and at what
times they babysit?

Do people have a tendency to choose a partner with a name starting with the same letter as their own?

Is there a significant difference in the number of left and right hand gloves found on the streets of Amsterdam?

Basic structure of a research paper

At least in some fields, which we now practice. Observe concentric structure:

- Intro: background, motivation, importance, past work.
- Definition of the problem. Hypothesis.
- Methodology: intro + procedure + details (data set, subjects...).
- Results
- Discussion
- Summary: conclusion and future directions

Writing and reading a research paper

What most readers will read/look at:

- The title
- The abstract
- Visualization
- Captions of figures
- Intro and summary

In two weeks time:

Think about paper structure and data types:

- Intro: General problem
→ *anecdotal evidence* and *available data*.
- Precise research question:
Hypothesis to be tested/rejected (H_0 and H_a).
- How to proceed?
Sample survey (observation) or experiment (intervention)?
- Pilot vs. “the real stuff” .

In two weeks time:

Define research question, in terms of:

What is your:

- Motivation? General problem? Operationalized research question?
- Population?
Parameter(s) of the population that interests you?
- Units?
Sample and sampling method?
- Explanatory variables, response/dependent variables?
Levels of the variables?

Send me a one-page summary (ppt or pdf).

Subsequently:

- What to measure on the sample?
- Statistic to be calculated?
- Best visualization?
- How to draw conclusions in order to answer research question?
- How to draw conclusions in order to contribute to general problem?

To read for next week:

- L. F. Bachman (2004), *Statistical Analyses for Language Assessment*. Chapter 6
- Moore & McCabe, chapters 1–5. Next week we cover chapter 2.

See you next week!