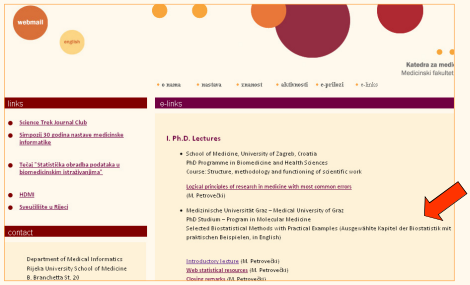
Medizinische Universität Graz
PhD Studium – Program in Molecular Medicine
Selected Biostatistical Methods with Practical Examples
June 2013

Introductory lecture

Prof. Mladen Petrovečki, MD, PhD

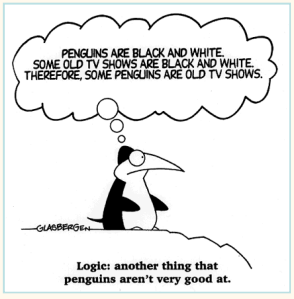
http://mi.medri.hr



Schedule

11.6. – Basic biostatistical terminology and methods
12.6. – Qualitative data, correlation & regression
13.6. – Testing the differences
14.6. – Meta-analysis & power analysis
15.6. – Exercises & discussion

Logic & reasoning



www.glasbergen.com/

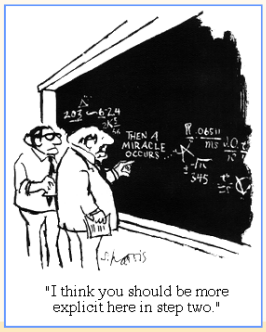
Types of logic

www.glasbergen.com/

Nonscientific procedures

- diligence
  (habit, attitude, manner, believe, momentum)
- authority
- intuition

Scientific procedures?



"I think you should be more explicit here in step two."

Argument, proof

Logic in science

- system
- models of the system
  - deterministic
  - probabilistic
- event probability → $P(E)$

$$0 \le P(E) \le 1$$

Probability

- mathematical calculation that something, event, will occur
- mathematic ⇨ probability theory
  - statistics
  - mathematics
  - scientific methodology
  - logic, philosophy
- reasoning about event feasibleness

Probability, calculation

- symbol – P
- $$P = \frac{\text{No. of expected events}}{\text{No. of all events}}$$
- values range 0 – 1:
  - 0 – impossible event
  - 1 – certain event

Probability vs. fortune

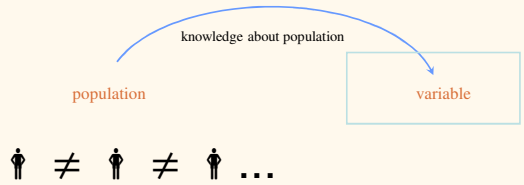
www.cartoonstock.com

Probability vs. coincidence

Probability vs. impossibility

www.christianforums.com/

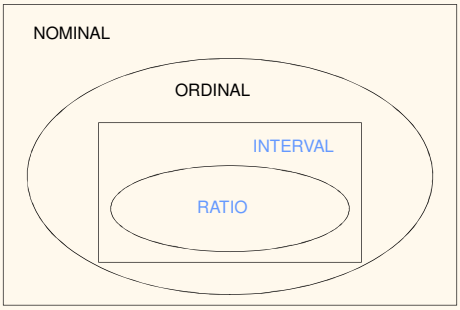
Measuring & Research



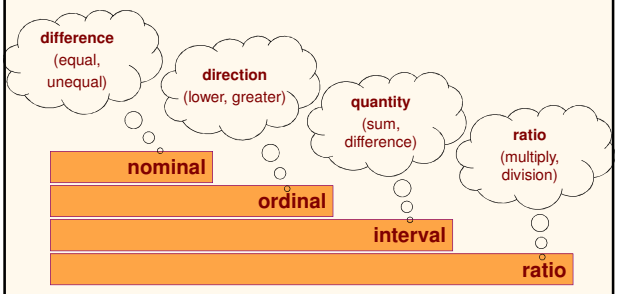
Variables in research

- all variables in research
- as many of them
- the end of research
- simple → complex (data)
- accuracy (numbers)
- measuring scales

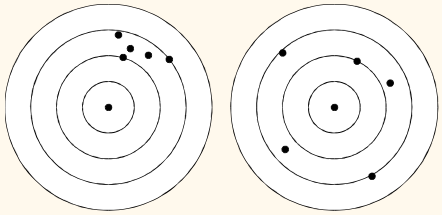
Measurement scales



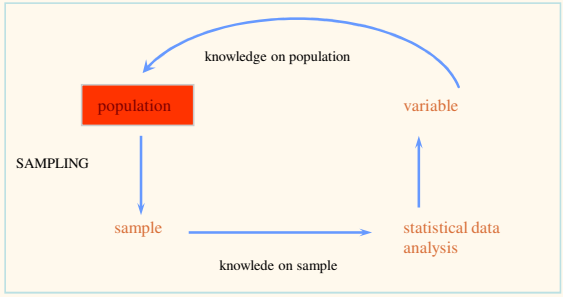
Measurement scales



Error



Population



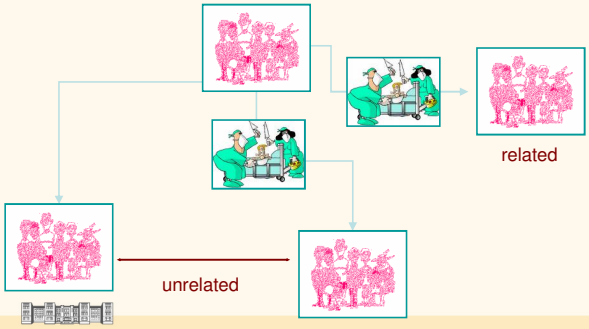
Sample

- part of population
  - what? who?
  - when?
  - where?
  - size

Sample

- representative
- measurable
- probabilistic
  - simple
  - system
  - stratified
  - cluster

Sample

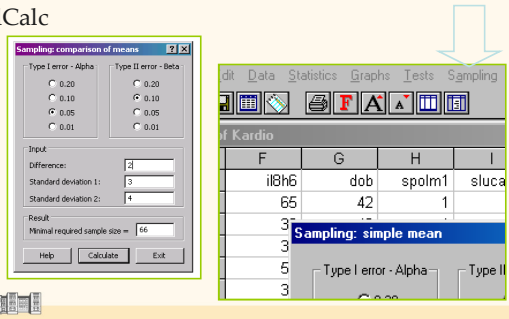


Sampling

www.statehousereport.com

Sampling

MedCalc

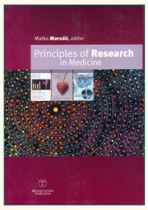


Bias (sampling)

Bias (sampling)

- Bias – systemic sampling error

- prevalence bias (Neyman)
- admittance rate bias (Berkson)
- answering rate bias
- etc.



Blinding

- single-blind
- double-blind
- triple-blind
- quadruple-blind

Bias, blinding ☺

Control group

- must have
- to be compared with experimental group
- Hawthorn's effect
  - research with no control group
  - subject changes behavior with a knowledge that is a part of experiment
  - subject feels better with knowledge to be a part of experiment

Hypothesis

http://biology.ucf.edu/~pascencio/images/Hypothesis.jpg

Statistical hypothesis

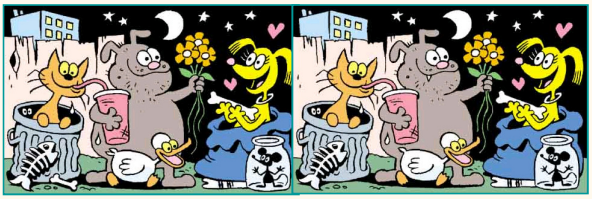
- elemental statement
- truth or not (false, lie)
- hypothesis testing → finding the truth

Ivana Brlic Mazuranic
How Quest Sought the Truth
(*Kako je Potjeh tražio istinu*)
http://www.bulaja.com/FAIRYTALES/

Statistical hypothesis

- truth ⇨ real object state
  probabilistic system:
  truth → probability
- significant ⇨ any occasion other that accidentally:
  probability → level of significance

Null-hypothesis



No difference

Null-hypothesis

No difference ≈ Not guilty

Testing the hypothesis

A. null-hypothesis
B. statistical test
C. level of significance
D. statistics calculation
E. conclusion

A. Hypothesis

- null – H_0 – no difference
- alternate – H_1 – difference exists
- only one can be truthful
- only one can be accepted, other will be rejected

B. Choosing the test

- measuring scales
- sample
  - size
  - related on unrelated samples
- data distribution
  - parametric
  - nonparametric
- no. of variables
- etc.

Statistical tests

Scale	One sample	Two		Three or more	
		related	unrelated	related	unrelated
Nominal	binomial	McNemar		Cohran	
	chi-square		Fisher		chi-sqr.
			chi-square/		
Ordinal	Kol.-Smirn.	Wilcoxon		Friedman	
			MW		p/median
			Moses		KW
Interval	...				
Ratio	...				

Paired & unpaired tests

Level of significance

- *P*
  - ➢ α if defined before statistics
  - ➢ α – probability of rejecting H_0 when H_0 = truth
- error α (type I error or false positive error)
- as less as possible
- default values, e.g. $P<0,05$

Statistical errors

Table 3-1. Making the conclusions – correct and incorrect concluding		
	Conclusion from statistical hypothesis test	
True situation	No difference (accept H_0)	Difference exist (reject H_0)
No difference (H_0)	Correct conclusion (no error)	Incorrect conclusion (α error or type I error)
Difference exist (H_1)	Incorrect conclusion (β error or type II error)	Correct conclusion (no error)

Deducting presupposes the comparison of two systems. Systems may be, in theory, equal (the same) or they may differ in any aspect. We do not know the actual state and therefore we investigate. Assumption is presented in a statistical hypothesis, in two ways; null (H_0: no difference) and alternative (H_1: difference exists). From the testing we draw a conclusion with which we prove that the difference exists or that it does not exist. Correct conclusions are when there is no actual difference of systems and we do not find it, or when there is actual difference and we find it. Incorrect conclusions are when the difference actually does not exist but we find it, as well as when the difference actually does exist but we do not find it. Types of errors are indicated next to incorrect conclusions.

D. Statistics

- computation...
- P = exact value
- three decimals

$P > 0.05$

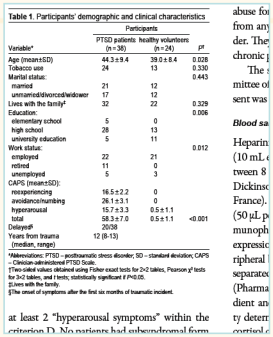
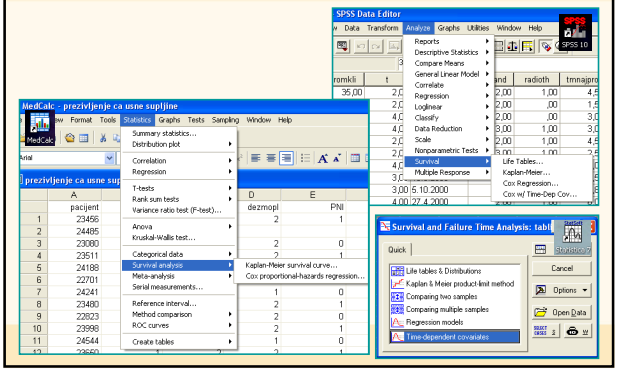


Table 1. Participants' demographic and clinical characteristics

Software



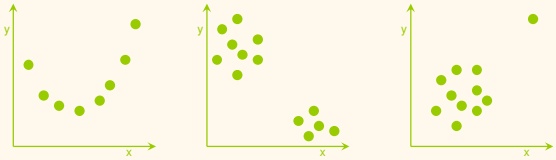
Conclusion (E)

- low P ⇨ low possibility to reject the truth
- conclusion:
  - $P < \alpha$
  - low probability that H_0 is true
  - reject (not accept) null hypothesis
  - accept alternate hypothesis
  - statement "..." is truth with P = ...

Yes & No in statistics

- hypothesis = ?
- calculation = ?
- correct data = ?
- all conditions for statistic valid = ?
- no limitations = ?

Example 1: "Not" in correlation



Example 2: "Not" with χ^2-test

lectures quality	students Zagreb	students other
well	10	31
bad	0	19
total	10	50

Example 3: Another "not"

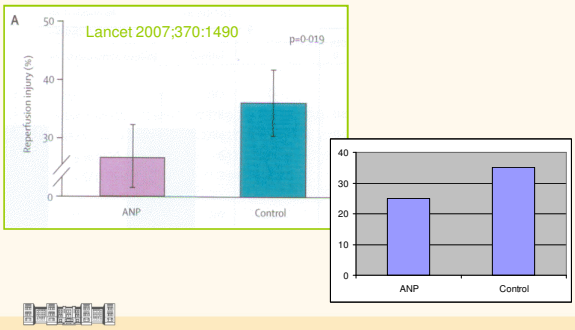
a predictor. All statistical tests were performed using the SAS software system and significance was determined when *P*-values were less than 0.05.

in Group I–II versus Group III was marginally significant ($P = 0.07$). However, when tests were

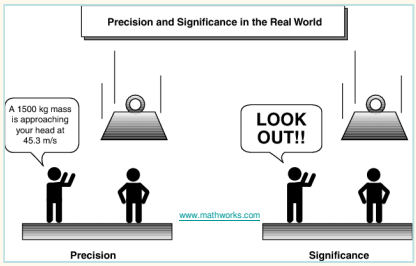
Lupus 2004;14:426

($P = 0.0007$) and a marginally significant increase in creatinine clearance ($P = 0.096$). There was no statistically significant longitudinal effect in serum creatinine levels.

Example 4: "Not" in graphs



Significance *vs.* accuracy



The last one: The truth

Prof. Mladen Petrovečki, MD, PhD

Department of Medical Informatics
Rijeka University School of Medicine
http://mi.medri.hr

Department of Clinical Laboratory Diagnosis
Immunology Division
Dubrava Clinical Hospital, Zagreb
www.kbd.hr/lab

mladenp@kbd.hr