

Medical statistics, informatics and telemedicine

Communication with a Statistician.

Sample Size. Data Tables. Survey Tips.

by: Dániel VERES

2021. November 26th

EBM: Evidence-Based Medicine

A series of conscientious, unambiguous, and **logical decisions** based on **evidences available at the time**, which serve the treatment of the given patient.

EBM: Evidence-Based Medicine

A series of conscientious, unambiguous, and logical decisions based on **evidences available at the time**, which serve the treatment of the given patient.

Clinical Studies
Your Diploma Thesis



WORK TOGETHER

Physician's (medic) tasks:

- What is the question
- How we could measure
- What could be the confounders
- Relevance (effect size)
- Feasible
- Ethical
- ...

Statistician's task:

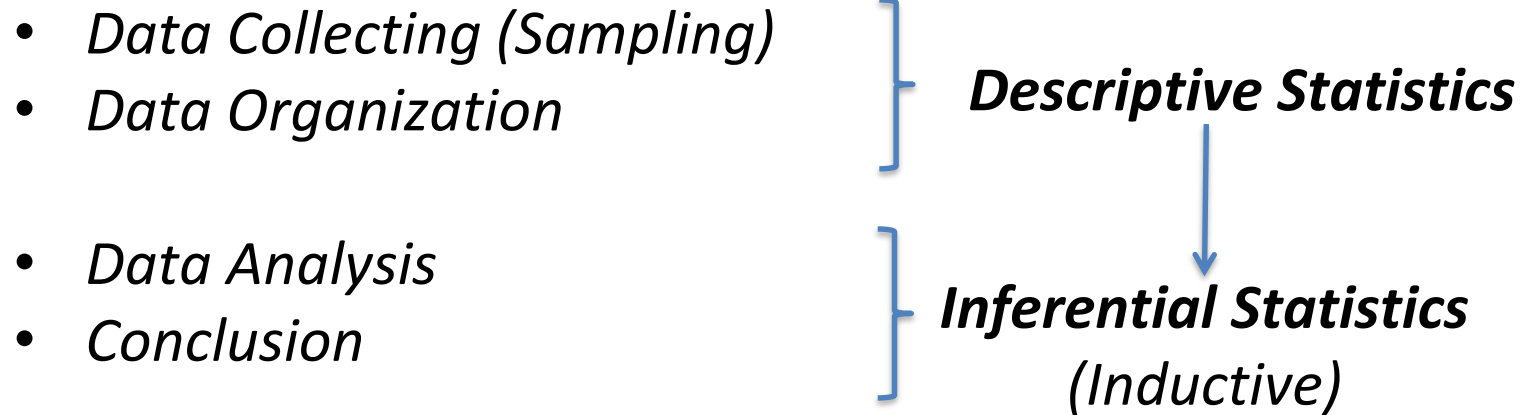
- The data can really answer that
- How could we perform calculations
- How could we analyse
- Estimate sample size
- Randomization
-
- ...

Physician and statistician together:

- Understand the results
- Interpret the results
- Further opportunities, usage

When to contact with the statistician?

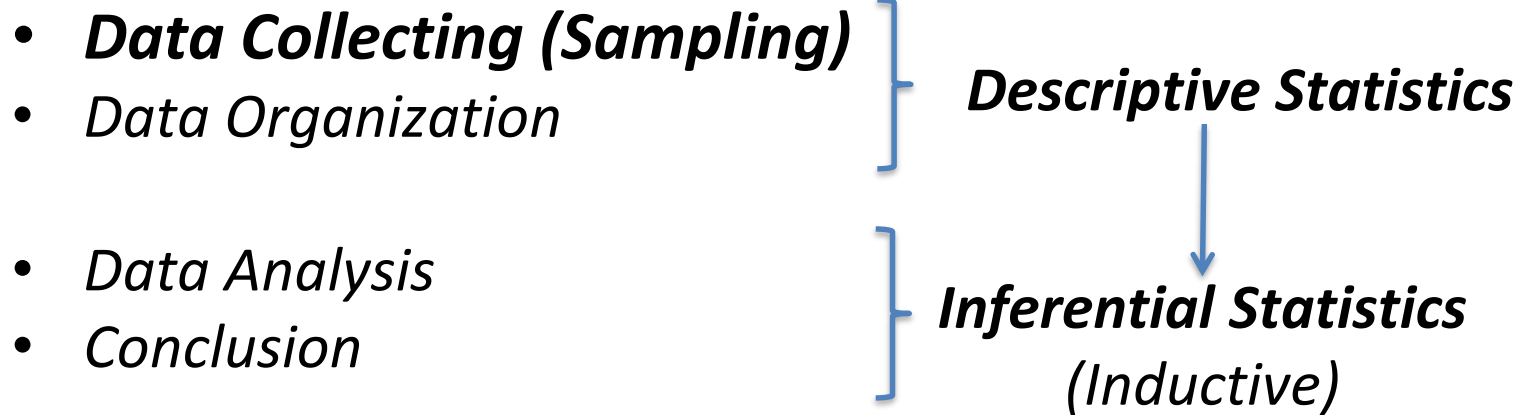
At the 2nd lecture:



As soon as possible...

When to contact with the statistician?

At the 2nd lecture:



Sample size?

When to contact with the statistician?

Biologist talks to statistician



Sample size?

<https://www.youtube.com/watch?v=Hz1fyhVOjr4>

If there is a difference can we recognize it (power)?

		In population (in reality) the null hypothesis is:	
		True	False
Decision on null hypothesis:	Accepting (Not rejecting)	Good decision	Error (type II) (β) (false negative result)
	Rejecting	Error (type I) (α) (false positive result)	Good decision power($1-\beta$)

Example: The probability of rolling six > 1/6?

If there is a difference can we recognize it?

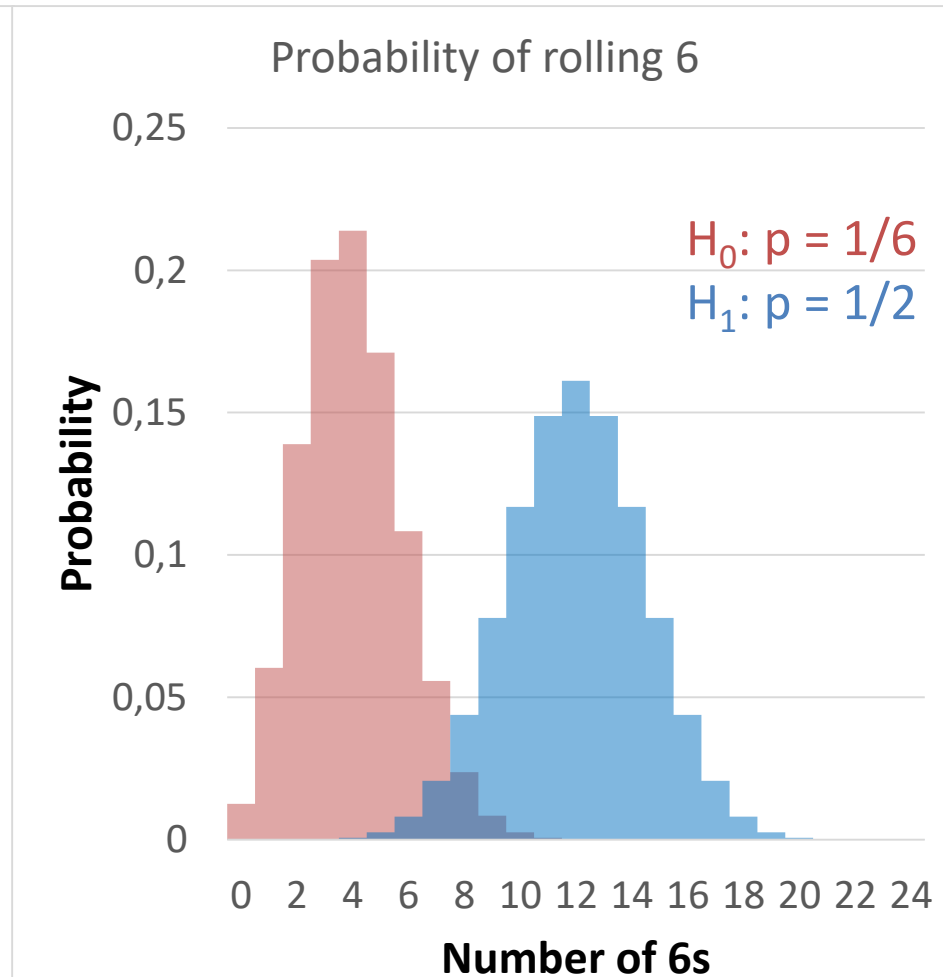
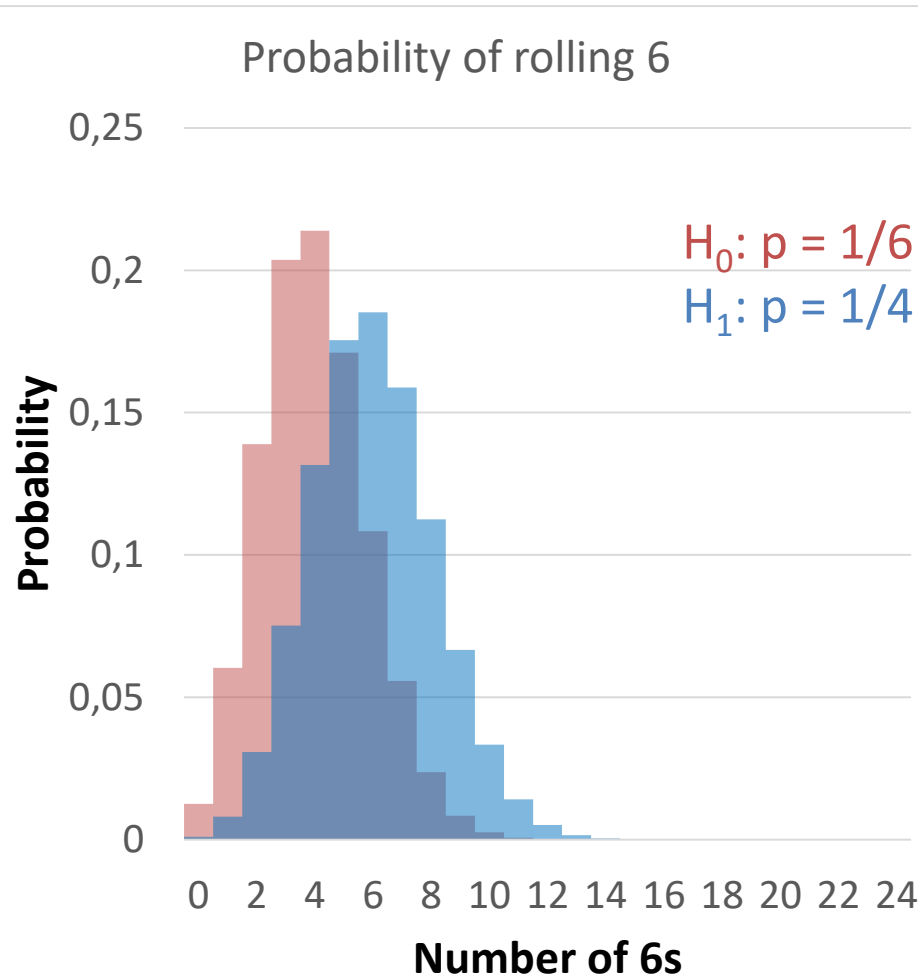
1. Effect size

Showing relevance:

Effect size – eg. Difference between means or standard deviations, difference between ratios, different distributions, etc.

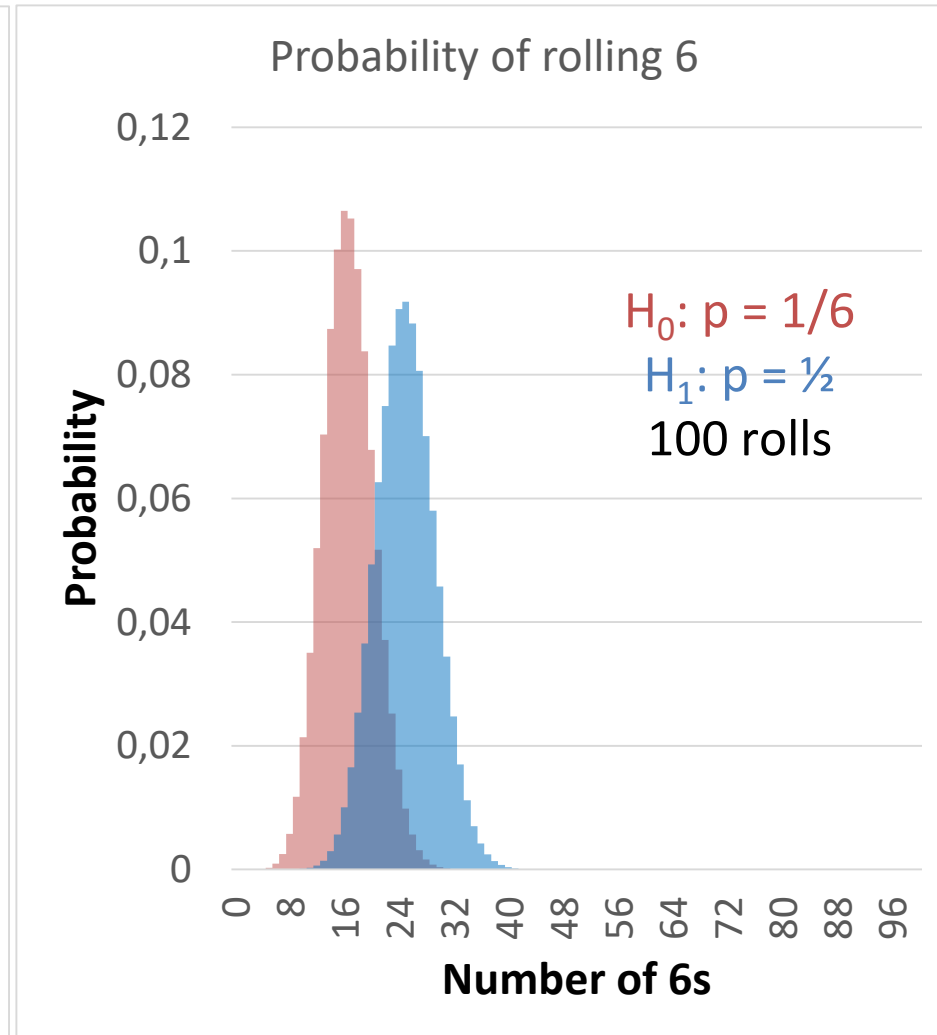
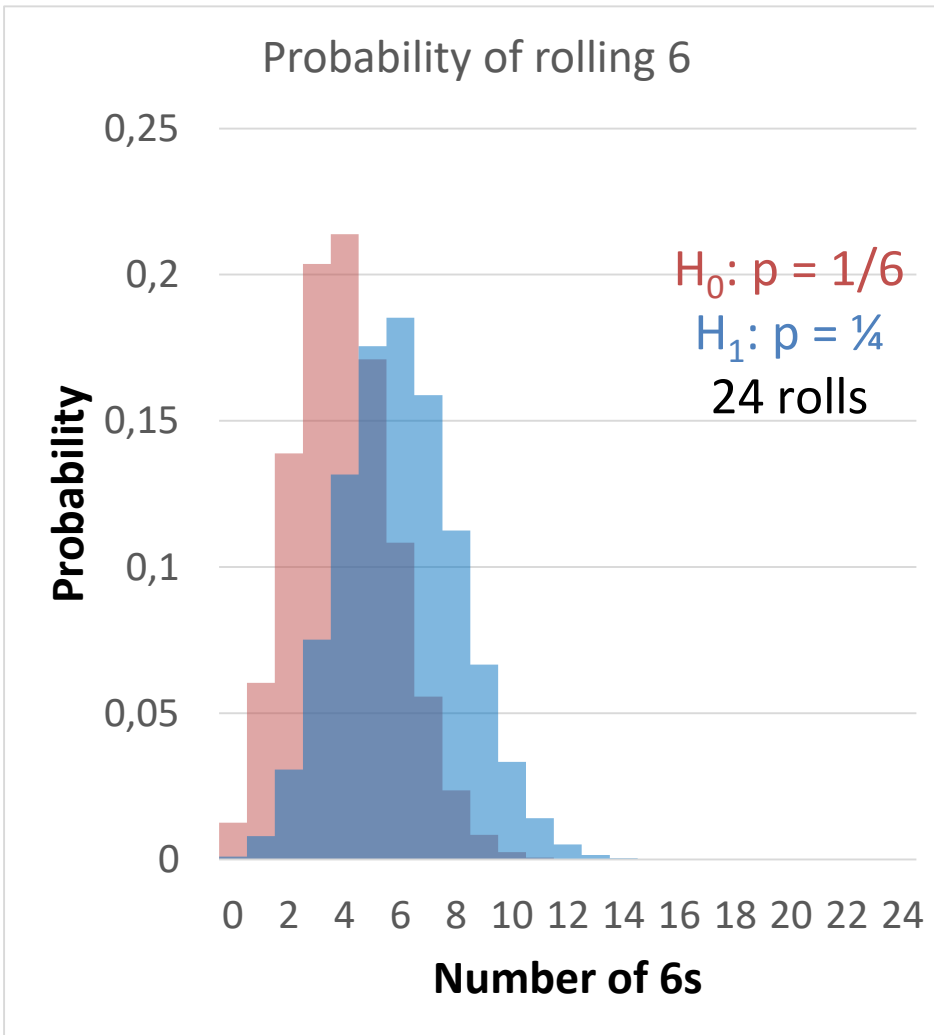
If there is a difference can we recognize it?

1. Effect size



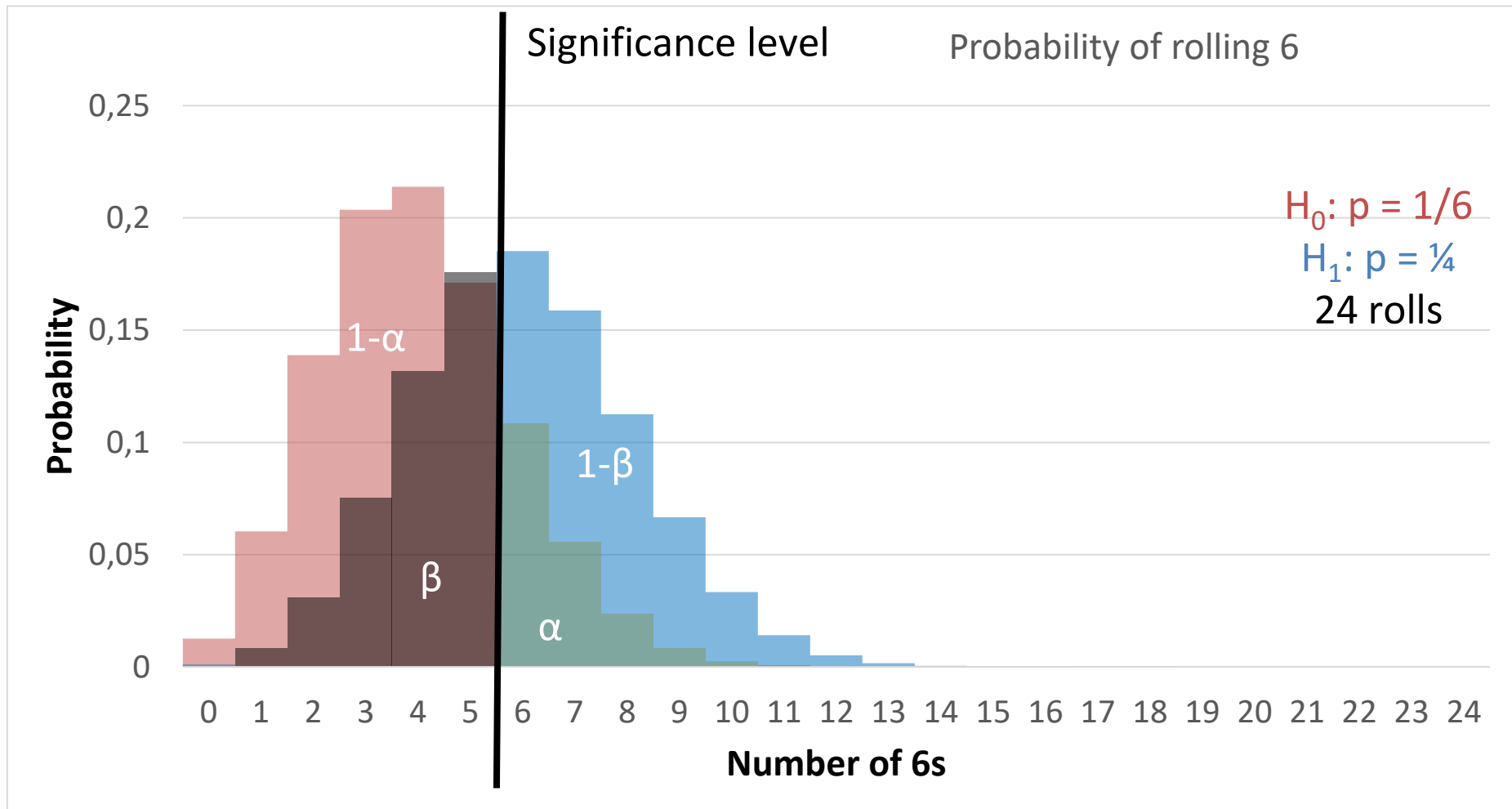
If there is a difference can we recognize it?

2. Sample size



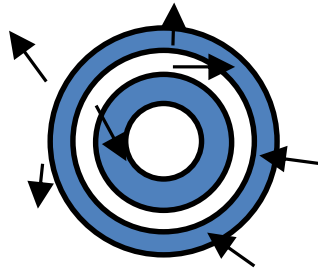
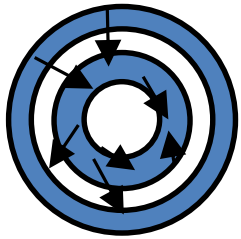
If there is a difference can we recognize it?

3. Significance level – threshold for decision (1st and 2nd type errors)

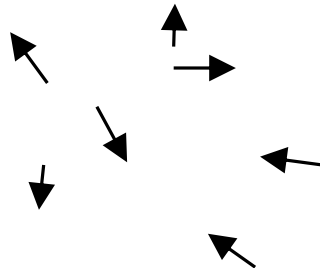
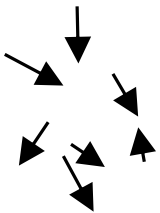


If there is a difference can we recognize it?

4. Variability

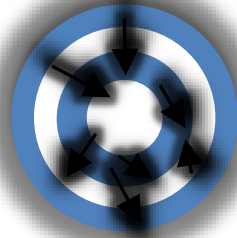
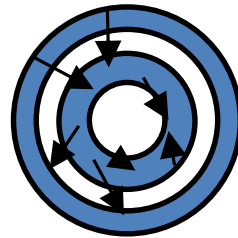


Where is the „bull eye“?



If there is a difference can we recognize it?

5. Type of hypothesis test – measuring scale

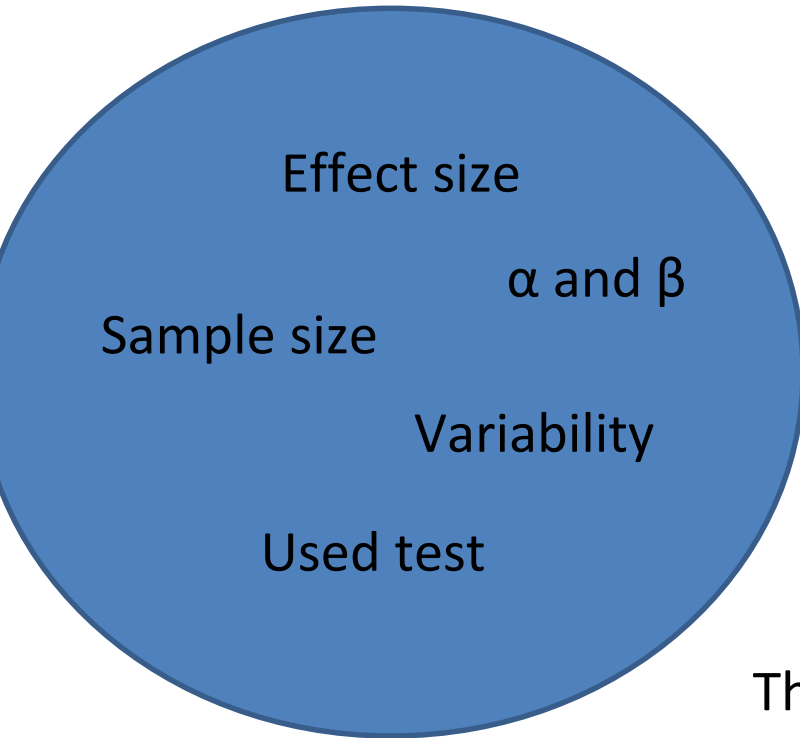


Higher scale tests are more powerfull!

Sin in statistics: using categorical scale if continous is available

If there is a difference can we recognize it?

For sample size physician (medic) should give:

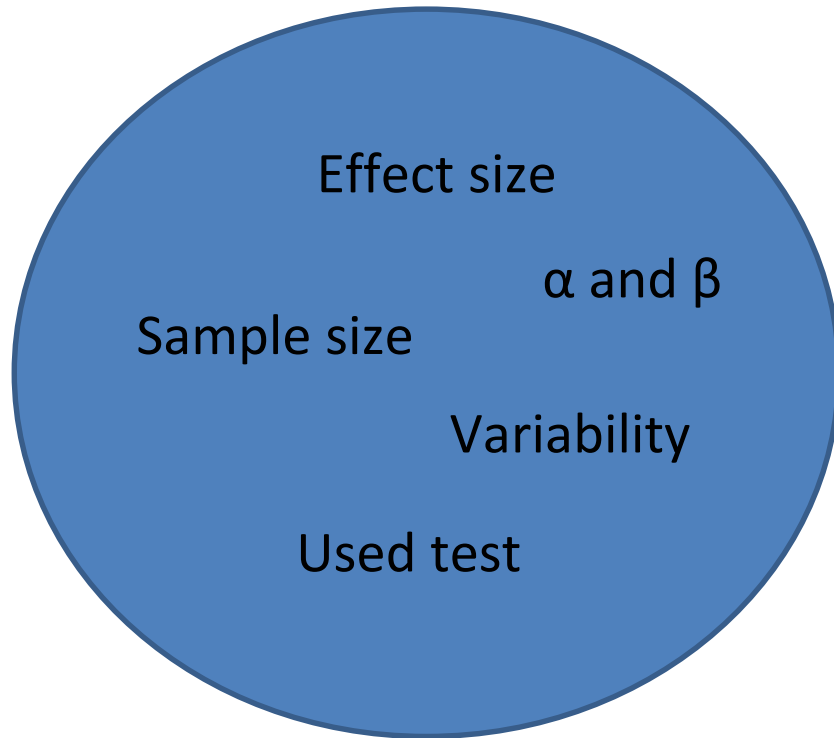


- Type I and II errors
- Effect size, variability: eg.:
 - A) mean or difference of means + SDs
 - B) ratios in the 2 group
no need for variability
difference of ratios not enough!
 - C) ...

The used test should be given by the statistician

They depend on each other.

If there is a difference can we recognize it?



Is the calculated sample size:

- **Feasible?**
- **Ethical?**

They depend on each other.

Relevant but not significant...

Possibilities:

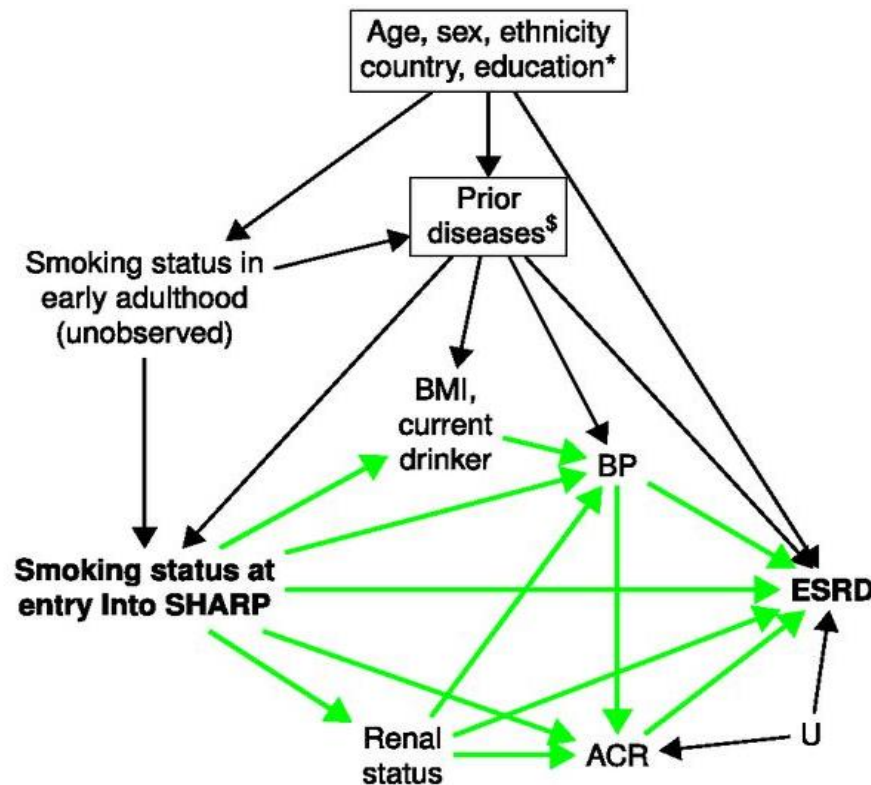
- small sample size
- large variability
- we could not measure it accurately
- less powerful statistical test
- violated assumptions for the test
- we were unlucky (sampling error)



Plan ahead!!

Physician's (medic) task

- **What are we interested in?**
Cause and effect, repeatability,
difference in means or in standard deviations,
drug effect or toxicology?
difference or no difference, not worse, not better
- **Previous knowledge: relevance + possible confounders**



Make a graph!

Only the clinician knows

- **Who is the analysis for? (How complex analysis we need)**
For the boss, thesis, tabloid or Nature article
- **Study design**
Observational (...), experimental test, case review...
Repeated measures, strata
- **Measured variables**
How was it measured (measurement error, precision),
units
- ...

Regulations

Personal and medical data usage – needs authority permission!
(your diploma thesis...)

<https://ett.aeek.hu/tukeb/>

In Hungary



Egészségügyi Tudományos Tanács
Medical Research Council

Tudományos és Kutatásetikai Bizottság

Keresés ...

TUKEB A Bizottság tagjai Az eljárásról **Kérelem** Jelentés Jogszabályok

Engedélyezett kutatási tervek

I.) Kérjük, hogy az OGYÉI vagy az országos tisztifőorvos által engedélyezhető kutatás engedélyezése iránti kérelmet az OGYÉI-nak illetve az országos tisztifőorvosnak (Nemzeti Népegészségügyi Központnak) nyújtsák be, az ETT-TUKEB-hez csak az ETT-TUKEB által engedélyezhető vizsgálat engedélye iránti kérelmet nyújtsák be. (Az ETT-TUKEB levelezési címe: 1051 Budapest, Széchenyi István tér 7-8. A borítékokra, a küldemények csomagolására feltétlenül írják rá, hogy “Egészségügyi Tudományos Tanács”. E-mail cím: tukeb@emmi.gov.hu).

II.) Az ETT-TUKEB-hez benyújtandó beavatkozással nem járó vizsgálat engedélyezése iránti kérelemmel kapcsolatban minden protokollt elektronikusan, ha lehet az Ügyfélkapun (epapir.gov.hu), illetve a Hivatali kapun keresztül küldjenek be.

Javasolt minden kérelmet a honlapról letölthető “[ÜRLAP ETT-TUKEB KUTATÁSI ENGEDÉLYHEZ](#)” űrlap kitöltésének segítségével elkészíteni, benyújtani. A kérelemhez csatolandó nyilatkozatok szintén megtalálhatók és letölthetők az ETT-TUKEB honlapján: [DOC](#) vagy [PDF](#) formájában.

A nyomtatványoknak a honlapon megadott formában történő kitöltése ajánlott, de nem kötelező jellegű, a pályázó az ajánlott mintától eltérhet.

Az “Előzetes áttekintő” című, [letölthető](#) dokumentum a kutatási terv kötelező eleme, és segítséget nyújthat a pályázat összeállításában. Hiánya esetén a protokoll nem kerül elbírálásra.

Scientific and Research
Ethics Committee

TUKEB

Részletes vizsgálati terv, amely tartalmazza az alább felsorolt főbb pontokat :

A vizsgálat célja

A kutatás tudományos megalapozottságát alátámasztó irodalmi hivatkozások megjelölése

A vizsgálat elrendezése

A vizsgálat módszertana, megszervezése

Publikációs elvek

Mik a vizsgálati alanyok bevonási és kizárási kritériumai?

Milyen módon kívánnak alanyokat toborozni?

Hány főt vonnak be a vizsgálatba?

Milyen statisztikai módszerrel dolgozzák fel az adatokat?

Mik a kutatás anyagi fedezetének forrásai?

Mellékletek:

Betegtájékoztató – intézeti fejléces papíron

Beteg-belegyűző nyilatkozat – intézeti fejléces papíron

Sample size

Statistical Analysis method

Data Table 1

Preparation is the responsibility of the physician (medic)!

Structure (in „excel“):

- should contain only the data
(no mean, sd... calculations)
- cases in rows, variables in columns
- 1 variable 1 column – not 2 for different outcomes
eg. use „sex” as variable
Do NOT; 2 variables as male and female (yes-no for both)
- all data for the same cases should be in one sheet
- do NOT leave empty row or column
- do NOT merge cells
- colouring looks nice, but meaningless

Data Table 2

- Do NOT use special characters (@!"*+-\$#/óéá...) neither for variable names and variable values
- use 1 language (recommended: english)
- name of variable should be short, clear, informative
- values of categorical variables should be short, clear, informative
- 1 cell: 1 type of data and 1 data
- do not use numbers and categories at same variable
do not include text in a numeric variable (eg. <60)
may empty it and generate new variable

Data Table 3

- use a variable for identify cases
(Do NOT use the name of patients)
- use everywhere the same and appropriate decimal separator
- I recommend to instead of a date, enter the time in days (hours ...) relative to a given date
- we clearly indicate whether the value of the data is 0 or missing
(or forgot to type ...)
- the labeling of the missing data should be clear (use different labels if the reason for the absence is different)
- **Do NOT categorize numerical variables!**
 - use the highest scale
- make an additional file where you detail the data table abbreviations

See more on practice

Survey

„Systematic way of data collection“

Surveys

Common aim

For assessing outcomes which are not easily measures or observed, or have complicated structure

- e.g. behaviour, health condition

We usually could not do a perfect one, only a not too bad.

Surveys

Most important

- **Think – common sense**
- Plan ahead
 - sample size, methods of data collection, ethic and law
- How others have done
 - there are several surveys for many things
- Try it

Other common sources of error

- target population distribution vs distribution of respondents (eg family)
- Distribution of respondents vs who you reach (eg online survey, but no net)
- to whom you reach vs complete it (answering)
- complete it - understood and “correctly” completed

Survey tips 1

- Ethical and legal test at the beginning
- Aim of the survey at the beginning
- Contact details in the beginning and in the end (if there is a question)
- **As short as possible**
- Question order:
 - first demographic description (easy to fill)
 - main questions
 - less important questions
 - „sensitive” questions (eg. income) (could has negative impact on fill)
 - final questions: ask if the survey was long, easy
- Different questionnaires regarding to the order should be considered
 - (+ swapping questions and answers)
 - kind of "randomization"

Survey tips 2

- The questions should precisely refer to what we are interested in precise definitions (previous year mean income instead of income) 1 situation (strong and fast)
- Offer *I don't know* or *Not relevant* options as answers
- Include close questions
- In multiple answer questions rather give Yes and No options (than the whole list of multiple options) as they do not read through the whole list (first one has larger chance)
- use a scale bar instead of a categorical scale



- **Do NOT categorize** if the answer could be a number (eg. sleeping time in hours instead of <5, 5-6, 6-7, >7)

Try it!

The process of creating a survey

1. Draft
2. Preliminary survey
3. Testing
- 1-3. Repetition
4. Final version

On smaller groups which are similar to the target populatio – from where you can get feedbacks for the following

- are the questions understandable
- Everybody understands the questions in the same way
- Are there any other possible answers