Principles of Biostatistics and Informatics

Lectures II.

2019. October 26.

VERES Dániel

Part I.
Repetition

Population and Sample

Population – we are interested in

The size of the **population** usually does not allow the examination of all of its elements.

Uncertainty!

Characteristics of the sample can be used to draw conclusions on the population.

Random ()

Sample – that we have…

Therefore, only a subset of the population is examined. That is what we call a **sample**.

We carry out measurements on the sample elements, then this data set (which is also called **sample**) will be characterized by graphs and numbers

Hipothesis tests

One way to handle *sampling („random") error*

Aim of hypothesis tests: **Statistical answer on YES/NO question**

Steps of a hypothesis test, example

Situation: We play a board game with a dice – we do not win...
This is a „wrong" dice?

Question: The probability of six-throw is different from 1/6, even bigger?

Null hypothesis: H_0: The probability of rolling 6 is 1/6.

Significance level: 10%

Sample: 6 times 6 out of 24 rollings. Differ from H0 only because random error?

Is the difference important at all?– RELEVANT? : 1/4 probability, that **1,5** times higher than 1/6 – YES it is

How much evidence *against H0* (indirect proof) based on the sample?
– p-value: 0,1995

6x or higher number of rolling 6, **IF H_0 IS TRUE**

Generally: the probability of having a corresponding (to the sample) or more extreme value if the null hypothesis is true

Steps of a hypothesis test, example

Situation: We play a board game with a dice – we do not win...
This is a „wrong" dice?

Question: The probability of six-throw is different from 1/6, even bigger?

Null hypothesis: H_0: The probability of rolling 6 is 1/6.

Significance level: 10%

Sample: 6 times 6 out of 24 rollings. Differ from H0 only because random error?

Is the difference important at all?– RELEVANT? : 1/4 probability, that **1,5** times higher than 1/6 – YES it is

How much evidence *against H0* (indirect proof) based on the sample?
– p-value: 0,1995

Decision: there is not enough evidence for reject H_0– accept H_0

		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-β)

Part II.
Frequently used Hypothesis tests

What am I curious about?

1. Location parameters: eg. Difference in expected values (mean, median)?

2. Scale parameters: difference in standard deviation, variance?

3. Distributions: difference in distributions?

4. Frequencies, probabilities: difference in probability (frequency, count)?

One sample Student t-test

What I'm curious about

Expected value of the sample is equal with a known population mean

Type of variable

1 numerical and continous

Assumption

Independent observations in the sample

distribution of **means** is normal:

normally distributed sample or large sample size (CLT)

Notes: Calculation:

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$$

Do NOT test normality with other hypothesis test (it increases Type I. error, „multiplicity" – see later)!

Use previous knowledge on the variable and make graphs.

Paired Student t-test

What I'm curious about

Two expected values in two groups are equals – in paired groups

Type of variables

1 numerical and continous,

1 categorical with 2 outcomes (binary):

usually called grouping variable („groups")

Assumptions

Independent observations in the groups,

Repeated measures on cases – means dependent, „paired groups"

the distribution of the difference of means is normal

normally distributed differences or large sample size

Notes:

calc.: same as one sample t-test (sample=difference of groups)

suitable to compare other location parameters (quantiles)

difference of the means = mean of the differences

2 sample Student t-test

What I'm curious about

Two expected values in two groups are equals.

Type of variable

1 numerical and continous,

1 categorical with 2 outcomes (binary)

Assumptions

Independent observations between and within groups

distribution of means is normal in each group:

distribution is normal in each group or large sample size

distribution of standard deviations are the same

Notes:

suitable to compare other location parameters (quantiles)

if we don't know the variances do not test (multiplicity) – use Welch test instead!

Welch test

What I'm curious about

Two expected values in two groups are equals.

Type of variable

1 numerical and continous,

1 categorical with 2 outcomes (binary)

Assumptions

Independent observations between and within groups

distribution of means is normal in each group:

distribution is normal in each group or large sample size

Notes:

suitable to compare other location parameters (quantiles)

not sensitive for different variances (*robust* for variance differences)

F-test

What I'm curious about

Two theoretical variances in two groups are equals.

Type of variable

1 numerical and continous,
1 categorical with 2 outcomes (binary)

Assumptions

Independent observations between and within groups
normal distribution in each group
large sample size is not enough here!

Notes

variance is a scale parameter

Wilcoxon matched-pairs signed-ranks test

What I'm curious about

The examined distribution is symmetrical to a given value.
Assuming that the distribution is symmetrical then we can examine the equivalence of expected values

Type of variables

1 numerical (or ordinal) and continous,
1 categorical with 2 outcomes (binary)

Assumptions

Independent observations in the groups,
Repeated measures on cases – „paired groups" (so 1 sample)
No assumption for normal distribution – „nonparametric test"

Notes

Most often, we ask about the medians, (but symmetry!)
Difference of medians is not equal with median of differences!
(Here we test the latter.)
Using ranks in calculation.

Mann-Whitney U test (Wilcoxon rank sum test)

What I'm curious about

The examined distributions are the same.
If we assume that the two distributions differ only in shift (similar width and shape of distribution), then we can examine the equivalence of expected values.

Type of variables

1 numerical (or ordinal) and continous,
1 categorical with 2 outcomes (binary)

Assumptions

Independent observations in the groups and between groups,
No assumption for normal distribution – „nonparametric test"

Notes

Most often, we ask about the medians, (but only shift differnce!)
Difference of medians is not equal with median of differences!
(Here we test the first.)
Using ranks in calculation.

Chi-square test for fitting

What I'm curious about

An unknown and a known distributions is equal.

Type of variable

1 categorical or categorized variable.

Assumptions

Independent observations
None of the „expected" frequencies smaller than 1 and maximum 20% smaller than 5.

Chi-square test for independence

What I'm curious about

Two (categorical/categorized) variables are depends on each other – same relative frequencies.

Type of variable

2 categorical variable

Assumptions

Independent observations

None of the „expected" frequencies smaller than 1 and maximum 20% smaller than 5.

Note

Reminder: P(A and B) = P(A)*P(B) then and only then:

A independent from B!!

Fisher test for independence

What I'm curious about

Two (categorical/categorized) variables are depends on each other – same relative frequencies.

Type of variable

2 categorical variable

Assumptions

Independent observations

(For big contingency tables it needs more calculation than chi-square.)

Notes

Pearson
Hypothesis test
Assumption: this is one of several experiment: Long-term goodness for type I. and II. errors! (since 1 sampling does not make it possible to decide „how much H0 is true")
Decision: yes/no with a given limit!
No meaning of p-**value**, only <> significance level! (distribution of P-value is uniform if H_0 is ture!)
P-value is an *error probability* (if H_0 is true, the probability of having a corresponding or more extreme sample)

„ Are we interested in whether the suspect is innocent or is it that we only send a few innocent prisoners in the long run?" (F.T.)

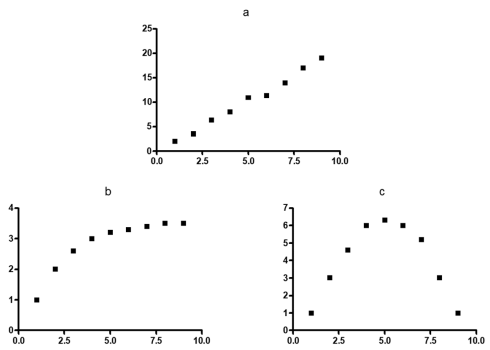
Both of them together are **NOT** working!!

Part III.
Correlation and Regression

Relation between variables

type of *relation*:

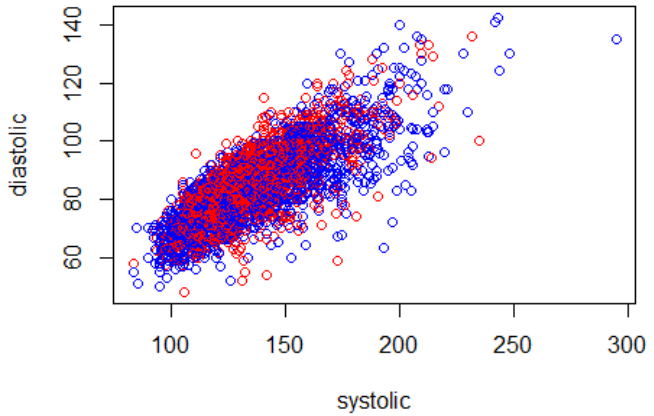
- monotonic
  - – positive
  - – negative
  - – linear positive
  - – …
- not monotonic
  - – parabolic
  - – …
- no relation



Correlation

Monotonic,

symmetric (changing together, not „dependent and independent”) relation of **2**, **random** (random error, not setted – like a dose) variable.



Correlation

Expressing **strength** of correlation by:

Correlation coefficients (r):

if **linear relation** is assumed: **Pearson** r

if **monotonic** (not necesserily linear): **Spearman rank** r,

Value of Correlation coefficients :

-1 to +1

negative: negative correlation

positive: positive correlation

closer to |1|: higher strength of correlation

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} = \frac{s_{xy}}{s_x s_y}$$

„Distance from middle” – both in y and in x direction

„Correlation” t-test (on Pearson r)

What I’m curious about

Two numerical variables are linearly correlated (r not=0).

Type of variable

2 numerical variable (X and Y)

Assumptions

Independent observations for pairs

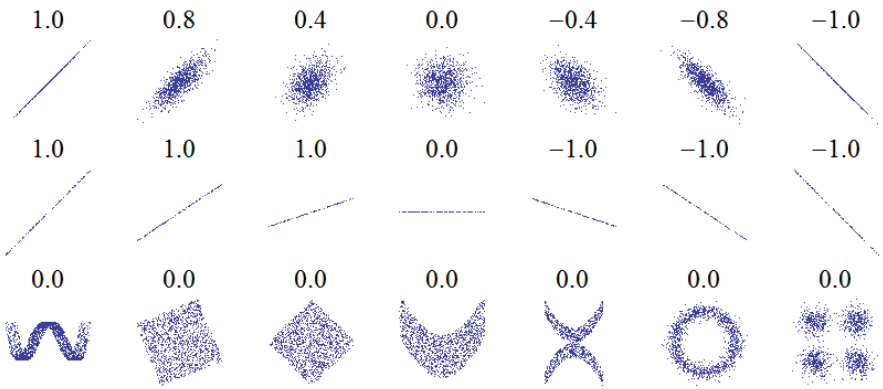
symmetrical, linear relation assumed

x and y are random variables

Notes

Notes

- *GRAPH!!!!;*
- *Correlation NOT egual with causality*
  ☺eg: http://www.fastcodesign.com/3030529/infographic-of-the-day/hilarious-graphs-prove-that-correlation-isnt-causation



Regression

Function relation (NOT symmetric) between a dependent (outcome, result, Y) variable and an independent (explanatory, predictor, X) variable(s). [Y is a random variable, X not necessarily]

Y depends on X – assumed **knowledge,** not a statistical assumption.

Questions:

- is there a (given kind of) relation? (statistical relation, not causality)
- what is the value of Y if X is:…? (estimation)
- what is the value of X if Y is:…?
- what is the best function that describe the relation?

Linear regression

Linear function relation assumed.

For 2 variables correlation and regression questions could be „transformed" to each other

Linear regression

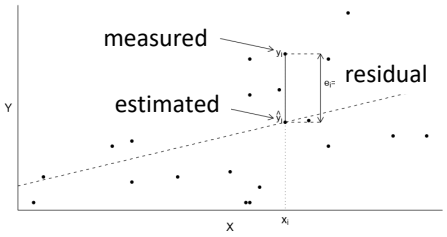
Estimation of the linear: OLS (Ordinary Least Square method)

Intercept — Slope

Linear function: $Y = \beta_0 + \beta_1 * X + \epsilon$

Error term; residuals: points-line **vertical** differences (difference of measured and estimated values)

Best line for OLS: where the sum of the square of vertical differences are the smallest



„Correlation” t-test (linear regression)

What I’m curious about

Y linearly depends on X.

Type of variable

2 numerical variable (X and Y)

Assumptions

Independent observations for pairs
linear relation assumed
x values measured with no error
residuals are normally distributed with constant variance
(it is a consequence if X and Y are both normally distributed)

Notes

results: 2 estimates: slope and intercept,
slope is the more important and that is tested (for 1 explanatory variable it has the same result as Pearson r testing)

Slope and R

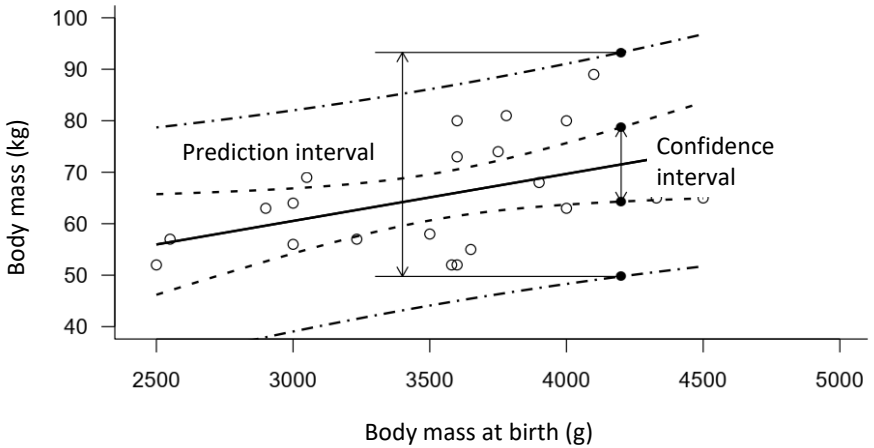
Slope

the change in Y if X changed by 1 unit

R^2 – coefficient of determination

how much of variability of Y explaned by X

Confidence and prediction intervals



Part IV.
Multiplicity and ANOVA

…Chocolate Helps Weight Loss.

„A team of German researchers had found that people on a low-carb diet lost weight 10 percent faster if they ate a chocolate bar every day. It made the front page of Bild, Europe's largest daily newspaper…"

„*statistically significant* benefits of chocolate that we reported are *based on the actual data*"
HOW??

…Chocolate Helps Weight Loss.

„…randomly assigned the subjects to one of three diet groups. One group followed a low-carbohydrate diet. Another followed the same low-carb diet plus a daily 1.5 oz. bar of dark chocolate. And the rest, a control group, were instructed to make no changes to their current diet.

„Our study included 18 different measurements—weight, cholesterol, sodium, blood protein levels, sleep quality, well-being, etc.—from 15 people."

…Chocolate Helps Weight Loss.

„…randomly assigned the subjects to one of three diet groups. One group followed a low-carbohydrate diet. Another followed the same low-carb diet plus a daily 1.5 oz. bar of dark chocolate. And the rest, a control group, were instructed to make no changes to their current diet.

„Our study included 18 different measurements—weight, cholesterol, sodium, blood protein levels, sleep quality, well-being, etc.—from 15 people."

…Chocolate Helps Weight Loss.

Usually used significance level: 5%, which means that there is just a 5 percent chance that our result is a random fluctuation if H0 is true. (Remember: p if H0 is true has a uniform distribution)

Error probability for a test:	p
No error for 1 test:	$1-p$
No error for k independent test:	$(1-p)^k$
Error at least 1 case for k independent test:	$1 - (1-p)^k$

If k = 18 it means 60% probability to have at least 1 error! (p=5%)

Called: multiplicity problem = alpha inflation =…

I Fooled Millions Into Thinking Chocolate Helps Weight Loss. Here's How.

- Multiplicity
  - ☺*eg: Chocolate Helps Weight Loss*
  - *https://io9.gizmodo.com/i-fooled-millions-into-thinking-chocolate-helps-weight-1707251800*

Multiplicity

- Multiplicity
  - *> 1 variable*
  - *>2 group* ***- ANOVA***

ANOVA: Analysis Of Variances (to compare means)
(see: ANOVA excel file)

ANOVA

What I'm curious about

At least 1 expected value in more groups is different.

Type of variable

1 numerical and continous,
1 categorical with >1 outcomes

Assumptions

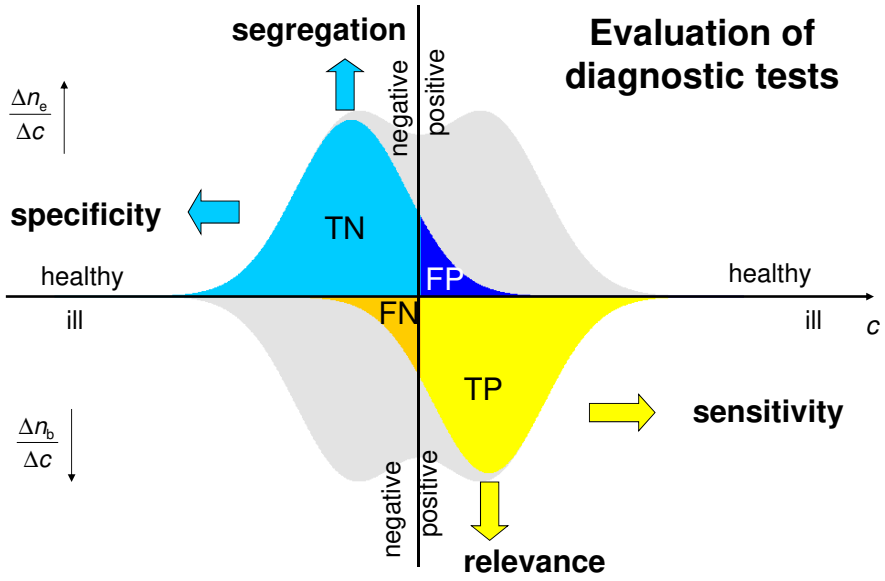
Independent observations between and within groups
distribution of means is normal in each group:
distribution is normal in each group or large sample size
standard deviations are the same

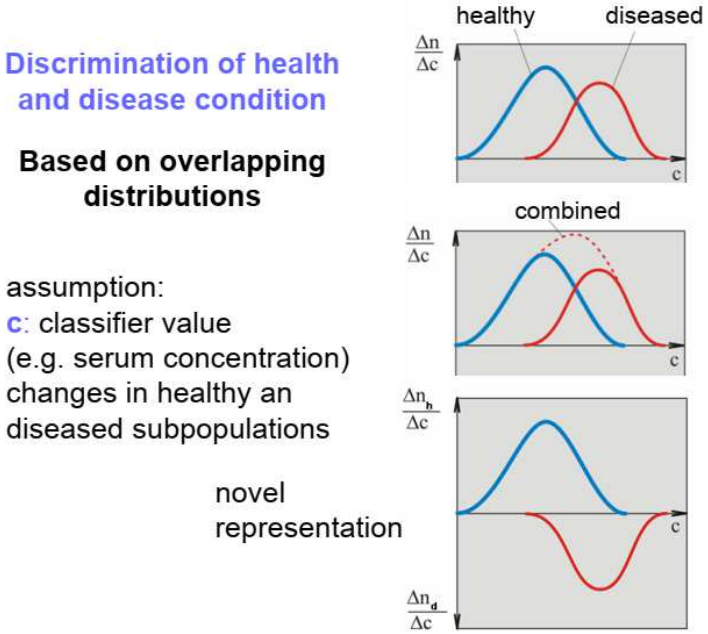
Notes:

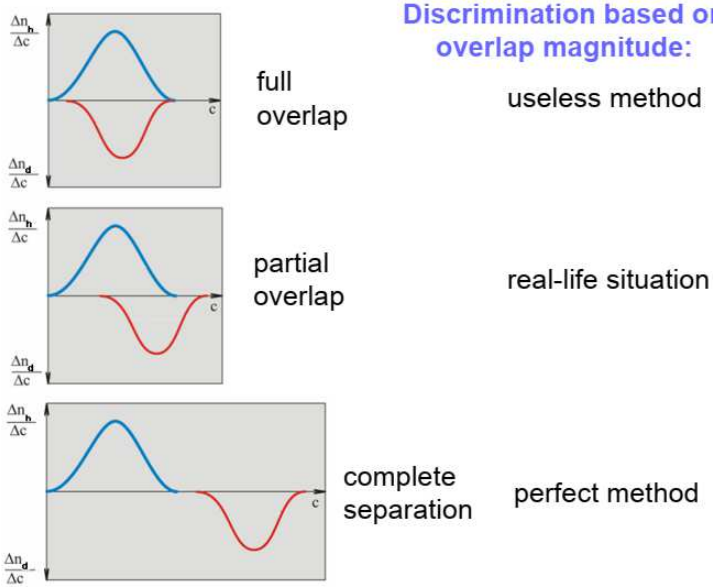
it is an F-test
it tells: at least 1 expected value is different
- which? – Post hoc tests

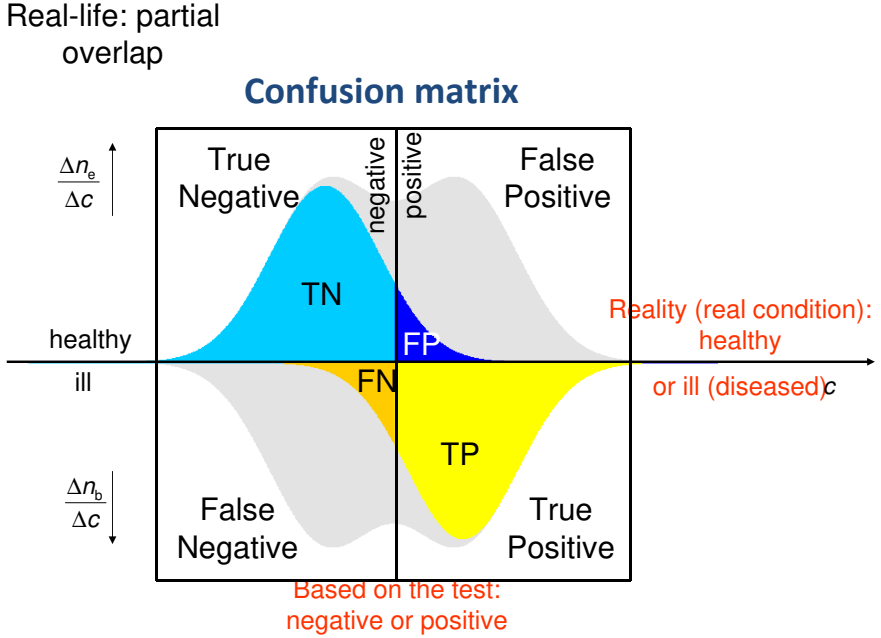
Part V.
Diagnostic tests

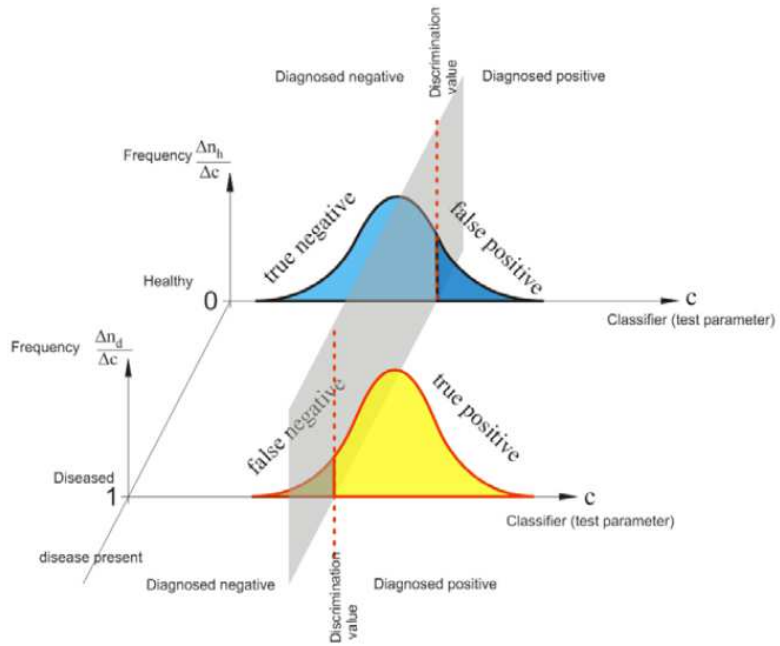
Slides from András Kaposi


Evaluation of diagnostic tests
segregation
negative
positive
specificity
TN
FP
FN
TP
healthy
ill
healthy
ill
c
sensitivity
relevance
negative
positive


Discrimination of health and disease condition
Based on overlapping distributions
assumption:
c: classifier value
(e.g. serum concentration)
changes in healthy an diseased subpopulations
novel representation
healthy
diseased
combined
c

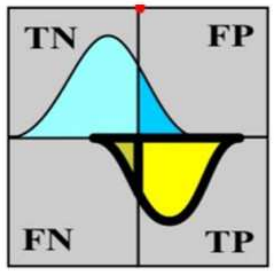

Discrimination based on overlap magnitude:
full overlap
useless method
partial overlap
real-life situation
complete separation
perfect method


Real-life: partial overlap
Confusion matrix
True Negative
False Positive
negative
positive
TN
FP
FN
TP
healthy
ill
Reality (real condition): healthy or ill (diseased)
False Negative
True Positive
Based on the test: negative or positive



Prevalence

= frequency of diseased in examined population
= probability prior to test
= a-priori-probablity

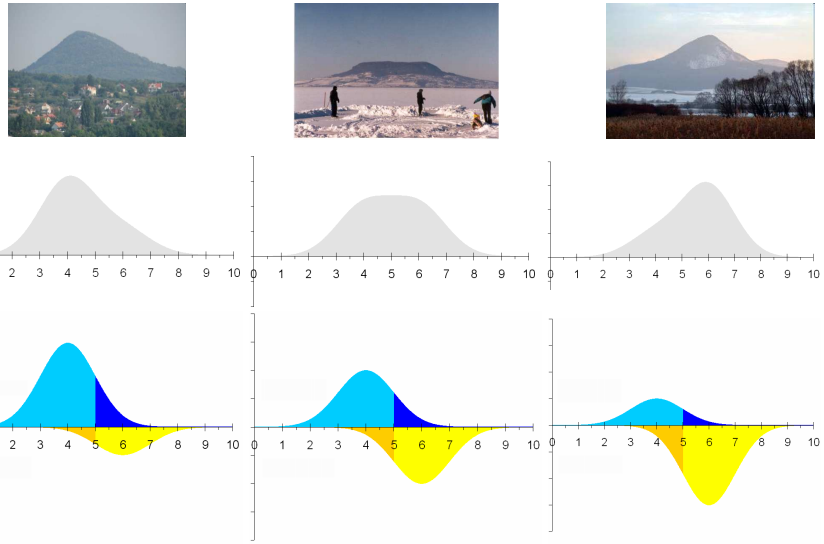


measure of how common the disease is

$$\frac{\nabla}{\blacktriangle} = \boxed{w} = \frac{\text{diseased}}{\text{total}} = \frac{\boxed{\text{TP} + \text{FN}}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}} = \frac{\text{de} - \text{sp}}{\text{se} - \text{sp}}$$

Incidence

Rate of new cases of disease/year/person in a given population
(e. g. per year, per 10 000 people)



w = 25% w = 50% w = 75%

Parameters of diagnostic „goodness"

The goodness of a test can be described in terms of the following diagnostic parameters

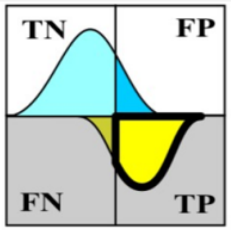
Sensitivity
Specificity
PPV, relevance
NPV, segregation

Every method must be compared with a reference-method (gold standard)

Gold standard: method known to always work; often autopsy

Diagnostic **sensitivity**

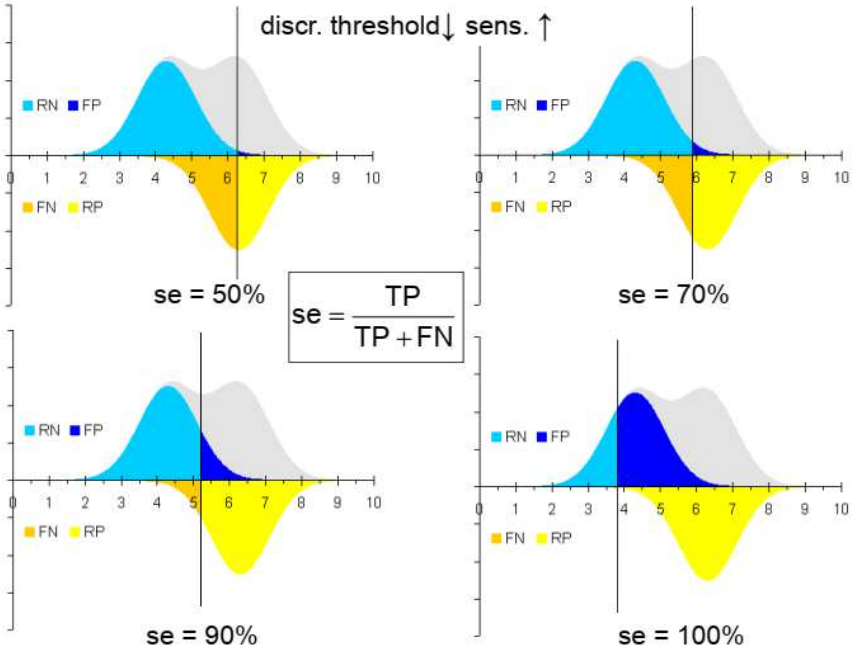
= positive within diseased
= true postive rate
= recall rate



probability that the test finds the diseased positive

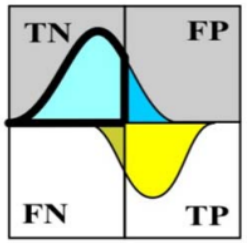
$$\frac{\text{diseased}}{\text{diseased}} = \boxed{\text{se}} = \frac{\text{true positive}}{\text{diseased}} = \boxed{\frac{TP}{TP+FN}} = p(\text{positive}|\text{diseased})$$

High-sensitivity tests are required:
In early diagnosis (screening) so that few patients remain unrecognized.
If the risk of disease is higher than the risk of treatment.



Diagnostic **specificity**

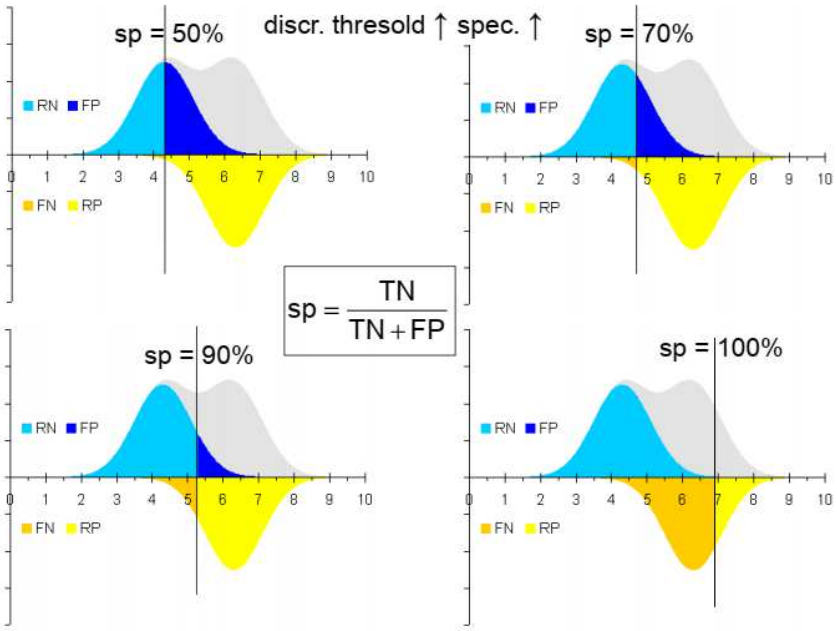
= negative among healthy
= true negative rate



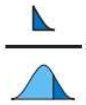
probability that the test finds a healthy negative

$$\frac{\text{healthy}}{\text{healthy}} = \boxed{\text{sp}} = \frac{\text{true negative}}{\text{healthy}} = \boxed{\frac{TN}{TN+FP}} = p(\text{negative}|\text{healthy})$$

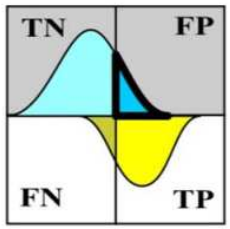
High-specificity tests are important:
When the false positive values have severe consequences (e.g. surgery).
When the risk of treatment is higher than the risk of disease.



Diagnostic False Positive Rate (Type-I error)

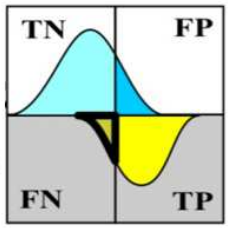


$$= 1 - \text{sp} = \frac{\text{FP}}{\text{healthy}} = \frac{\text{FP}}{\text{TN} + \text{FP}} = p(\text{positive}|\text{healthy})$$



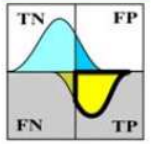
Diagnostic False Negative Rate (Type-II error)

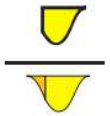
$$= 1 - \text{se} = \frac{\text{FN}}{\text{diseased}} = \frac{\text{FN}}{\text{FN} + \text{TP}} = p(\text{negative}|\text{diseased})$$



Horizontal rates are independent of prevalence

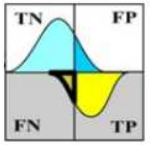
sensitivity (se)

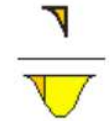




$$\text{se} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

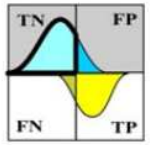
false negative rate (1-se)

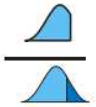




$$1 - \text{se} = \frac{\text{FN}}{\text{FN} + \text{TP}}$$

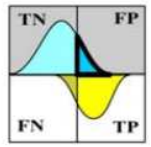
specificity (sp)

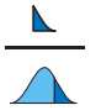




$$\text{sp} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

false positive rate (1-sp)



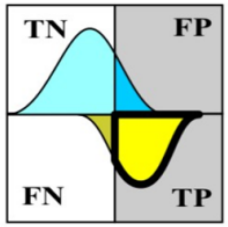


$$1 - \text{sp} = \frac{\text{FP}}{\text{TN} + \text{FP}}$$

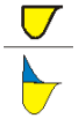
Predictive values (vertical rates)
a-posteriori-probabilities; they depend strongly on prevalence

Positive predictive value

= PPV
= predictive value positive
= PVP
= diagnostic **relevance**
= diseased among positive



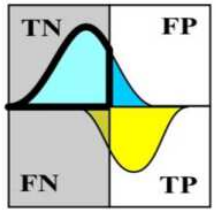
probablity of disease if test is positive



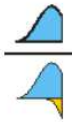
$$= \boxed{\text{PPV}} = \frac{\text{TP}}{\text{positive}} = \boxed{\frac{\text{TP}}{\text{TP} + \text{FP}}} = p(\text{diseased}|\text{positive}) = \frac{\text{se} \cdot \text{w}}{\text{se} \cdot \text{w} + (1 - \text{sp}) \cdot (1 - \text{w})}$$

Negative predictive value

= NPV
= predictive value negative
= PVN
= diagnostic segregation
= healthy among negatives



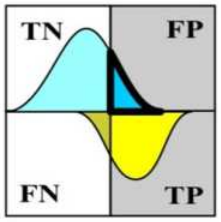
probabilty of health if test is negative



$$= \boxed{\text{NPV}} = \frac{\text{TN}}{\text{negative}} = \boxed{\frac{\text{TN}}{\text{TN} + \text{FN}}} = p(\text{healthy}|\text{negative}) = \frac{\text{sp} \cdot (1 - \text{w})}{\text{sp} \cdot (1 - \text{w}) + (1 - \text{se}) \cdot \text{w}}$$

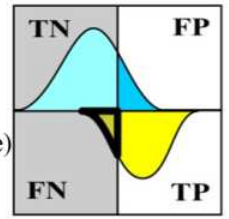
False alarm rate

$$= 1 - \mathrm{PPV} = \frac{\mathrm{FP}}{\text{positive}} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TP}} = p(\text{healthy}|\text{positive})$$



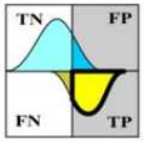
False reassurance rate

$$= 1 - \mathrm{NPV} = \frac{\mathrm{FN}}{\text{negative}} = \frac{\mathrm{FN}}{\mathrm{FN} + \mathrm{TN}} = p(\text{diseased}|\text{negative})$$



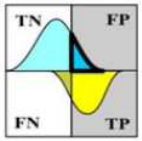
Vertical rates are dependent on prevalence

positive predictive value (PPV)



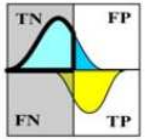
$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{FP} + \mathrm{TP}}$$

false alarm rate (1-PPV)



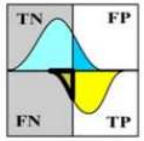
$$1 - \mathrm{PPV} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TP}}$$

negative predictive value (NPV)



$$\mathrm{NPV} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FN}}$$

false reassurance rate (1-NPV)

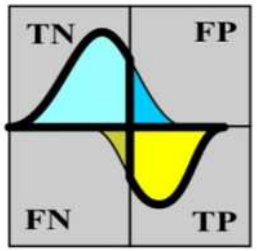


$$1 - \mathrm{NPV} = \frac{\mathrm{FN}}{\mathrm{TN} + \mathrm{FN}}$$

Diagnostic **efficacy/efficiency**

also called as **accuracy**

= probability of correct diagnosis



$$= \mathrm{de} = \frac{\mathrm{TP} + \mathrm{TN}}{\text{total}} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TN} + \mathrm{FP} + \mathrm{FN} + \mathrm{TP}} = \mathrm{se} \cdot \mathrm{w} + \mathrm{sp} \cdot (1 - \mathrm{w})$$

often: discrimination threshold is chosen so that **de** is maximized

Summary table

sensitivity	se	$\frac{\mathrm{VP}}{\mathrm{VP} + \mathrm{ÁN}}$	$p(P\|B)$	Positive among ills	TPR (True Positive Rate)
specificity	sp	$\frac{\mathrm{VN}}{\mathrm{VN} + \mathrm{ÁP}}$	$p(N\|E)$	Negative among healthy	TNR (True Negative Rate)
false neagative rate	1-se	$\frac{\mathrm{ÁN}}{\mathrm{VP} + \mathrm{ÁN}}$	$p(N\|B)$	Negative among ills	FNR
false positive rate	1-sp	$\frac{\mathrm{ÁP}}{\mathrm{VN} + \mathrm{ÁP}}$	$p(P\|E)$	Positive among healthy	FPR
relevance	PPV	$\frac{\mathrm{VP}}{\mathrm{VP} + \mathrm{ÁP}}$	$p(B\|P)$	Ill among positives	
segregation	NPV	$\frac{\mathrm{VN}}{\mathrm{VN} + \mathrm{ÁN}}$	$p(E\|N)$	Healthy among negatives	
false alarm rate	1-PPV	$\frac{\mathrm{ÁP}}{\mathrm{VP} + \mathrm{ÁP}}$	$p(E\|P)$	Healthy among positives	
false reasurrance rate	1-NPV	$\frac{\mathrm{ÁN}}{\mathrm{VN} + \mathrm{ÁN}}$	$p(B\|N)$	Ill among healthy	

Prevalence independent

a-posteriori probabilities

Effect of prevalence

NPV = 90%

case1: ***w* = 50%**

			Test	
			negative	positive
sp = 90%	Gold-standard	healthy	90	10
		diseased	10	90

se = 90%

(de = 90%)

PPV = 90%

NPV = 99%

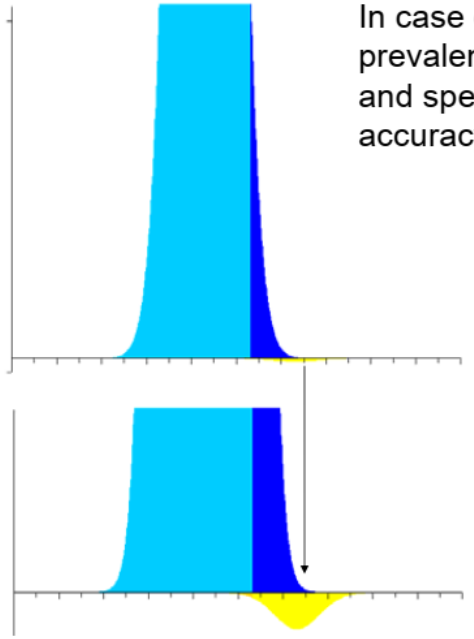
Case 2: ***w* = 10%**

			Test	
			negative	positive
sp = 90%	Gold-standard	healthy	810	90
		diseased	10	90

se = 90%

(de = 90%)

PPV = 50%



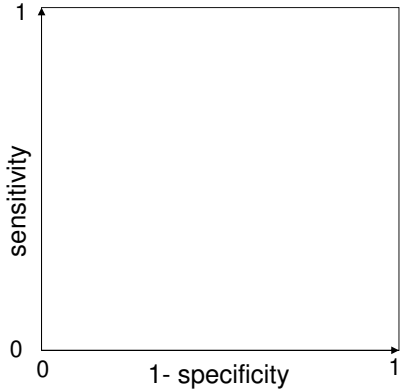
In case of very small prevalence a highly sensitive and specific test have low accuracy (PPV).

prevalence = 0.1 %
sensitivity = 98 %
specificity = 98 %

PPV = 4 %

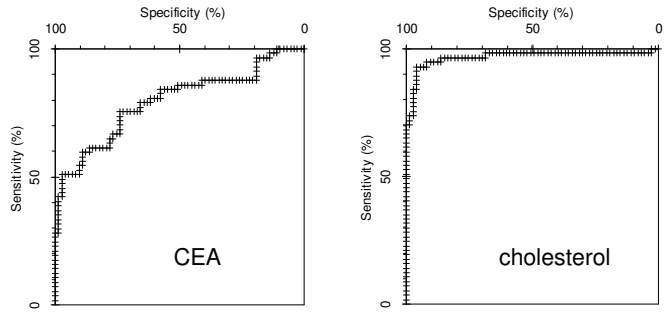
Comparison of diagnostic tests: ROC curves

ROC: receiver-operator (operating) characteristic



Hasüregi folyadékgyülemek daganatdiagnosztikája

Increased CEA and/or cholesterol concentrations in ascites are diagnostic markers for carcinomatosis



Which method is better? What discrimination threshold should be used?

Gulyás M, Kaposi AD, Elek G, Szollár LG, Hjerpe A, Value of carcinoembryonic antigen (CEA) and cholesterol assays of ascitic fluid in cases of inconclusive cytology, J Clinical Pathology 2001 (54) 831-835

Part VI.
Concept of Information

Concept of Information

Less frequent events (events with smaller probability) has **higher information** content.

Single event: $I = \log_2\left(\frac{1}{p}\right)$

Part VII.
Clinical studies

Evidence Based Medicine (EBM)

A series of conscientious, unambiguous and logical decisions based on the evidence currently available that serves the patient's personalized treatment.

Known results of „statistical trials”
Personal knowledge, clinical practice
„Common sense”

Biostatistics – why to learn?

- „ To decide whether we should believe in something we are reading or to see where the mistake is, that is to say, do not fall so easily into statistical „juggling", artifacts and mistakes. (see excel – panacea,…)
- „To judge better whether we were lucky or not – or none of them…."
- „ To judge better what is worth , whether it is worth for risking it…"

- „ So that we can do our best to design and evaluate our own statistics in our work (diploma…)."
- „I got an interested, unexpected result? I just discovered something or just the game of chance I see?"
- „ To make our results more understandable and effective, we can highlight the essence. "
- „ To have a clear understanding of the literature. "

(J. Reiczigel)

Known results of „statistical trials" – how they will be?

- We collecting data and analysing them.

Known results of „statistical trials" – how they will be?

- ~~We collecting data and analysing them.~~
- HOW TO GET DATA? – first PLAN it!

Considerational considerations :

- What is the aim, the question?
- What „mistakes" should be considered?
- How much should the sample size be?
- What methods can be applied?
- Which sampling techniques are available?
- ... So THEN collect data…

Known results of „statistical trials" – how they will be?

- ~~We deal with existing, existing data that we have collected.~~
- HOW TO GET DATA?

Considerational considerations :

- What is the aim, the question?
- What „mistakes" should be considered?
- What methods can be applied?
- Which sampling techniques are available?
- How much should the sample size be?
- ... So THEN collect data…

- **The most sophisticated, most accurate data analysis does not compensate for a poorly planned, designed or executed data collection or survey !!!!!!!!**

What is the aim?

Is there any difference?
Is there any relation (correlation)?
Is there any effect?
…

Aim + Relevant?

Is there any difference?
Is there any relation (correlation)?
Is there any effect?
…

There may be a „difference" – but is it relevant (clinically important)?
Based on clinical practice – that is NOT A STATISTICAL QUESTION – but it is really IMPORTANT

– How much is the difference…? EFFECT SIZE
…

Effect size

There could be a difference, but is it relevant?
– How much is the „difference"? : How to express it?

Difference between means, medians; ratio of means, medians
How much (how many times) does it change in another group?

Correlation, determination coefficient
To what extent does the change in y affected by the change in x?
slope
If the x (independent) variable increased by 1 unit what will be the average change in the y (dependent) variable.

Odds ratio, Risk ratio
How many times does the odds or risk increased if the risk factor present?
…

Error 1 – why? Significance

Problem: we couldn't examine everybody (the population)!
Solution: sampling – but…
…sampling error (chance)… Hypothesis testing!
The observed effect could be by chance?
Is the „difference" significant?
(*see excel, radiation)

Effect size and „significance” together

Confidence interval!

(NOTE: learn this carefully – we like to ask it in the exam!)

If there is a difference can we recognize it?

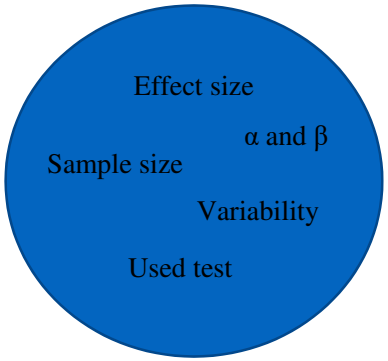
		In population (in reality) the null hypothesis is:	
		True	False
Decision on null hypothesis:	Accepting (Not rejecting)	Good decision	Error (**type II**) (β) (false negative result)
	Rejecting	Error (**type I**) (α) (false positive result)	Good decision **power(1-β)**

(NOTE: learn this strictly – we like to ask it in the exam!)

Power

Is there a difference can we recognize it?
- it can be recognized easier, if:
  - higher sample size
  - high effect size
  - (may be others: small SD…)
  - high power hypothesis test:
    - higher scale!!!
    - normal distribution
    - „effectively” paired

Power **(1-β)**



They depend on each other.

Relevant, but not significant…

Reasons:

small power:
- small sample size (limitation: money, ethical issues)*
- large variability
- less powerfull statistical test

we could not measure it accurately
violated assumptions for the test

Plan ahead!!

we were unlucky (sampling error)

other errors

-*Ask yor statisticians...

(☺ eg: https://www.youtube.com/watch?v=PbODigCZqL8)

Other errors

Effect size based on the sample (the estimate)

True effect size (the estimand)

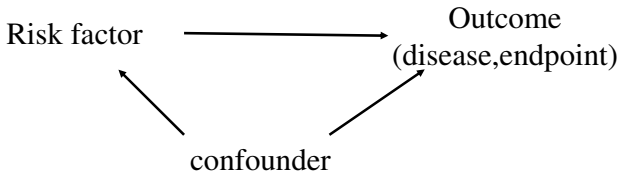
Errors

Bias:
Selection bias
Information bias
confounding bias

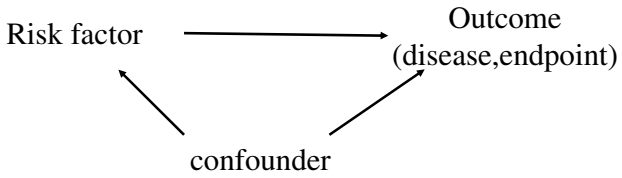
Sampling (random) error

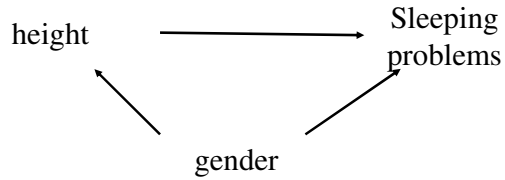
(NOTE: learn this strictly – we like to ask it in the exam!)

Confounding bias



Confounding bias





Most common confounders: gender, age – always think about them!

(NOTE: learn this strictly – we like to ask it in the exam!)

Selection bias, Information bias

Selection bias:
There is a difference between the selected and not selected individuals, or difference between assignment to groups (erroneous selection with respect to an outcome influencing parameter)

tipical: age, gender different in the groups
different population
different follow-up time

Information bias:
erroneous data collection about or from subjects (which affects the outcome)

tipical: recall bias
more careful monitoring for diseased, young

(NOTE: learn this strictly – we like to ask it in the exam!)

Which methods?

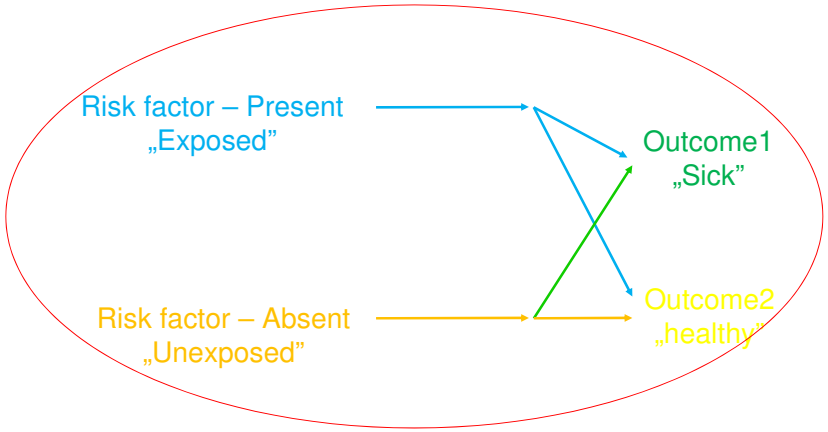
Main type of Clinical Studies (Study Designs):

Observation: no intervention, just observation
Cross-sectional studies – at a given time
Case-control studies – pro-/retrospective
Cohort – pro-/retrospective

Experimental: intervention („treating")
main type: randomized controlled and clinical trials

(NOTE: learn this strictly – we like to ask it in the exam!)

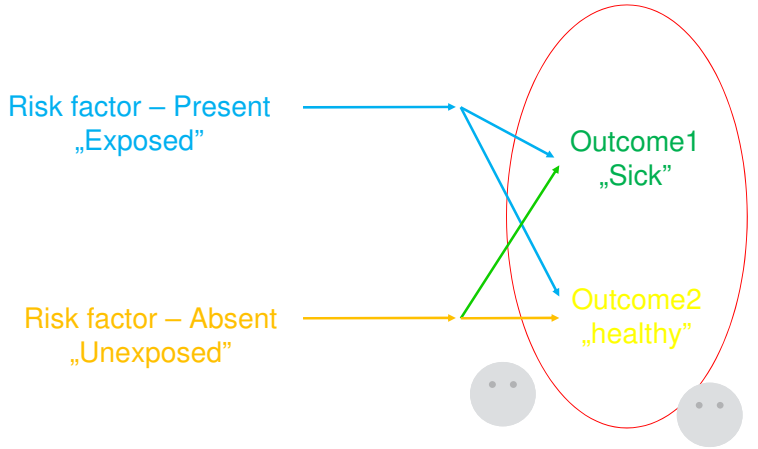
Cross-sectional



Risk factor and outcomes observed at the same time

Case-control

Selection based on the outcomes.

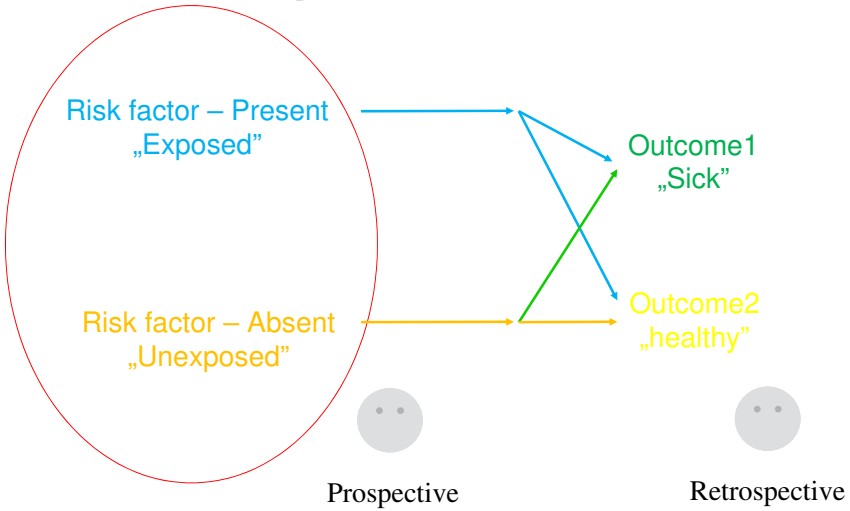


Propective (rare)

Retrospective

Cohort

Selection based on the exposure.



Note

Prevalence - Measures the existence of a disease (or exposure, treatment) at a given point in time (point prevalence)

Incidence - Measures occurence of a disease in a population over a specified period in time (**new cases**!)

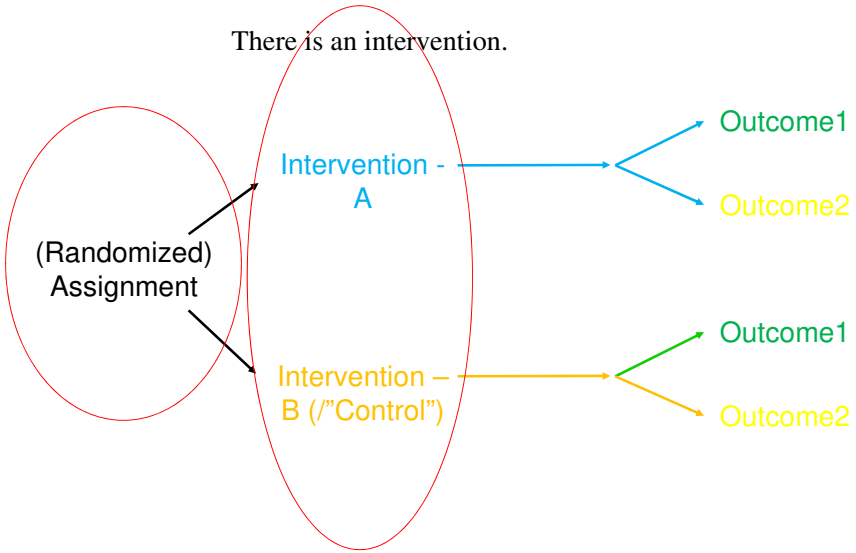
Acute disease: prevalence ~ incidence

chronic: prevalence ? incidence

(NOTE: learn this strictly – we like to ask it in the exam!)

Experimental (RCT)

There is an intervention.



Summary table of study design

	Cross-sedtional	Case-control	Cohort	RCT
Property	Selection at a given time point	Selection based on: outcome (case/control)	Selection based on: risk factor (exposure)	There is an intervention
Advantages	Logistically easier and faster Cheap	Good for rare disease, Logistically easier and faster Less expensive	Good for rare risk factor	Reduced bias
Dis-advantages	No causality	Hard to select controls – selection, information bias (eg. recall bias)	Long follow-up time Information bias (eg. recall bias)	Expensive, logistically hard and slow

(NOTE: learn this strictly – we like to ask it in the exam!)

...

What's the difference between a physicist, a mathematician, and a statistician?

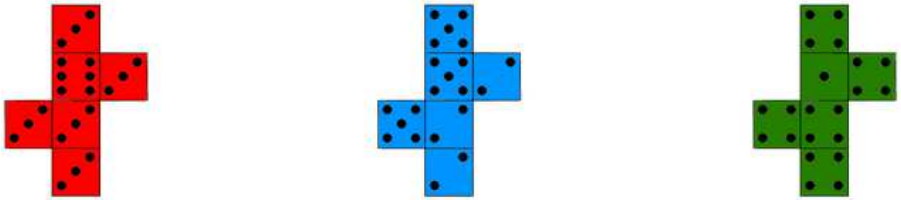
The physicist calculates until he gets a correct result and concludes that he has proven a fact.

The mathematician calculates until he gets a wrong result and concludes that he has proven the contrary of a fact.

The statistician calculates until he gets a correct result about an obviously wrong proposition and concludes NOTHING, because the explanation is the task of the scientist who consulted the statistician.

Source of stat jokes: http://www.ilstu.edu/~gcramsey/Gallery.html

Non-transitivity



*see excel file

Links:

https://en.wikipedia.org/wiki/Nontransitive_dice

http://singingbanana.com/dice/article.htm

https://plus.maths.org/content/taxonomy/term/789