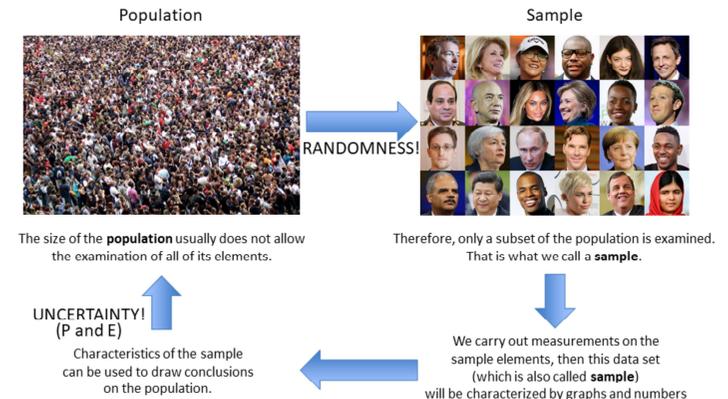


Biophysics I. for dentistry students

Lecture 2nd:
Biostatistics II.
2019. September 16.
Dániel Veres

Repetition: Population and Sample (Problematics of inferential statistics)



As we mentioned before, statistics examines random mass phenomena. This means that during examination of a phenomenon many, if not infinitely many measurements would be possible. The set containing the outcomes of all these theoretically possible measurements is called **population**. Theoretically, the complete understanding of a variable would require the execution of all the possible measurement, but of course it is not possible.

Consequently, we only observe a subset of the population, which is called **sample**. The most evident way of generating this subset is **random selection**.

We carry out measurements on the sample, the set of measurement results is also called **sample**. (That is: in less precise way the sample may be a group of students [individuals, objects] of the university as a population. In more precise way, the population is the height of all people at the university, the sample is the set of height values for

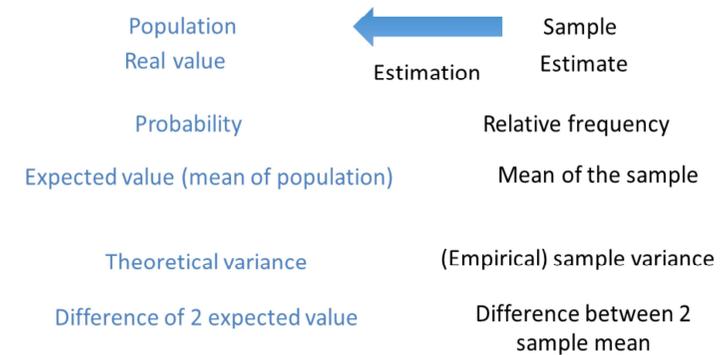
a group that was actually measured.)

The sample might be characterized graphically or numerically as we learned in the last lecture, then the properties learned that way may be extrapolated to the population. E.g. if 25% of people in a group have blood type "A", we may expect the same from the whole population. Since the sample is chosen randomly, it will not necessarily represent the population, the frequency of occurrence of different values within the population perfectly. As a result, every conclusion drawn from a sample carries a burden of **uncertainty**.

What is the quantity of the uncertainty? How to define?

We are using 2 term for it: probability (P) and error E.

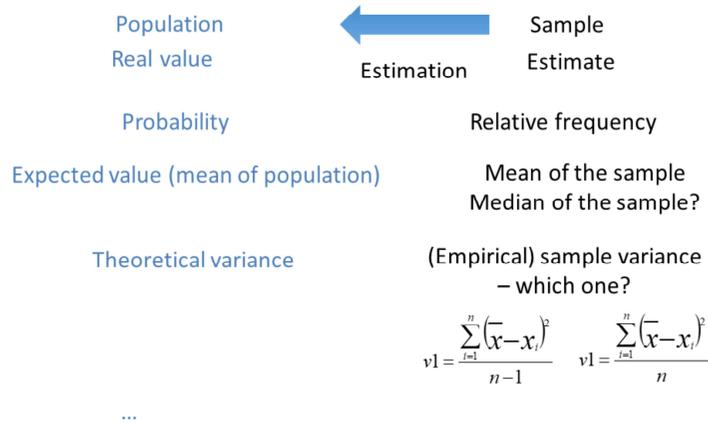
Estimation



So in the inferencial statistics we make estimates and analyze them.

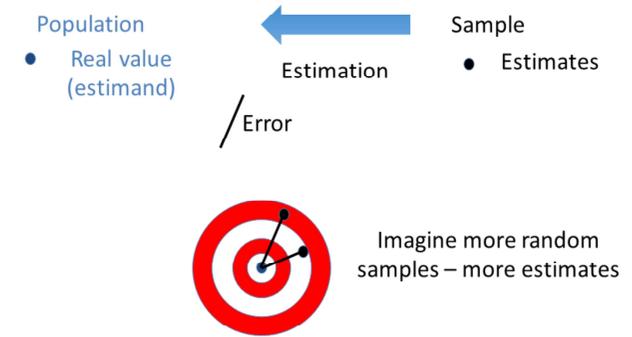
For example, the probability of a given variable outcome in a population is estimated by the relative frequency of the sample. The expected value of the variable is estimated by the mean of the sample, and the variance is estimated by the sample (so called empirical) variance. The difference between the averages of two populations can be estimated by the difference between the sample means (eg whether drug is different - different means).

Estimation



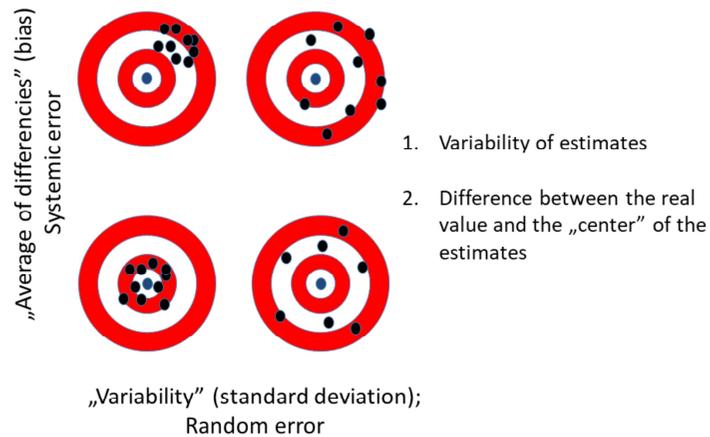
But why don't we estimate the expected value with the median, why is there an empirical standard deviation with (n-1) instead of n (n is the sample size)? (If we were to "translate" the theoretical variance, then it would have to be n.) For the answer we have to consider the following.

Error



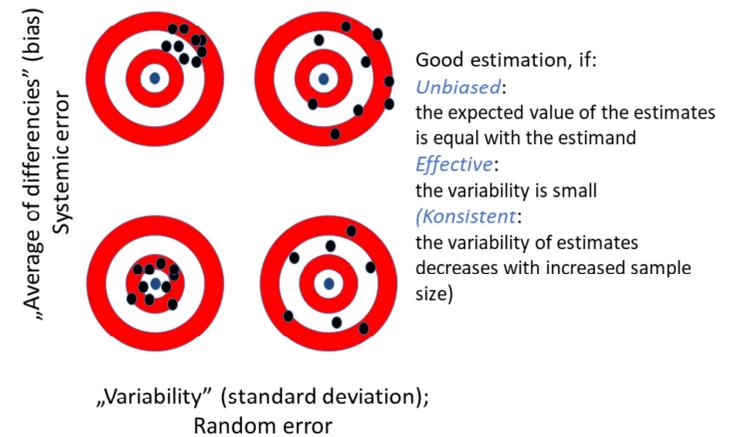
Expected value, standard deviation, difference in the expected value of two populations - all measures of efficacy are estimated parameters. The question is how big a mistake can we make? Imagine estimating a given true value with multiple samples - the samples come from the population by random sampling. The difference between the real value and the estimate is the error.

Error – 2 dimension



The error „has 2 factors“: random error and systematic error. The random error is the variance of the estimates, while the systematic error is characterized by the difference between the „center“ of the estimates and the estimated value.

Error – 2 dimension



The estimates is good, if:

1. *Unbiased*
 2. *Effective*
 3. *Consistent*
- (+ 1 not mentioned...)

Notes

- The mean is an unbiased, consistent and most effective estimate for estimating the expected value
- In the estimation of variance, the division by $n-1$ is unbiased, while division by n will be biased

A few notes

- The mean is an unbiased, consistent and most effective estimate for estimating the expected value
- In the estimation of variance, the division by $n-1$ is unbiased, while division by n will be biased

Confidenc intervals

But we have only 1 estimate and we don't know the real value!
– Could we measure the error?

YES, the random error (sampling error) could be estimated based on the sample! It is called the standard error of the estimate.

HOW?

We have only 1 sample, therefore 1 estimate and we don't know the real value – could we measure the error in this case?

YES! But only the random part! We can estimate random error based on the sample.

Confidence intervals

But we have only 1 estimate and we don't know the real value!

– Could we measure the error?

The variability (random error, sampling error) could be estimated based on the sample!

Example:

The standard error of the mean (shortly standard error, SEM)

Reminder: CLT: Central limit theorem (on sampling): for given conditions, sampling with large sample size (n) the distribution of the sample means is normal with:

$$\text{Var}_{\text{normal}} = \frac{\text{Var}_{(\text{original})}}{n}$$

Let's show an example based on the mean (estimate of expected value).

Reminder: central limit theorem (CLT)

Confidence intervals

1. Therefore the standard error of the mean (the variability of estimating the mean): is the square root of the nth part of the sample variance.
2. We can construct a range, that contains the mean with a given probability: called confidence interval of the mean.
the given probability: called confidence level

In the case of the mean (because of normal estimation – see CLT) eg.:
The limits of the 95% confidence interval of the mean are:

$$\sim \bar{x} \pm 2 * SEM$$

We can construct confidence intervals for other estimats too!
It shows the value of the estimate, its error and its confidence interval

Based on the CLT we have 2 conclusion:

1. the standard error of the mean (the variability of estimating the mean): is the square root of the nth part of the sample variance.
2. We can construct a range, that contains the mean with a given probability: called confidence interval of the mean. The given probability: called confidence level.

For the mean eg. : The limits of the 95% confidence interval of the mean are mean+/-2*SEM.

We can construct confidence intervals for other estimats too! (Nearly for all estimates).

It shows the value of the estimate, its error and its confidence interval.

Hypothesis tests

Sampling („random”) error with an other method

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

HOW to prove a statement?

Direct proof: prove for all cases that the statement is true.

eg: sum of 1,2...n : $(n+1)*(n/2)$ – proving with induction

Indirect proof: assume the opposite statement and I prove that is false.

One main tool in statistics to make decisions with a given error probability is hypothesis tests.

In this chapter I gave detailed description in the slides, therefore I comment it shortly.

In hypothesis tests we would like to give an answer for a YES/NO question. An answer is a statement – but how to prove a statement?

In math we have 2 possibilities:

Direct/indirect proving.

Indirect proof

We have a box containing 100 marbles. Each of them are either red or white.

Case #1: hypothesis (H): all of them are white.

Experiment: We randomly take a marble out of the box.

Our observation: It is red.

Conclusion: The probability of our observation *given our hypothesis* is 0: Our hypothesis is for 100% sure wrong.

Case #2: hypothesis (H): 99 are white and one is red.

Experiment: We randomly take a marble out of the box and put it back; we do this 5 times.

Our observation: All of them are red.

Conclusion: Our hypothesis is for *almost* 100% sure wrong: The probability of our observation *given our hypothesis* is $0.01^5 = 10^{-10}$: practically impossible.

Case #3: hypothesis (H): 50 are white and 50 are red.

Experiment: We randomly take a marble out of the box and put it back; we do this 5 times.

Our observation: All of them are red.

Conclusion: Now we are not sure what to do: The probability of our observation given our hypothesis is $0.5^5 = 0.03125$: low but not that unlikely...

Case #4: hypothesis (H): all of them are red.

Experiment: We randomly take a marble out of the box and put it back; we do this 5 times.

Our observation: All of them are red.

Conclusion: The probability of our observation *given our hypothesis* is $1^5 = 1$. Are we sure what to do now?

In hypothesis tests we would like to use indirect proving, therefore execute the next theoretical experiments.

Indirect proof

We have a box containing 100 marbles. Each of them are either red or white.

Case #1: hypothesis (H): all of them are white.

Experiment: We randomly take a marble out of the box.

Our observation: It is red.

Conclusion: The probability of our observation given our hypothesis is 0: Our hypothesis is for 100% sure wrong.

Impossible event

The hypothesis is false

Falsification works

Verification does not work

The hypothesis is true

Sure event

Case #4: hypothesis (H): all of them are red.
Experiment: We randomly take a marble out of the box and put it back; we do this 5 times.
Our observation: All of them are red.
Conclusion: The probability of our observation given our hypothesis is $1^5 = 1$. Are we sure what to do now?

Falsification: refutation of the statement
Verification: supporting the statement

Indirect proof

What about #2 and #3?

Mathematical Logic:

We have a hypothesis (H).

If H is true, E event cannot occur.

E occurs.

So H is not true.

As we saw it before, a hypothesis can only be rejected.

Statistical Logic:

We have a hypothesis (H).

If H is true, F event is very unlikely to occur.

F occurs.

So we reject H . But we are not 100% sure if H is not true.

In this case a hypothesis cannot even be rejected with 100% certainty.

WE HAVE TO ESTIMATE THIS PROBABILITY

In math the falsification is clear – but in stat it works with a given probability.

What kind of questions can we test?

Y/N ...must be a yes/no (a.k.a. dichotomous or polar) question. **5W1H**

- Is the 5-years survival rate (i.e. probability) for myeloma 50%? ✓
- Does the total blood cholesterol level of Cushing's syndrome patients differ from the general 200 mg/dL population mean? ✓
- What is the 5-years survival rate for myeloma? ✗
- What is the expected value of total cholesterol level in Cushing's syndrome patients? ✗

...must refer to a set of observations, not to individual cases.
(And the question is aimed at a population, not a sample.)

- Is the 5-years survival rate for myeloma 50%? ✓
- Will this myeloma patient survive for 5 years? ✗

...must have at least one unambiguous answer.

- Is the 5-years survival rate for myeloma 50%? ✓
- Is the 5-years survival rate for myeloma less than 50%? ✗

...

What kind of answers can we test?

We have two answers for our question:

The null hypothesis (H_0)

- **Unambiguous:** can be realized in only one way. It contains some form of =.
The 5-years survival rate for myeloma is 50%.

- Represents the current well-established, generally **accepted scientific knowledge**,
The total blood cholesterol level of Cushing's syndrome patients is same as the population mean
or something that is the **most trivial** with the least assumptions (Occam's razor).
The probability of landing on heads in a coin tossing experiment is 50%.
- It is **not** necessarily the negative answer to the question.

The alternative hypothesis (H_1)

- Typically can be realized in more than one way.
The 5-years survival rate for myeloma is not 50%.
(can be a little more, a lot less etc.)
- Represent a **new statement** challenging the current scientific consensus,
The total blood cholesterol level of Cushing's syndrome patients differs from the population mean
or a set of all the **not-so-trivial** answers needing more or special assumptions.
The probability of landing on heads in a coin tossing experiment is other than 50%.
- It is typically **complementary to H_0** (i.e., its negation).
 $H_1 = \text{not } H_0$

...

Sampling („random“) error

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

Starting point: create a specific statistical question and answers:

H_0 : null hypothesis – „random“ error only

H_a (or H_1): alternative hypothesis – not H_0

Decision is based on: role of „randomness“ if H_0 true (sampling error)

A sample is that could contradict H_0

So we have a question and the hypotheses.

The H_0 is the more important: if H_0 is true in the reality we could see something else in the sample because of the sampling error (we have a sample we did not measure everybody).

Our aim is to answer the question with a decision: accepting or rejecting the H_0 .

Sampling („random“) error

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

Starting point: create a specific statistical question and answers:

H_0 : null hypothesis – „random“ error only

H_a (or H_1): alternative hypothesis – not H_0

Decision is based on: role of „randomness“ if H_0 true (sampling error)

A sample is that could contradict 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) ($1-\beta$)

Let's look a table on H_0 what could be the results of our decision. In reality (that we don't know) we have to options: the H_0 is true or false. Or decision could be to accept or to reject the H_0 .

If we accept the true H_0 or we reject the false H_0 we make a good decision. The latter one called power: it gives the probability to find (reject H_0) if the H_0 is false.

If we reject the true H_0 we make a wrong decision called type I. error. This probability – rejecting H_0 if H_0 is true – symbolized by alfa.

If we accept the false H_0 we make a wrong decision called type II. error. This probability – accepting H_0 if H_0 is false – symbolized by beta.

An example – first steps

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

What is the question??!! (and what is it about):

All of the sides has the same probability? (for this dice)
The probability of six-throw is different from 1/6, even bigger?
– let's use this question

Null hypothesis ??!!:

The probability of rolling 6 is 1/6 or less – hmm (*multiple hypothesis*)
Let's use the worst according to alternative hypothesis
(since if only one element of the multiple hypothesis can "conform" to the null hypothesis, there is no evidence of rejection in the full range of hypothesis):
 H_0 : The probability of rolling 6 is 1/6.
 H_a : greater than 1/6

Let's see an example for a hypothesis test.

An example – next steps

What is the 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.

How much evidence do we need for saying it is differ?

– Significance level??!!:

Given by authority... Used in the literature--- **STARTING VALUE**, BUT smaller/larger, more/less frequent side effects?, cheaper/ more expensive?...

here: we play this game every weekend, it's not expensive to change the die
– let's use 10% instead of 5% (smaller evidence enough to reject H_0)

Collecting evidence – the sample ??!! (how, how much... ask your statistician):

results: 6 times 6 out of 24 rollings

I really answer the question I have created?

Do I need to modify the question? (eg. the population of interest)

Significance level: the limit where we accept or reject H_0 .

An example – additional steps

What is the 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.

Is the difference important at all?– relevant ???!:

6 times 6 out of 24 rollings, it is $6/24 = 1/4$ probability that 1,5 times higher than 1/6 – YES it is

How much evidence – p-value ???!:

How to calculate it?, (unbiased, effective, consistent...)

Important aspects: the question, type of variables (measuring scale), measurement conditions

We need for calculation:

Null distribution – if null hypothesis is true, what is the probability of all theoretical samples we could have?

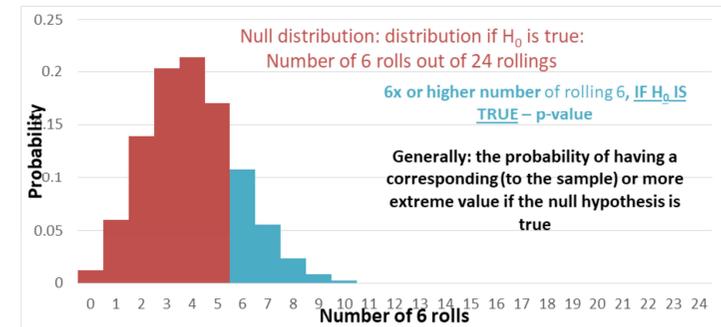
The „position“ of our sample in this distribution:

test statistics (or statistics, eg: frequency, probability , t-value , p-value...)

Relevant: clinically important effect.

Calculation in the „background“

Using binomial test!



...

An example - last steps

What is the 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.

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

How much evidence? – 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) ($1-\beta$)

...

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 continuous

Assumption

Independent observations

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“)! Use previous knowledge on the variable and make graphs.

Let's look a few hypothesis tests that you will learn in practices. I highlight here which hypothesis test could be used for a given problem, given variable (measuring scale), and what condition ave to be fulfilled.

(CLT: central limit theorem)

Paired Student t-test

What I'm curious about

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

Type of variable

1 numerical and continuous, 1 binary („groups“)

Assumptions

Independent observations in the groups, paired groups
the distribution of the difference of means is normal
normally distributed differences or large sample size

Notes:

paired test usually has higher power
suitable to compare other location parameters (quantiles)
difference of the means = mean of the differences

Mathematically this is the same as the one sample t-test, if we use the difference of paired values as a dataset.

2 sample Student t-test

What I'm curious about

Two expected values in two groups are equal

Type of variable

1 numerical and continuous, 1 binary („groups“)

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 continuous, 1 binary („groups“)

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)

...

Chi-square test for independence

What I'm curious about

Two variables are depends on each other

Type of variable

2 categorical variable

Assumptions

Independent observations

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

...

„Correlation” t-test (Pearson linear regression)

What I'm curious about

Two variables are (linearly) depends on each other

Type of variable

2 numerical variable (X and Y)

Assumptions

Independent observations for pairs

linear relation assumed

x values measured with no error

y-s have a normal distribution at each x

y-s have same variance at each x

Notes

2 estimates: slope and intercept, slope is important and tested

...

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>)

It is a common problem that we get a relevant (clinically important), but not significant (not rejected H0) result – that means no effect? or...

+We used too small sample size

+ the variability of the variables are too large

+ wrong, or wrongly used hyptohesis test was performed

+ low measuring scale

+ we were unlucky

+ ...

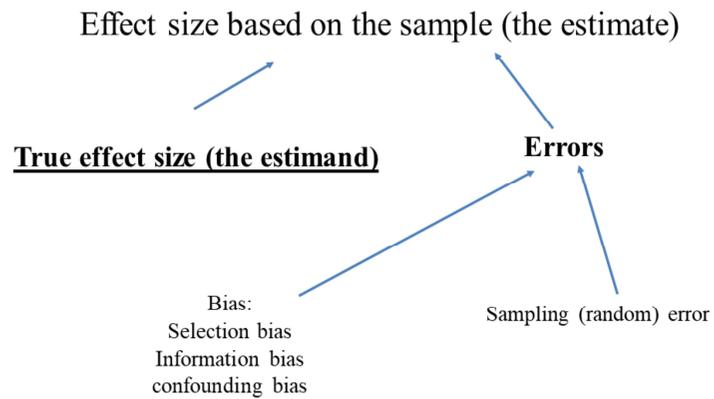
To avoid this problems it is very important to plan the experiment before perform data collection.

A typical big question is the sample size.

Sample size video:

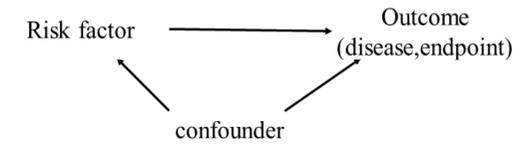
<https://www.youtube.com/watch?v=Hz1fyhVOjr4>

Other errors



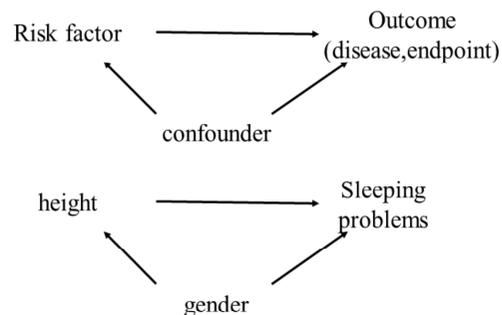
Until now, we talked about the error because sampling. But there are other source of errors called bias (systemic error). We usually classify them into 3 category:
Confounding
Selection bias
Information bias

Confounding bias



Shortly, confounding bias occur when the effect on the outcome variable is described, interpreted by a variable (here called risk factor), but in reality the effect caused (modified) by an other variable (a confounder).

Confounding bias



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

Eg. We found that higher people has sleeping problem with higher probability and therefore we interpret that saying that higher body high increase the sleeping problems. But in reality the reason is not the high: the reason is the gender – men has more probable sleeping problem. If we examine men and women separately we could not detect the high effect on sleeping problem.

(A variable could be a confounder if it has „effect” both on the risk factor and the outcome)

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)

typical: 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)

typical: recall bias
more carefull monitoring for diseased, young

Selection bias examples:

AGE: Knowing that gender has an effect on sleeping problems, the proportion of women in a „new drug group” is different than in the „placebo group”.

LOSS OF FOLLOW UP: – we know that AIDS is more frequent in i.v. drug users and homosexuals: those who has AIDS will more often "quit" the follow-up, than those who do not get AIDS, as well as IDU users more often „disappear”, than homosexuals

DIFFERENT POPULATION: Fracture in Women and Nutrition Relationships: We choose bone trauma from a trauma class, control of the hospital's internal medicine (But there are other more frequent illnesses in the internal medicine , eg diabetes is more common that has an effect on fracture !!)

Information bias examples:

RECALL BIAS: Parents of children diagnosed with cancer may be more likely to recall infections earlier in the child's

life than parents of children without cancer.

Good source:

Catalogue of Bias Collaboration, Spencer EA, Brasseley J, Mahtani K., 2017. <https://catalogofbias.org/>

Test Questions

- What does systematic and random error mean?
- What does unbiased estimation mean?
- What does effective estimation mean?
- What does consistent estimation mean?
- What does confidence interval mean?
- What is the aim of a hypothesis test?
- Give the criteria of a good question in a hypothesis test.
- Give the criteria of a good null hypothesis in a hypothesis test.
- What does significance level mean?
- What does relevant mean?
- Define first type error.
- Define second type error.
- Define the p-value in a hypothesis tests.
- What does confounding bias mean?
- What does selection and information bias mean?

The following questions may be answered using lecture material, consultation with practice teacher, or your own investigation (on the library or the internet). These test questions are examples for questions that may occur in the midterm and exam tests.