

# Biostatistics for the biomedical profession

## Lecture 7

BIMM34

Karin Källen & Linda Hartman

September 2016

# Today

- **Repetition**
- Lecture 6:
  - Power calculations
  - Probability, proportions, CI for proportions
  - 2x2 Table. Chi2-test, Fisher exact test
- **Lecture 7:**
- Sensitivity, Specificity
- Positive predictive value, Negative predictive value
- ROC-curves

## Exercise:

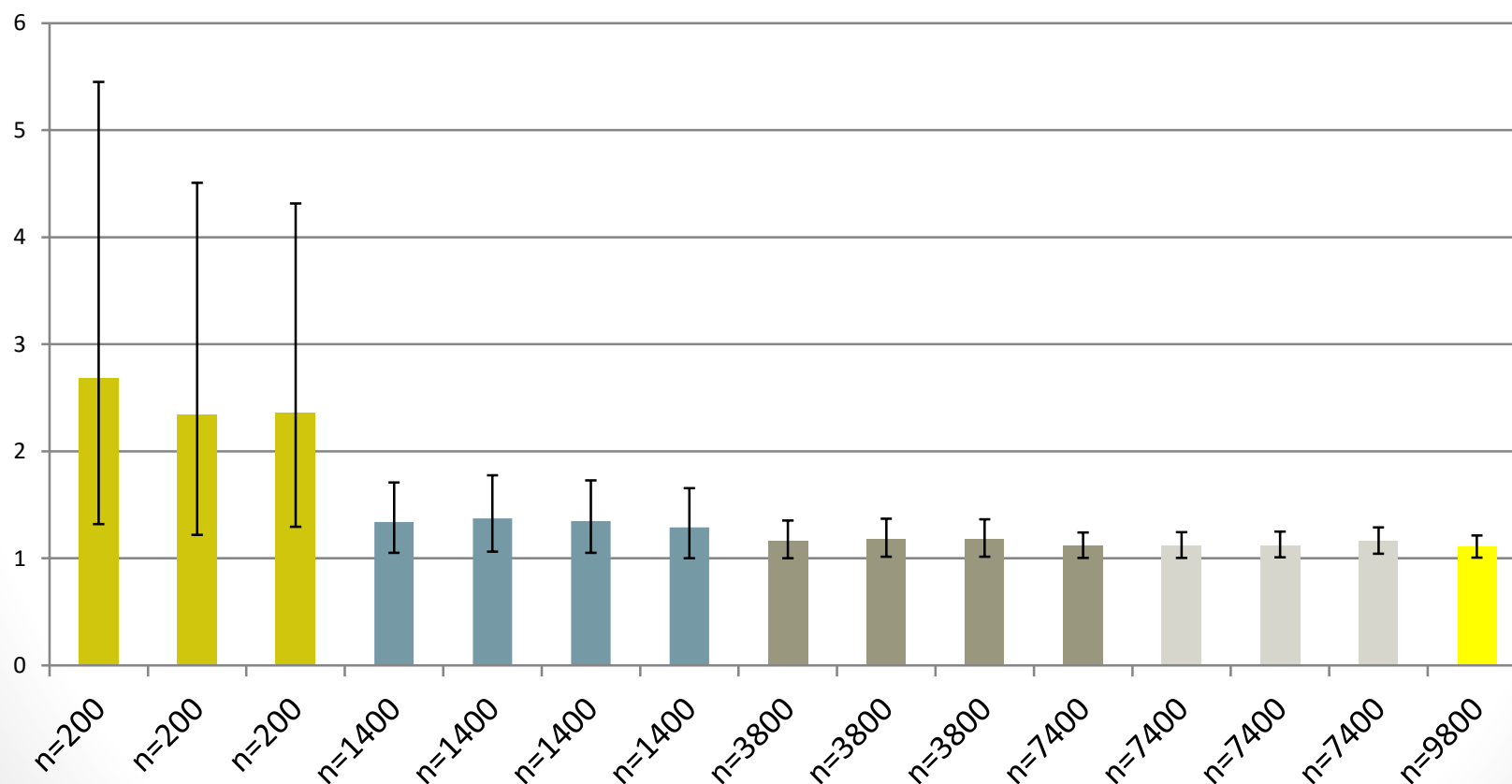
Combine the statistical terms with the correct common phenomena (in common language)

- Type I error
- Type II error
- Confounding

- Non-causal association
- Mass-significance
- Lack of power

So.... Which studies are most likely to be published, and to reach the headlines in the mass-media?

Significant ORs in 100 trials (case/controls 1:1) by sample size



Thus, for the sake of good science:

Be sure to perform adequate power analyses when designing a study

Repetition:

# Hypothesis testing.

- $H_0$  The hypothesis of no difference
- $H_1$  The hypothesis of a true difference

Example: Comparing the means of two samples;  $\mu_A$  = mean of sample A  
 $\mu_B$  = mean of sample B

- **Two-tailed test:**

- $H_0: \mu_A = \mu_B \quad H_1: \mu_A \neq \mu_B$

The most appropriate approach with few exceptions

- **One-tailed test (specifies the direction of the difference in advance)**

- $H_0: \mu_A \leq \mu_B \quad H_1: \mu_A > \mu_B$

- or

- $H_0: \mu_A \geq \mu_B \quad H_1: \mu_A < \mu_B$

# The t-test

Uses the fact that the difference between two means that both comes from normally distributed data, will follow a normal distribution with expected mean=0, and expected standard deviation SE.

$$t = \frac{\text{Difference between group means}}{\text{Standard error of difference in means}} = \frac{\bar{X}_1 - \bar{X}_0}{SE}$$

$$SE = s\sqrt{1/n_0 + 1/n_1}; s = \sqrt{\left[ \frac{(n_1 - 1)s_1^2 + (n_0 - 1)s_0^2}{n_1 + n_0 - 2} \right]}$$

# Elements of statistical inference, hypothesis testing, two-sided

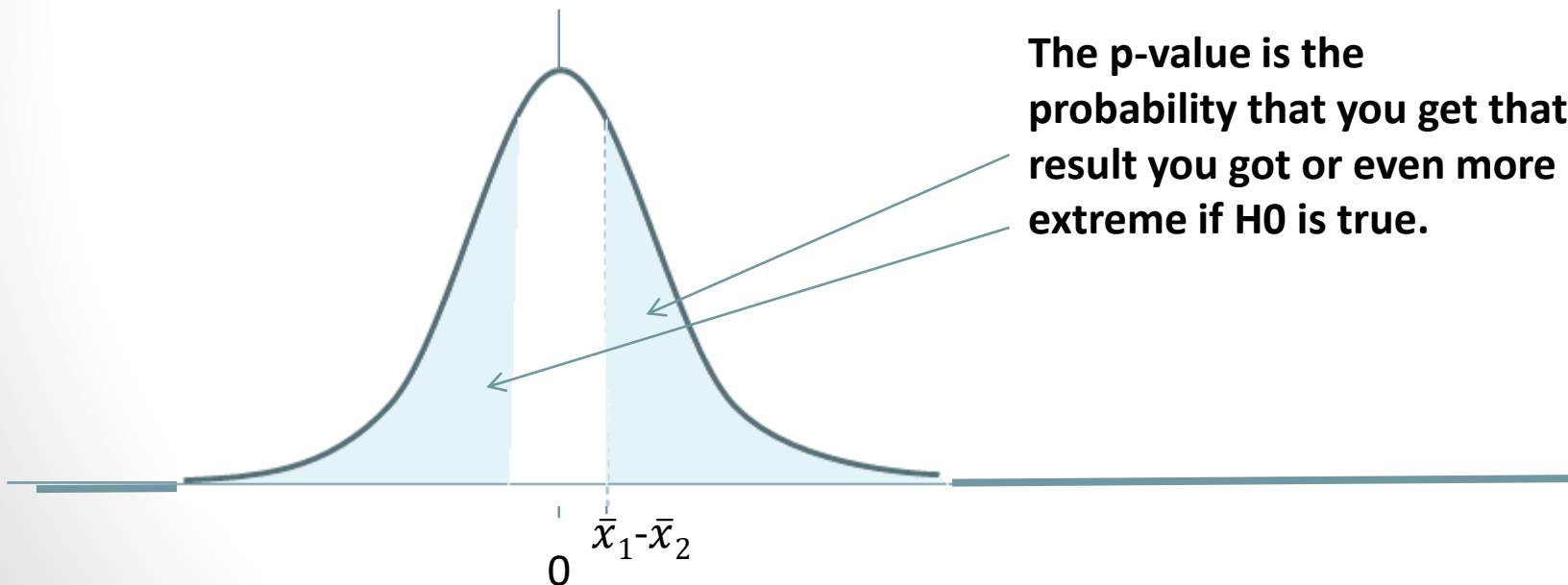
$H_0: \mu_1 = \mu_2$ , i.e.  $\mu_1 - \mu_2 = 0$

$H_1: \mu_1 \neq \mu_2$ , i.e.  $\mu_1 - \mu_2 \neq 0$

\_\_\_\_\_ Expected distribution of  $\bar{x}_1 - \bar{x}_2$

----- Sample mean,  $\bar{x}_1 - \bar{x}_2$

- Large p-value  $\rightarrow$  Data probable if  $H_0$  is true
- Thus:  $H_0$  is not rejected (= considered to be true until further evidence)





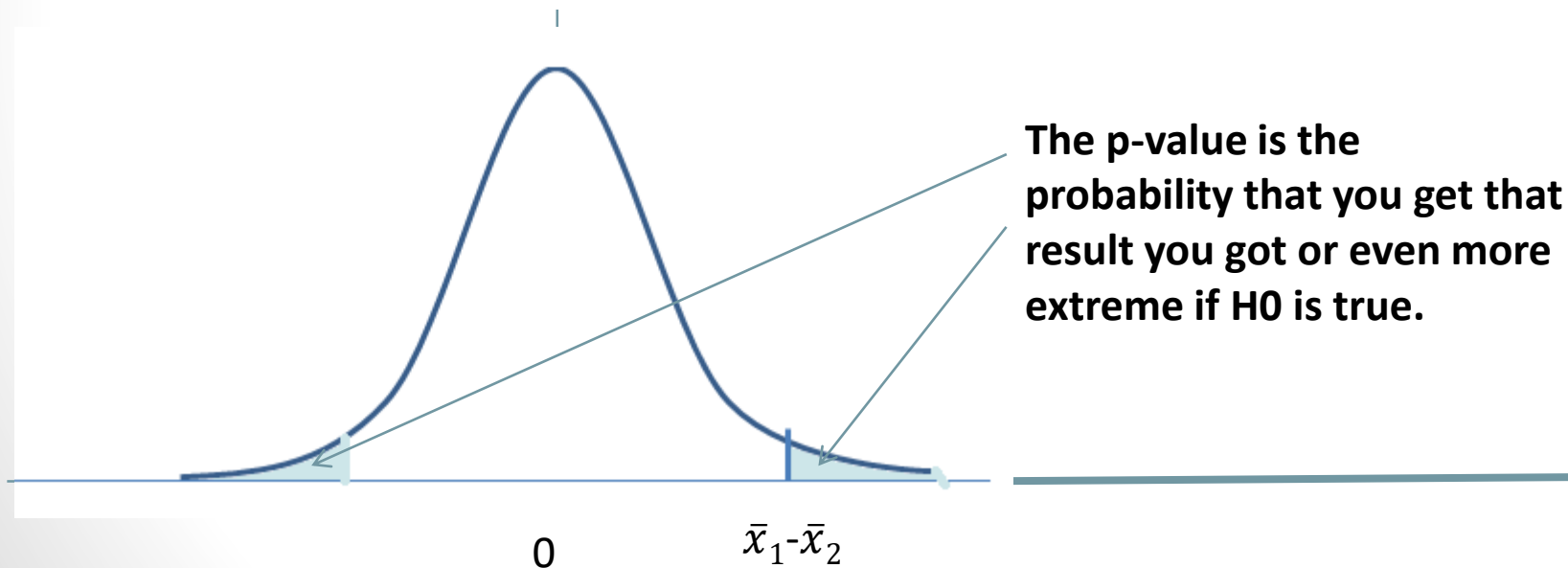
# Elements of statistical inference, hypothesis testing, two-sided

$H_0: \mu_1 = \mu_2$ , i.e.  $\mu_1 - \mu_2 = 0$

$H_1: \mu_1 \neq \mu_2$ , i.e.  $\mu_1 - \mu_2 \neq 0$

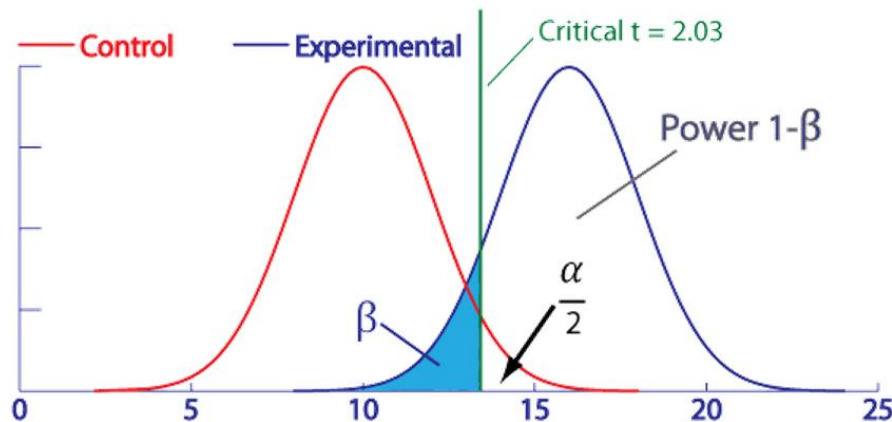
\_\_\_\_\_ Expected distribution of  $\bar{x}_1 - \bar{x}_2$   
- - - - Sample mean,  $\bar{x}_1 - \bar{x}_2$

- Data unlikely if  $H_0$  is true.
- Thus:  $H_0$  is rejected (=considered to be false)



# Statistical hypothesis testing

**Statistical power = The probability *a priori* that  $H_0$  will be rejected, given that  $H_1$  in fact is true**



$H_0$ : Null hypothesis = no difference  
 $H_1$  = Alternative hypothesis = difference

	$H_0$ True	$H_1$ True
$H_0$ not rejected	OK	Falsely negative (Type 2-error)
$H_0$ rejected	Falsely positive (Type 1-error)	OK Statistical power

## Example:

### The power of a study investigating the effect of a treatment, where there is a 'true' effect

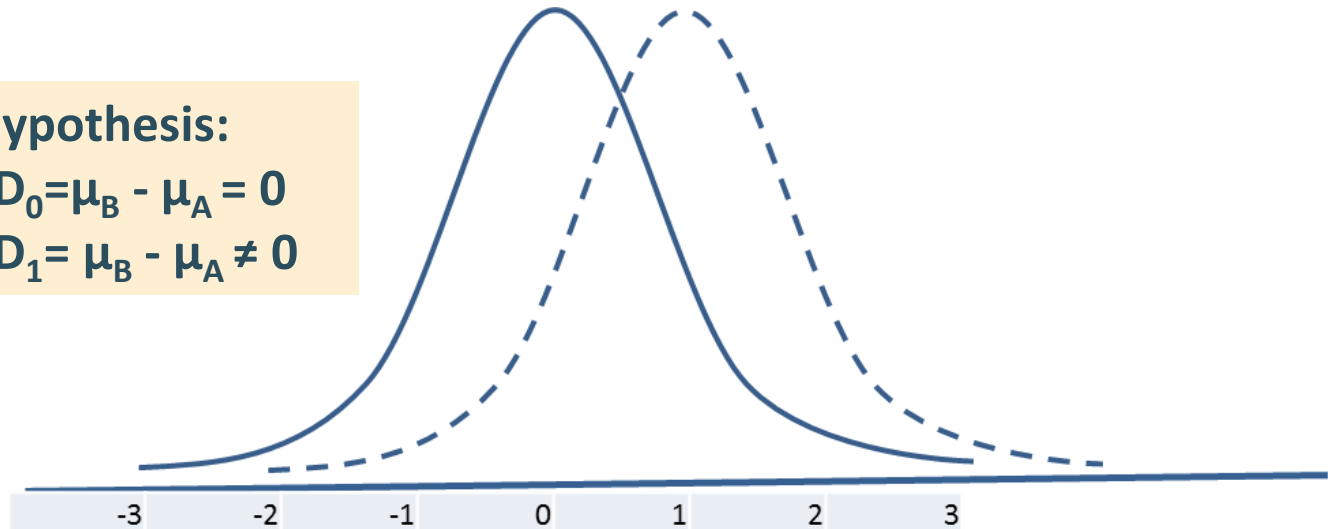
Suppose that we have two groups, A and B, both with  $\sigma = 1$ , where  $\mu_A = 0$  and  $\mu_B = 1$

'True' distribution of non-treated group (A) —  
'True' distribution of treated group (B) - - -

#### Formulate the Hypothesis:

$H_0: \mu_A = \mu_B$ , i.e.  $D_0 = \mu_B - \mu_A = 0$

$H_1: \mu_A \neq \mu_B$ , i.e.  $D_1 = \mu_B - \mu_A \neq 0$



#### According to the null hypothesis:

Suppose we take a sample a and b from groups A and B,

$a \in N(0,1)$  and  $b \in N(0,1)$

the expected difference  $D_0 \in N(0, SE^2)$ ;

$SE = 1 * \sqrt{1/na + 1/nb}$

#### The truth is that:

$b \in N(1,1)$

Then, the 'true' difference

$D_{\text{True}} \in N(1, SE^2)$ ;

$SE = (1 * \sqrt{1/na + 1/nb})$

## Example: The power of a study investigating the effect of a treatment, where there is a 'true' effect

Suppose that we have two groups, A and B, both with  $\sigma = 1$ , where  $\mu_A = 0$  and  $\mu_B = 1$

### Formulate the Hypothesis:

$H_0: \mu_A = \mu_B$ , i.e.  $D_0 = \mu_B - \mu_A = 0$

$H_1: \mu_A \neq \mu_B$ , i.e.  $D_1 = \mu_B - \mu_A \neq 0$

### According to the null hypothesis:

Suppose we take a sample  $a$  and  $b$  from groups A and B,  
 $a \in N(0,1)$  and  $b \in N(0,1)$

the expected difference  $D_0 \in N(0, SE^2)$ ;  $SE = 1 * \sqrt{1/na + 1/nb}$

### The truth is that:

$b \in N(1,1)$

Then, the 'true' difference  $D_{\text{true}}$

$\rightarrow D_{\text{True}} \in N(1, SE^2)$ ;  $SE = 1 * \sqrt{1/na + 1/nb}$

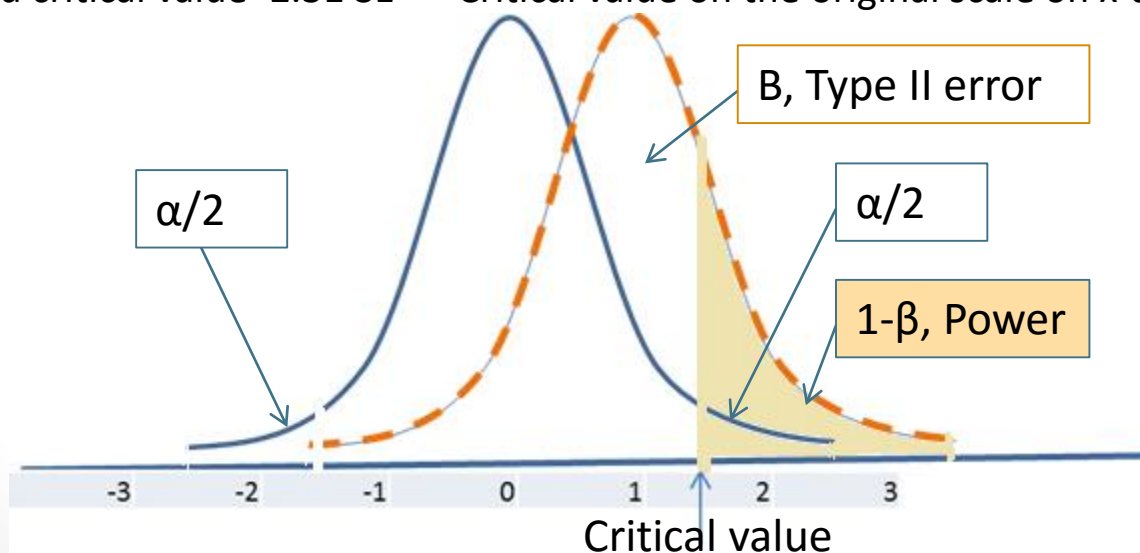
• Now, assume that we pick a sample of 5 individuals from each group, then

• The  $DF = n_a - 1 + n_b - 1 = 8$ . The standardized 'critical value' for  $t(8) = 2.31$

$$D_0 \in N\left(0, \sqrt{1/5 + 1/5}\right)^2 \Rightarrow D_0 \in N(0, 0.63^2)$$

$$D_{\text{True}} \in N(1, 0.63^2)$$

Standardized critical value = 2.31 SE  $\Rightarrow$  Critical value on the original scale on x-axis =  $2.31 * .63 = 1.46$



Exercise:

Look at the figure,  
How large is (approx) the  
chance that our sample  
mean is above the critical  
value?

Perhaps 25% ?

# Example continued...

Let us check the power of the previous example using PS Power

Power and Sample Size Program: Main Window

File Edit Log Help

Survival **t-test** Regression 1 Regression 2 Dichotomous Mantel-Haenszel Log

[Studies that are analyzed by t-tests](#)

**Output**

[What do you want to know?](#) Power

**Design**

[Paired or independent?](#) Independent

**Input**

$\alpha$  0.05  $\sigma$  1

$n$  5  $m$  1

Calculate

Graphs

**Description**

PS version 3.1.2

Logging is enabled.

Copy to Log Exit

# Example continued...

Let us check the power of the previous example using PS Power

PS Power and Sample Size Program: Main Window

File Edit Log Help

Survival t-test Regression 1 Regression 2 Dichotomous Mantel-Haenszel Log

[Studies that are analyzed by t-tests](#)

**Output**

[What do you want to know?](#) Power

[Power](#) .247

**Design**

[Paired or independent?](#) Independent

**Input**

$\alpha$  .05  $\delta$  1

$n$  5  $\sigma$  1

$m$  1

Calculate

Graphs

**Description**

We are planning a study with 5 experimental subjects and 5 control subjects. In a previous study the response within each subject group was normally distributed with standard deviation 1. If the true difference in the experimental and control means is 1, we will be able to reject the null hypothesis that the population means of the experimental and control groups are equal with probability (power) .247. The Type I error probability associated with this test of this null hypothesis is 0,05.

PS version 3.1.2

Copy to Log Exit

Logging is enabled.

# Example, continued

Now, let us use PS Power to calculate the sample size needed to have 80% power to detect that the 'true' effect of the treatment exists.

PS Power and Sample Size Program: Main Window

File Edit Log Help

Survival t-test Regression 1 Regression 2 Dichotomous Mantel-Haenszel Log

[Studies that are analyzed by t-tests](#)

**Output**

[What do you want to know?](#) Sample size

**Design**

[Paired or independent?](#) Independent

**Input**

$\alpha$  .05  $\delta$  1  $\sigma$  1  $m$  1

[power](#) .8

Calculate

Graphs

**Description**

PS version 3.1.2

Logging is enabled.

Copy to Log Exit

# Example, continued

Now, let us use PS Power to calculate the sample size needed to have 80% power to detect that the 'true' effect of the treatment exists.

Power and Sample Size Program: Main Window

File Edit Log Help

Survival t-test Regression 1 Regression 2 Dichotomous Mantel-Haenszel Log

[Studies that are analyzed by t-tests](#)

**Output**

[What do you want to know?](#) Sample size

[Sample Size](#) 17

**Design**

[Paired or independent?](#) Independent

**Input**

$\alpha$  .05  $\delta$  1

$\sigma$  1

[power](#) .8  $m$  1

Calculate

Graphs

**Description**

We are planning a study of a continuous response variable from independent control and experimental subjects with 1 control(s) per experimental subject. In a previous study the response within each subject group was normally distributed with standard deviation 1. If the true difference in the experimental and control means is 1, we will need to study 17 experimental subjects and 17 control subjects to be able to reject the null hypothesis that the population means of the experimental and control groups are equal with probability (power) 0.8. The Type I error probability associated with this test of this null hypothesis is 0.05.

PS version 3.1.2

Copy to Log Exit

Logging is enabled.



## Example: The power of a study investigating the effect of a treatment, where there is a 'true' effect

Suppose that we have two groups, A and B, both with  $\sigma = 1$ , where  $\mu_A = 0$  and  $\mu_B = 1$

### Formulate the Hypothesis:

$H_0: \mu_A = \mu_B$ , i.e.  $D_0 = \mu_B - \mu_A = 0$

$H_1: \mu_A \neq \mu_B$ , i.e.  $D_1 = \mu_B - \mu_A \neq 0$

### According to the null hypothesis:

Suppose we take a sample a and b from groups A and B,  
 $a \in N(0,1)$  and  $b \in N(0,1)$

the expected difference  $D_0 \in N(0, SE^2)$ ;  $SE = 1 * \sqrt{1/na + 1/nb}$

### The truth is that:

$b \in N(1,1)$

Then, the 'true' difference  $D_{\text{true}}$

$\rightarrow D_{\text{True}} \in N(1, SE^2)$ ;  $SE = 1 * \sqrt{1/na + 1/nb}$

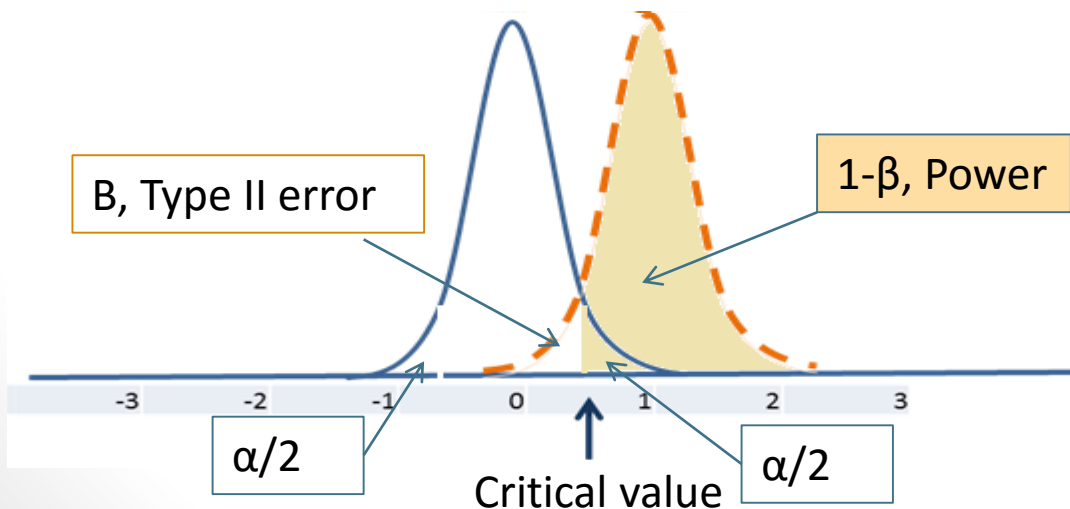
• Now, assume that we pick a sample of 17 individuals from each group, then

• The  $DF = n_a - 1 + n_b - 1 = 32$ . The standardized 'critical value' for  $t(32) = 2.04$

$D_0 \in N\left(0, \left(\sqrt{1/17 + 1/17}\right)^2\right) \Rightarrow D_0 \in N(0, 0.34^2)$

$D_{\text{True}} \in N(1, 0.34^2)$

Standardized critical value = 2.04 SE  $\Rightarrow$  Critical value on the original scale on x-axis =  $2.04 * .34 = 0.70$

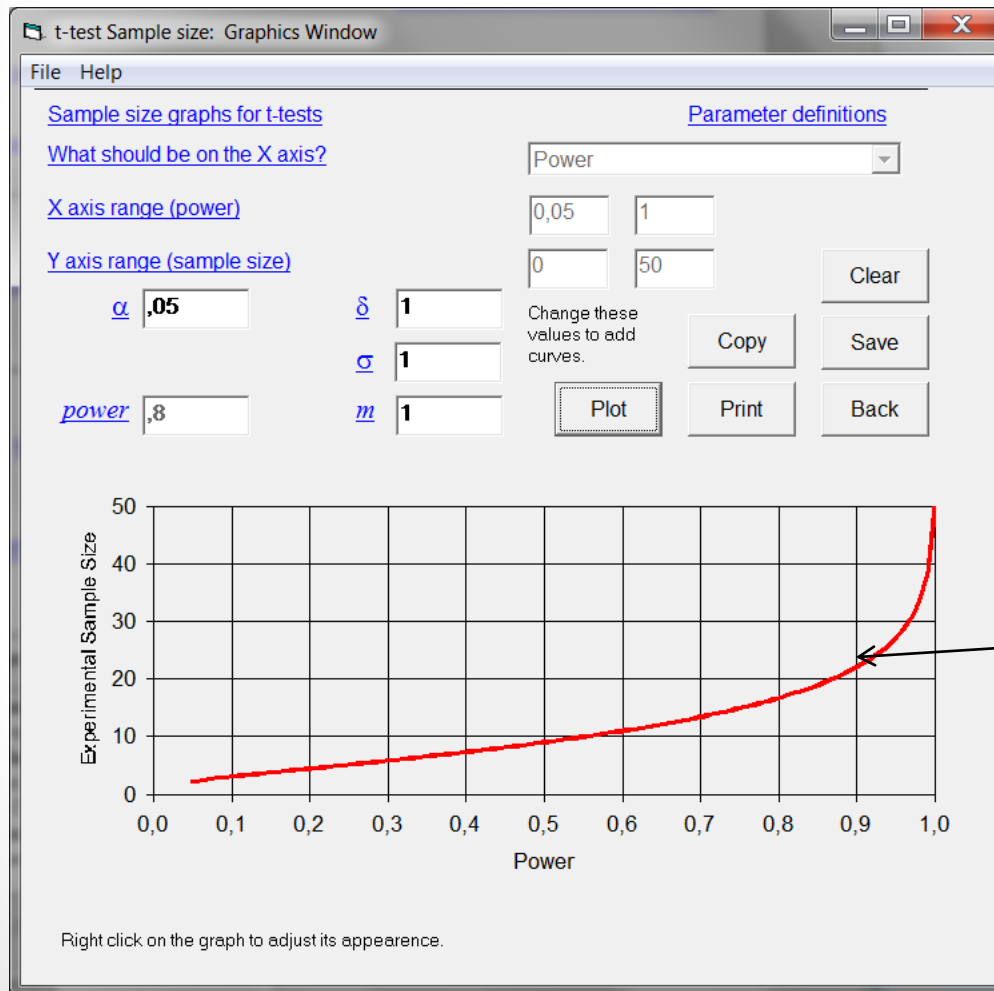


### Exercise:

Look at the figure,  
Does the PS power calculation of sample size seem correct (would you say that 80% of the area of the  $D_{\text{true}}$ -distribution is above the critical value)?

Yes!!!

Using the 'graph'-button in PS power, we can also create a graph showing i.e. the power by sample size, given a certain difference between means.



Exercise:

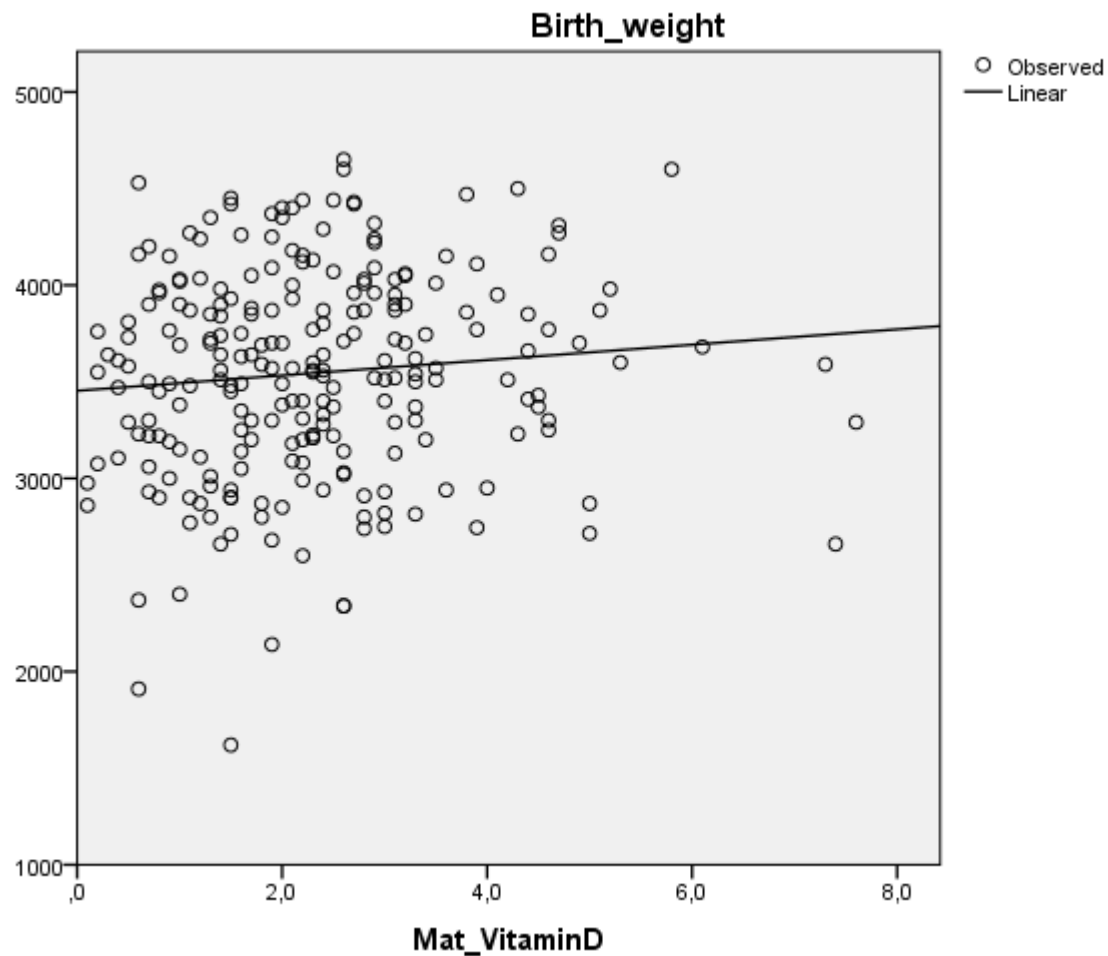
Approximately, how large should the Samples be in order to have 90% chance to detect a true Difference of 1.?

Approx 22

Now, let's use PS Power to estimate the needed sample size in order to detect an assumed 'true' linear association

Let's assume that we will design a study in order to investigate a possible linear relationship between maternal vitamin D levels during pregnancy and child's birth weight.

Assume that we have access to the data from a pilot study..... (the Birth data set).



... a linear association between maternal vitamin D levels and child's Birth weight seems possible, but not convincing....

Let's proceed anyway

In order to perform power-analyses for studies using linear regression, we will have to Estimate the SD or the regression residuals.

The screenshot displays the IBM SPSS Statistics Data Editor interface with the 'Analyze' menu open. The 'Linear...' option under the 'Regression' submenu is highlighted with a red circle. Three sub-dialogs are open in the foreground:

- Linear Regression:** The 'Dependent' variable is 'Birth\_weight' and the 'Independent(s)' variable is 'Mat\_VitaminD'. The 'Statistics...' button is circled in red.
- Linear Regression: Plots:** The 'DEPENDNT' list contains '\*ZPRED', '\*ZRESID', '\*DRESID', '\*ADJPRED', '\*SRESID', and '\*SDRESID'. The 'Y' variable is '\*ZRESID' and the 'X' variable is '\*ZPRED'. Both are circled in red.
- Linear Regression: Statistics:** The 'Regression Coefficients' section has 'Estimates' and 'Confidence intervals' checked. The 'Descriptives' checkbox is circled in red.

The background data editor shows a dataset named 'BIMM\_Births.sav' with columns for 'Birth\_weight' and 'Head\_c'.

# SPSS-output

**Descriptive Statistics**

	Mean	Std. Deviation	N
Birth_weight	3545,32	539,144	237
Mat_VitaminD	2,310	1,3469	237

**Coefficients<sup>a</sup>**

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
1 (Constant)	3453,517	69,436		49,737	,000	3316,721	3590,312
Mat_VitaminD	39,745	25,984	,099	1,530	,127	-11,445	90,936

a. Dependent Variable: Birth\_weight

**Residuals Statistics<sup>a</sup>**

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	3457,49	3755,58	3545,32	53,531	237
Residual	-1893,135	1093,146	,000	536,480	237
Std. Predicted Value	-1,641	3,928	,000	1,000	237
Std. Residual	-3,521	2,033	,000	,998	237

a. Dependent Variable: Birth\_weight

Now, we just fill out the required values in PS Power, using the 'Regression 1' -Tab

The required sample size in order to have 80% power is N=790

Power and Sample Size Program: Main Window

File Edit Log Help

Survival t-test Regression 1 Regression 2 Dichotomous Mantel-Haenszel Log

Studies that are analyzed by linear regression (1 treatment)

**Output**

What do you want to know? Sample size

Sample size 790

**Design**

Investigator chooses treatment level? No

How is  $\sigma$  obtained? Entered directly

**Input**

$\alpha$  .05  $\sigma$  536.48 Calculate

power .8  $\sigma_x$  1.3469 Graphs

$\lambda$  39.745

**Description**

We are planning a study of subjects in which we will regress their values of yvar against xvar. Prior data indicate that the standard deviation of xvar is 1,3469 and the standard deviation of the regression errors will be 536,48. If the true slope of the line obtained by regressing yvar against xvar is 39,745, we will need to study 790 subjects to be able to reject the null hypothesis that this slope equals zero with probability (power) 0,8. The Type I error probability associated with this test of this null hypothesis is 0,05.

PS version 3.1.2

Logging is enabled.

Copy to Log Exit



# The statistical power depends on:

- The study design
- Sample size
- The size of the difference between groups
- The size of the variance within the groups
- The chosen level of significance  $\alpha$
- Sources of errors
- The number and importance of confounders
- ...

# Today

- Repetition
- Lecture 6:
  - Power calculations
  - **Probability, proportions, CI for proportions**
  - **2x2 Table. Chi2-test, Fisher exact test**
- Lecture 7:
- Sensitivity, Specificity
- Positive predictive value, Negative predictive value
- ROC-curves

# Lecture 6 repetition....

## Binary data

- **Binary data:**
- Probability, binomial distribution (don't worry – just recognize)
  - Proportions, confidence interval for proportions, rates
  - 2 x 2 tables, Chi-squared test, Fisher Exact test

# Binary data

- When investigating binary data (e.g. alive/dead, sick/healthy), we could not assume that the outcomes follow a normal distribution. All subjects have either the value =1 or =0.
- Another set of statistical methods is required to analyze these data. Let's start with a quick look at the binomial distribution.

# Combinations and permutations

- Discuss: How to calculate the number of different pokerhands that could be dealt?
- For the first card, there is 52 possibilities, for the second 51, etc...
- Thus five cards:  $52 \cdot 51 \cdot 50 \cdot 49 \cdot 48$  / the number of different orders
- The number of different orders: the first card could have 5 places, the second 4...etc
- **Answer:  $52 \cdot 51 \cdot 50 \cdot 49 \cdot 48 / (5 \cdot 4 \cdot 3 \cdot 2)$  different pokerhands**

## General equation:

Let  $n$ =the total number of objects in the set

Let  $r$ =the number of objects that will be drawn

Let  $n!$  ( $n$  faculty) =  $n \cdot (n-1) \cdot (n-2) \cdot \dots \cdot (n-r+1)$

Let  $p_r^n$  = number of permutations of  $n$  objects taken  $r$  at a time

Then

the number of combinations ( $c_r^n$ )(ignoring the order of the  $r$  objects) are:

$$C_r^n = n! / (r! \cdot (n-r)!)$$

Do this exercise at home....

# Combinations and permutations

- Discuss: How to estimate the chance of having at least one girl, if you're planning for four children? (suppose that the chance of having a girl is 0.5 at each delivery)

Answer:  $1 - p(\text{having no girl})$

Where

$$P(\text{having no girl}) = 0.5 * 0.5 * 0.5 * 0.5 = 0.5^4 = 0.0625$$

Thus

$$P(\text{having at least one girl}) = 1 - 0.0625 = 0.9375$$

- Discuss: If you have three boys: How big is the chance that your fourth child will be a girl?
- Answer: Random events do not remember.....  $p=0.5$

# The binomial distribution

- The binomial distribution show the probabilities of different outcomes for a series of random events, each of which can have only one of two values.
- Let
- $p$ =the probability on each try of the outcome of interest.
- $q=(1-p)$
- $P$ =the probability for the outcome exact  $r$  times

- Then

- $P = \frac{n!}{r!(n-r)!} \cdot p^r q^{n-r}$

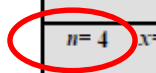
And

$$P(X \leq x) = \sum_{r=0}^x C_r^n p^r (1-p)^{n-r}$$

### Tables of the Binomial Cumulative Distribution

The table below gives the probability of obtaining at most  $x$  successes in  $n$  independent trials, each of which has a probability  $p$  of success. That is, if  $X$  denotes the number of successes, the table shows

$$P(X \leq x) = \sum_{r=0}^x C_r^n p^r (1-p)^{n-r}$$



$p=$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
$n=2$ $x=0$	0.9801	0.9604	0.9409	0.9216	0.9025	0.8836	0.8649	0.8464	0.8281	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500
$x=1$	0.9999	0.9996	0.9991	0.9984	0.9975	0.9964	0.9951	0.9936	0.9919	0.9900	0.9775	0.9600	0.9375	0.9100	0.8775	0.8400	0.7975	0.7500
$x=2$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$n=3$ $x=0$	0.9703	0.9412	0.9127	0.8847	0.8574	0.8306	0.8044	0.7787	0.7536	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1664	0.1250
$x=1$	0.9997	0.9988	0.9974	0.9953	0.9928	0.9896	0.9860	0.9818	0.9772	0.9720	0.9393	0.8960	0.8438	0.7840	0.7183	0.6480	0.5748	0.5000
$x=2$	1.0000	1.0000	1.0000	0.9999	0.9999	0.9998	0.9997	0.9995	0.9993	0.9990	0.9966	0.9920	0.9844	0.9730	0.9571	0.9360	0.9089	0.8750
$x=3$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$n=4$ $x=0$	0.9606	0.9224	0.8853	0.8493	0.8145	0.7807	0.7481	0.7164	0.6857	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915	0.0625
$x=1$	0.9994	0.9977	0.9948	0.9909	0.9860	0.9801	0.9733	0.9656	0.9570	0.9477	0.8905	0.8192	0.7383	0.6517	0.5630	0.4752	0.3910	0.3125
$x=2$	1.0000	1.0000	0.9999	0.9998	0.9995	0.9992	0.9987	0.9981	0.9973	0.9963	0.9880	0.9728	0.9492	0.9163	0.8735	0.8208	0.7585	0.6875
$x=3$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9999	0.9995	0.9984	0.9961	0.9919	0.9850	0.9744	0.9590	0.9375
$x=4$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$n=5$ $x=0$	0.9510	0.9039	0.8587	0.8154	0.7738	0.7339	0.6957	0.6591	0.6240	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503	0.0313
$x=1$	0.9990	0.9962	0.9915	0.9852	0.9774	0.9681	0.9575	0.9456	0.9326	0.9185	0.8352	0.7373	0.6328	0.5282	0.4284	0.3370	0.2562	0.1875
$x=2$	1.0000	0.9999	0.9997	0.9994	0.9989	0.9980	0.9968	0.9955	0.9937	0.9914	0.9724	0.9421	0.8965	0.8360	0.7648	0.6826	0.5924	0.5000

Discuss:

- How big is the chance to have less than 2 accidents with the bike during 4 rides, if there is a 10% accident risk at each ride?

Answer: 0.95

- How big is the chance of having no accidents on three rides, if there is a 5% risk for accident at each ride? Answer: 0.86



Luckily, if the sample size is large, we could use normal approximation to obtain confidence intervals for proportions

Let

- $n$ =total number of observations
- $p$ =probability for event
- $q=1-p$

If  $n \cdot p \cdot q > 10$  (golden rule of thumb), we can use normal approximation to estimate the spread of the mean of number of events.

- Normal approximation for binomial distribution:
  - Mean =  $n \cdot p$
  - Variance of mean =  $n \cdot p \cdot q$
  - SD of mean =  $\sqrt{\text{variance}} = \sqrt{n \cdot p \cdot q}$

Confidence interval for a sample proportion using normal approximation:

Assume that :

- $x$ =number of events;  $p=x/n$
- Then, the variance for  $p=x/n$  will be:  $1/n^2 \cdot \text{Var}(x) = (1/n^2) \cdot n \cdot p \cdot q = \frac{p \cdot q}{n}$
- And the SE of  $p$ :  $\sqrt{\frac{p \cdot q}{n}}$

Thus, using normal approximation,

a 95%CI for a proportion  $x/n$  could be estimated:  $p = x/n \pm 1.96 \cdot \sqrt{\frac{p \cdot q}{n}}$

# Confidence interval for a sample proportion (normal approximation)

Let  $x$ = observed number of events; Let  $n$ =total number of observations

Then  $p = x/n$  and  $q=1-p$

Using normal approximation,

a 95%CI for a proportion  $x/n$  could be estimated:  $p = x/n \pm 1.96 * \sqrt{\frac{p*q}{n}}$

Exercise:

Compute a 95%CI for the proportion of obese men in a certain population if  $x$ =obese men=100, and total  $n$ =400.

$$p = x/n = 100/400 = 0.25 = 25\%$$

$$q = (1-p) = 0.75$$

$$SE = \sqrt{(0.25 * 0.75 / 400)} \approx 0.022$$

$$95\%CI = p - 1.96 * SE \text{ to } p + 1.96 * SE = 0.25 - 1.96 * .022 \text{ to } 0.25 + 1.96 * 0.022$$

$$p \text{ with } 95\%CI \approx 0.250 (0.207 - 0.293) \text{ or } 25.0\% (20.7\% \text{ to } 29.3\%)$$

# The most basic 2 x 2 contingency table

	Exposed	Non-exposed
Cases	a	b
Non-cases	c	d

- **Main methods to investigate whether there is a difference between cases and non-cases:**
- **Chi-squared test** (tests if the number of exposed equally distributed over cases and non-cases), **Fisher Exact test**
- **Calculate Risk Ratios** (Comparing proportions)
- **Calculate Odds Ratios** (Comparing odds)

# The most basic 2 x 2 contingency table

	Exposed	Non-exposed	
Cases	a	b	$m_1 = a + b$
Non-cases	c	d	$m_2 = c + d$
	$n_1 = a + c$	$n_2 = b + d$	$N = n_1 + n_2 = m_1 + m_2$

- **The Chi-squared test** (tests if the number of exposed equally distributed over cases and non-cases)
- $H_0$ : The proportion of exposures is the same among cases and controls
- $H_1$ : The proportion of exposures is not the same among cases and controls (there is an association between exposure and the outcome).

- The Chisquared statistic:  $\chi^2 = \sum \frac{[O_i - E_i]^2}{E_i}$ 
  - To be compared with the  $\chi^2$  (df) distribution, where
  - $O_i$ =observed frequencies (a,b,c,d)
  - $E_i$ =expected frequencies ( $m_1 * n_1 / N$ ;  $m_1 * n_2 / N$ ;  $m_2 * n_1 / N$ ;  $m_2 * n_2 / N$ )
  - Df=degrees of freedom=(r-1) \* (c-1), where r=n rows; c=n columns

# The most basic 2 x 2 contingency table

	Exposed	Non-exposed	
Cases	10	10	$m_1=20$
Non-cases	10	70	$m_2=80$
	$n_1=20$	$n_2=80$	$N=100$

- **Exercise:**
- 1. Calculate expected frequencies
- 2. Calculate the  $\chi^2$  statistic

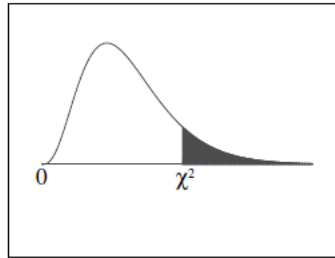
$$\chi^2 = \sum \frac{[O_i - E_i]^2}{E_i}$$

$$\chi^2 = (10-4)^2/4 + (10-16)^2/16 + (10-16)^2/16 + (70-64)^2/64 = 36/4 + 16/16 + 16/16 + 36/64 = 11.56$$

	Exposed	Non-exposed
Cases	$20 \cdot (20/100)$ 4	$20 \cdot (80/100)$ 16
Non-cases	$80 \cdot (20/100)$ 16	$80 \cdot (80/100)$ 64

- 3. Is there a significant association?

## Chi-Square Distribution Table



The shaded area is equal to  $\alpha$  for  $\chi^2 = \chi^2_{\alpha}$ .

$df$	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188

Exercise (continued):

Compare the obtained Chi-squared statistic with the Chi-squared distribution

11.56 >> 3.84 : There was a strong statistically significant heterogeneity

# The Yates' correction

When numbers are small (any expected cell frequency <5), the  $\chi^2$  statistic will be instable.

One easy solution is to use the Yates' correction:

$$\chi^2 = \sum \frac{[|O_i - E_i| - .5]^2}{E_i}$$

# The Fischer's Exact test

	Exposed	Non-exposed	
Cases	a	b	$m_1 = a + b$
Non-cases	c	d	$m_2 = c + d$
	$n_1 = a + c$	$n_2 = b + d$	$N = n_1 + n_2 = m_1 + m_2$

A more exact way of testing a 2x2 contingency table than the Chi-squared test, especially when the numbers are small

- $p = \frac{(a+b)! (c+d)! (a+c)! (b+d)!}{N! a! b! c! d!}$



# A 2x2 contingency table-output from SPSS

SGA * Smoking_yes_no Crosstabulation					
			Smoking_yes_no		Total
			,00	1,00	
SGA	,00	Count	174	60	234
		% within SGA1	74,4%	25,6%	100,0%
	1,00	Count	9	6	15
		% within SGA1	60,0%	40,0%	100,0%
Total		Count	183	66	249
		% within SGA1	73,5%	26,5%	100,0%

SGA=Small for Gestational Age

# Output from the SPSS 2x2 contingency table, continued

Chi-Square Tests					
	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1,492 <sup>a</sup>	1	,222		
Continuity Correction <sup>b</sup>	,846	1	,358		
Likelihood Ratio	1,377	1	,241		
Fisher's Exact Test				,234	,177
Linear-by-Linear Association	1,486	1	,223		
N of Valid Cases	249				
a. 1 cells (25,0%) have expected count less than 5. The minimum expected count is 3,98.					
b. Computed only for a 2x2 table					

Thus, no significant association could be detected between maternal smoking and SGA.

Does that mean that no association exists?

No! Perhaps did our study have too low power to have a chance to detect a true association

# The most basic 2 x 2 contingency table

	Exposed	Non-exposed
Cases	a	b
Non-cases	c	d

- The Chi-squared test, and the Fisher Exact test just investigate whether there is a heterogeneity between cases and non-cases, or not.
- **No effect measurements are provided!**

**Two main measurements for comparing proportions between groups:**

**Risk Ratios** (Comparing proportions)

**Odds Ratios** (Comparing odds)

# The most basic 2 x 2 contingency table

	Exposed	Non-exposed
Cases	a	b
Non-cases	c	d

**Risk Ratio:**  $\frac{a/(a+c)}{b/(b+d)}$

**Odds Ratio:**  $\frac{a/c}{b/d} = \frac{a * d}{c * b}$

Exercise:

What is the expected value for

a) Risk Ratio (RR) **1.0**

b) Odds Ratio (OR) **1.0**

Under the  $H_0$  hypothesis (hypothesis of no effect)?

# Today

- Repetition
- Lecture 6:
  - Power calculations
  - Probability, proportions, CI for proportions
  - 2x2 Table. Chi2-test, Fisher exact test, Odds Ratio, Risk Ratio

## Lecture 7:

- Sensitivity, Specificity
- Positive predictive value, Negative predictive value
- ROC-curves

- Sensitivity: The proportion of detected cases
- Specificity: The proportion of detected non-cases
- 1-specificity: The false positive rate
- Positive predictive value: The proportion of cases among detected
- Negative predictive value: The proportion of non-cases among non-detected.

# The most basic 2 x 2 contingency table

	Positive test	Negative test
Cases	a	b
Non-cases	c	d

- **Exercise: What are the formulas for:**
- Sensitivity (proportion of detected cases)
- Specificity (proportion non-detected non-cases)
- 1-specificity (proportion false positive)
- Positive predictive value (prob case given pos test)
- Negative predictive value (prob noncase given neg test)

- $a/(a+b)$
- $d/(d+c)$
- $1-d/(d+c)=c/(d+c)$
- $a/(a+c)$
- $d/(b+d)$

2x2 contingency table showing neurodevelopmental disability at 2 and 6 years, respectively, among children born extremely premature (<27 weeks)

The question is: Can the results at 2 years of age predict the results at 6 years?

	Disability at 2 years (columns)		Total N
Disability at 6 years (rows)	Disability	No disability	
Disability	71	74	145
No disability	44	244	288
Total N	115	318	433

Exercise: Decide the

- Sensitivity
- Specificity
- False positive rate
- Positive predictive value
- Negative predictive value

- $a/(a+b)$

- $d/(d+c)$

- $1-d/(d+c)=c/(d+c)$

- $a/(a+c)$

- $d/(b+d)$

- Sensitivity= $a/n_1=71/145=49.0\%$

- Specificity= $d/n_2=244/288=84.7\%$

- False pos rate= $c/n_2=44/288=15.3\%$

- PPV= $a/m_1=71/115=61.7\%$

- NPV= $d/m_2=244/318=76.7\%$

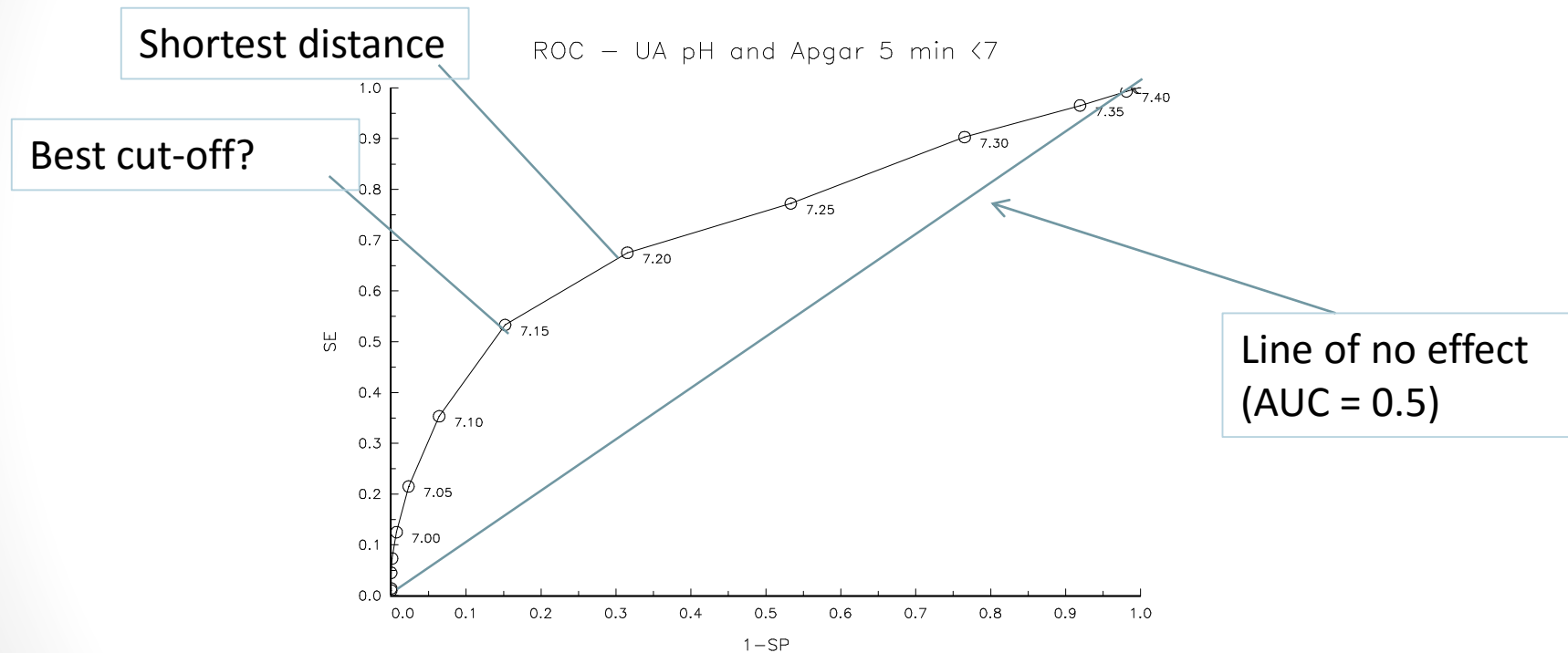
Note: Using the formula for 95%CI for proportions, you could also easily compute CI for the estimates (for large samples).



# Receiver operating characteristic (ROC) curves

- A **ROC curve**, is a [graphical](#) plot which illustrates the performance of a [binary classifier](#) system as its discrimination threshold is varied. It is created by plotting the fraction of [true positives](#) out of the total actual positives (TPR = true positive rate) vs. the fraction of [false positives](#) out of the total actual negatives (FPR = false positive rate), at various threshold settings

# Example. The corresponding ROC-curve. pH in umbilical cord's ability to predict low Apgar Score



Example. Table for producing an ROC-curve.  
pH in umbilical cord's ability to predict low Apgar Score

The cut-off point where the distance between the curve and the 'optimal point is smallest

Cutoff	Exp cases	Oexp cases	Exp ref	Oexp ref	Se	1-Sp	PPV	NPV	Distance	Area
6.8	3	286	2	43260	.01		0 .6	.99		
6.85	4	285	5	43257	.014		0 .444	.993	.986	0
6.9	13	276	17	43245	.045		0 .433	.994	.955	0
6.95	21	268	64	43198	.073	.001	.247	.994	.927	0
7.00	36	253	287	42975	.125	.007	.111	.994	.875	.001
7.05	62	227	1008	42254	.215	.023	.058	.995	.786	.003
7.1	102	187	2763	40499	.353	.064	.036	.995	.65	.015
7.15	154	135	6595	36667	.533	.152	.023	.996	.491	.054
7.2	195	94	13606	29656	.675	.315	.014	.997	.452	.152
7.25	223	66	23063	20199	.772	.533	.01	.997	.58	.31
7.3	261	28	33102	10160	.903	.765	.008	.997	.771	.504
7.35	279	10	39763	3499	.965	.919	.007	.997	.92	.648
7.4	287	2	42426	836	.993	.981	.007	.998	.981	.709
7.45	289	0	43118	144		1 .997	.007		1 .997	.724
7.5	289	0	43249	13		1	1 .007		1	1 .727

The area under the curve.

2016-09-29

# The ROC-curve showing the ability of maternal BMI to predict BW>4000g

## Area Under the Curve

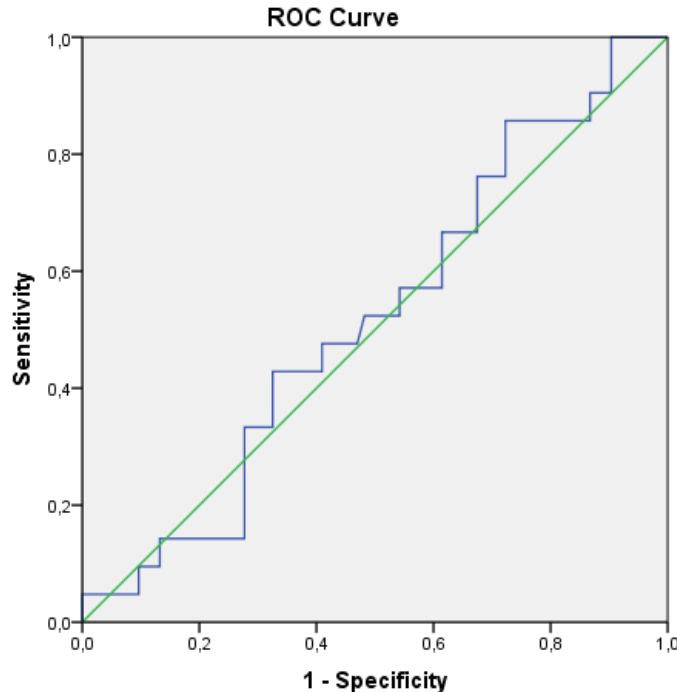
Test Result Variable(s): Maternal\_BMI

Area	Std. Error <sup>a</sup>	Asymptotic Sig. <sup>b</sup>	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,518	,067	,796	,386	,651

The test result variable(s): Maternal\_BMI has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5



Impressive? Significant?

No!

# Today

- **Today**

- Repetition
- Lecture 5:
  - Power calculations
  - Probability, proportions, CI for proportions
  - 2x2 Table. Chi2-test, Fisher exact test

## Lecture 6:

- Sensitivity, Specificity
- Positive predictive value, Negative predictive value
- ROC-curve

- **Next lecture**
- **Reliability**
- **Association versus agreement**
- **Bland-Altman plots**
- **Limits of agreement**
- **ICC**
- **Cohen's Kappa**