



# Statistics

## Computer Science

Faculty of Electrical Engineering, Automatics, Computer Science and  
Biomedical Engineering

Agnieszka Obłąkowska-Mucha  
Tomasz Szumlak

**Faculty of Physics and Applied Computer Science**  
**AGH University of Krakow**



More on Confidence Interval and Hypothesis Testing  
(test of two samples)  
including  
p value and power of a test

# Examples – C.I. for the difference in means

- **Ex.** A sample of 150 brand A light bulbs showed a mean life-time of 1400 hours and a standard dev. of 120 h. A sample of 200 brand B light bulbs showed the corresponding values of 1200 h and 80 h. Find (a) 95% and (b) 99% *C. I.* for the difference of the mean life-times of the populations of brands A and B.

a) The confidence interval for the difference in means can be written as:

$$\bar{L}_A - \bar{L}_B \pm z_c \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} = 1400 - 1200 \pm 1.96 \sqrt{\frac{120^2}{150} + \frac{80^2}{100}} = 200 \pm 24.8 \text{ h}$$

We can be 95% confident that the difference in population means lies between 175 and 225 h

b) And for the 99% *C. I.*:

$$\bar{L}_A - \bar{L}_B \pm z_c \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} = 200 \pm 1.96 \sqrt{\frac{120^2}{150} + \frac{80^2}{100}} = 200 \pm 32.6 \text{ h}$$

# Two population means

- The two independent samples are simple random samples from two distinct populations.
- For the two distinct populations:
  - ✓ if the sample sizes are small, the distributions are important (should be normal)
  - ✓ if the sample sizes are large, the distributions are not important (need not be normal)

$$\bar{X}_1 \rightarrow N\left(\mu, \frac{\sigma_1}{\sqrt{n_1}}\right) \quad \bar{X}_2 \rightarrow N\left(\mu, \frac{\sigma_2}{\sqrt{n_2}}\right) \quad \bar{X}_1 - \bar{X}_2 \rightarrow N(0, \sigma(\bar{X}_1 - \bar{X}_2))$$

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_{\bar{X}_1} - \mu_{\bar{X}_2})}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$\sigma(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

# Hypothesis – differences of means

- **Ex.** Two classes attended the same course and took an exam. The mean mark for the first class (40 students) was 74 points with a standard deviation of 8, the second (50 students) scored the mean of 78 points with the standard deviation of 7. Can we claim that one of the class is significantly better than other? We set the  $\alpha = 0.05$

- ✓  $H_0: \mu_1 = \mu_2$  - no significant difference, just fluctuation
- ✓  $H_1: \mu_1 \neq \mu_2$  - the second class is better
- ✓ The test statistics is the difference of means

$$\mu_{\bar{X}_1 - \bar{X}_2} = 0, \sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{64}{40} + \frac{49}{50}} = 1.606$$

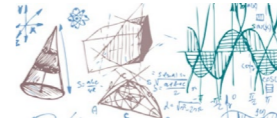
$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_{\bar{X}_1 - \bar{X}_2})}{\sigma_{\bar{X}_1 - \bar{X}_2}} = \frac{74 - 78}{1.606} = -2.49$$

$$z_{1,2} = \pm 1.96 \rightarrow H_0 \text{ - rejected @ 5\%}$$

- ✓ p-value:  $P(Z \leq -2.49) + P(Z \geq 2.49) = 0.0128$

we can use z-scores for normal distribution (big sample)

# CI – differences of means



**Ex.** A sample of 150 brand A light bulbs showed a mean life-time of 1400 hours and a standard dev. of 120 h. A sample of 200 brand B light bulbs showed the corresponding values of 1200 h and 80 h. Find (a) 95% and (b) 99% *C.I.* for the difference of the mean life-times of the populations of brands A and B.

a) The confidence interval for the difference in means can be written as:

$$\bar{L}_A - \bar{L}_B \pm z_c \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} = 1400 - 1200 \pm 1.96 \sqrt{\frac{120^2}{150} + \frac{80^2}{100}} = 200 \pm 24.8 h$$

We can be 95% confident that the difference in population means lies between 175 and 225 h

b) And for the 99% *C.I.*:

# CI – differences of means



- What would be a change if  $\sigma$  is unknown?
- The test statistics is the same: the difference of means

$$\bar{X}_1 - \bar{X}_2 \rightarrow N(0, \sigma(\bar{X}_1 - \bar{X}_2))$$

- but now  $\sigma(\bar{X}_1 - \bar{X}_2)$  is estimated:

$$\sigma(\bar{X}_1 - \bar{X}_2) \rightarrow \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$s_1^2 = \frac{1}{n_1} \sum_{i=1}^{n_1} (X_i - \bar{X})^2 \frac{1}{n_1}$$

- and test startistic is t-score:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_{\bar{X}_1} - \mu_{\bar{X}_2})}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1} \frac{s_1^2}{n_1} + \frac{1}{n_2 - 1} \frac{s_2^2}{n_2}}$$

- and we need to use  $t$ -Student distribution with  $\nu$  degrees of freedom

# Examples – C.I. - proportion



- **Ex.** A sample poll of 100 voters has been chosen at random in a given district. The result indicated that 55% of them were supporting a party A. Find (a) 95% and (b) 99% c.i. for the proportion of all voters supporting this party.

- a) The c.i. for the population  $p$  is defined as:  $P \pm z_c \sigma_P = P \pm z_c \sqrt{\frac{p(1-p)}{n}}$ . In order to estimate parameter  $p$  we use the sample proportion (measurement!)

$$P \pm z_{0.975} \sqrt{\frac{p(1-p)}{n}} = 0.55 \pm 1.96 \sqrt{\frac{0.55 \cdot 0.45}{100}} = 0.55 \pm 0.10$$

- b) And the 99% c.i.:

$$P \pm z_{0.995} \sqrt{\frac{p(1-p)}{n}} = 0.55 \pm 2.58 \sqrt{\frac{0.55 \cdot 0.45}{100}} = 0.55 \pm 0.13$$



# Examples – C.I. – two proportions



- **How we can compare two independent Population Proportion?**

- There are a few assumptions that must be fulfilled:

- samples should be random and independent

- number of successes  $> 5$

- ✓ This enables using the normal dist, approximation.

- ✓ The pooled proportion RV:

$$P = \frac{x_A + x_B}{n_A + n_B}$$

- ✓ the difference:

$$\Delta_P = P'_A - P'_B$$

$$\Delta_P \rightarrow N(0, \sigma_\Delta)$$

$$\sigma_\Delta = \sqrt{P(1 - P) \left( \frac{1}{n_A} + \frac{1}{n_B} \right)}$$

- ✓ and test statistics (z-score):

- $H_0: P_A = P_B$ , so  $\Delta_P = 0$

- $H_a: P_A \neq P_B$ , so  $\Delta_P \neq 0$

$$Z = \frac{(P'_A - P'_B) - (P_A - P_B)}{\sqrt{P(1 - P) \left( \frac{1}{n_A} + \frac{1}{n_B} \right)}}$$

# Test of two proportions



**Ex.** Two types of valves are being tested to determine if there is a difference in pressure tolerances. Fifteen out of a random sample of 100 of Valve A cracked under 4,500 psi. Six out of a random sample of 100 of Valve B cracked under 4,500 psi. Test at a 5% level of significance.

$$P = \frac{x_A + x_B}{n_A + n_B} = \frac{15 + 6}{100 + 100} = 0.105 \qquad P'_A = \frac{15}{100} \qquad P'_B = \frac{6}{100}$$

the difference:

$$\Delta_P = P'_A - P'_B = 0.21 \qquad \Delta_P \rightarrow N(0, \sigma_\Delta)$$

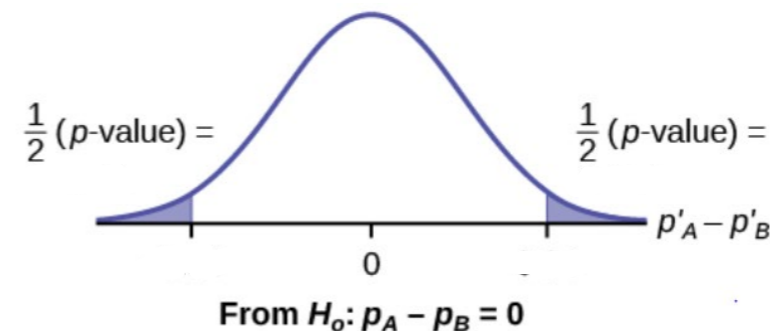
and test statistics (z-score):

$$Z = \frac{(P'_A - P'_B) - (P_A - P_B)}{\sqrt{P(1 - P) \left( \frac{1}{n_A} + \frac{1}{n_B} \right)}}$$

$$\sigma_\Delta = \sqrt{P(1 - P) \left( \frac{1}{n_A} + \frac{1}{n_B} \right)}$$

$$H_0: P_A = P_B, \text{ so } \Delta_P = 0$$

$$H_a: P_A \neq P_B, \text{ so } \Delta_P \neq 0$$



The p-value is 0.0379, so we can reject the null hypothesis. At the 5% significance level, the data support that there is a difference in the pressure tolerances between the two valves.

# Examples – C.I. for variance

$$\chi_{0.025}^2 \leq \frac{(n-1)\hat{S}^2}{\sigma^2} \leq \chi_{0.975}^2$$

- **Ex.** The standard deviation of heights of 16 male students chosen at random in a school of 1000 male students yielded:  $s = 5.9 \text{ cm}$ . Find (a) 95% and (b) 99% c.i. of the standard deviation for all male students at this school. Assume that height is normally distributed.

✓ The critical values in this case are:  $S \frac{\sqrt{n}}{\chi_{0.975}}$  and  $S \frac{\sqrt{n}}{\chi_{0.025}}$ . From a table for  $\nu = 15$ :

$$\chi_{0.975}^2 = 27.5 \rightarrow \chi_{0.975} = 5.24 \text{ and } \chi_{0.025}^2 = 6.26 \rightarrow \chi_{0.025} = 2.50$$

$$C.I._{0.975}^{(s)} = S \frac{\sqrt{n}}{\chi_{0.025}} < s < S \frac{\sqrt{n}}{\chi_{0.975}} = 4.9 < s < 9.4 \text{ cm}$$

✓ The 99% c.i.:

$$C.I._{0.995}^{(s)} = S \frac{\sqrt{n}}{\chi_{0.005}} < s < S \frac{\sqrt{n}}{\chi_{0.995}} = 4.1 < s < 11.0 \text{ cm}$$

## Examples – test of variance

$$\chi_{0.025}^2 \leq \frac{(n-1)\hat{S}^2}{\sigma^2} \leq \chi_{0.975}^2$$

**Ex.** With individual lines at its various windows, a post office finds that the standard deviation for normally distributed waiting times for customers on Friday afternoon is 7.2 minutes. The post office experiments with a single, main waiting line and finds that for a random sample of 25 customers, the waiting times for customers have a standard deviation of 3.5 minutes.

With a significance level of 5%, test the claim that **a single line causes lower variation among waiting Times (shorter waiting times) for customers.**

**Ex.** A SCUBA instructor wants to record the collective depths each of his students dives during their checkout. He is interested in how the depths vary, even though everyone should have been at the same depth. He believes the standard deviation is three feet. His assistant thinks the standard deviation is less than three feet. If the instructor were to conduct a test, what would the null and alternative hypotheses be?

# Examples – test of variance

$$\chi^2_{0.025} \leq \frac{(n-1)\hat{S}^2}{\sigma^2} \leq \chi^2_{0.975}$$

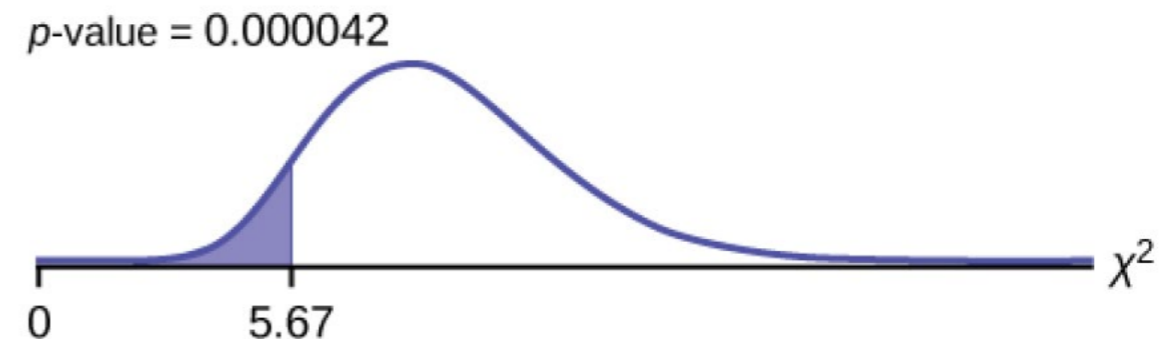
**Ex.** With individual lines at its various windows, a post office finds that the standard deviation for normally distributed waiting times for customers on Friday afternoon is 7.2 minutes. The post office experiments with a single, main waiting line and finds that for a random sample of 25 customers, the waiting times for customers have a standard deviation of 3.5 minutes.

With a significance level of 5%, test the claim that **a single line causes lower variation among waiting Times (shorter waiting times) for customers.**

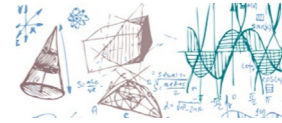
- $H_0: \sigma^2 = 7.2^2$
- $H_a: \sigma^2 < 7.2^2$

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(25-1)(3.5)^2}{7.2^2} = 5.67$$

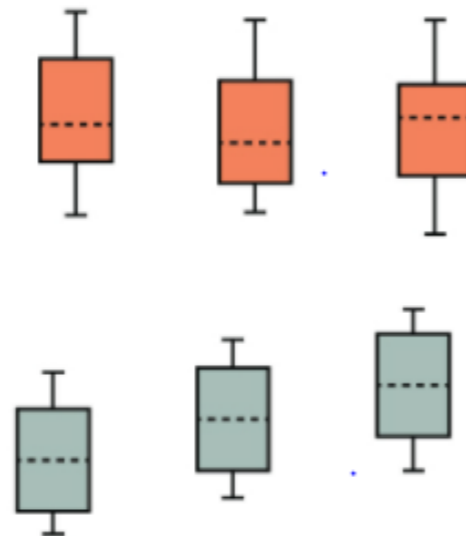
$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(25-1)(3.5)^2}{7.2^2} = 5.67$$



# Analysis of variance



- In case of comparing the mean values of two populations we introduced how the sampling distribution of differences (and sums...) of means can be obtained. And what about the distribution of **differences of variances** that can be denote as:  $S_1^2 - S_2^2$ ?
- For hypothesis tests comparing averages between more than two groups: "Analysis of Variance" (ANOVA).
- The null hypothesis is simply that all the group population means are the same. The alternative hypothesis is that at least one pair of means is different.



$H_0: \mu_1 = \mu_2 = \mu_3$  and the three populations have the same distribution if the null hypothesis is true. Differences are due to random variation.

If the null hypothesis is false, then the variance of the combined data is larger which is caused by the different means. Differences are too large to be due to random variation.

# C.I. for variance ratios

- In case of comparing the mean values of two populations we introduced how the sampling distribution of differences (and sums...) of means can be obtained. And what about the distribution of **differences of variances** that can be denote as:  $S_1^2 - S_2^2$ ?
- It turns out, that such distribution is **surprisingly** troublesome! Instead we can use **another statistic**:  $S_1^2/S_2^2$ . We conclude that if the ratio is small/large the difference between variance is large, conversely if the ratio is close to unity the difference should be small, and we have a theorem...
- **Theorem 1.** Consider, we have two random and independent samples, of size  $n$  and  $m$  respectively, drawn from two normal populations with variances  $\sigma_1^2$  and  $\sigma_2^2$ . It can be shown that the statistic

$$F = \frac{\frac{mS_1^2}{(m-1)\sigma_1^2}}{\frac{nS_2^2}{(n-1)\sigma_2^2}} = \frac{\frac{\hat{S}_1^2}{\sigma_1^2}}{\frac{\hat{S}_2^2}{\sigma_2^2}}$$

has the F distribution with  $m - 1, n - 1$  degrees of freedom

## C.I. for variance ratios

- Let's repeat our scheme again. We denote by  $F_{0.01}$  and  $F_{0.99}$  the values of F R.V. for which 1% of the area lies in each tail of the F distribution, then with 98% confidence we have the following

$$F_{0.01} \leq \frac{\frac{\hat{S}_1^2}{\sigma_1^2}}{\frac{\hat{S}_2^2}{\sigma_2^2}} \leq F_{0.99}$$

- So, a 98% C.I. for the variance ratio  $\sigma_1^2/\sigma_2^2$  of two populations is given by:

$$\frac{1}{F_{0.99}} \frac{\hat{S}_1^2}{\hat{S}_2^2} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{1}{F_{0.01}} \frac{\hat{S}_1^2}{\hat{S}_2^2}$$

- The respective critical values are tabulated. The way we read the tables are a bit difficult but we do that during the tutorial, so don't worry!



## Examples – C.I.

- **Ex.** Two samples of sizes 16 and 10 are drawn from two normal populations. If their variances are found to be 24 and 18, respectively, find (a) 98% and (b) 90% confidence limits for the ration of the variances

✓ First we can find the modified sample variances:

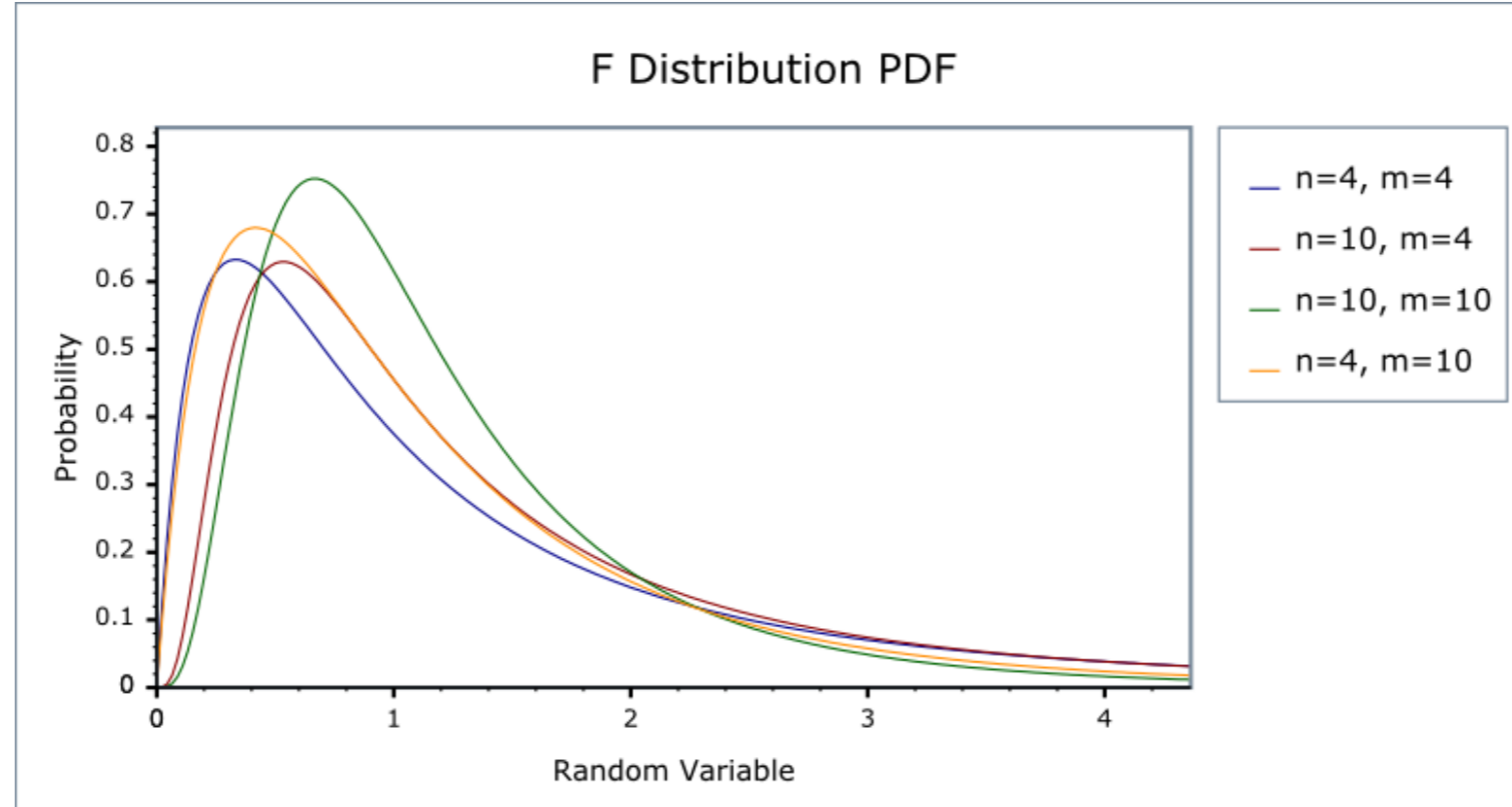
$$\hat{s}_1^2 = \frac{m}{m-1} s_1^2 = \frac{16}{15} \cdot 24 = 25.2$$

$$\hat{s}_2^2 = \frac{n}{n-1} s_2^2 = \frac{10}{9} \cdot 18 = 20$$

- ✓ We need to use F distribution. For the 98% confidence interval we have:  $F_{0.99} = 4.96$  for  $v_1 = 16 - 1 = 15$  and  $v_2 = 10 - 1 = 9$ . And  $F_{0.01} = 1/3.89$  for the same values of degrees of freedom. So, the c.i.:

$$C.I._{0.98}^{s^2} = \frac{1}{4.96} \frac{25.2}{20} < \frac{\sigma_1^2}{\sigma_2^2} < 3.89 \frac{25.2}{20} = 0.28 < \frac{\sigma_1^2}{\sigma_2^2} < 4.90$$

# Examples – C.I.



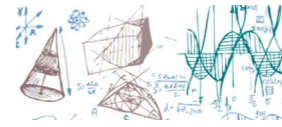
# Type I and type II error discussion

## Type I Error



## Type II Error





# What can happen?

- No matter what we do, in real life we are never going to know for certain that we made the right choice or we in fact made one of the errors
- Remember Type I is to **REJECT** the null when it is true and Type II is **NOT TO REJECT** the null when it is false in reality
- Let's start from a generic (and easy) example, then we try to make a number of general statements and we get back to some more examples...

- Say, we are testing (for normal distribution) the following:

$$H_0: \mu = 14 \quad H_1: \mu > 14$$

- After applying our „procedure” we decided to reject the hypo at significance level  $\alpha = 0.05$
- Either of the two happened: **the null is wrong** and we did good or the **null is true** and we made a Type I error



# What can happen?

- After applying our „procedure” we decided to accept the hypo at significance level  $\alpha = 0.05$
- Either of the two happened: **the null is ok** and we make a good decision or the **null is false** and we made a Type II error
- Is it a bit confusing...? Yes, and we can summarise this using the CONFUSION MATRIX – now you know why we use this name...

### Ground truth

Conclusions using data



	$H_0$ is false	$H_0$ is true
Reject	Super	Type I error
Do not reject	Type II error	Super

Note, using data we in fact make calculations, the ground truth remains, however, unknown...

# A trial example



- This is a classic example: criminal trial

$H_0$ : not guilty (no crime),

$H_1$ : guilty (criminal)

- In a typical jurisdiction system we have a rule „not guilty until proven otherwise”, so we need to have very strong evidences to convict
  - ✓ Type I: rule guilty when in reality a person is innocent
  - ✓ Type II: not guilty when a person committed a crime
- Definitely, we want to make sure, that we never convict an innocent person (well, in real life we do...)
- What we can do – we should decrease the possibility of making Type I error as much as possible

# A trial example

$H_0$ : not guilty (no crime),

$H_1$ : guilty (criminal)

decision	truth	
	guilty	not guilty
not guilty	Type II error	correct
guilty	correct	Type I error



<https://oursite-easyread.wvda.org.au/>

- Ok, it is time to write down some equations!

# Significance and power



- The probability of **Type I error** is called the **significance level** of the statistical test (we know that already), we write this as

$$P(\text{Type I error} | H_0) = \alpha$$

- Usually we start the test procedure by setting the value of  $\alpha$
- Now, we call the probability of making a Type II error as  $\beta$ 
  - ✓ There is a complication – it is much harder to define this error, it depend on the true value of the testing parameter, the size of the data sample and on the significance level itself!
- The **POWER** of a statistical test is **the probability of rejecting the null hypothesis when it is false**, we write

$$\mathcal{R} = 1 - P(\text{Type II error}) = 1 - \beta$$

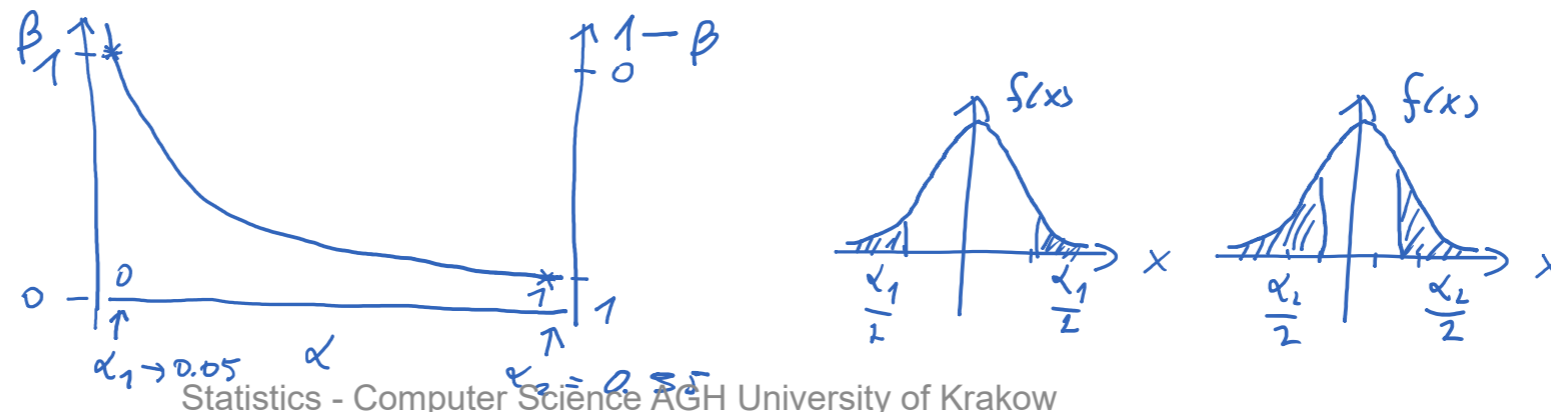
- The power depends on  $\beta$  – and we still do not know how to calculate it...



# Interplay of $\alpha$ and $\beta$



- Ok, we still build up the suspense..., Before we show all there is about  $\beta$  let's get some intuitive understanding
- So, why not pick-up some very, very tiny value for the significance? The critical region would be very small, thus we would make almost impossible to reject the null hypo. We also say, **that the efficiency of the selection is very high**, but the probability of making the Type II error will also be very high (or we say, **the purity of the selection is low**)
- Conversely, for the large values of  $\alpha$  the probability of Type II errors will go down
- Also, from this discussion you should start to understand, that the dependency is not trivial (i.e., it is not just some linear function or such)



# Formal description



- We should have acquired enough intuition by now and can introduce more formal definitions regarding errors and the test power
- First things first: every hypothesis, in one way or another, assumes some knowledge about the underlying probability density function

$$P.D.F \equiv f(\vec{x}; \vec{\lambda})$$

- If we know all of the model parameters, we say we are testing a **simple hypothesis**. Otherwise, we call it a **composite hypothesis**
- At the end of the day, different hypothesis will propose different values of parameters, we can make it clear by writing:

$$H_0 = H_0(\vec{\lambda} = \vec{\lambda}_0), H_i = H_i(\vec{\lambda} = \vec{\lambda}_i)$$

- Since, the null hypothesis makes a statement regarding a P.D.F. we can also make predictions for the probability of observing a point (our data sample):  $X^{(s)} = (X_1, X_2, \dots, X_n)$
- The first step now, is to pick up a value of the test significance  $\alpha$  and based on this value to define the critical region  $S_c$  of the test

$$P(X \in S_c | H_0) = \alpha$$

# Formal description



- So, for a given null hypothesis, the critical region is defined by requesting that the probability of observation of a sample  $X^{(s)}$  within the region is  $\alpha$ . If  $X^{(s)}$  is inside the critical region we should reject the null
- **To make things more practical** (we did mention this when discussing estimation) we do not use the data points from a sample explicitly but rather **define an appropriate test statistic** (random function)

$$T = T(X^{(s)}): X^{(s)} \rightarrow T(X^{(s)})$$

- Also, a second map is necessary (critical regions)

$$S_c(X^{(s)}) \rightarrow C(X^{(s)})$$

- And now, we reject the null if  $T$  belongs to  $C$
- The Type I error is defined then by  $P(X^{(s)} \in S_c | H_0) = \alpha$  (rejected null and it is true), this we already know, **but with all these considerations we can now also define Type II** ( $X^{(s)}$  not in the critical region but null is false)

$$P(X^{(s)} \notin S_c | H_1) = \beta$$

- Finally!

# Formal description



- Again, using the interplay between Type I and Type II we can define the critical region of the test in a better way. The driving idea is that for a particular value of  $\alpha$  **the critical region is defined by requiring the probability  $\beta$  (Type II) to be minimum**
- With the critical region defined, we can now clearly define the **probability of rejecting the null as a function of „true” hypothesis**, and we call it the power of the test

$$\mathcal{R}(S_c, \vec{\lambda}) = P(X \in S_c | H) = P(X \in S_c | \vec{\lambda})$$

- The complementary probability is called the operating characteristic function of the test, we interpret it as the **probability of accepting the null**

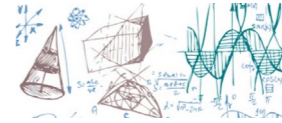
$$OC(S_c, \vec{\lambda}) = 1 - \mathcal{R}(S_c, \vec{\lambda})$$

- We can also easily find

$$\mathcal{R}(S_c, \lambda_0) = \alpha, \mathcal{R}(S_c, \lambda_1) = 1 - \beta, OC(S_c, \lambda_0) = 1 - \alpha, OC(S_c, \lambda_1) = \beta$$

- The most powerful test:  $max\{\mathcal{R}(S_c, \lambda_1) = 1 - \beta\}$

# Examples!



- **Ex. 1** Say, we consider a sampling from a normal distribution with known variance  $\sigma = 18$ , the size of the data sample  $n = 36$ . Our hypotesis

$$H_0: \mu = 40, H_1: \mu < 40$$

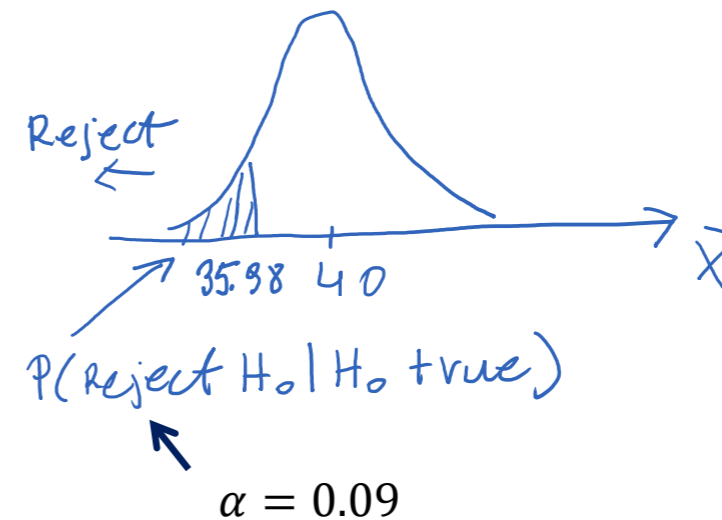
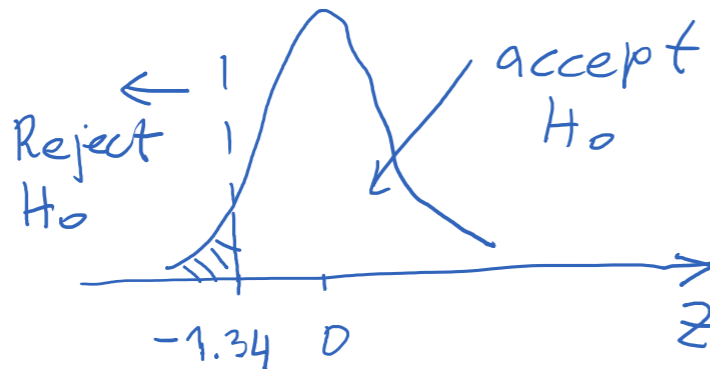
- We set the significance of the test:  $\alpha = 0.09$  and our test statistics is

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

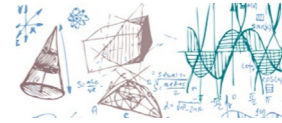
$$\bar{X} = \mu_0 + Z \frac{\sigma}{\sqrt{n}} = 40 + \frac{18}{\sqrt{36}}(-1.34)$$

- The critical region

$$\bar{X} = 35.98$$



# Examples!

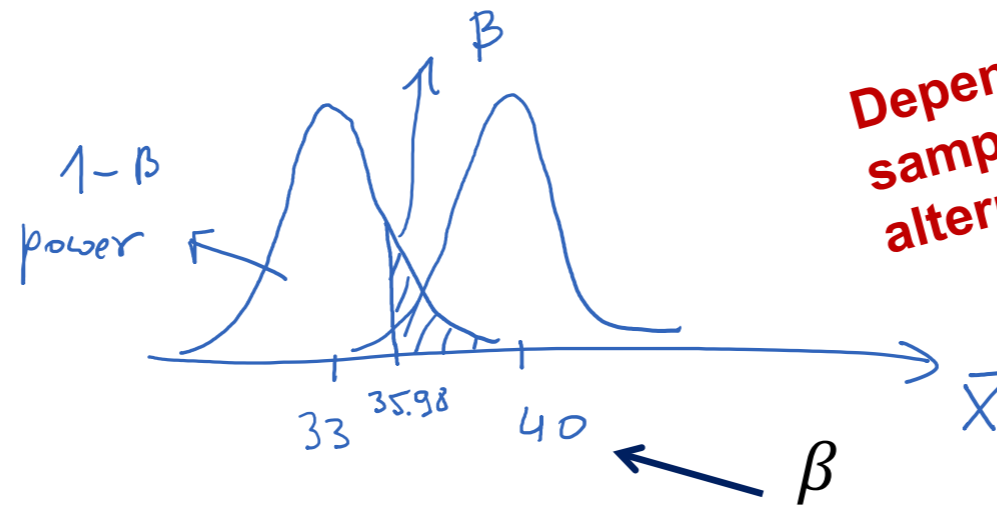


- Now, what if the true mean is  $\mu = 33$ ? What is the Type II error in this case?

$$P(\text{Accept } H_0 | \mu = 33)$$

- We reject the  $H_0$  hypo if the mean is less than 35.98, so the Type II error will occur with the probability

$$P(\bar{X} > 35.98 | H_1) = P(\bar{X} > 35.98 | \mu = 33)$$



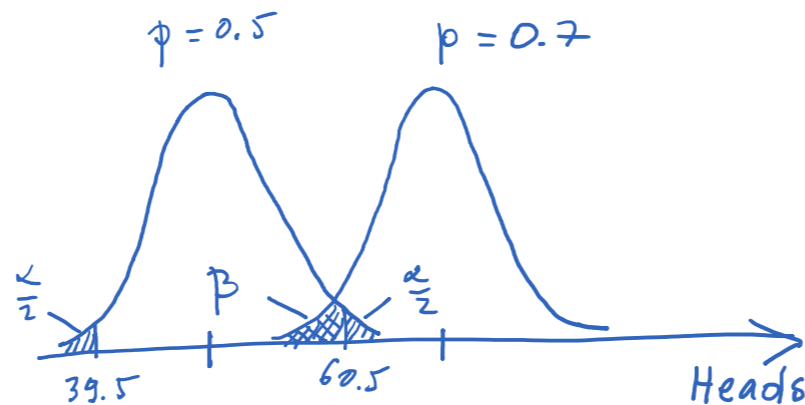
**Depends on the sample size and alternative!**

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{35.98 - 33}{18/6} = \frac{2.98}{3} = 0.99 \rightarrow P(Z > 0.99 | H_1) = 0.16$$

# Examples!



- Note, that when we calculate power and probability of the Type II error we use the alternative hypothesis – we say the calculations are done under the alternative hypothesis
- **Ex. 2** Let's go back to the example with the coin (last lecture) we mentioned the possibility of accepting it if the true probability of heads is  $p = 0.7$ . This time we calculate the exact value of the probability of accepting the null hypo (i.e., the coin is fair)
  - ✓ Our hypos are:  $H_0: p = 0.5, H_1: p = 0.7$
- The numbers we worked out: for the sample of **100 tosses**, we accept the null if we get between **39.5 and 60.5 heads**. This gave us, in turn, the significance  $\alpha = 0.0358$ .



# Examples!



- The right-hand side distribution represent 100 tosses with the probability of getting head set to 0.7
- Let's work out the probability of the Type II error

$$\mu_1 = np = 100 \cdot 0.7 = 70, \sigma_1 = \sqrt{npq} = 4.58$$

- Now, we need to translate the critical points of the null hypo into Z-score for the alternative hypo

$$z_{+/-}^{(1)} = \frac{h_{+/-} - \mu_1}{\sigma_1} \rightarrow z_+^{(1)} = \frac{60.5 - 70}{4.58} = -2.07, z_-^{(1)} = \dots = -6.66$$

$$\beta = P(-6.66 \leq z_1 \leq -2.07 | H_1) = 0.019$$

- There are slim chances, we accept the null (the coin is fair) when it is loaded with the probability of getting heads 0.7
- There is more to this example... When discussing the *C.I.* we just picked two values for heads and checked the probability – from this we were able to get the significance level – **we say the decision rule was given**



# Examples!



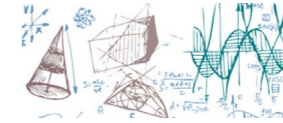
- We can also chose other strategies
  - ✓ **We could decide on the significance level** (the most popular way) and work out the decision rule and the Type II error
  - ✓ **We choose both  $\alpha$  and  $\beta$**  and then get a decision rule
- **Ex. 3** Parametric scan of Type II and the test power as a function of heads probability – for the sake of this example we choose:  $p = 0.4, 0.6, 0.8, 0.9$  ( $H_2, H_3, H_4, H_5$ )
- Can do the calculations as previously: evaluate the  $\mu_i, \sigma_i$  first and then  $z_{+/-}^{(i)}$  using the critical points of the null hypo

$$\mu_1 = 40, \sigma_1 = 4.90, z_{-}^{(2)} = \frac{39.5 - 40}{4.90} = -0.10, z_{+}^{(2)} = 4.18$$

$$\mu_2 = 60, \sigma_2 = 4.90, z_{-}^{(2)} = \frac{39.5 - 60}{4.90} = -4.18, z_{+}^{(2)} = 0.10$$

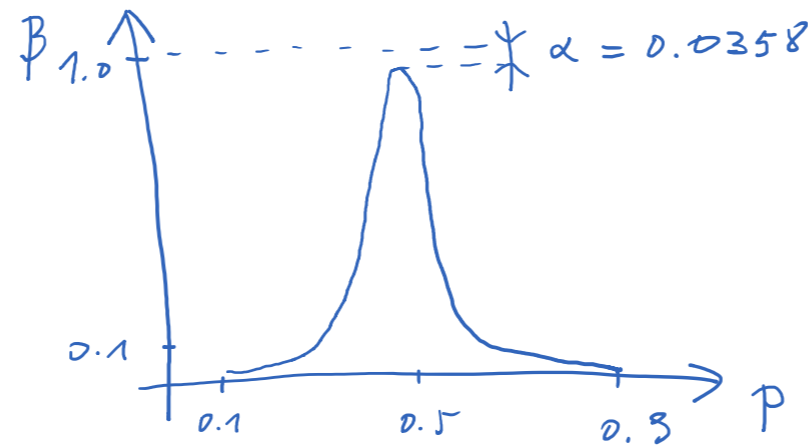
$$\mu_3 = 80, \sigma_3 = 4.0, z_{-}^{(3)} = \frac{39.5 - 80}{4.90} = -10.12, z_{+}^{(2)} = -4.88$$

# Examples (OC curve)

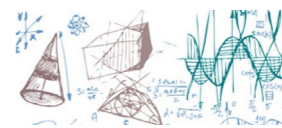


- Ok, you see the idea, we could add a few points and use the symmetry principle ( $p = 0.4 \equiv p = 0.6$ ). For all intents and purposes the probability for  $p$ : 0.1, 0.2, 0.8, 0.9 is very close to 0. Let's make a table

$p$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\beta$	0.000	0.000	0.019	0.54	0.964	0.54	0.019	0.000	0.000

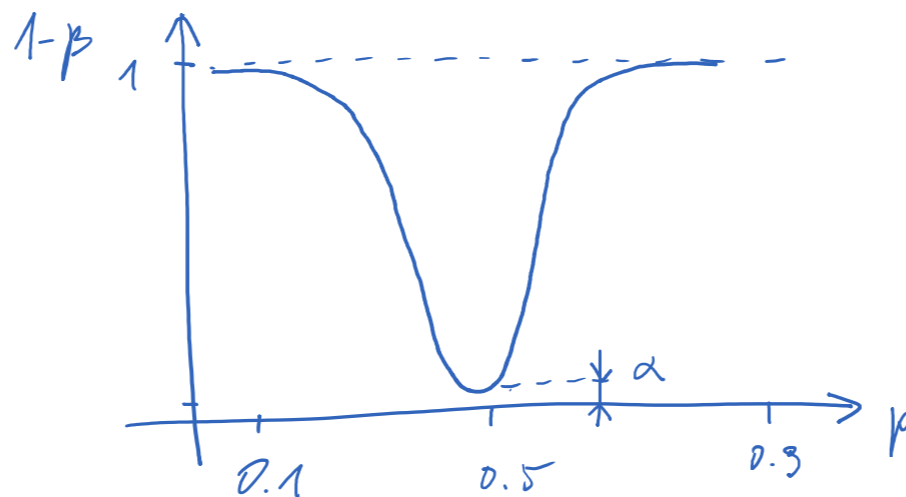


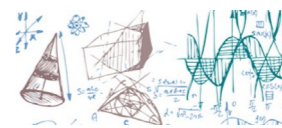
- The graph like that is called the **operating characteristic curve** of the decision rule. The faster the rise the better is the decision rule for rejecting „wrong” hypothesis



# Examples (Power curve)

- We can also plot the power of the statistical test versus the parameter – this carries actually the same information as the OC curve – we can judge the ability of a test to reject hypos which are „false” (w.r.t. the null)
- For our example, we can use the table from the previous slide to workout the values of the power





# Examples

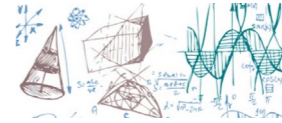
- **Ex. 4** Climbing ropes manufacturer is testing a new technology that allegedly increases the breaking strength. Assume that the „normal” b.s.  $\mu_{bs-0} = 300 \text{ kg}$  and the  $\sigma_0 = 24 \text{ kg}$ . First we design the decision rule assuming  $\alpha = 0.01$  and  $n = 64$ . Then we evaluate the probability of accepting the old technology when in fact there is an increase in the b.s.  $\mu_{bs-1} = 310 \text{ kg}$
- We are interested in a one-tailed test of course, for the given significance level we find the critical z-score to be  $z_c = 2.33$

$$z_c = \frac{\bar{X} - \mu_{bs-0}}{\sigma_0/\sqrt{n}} \rightarrow \bar{X} = \mu_{bs-0} + \frac{\sigma_0}{\sqrt{n}} z_c = 307.0$$



- ✓ Reject the  $H_0: \mu_{bs-0} = 300 \text{ kg}$  if  $\bar{X} > 307.0 \text{ kg}$
- ✓ Otherwise accept it

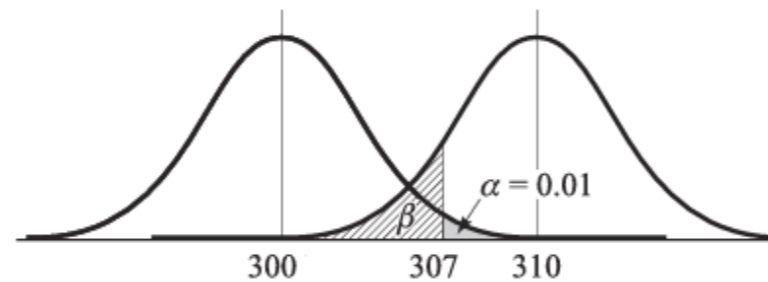
# Examples



- Now, the Type II error ( $\beta$ ), using  $H_1: \mu_{b_{S-1}} = 310 \text{ kg}$

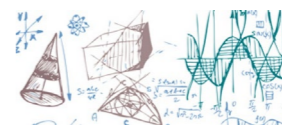
$$z_c^{(1)} = \frac{307 - 310}{3} = -1.0$$

- We assumed the same sample size and variance for constructing the alternative sampling distribution



- We can make a scan and plot the OC curve for our example

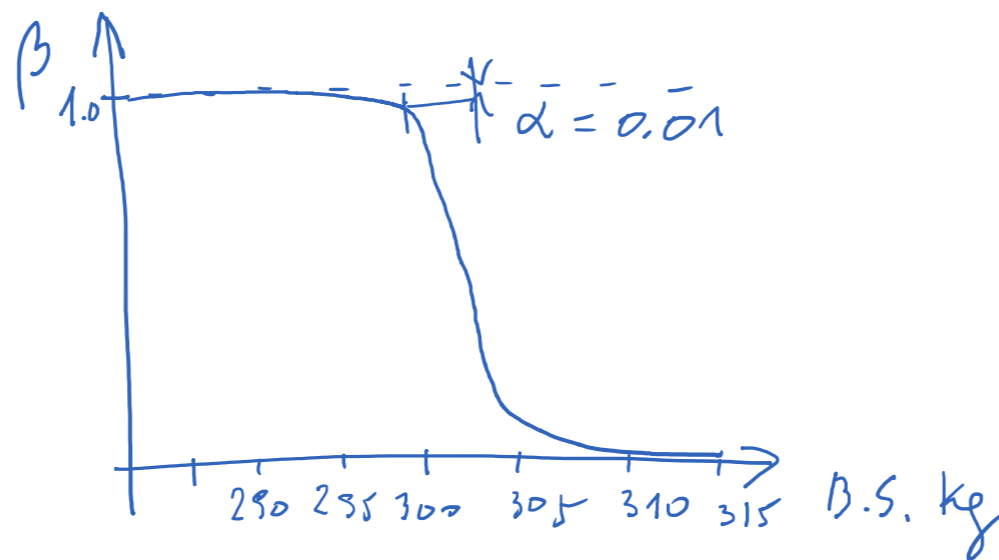
$$\beta = P(\bar{X} < 307.0 | H_1) = P(z < -1.0 | H_1) = 0.16$$



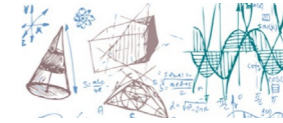
# Examples

- The scan outcome is in the table (strongly recommend to follow the calculations on your own!)

B.S.	290	295	300	305	310	315	320
$\beta$	1.000	1.000	0.990	0.749	0.159	0.004	0.000



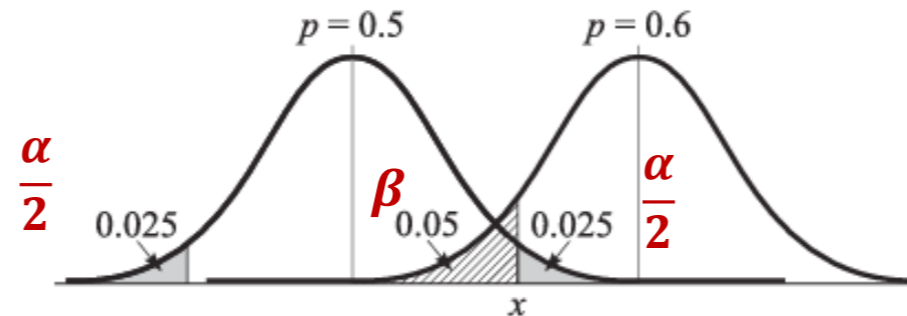
# Evaluating sample size



- **Ex. 5** Back again to the coin example, this time we want to work out the sample size given the null and alternative and both error types.

Our null will be  $H_0: p = 0.5$  and alternative  $H_1: p \neq 0.5$ . This time, however, we set that the max difference can be  $\Delta p = 0.1$ . The  $\alpha = 0.05$  at the most and  $\beta = 0.05$  at the most. What is  $n_{min} = ?$

- The restrictions coming from the limits set on the errors can be represented using graphs, here  $x$  is the number of heads observed



- We need to evaluate:  $P(x > x_c | H_0) = \frac{\alpha}{2}$ ,  $P(x < x_c | H_1) = \beta$

# Evaluating sample size



- We have the following: 
$$\frac{x_c - np_0}{\sqrt{np_0q}} = \frac{x_c - 0.5 \cdot n}{0.5 \cdot \sqrt{n}} \rightarrow \frac{\alpha}{2}$$

$$\frac{x_c - np_1}{\sqrt{np_1q}} = \frac{x_c - 0.6 \cdot n}{0.49 \cdot \sqrt{n}} \rightarrow \beta$$

- Note, by making these assumptions we automatically assert that the max Type II error for cases where the p is outside the range (0.4, 0.6) will be equal or smaller than 0.05! Using the normal approximations we have

$$\frac{x_c - 0.5 \cdot n}{0.5 \cdot \sqrt{n}} = 1.96 \rightarrow x_c = 0.5 \cdot n + 0.98 \cdot \sqrt{n}$$

$$\frac{x_c - 0.6 \cdot n}{0.49 \cdot \sqrt{n}} = -1.645 \rightarrow x_c = 0.6 \cdot n - 0.806 \cdot \sqrt{n}$$

$$n = 318.98 \rightarrow n_{min} = 319, x_c = 177$$

- Accept the null hypo if the number of heads are in range:  $159.5 \pm 17.5$