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

Confidence

- Statistical statements regarding R.Vs. and probability should always be interpreted in terms of model parameters and confidence

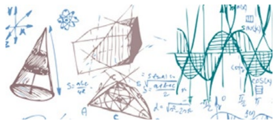
- **We express the confidence using fractional numbers** (%). So, we could say, for instance, a $\kappa\%$ confidence interval for parameter θ (based on an actual observation) is the interval from θ_{-} to θ_{+}, where $\kappa\% \rightarrow 99\%, 95\%, 90\%, \ldots$

- Its meaning is as follow: if we observe an event with the prob. of 95% we say it is reasonable, on the other hand if this is just 5% it should be considered **unlikely**

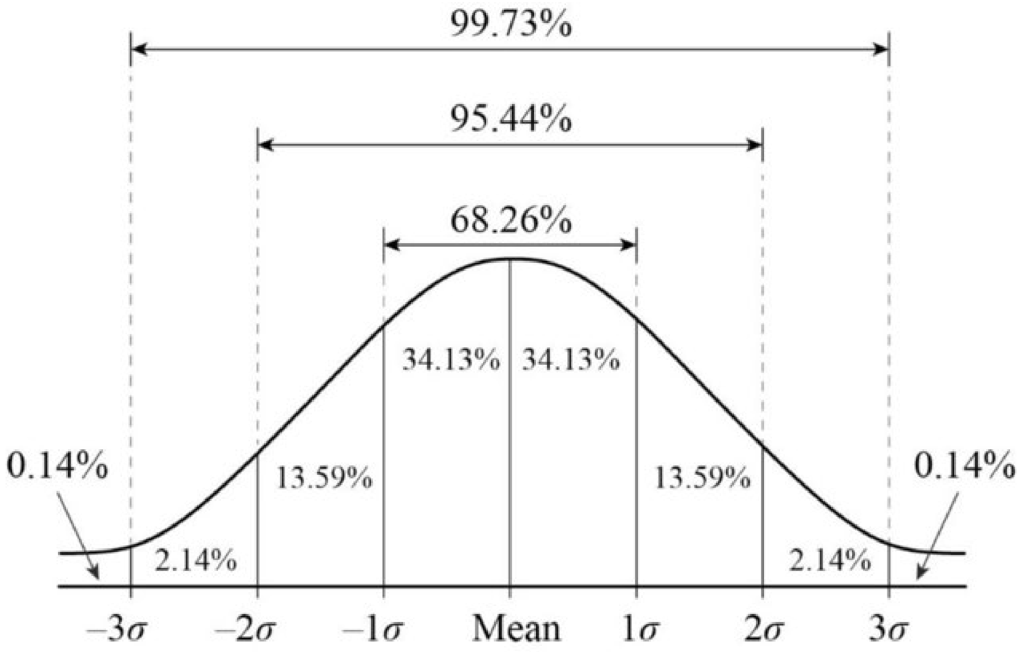
- So, what left now is to evaluate the confidence interval, we reserve for example **5%** of probability for „strange” events and consider both cases too-low-strange and too-high-strange

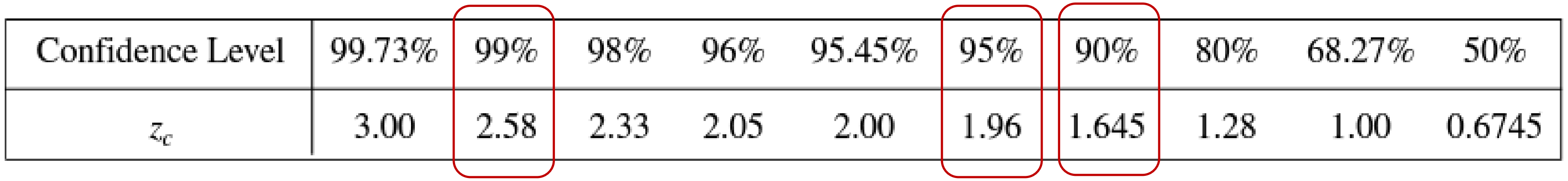
- This is, so called, **two tailed** or two-sided confidence interval and we have reserved **2.5%** probability for very high and very low results

$C.I.$ for the normal distribution



- We already know a lot about evaluating probabilities using the normal distribution

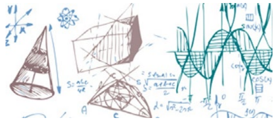




Confidence Level	99.73%	99%	98%	96%	95.45%	95%	90%	80%	68.27%	50%
z_c	3.00	2.58	2.33	2.05	2.00	1.96	1.645	1.28	1.00	0.6745

z-score for CL

$C.I.$ for the normal distribution



- Using the plot or the table from the previous slide we write for the critical values $\boldsymbol{z_c} = \pm\mathbf{1.96}$, which corresponds to the confidence level of 95%:

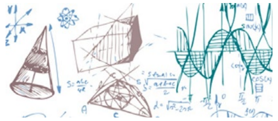
$$P\left(-1.96 \leq \frac{\bar{X} - \boldsymbol{\mu}}{\sigma/\sqrt{n}} \leq 1.96\right) = 0.95$$

- As usual, there are some tricks… For instance if we knew the distribution variance (remember the normal model has two parameters!) we could immediately solve these inequalities

$$P\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} \leq \boldsymbol{\mu} \leq \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$$

- This is a random interval, defined around the sample mean, which contains the unknown population mean with the probability of 95%. So, the 95% C.I. for μ is given by

$$C.I._{95\%}^{\mathcal{N}} = \left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$



$C.I.$ for the normal distribution

- A general formula that can be applied for the normal distribution for its mean is then

$$C.I.^{\mathcal{N}}_{100\cdot(1-\alpha)\%} = \left(\bar{X} - z_c\frac{\sigma}{\sqrt{n}}, \bar{X} + z_c\frac{\sigma}{\sqrt{n}}\right)$$

- Nice, but… what if we do not know the distribution variance (and we usually do not)? The most sensible approach would be to use the sample variance to estimate σ^2

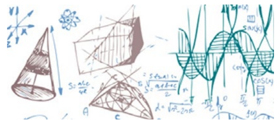
$$S^2 = \frac{1}{n-1}\sum\nolimits_i (X_i - \bar{X})^2 \to E[S^2] = \sigma^2$$

- We define a new $\mathrm{R.V.}\ T$

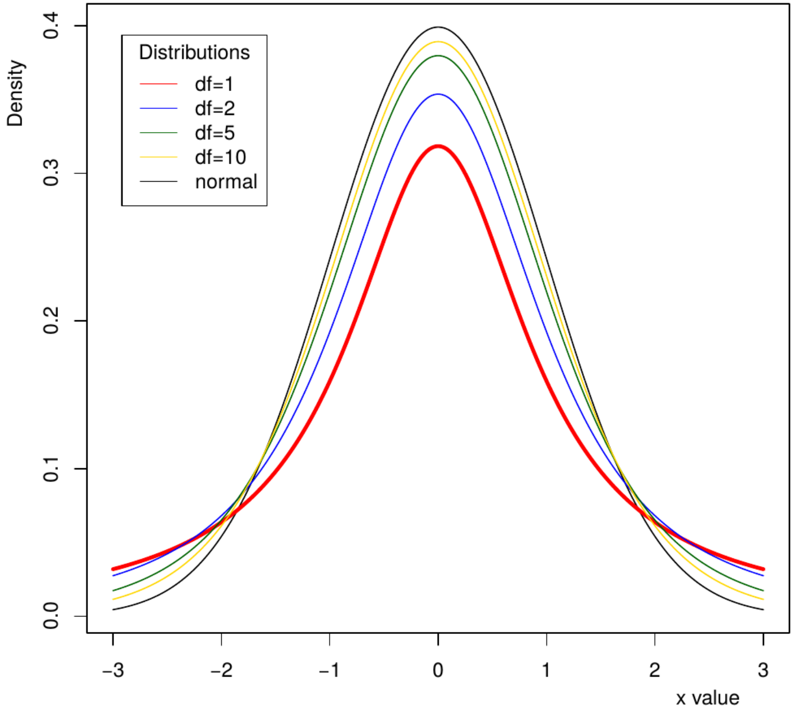
$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \to P\left(-t \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t\right) = 1 - \alpha$$

- The R.V. T follows the Student’s t-distribution (actually there is a whole family of distribution) $T \sim t(\nu)$

t-distribution



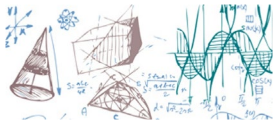
- t-distribution is similar to the normal one (obviously!)

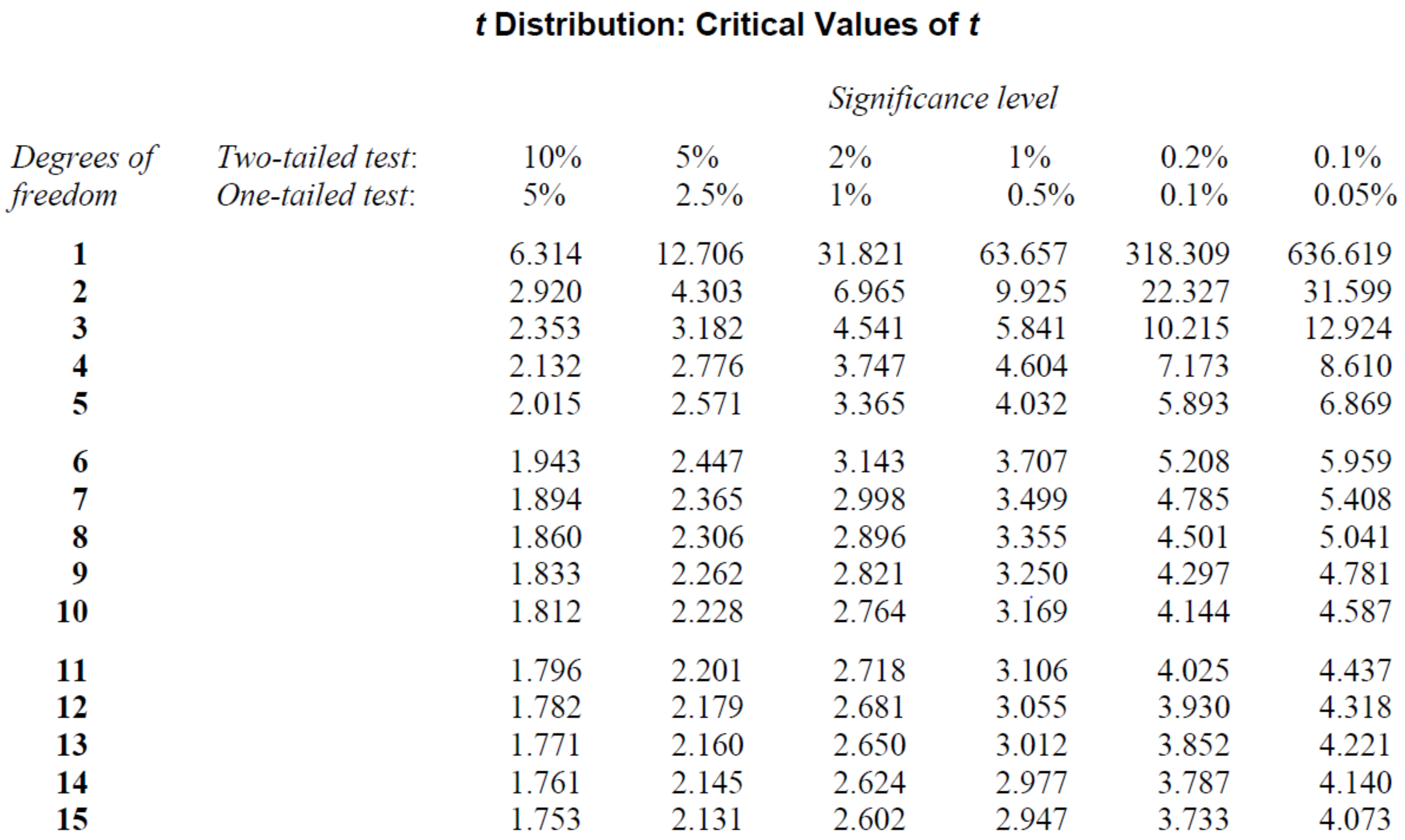


	50%	90%	95%	99%	99.9%
DF=5	0.73	2.02	2.57	4.03	6.87
DF=10	0.70	1.81	2.23	3.17	4.59
DF=20	0.69	1.72	2.09	2.85	3.85
DF=30	0.68	1.70	2.04	2.75	3.65
DF=50	0.68	1.68	2.01	2.68	3.50
(Normal)	0.67	1.64	1.96	2.58	3.29

- The larger the ν the more resemblance to the normal curve
- We use tables to evaluate the critical values t_c for a given confidence levels, let's continue on the next slide…

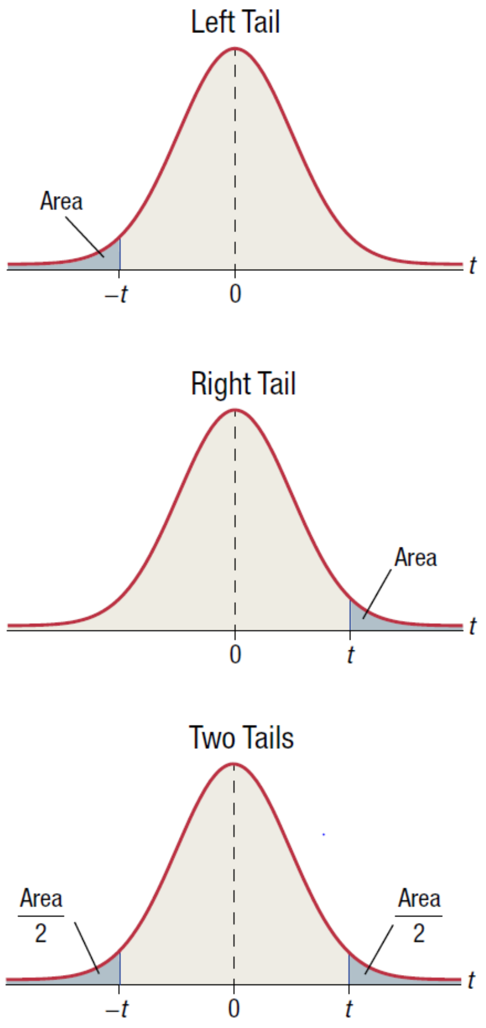
t-distribution

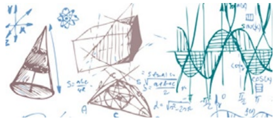




t Distribution: Critical Values of t

Degrees of freedom		*Significance level*					
	Two-tailed test:	10%	5%	2%	1%	0.2%	0.1%
	One-tailed test:	5%	2.5%	1%	0.5%	0.1%	0.05%
1		6.314	12.706	31.821	63.657	318.309	636.619
2		2.920	4.303	6.965	9.925	22.327	31.599
3		2.353	3.182	4.541	5.841	10.215	12.924
4		2.132	2.776	3.747	4.604	7.173	8.610
5		2.015	2.571	3.365	4.032	5.893	6.869
6		1.943	2.447	3.143	3.707	5.208	5.959
7		1.894	2.365	2.998	3.499	4.785	5.408
8		1.860	2.306	2.896	3.355	4.501	5.041
9		1.833	2.262	2.821	3.250	4.297	4.781
10		1.812	2.228	2.764	3.169	4.144	4.587
11		1.796	2.201	2.718	3.106	4.025	4.437
12		1.782	2.179	2.681	3.055	3.930	4.318
13		1.771	2.160	2.650	3.012	3.852	4.221
14		1.761	2.145	2.624	2.977	3.787	4.140
15		1.753	2.131	2.602	2.947	3.733	4.073





$C.I.$ for t-distribution

- Start with some formalities… If we draw a sample of size n from a normal distribution with the mean μ, the $R.V.\ T$

$$T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(\nu = n-1)$$

- Where $\bar{X}$ is the sample mean and S its standard deviation

$$P\left(-t_c \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq t_c\right) = 1-\alpha$$

$$P\left(\bar{X} - t_c\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_c\frac{S}{\sqrt{n}}\right) = 1-\alpha$$

- And the $C.I.$ is centred about the sample mean, which contains the true unknown population parameter μ with probability $1-\alpha$

$$C.I._{100\cdot(1-\alpha)}^{t} = \left(\bar{X} - t_c\frac{S}{\sqrt{n}}, \bar{X} + t_c\frac{S}{\sqrt{n}}\right)$$

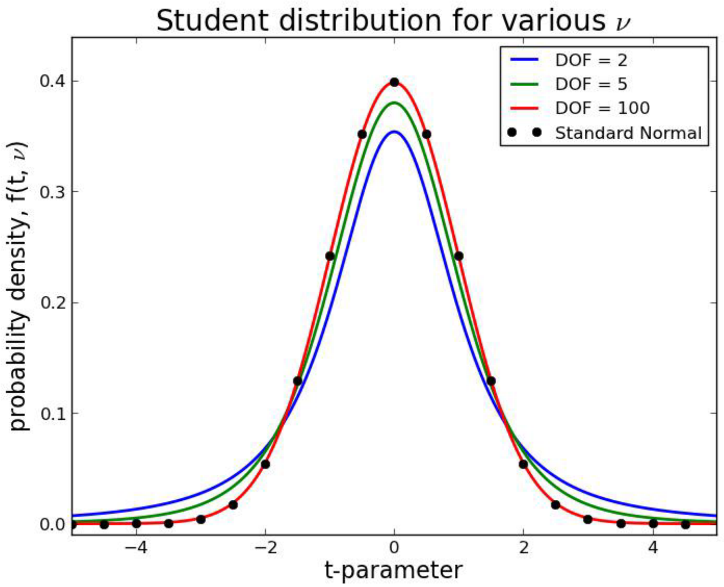
Examples – C.I.

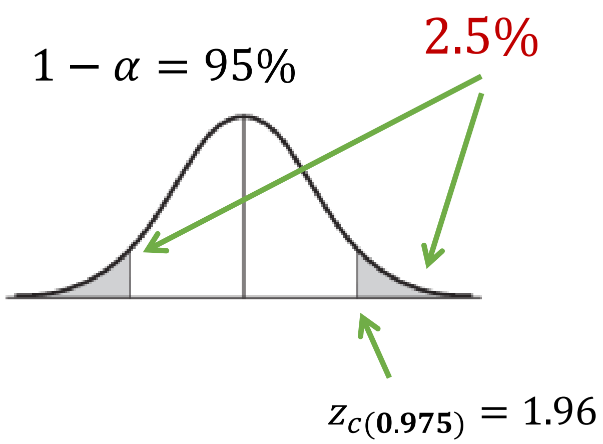
Ex. The 95% critical values (two tailed) for the normal distribution are given by $\boldsymbol{z_{c(0.975)}} = \pm\mathbf{1.96}$. What are the corresponding $t_{c(0.975)}$ for the t distribution with:

(a) $\upsilon = 9$,

(b) $\upsilon = 20$,

(c) $\upsilon = 60$





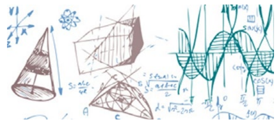
From the t distribution table with integrated probabilities we find:

a) $t_{c(0.975)}^{\upsilon=9} = \pm 2.26,$

b) $t_{c(0.975)}^{\upsilon=20} = \pm 2.09$

c) $t_{c(0.975)}^{\upsilon=60} = \pm 2.0$

Examples – C.I.



- ***Ex.*** A sample of $\mathrm{N} = 10$ measurements of the diameter of a ball bearing gave a mean $\bar{d} = 10.731\ cm$ and a standard deviation $s = 0.147\ cm.$ Find (a) 95% and (b) 99% c.i. for the actual diameter.

  ✓ In this case the confidence interval is given by: $C.I. = \bar{X} \pm t_{0.975}\frac{S}{\sqrt{N-1}}$. And since the sample size $N = 10$, the number of degrees of freedom is: $\upsilon = 10 - 1 = 9$:

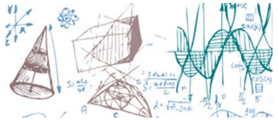
$$C.I._{0.975}^{(d)} = 10.731 \pm 2.26\frac{0.147}{\sqrt{9}} = 10.731 \pm 0.11\ cm$$

  ✓ Now, the 99% confidence interval: $t_{c(0.995)} = 3.25$ for $\upsilon = 10 - 1 = 9$:

$$C.I._{0.995}^{(d)} = 10.731 \pm 3.25\frac{0.147}{\sqrt{9}} = 10.731 \pm 0.159\ cm$$

  ✓ Note the precision of the measurement – we use a very accurate device (but that makes perfect sense – ball bearing may be a critical component of an aircraft engine).

Examples – C.I.



- ***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!)

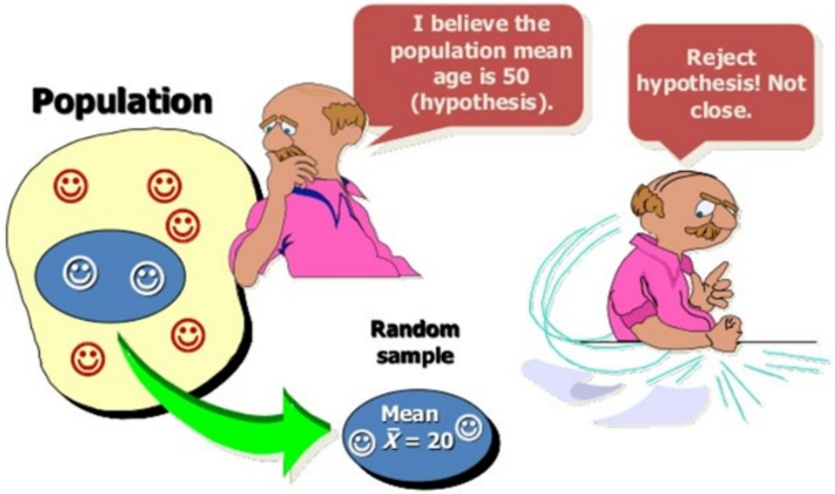
  $$P \pm z_{0.975}\sqrt{\frac{p(1-p)}{n}} = 0.55 \pm 1.96\sqrt{\frac{0.55 \cdot 0.45}{100}} = 0.55 \pm 0.10$$

  b) And the 99% c.i.:

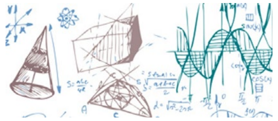
  $$P \pm z_{0.995}\sqrt{\frac{p(1-p)}{n}} = 0.55 \pm 2.58\sqrt{\frac{0.55 \cdot 0.45}{100}} = 0.55 \pm 0.13$$

Tests of Hypotheses

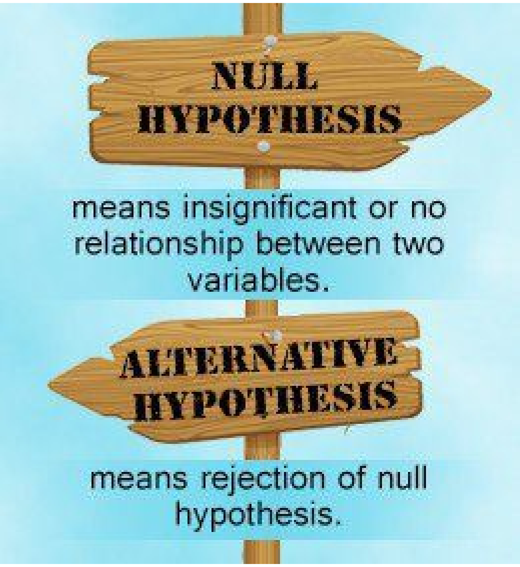
- **Statistical decisions** – the bread and butter of statistical reasoning. We need to make a decision about populations using collected samples.
- Using this approach we can check if a new medicine is really helping in curing a disease, if new educational system is better than the old etc.
- The first step of such mathematical procedure is preparing **assumptions**. They may be true or false, depending on the reasoning results, and are called **statistical hypotheses**.
- Formally, they are statements concerning the probability distributions describing respective populations. For instance, if we are investigating a coin that may be loaded, we can formulate first a hypothesis that the coin is fair: $p = 0.5$, where p is the probability of getting a head (tail). This we call the **null hypothesis**: $\boldsymbol{H_0}$.
- Now, any other hypothesis that is different from the null one is called an **alternative hypothesis**. We denote it by $\boldsymbol{H_1}$.


Population
I believe the population mean age is 50 (hypothesis).
Reject hypothesis! Not close.
Random sample
Mean $\bar{X} = 20$

Tests of Hypotheses



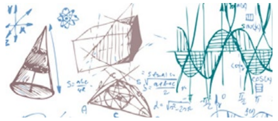
- The results of statistical reasoning **are not deterministic**! Repeating the **same experiment we may get different results** and draw different conclusions! It is possible – this is probability.
- The best we can do is the following:
  - We formulate a null hypothesis and say that **it is true**.
  - Next, we make an experiment and obtain a random sample.
  - If the results differ from those expected under H_0 on the basis of pure chance taking into account sampling theory, we would say the observed result **differ significantly** from **expectations**.
- Now, we have reason to **reject the null hypothesis** or not accept it on the basis of the **evidence**.



means insignificant or no relationship between two variables.

means rejection of null hypothesis.

Tests of Hypotheses



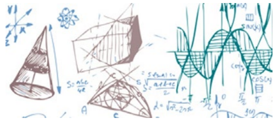
- Say, we tossed a coin 20 times and we observed 16 heads. This is a clear evidence to reject the H_0 that the coin is fair. However, we may be wrong! It is possible to get 16 heads, but the probability is low.
- Such procedure that enables us to decide whether to accept or reject hypotheses are called **tests of hypotheses or decision rules.**

Since the null and alternative hypotheses are contradictory, you must examine evidence to decide if you have enough evidence to reject the null hypothesis or not. The evidence is in the form of sample data.

H_0: **The null hypothesis:** It is a statement of no difference between the variables—they are not related. This can often be considered the status quo and as a result if you cannot accept the null it requires some action.

H_a: **The alternative hypothesis:** It is a claim about the population that is contradictory to H_0 and what we conclude when we reject H_0. This is usually what the researcher is trying to prove.

Type I and Type II errors

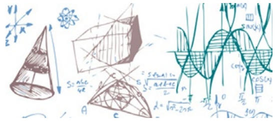


BUT we have a problem now:

- we could heve rejected a hypothesis which happens to be true. This kind of error is called **Type I**.
- we may have accepted a hypothesis which is actually not true, we say we made **Type II** error.

- In both cases we failed and error in judgement was made – so, it is bad…
- Whatever testing procedure we assume, we need to take necessary steps towards minimisation the errors of decision.
- It is not obvious and not trivial – decreasing one type of errors is accompanied by an increase in the other type.
- Very often in practice one type of error may be **much more serious** than the other (e.g. judicial system). The only way to limit both types of errors is to increase the sample size – which may not be easy or even possible

Type I and Type II errors

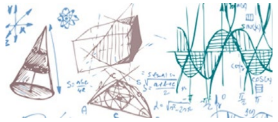


- Imagine that we rejected a hypothesis and it happens to be true. This kind of error is called **Type I**.
- If we accepted a hypothesis which is not true, we say we made **Type II** error.
- In both cases we failed and error in judgement was made – so, it is bad…(remember covid tests or cancer markers?)

		PREDICTED	
		sick	not sick
TRUE	sick	10	2
	not sick	1	8

		PREDICTED	
		sick	not sick
TRUE	sick	True positive	False Negative
	not sick	False Positive	True Negative

Type I and Type II errors



- Imagine that we rejected a hypothesis and it happens to be true. This kind of error is called **Type I**.
- If we accepted a hypothesis which is not true, we say we made **Type II** error.
- In both cases we failed and error in judgement was made – so, it is bad…(remember covid tests or cancer markers?)

Confusion Matrix – common tool for decision quality assesment

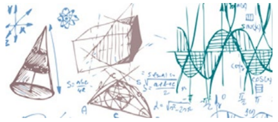
PREDICTED

		sick	not sick
TRUE	sick	☺	False Negative
	not sick	False Positive	☺

Type I (miss) – failing to assert true

Type II (false hit) – asserting somethig that is not there…

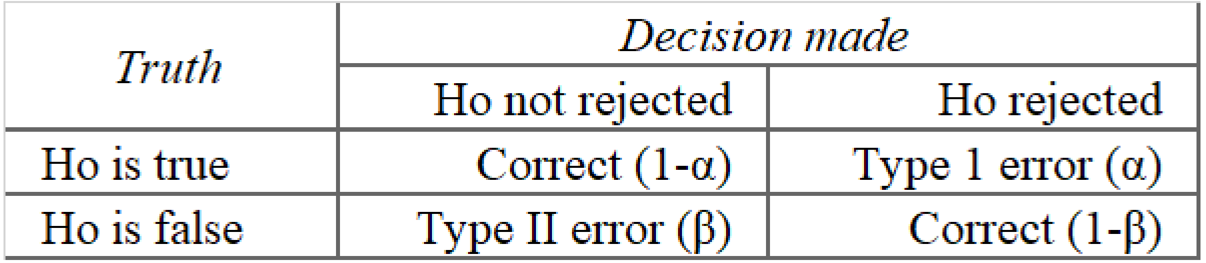
Level of significance



- When testing a hypothesis, the maximum probability with which we would be willing **to risk any Type I errors** is called the **level of significance** of the test.
- Good practice states that this value should be chosen before any sample is collected – we do not wish to bias our results in any way!
- Usually, we choose a **level of significance** of 0.05 or 0.01, other values can also be used.
- Say, we chose the S.L. to be 0.05, i.e., there is 5% chance to reject the hypothesis that should be accepted.
- Conversely, whenever the null hypothesis is true, we are about 95% confident that we would make the correct decision.
- We also say, that the null hypothesis has been rejected at a 0.05 level of significance, which means that we could make a wrong decision with probability of 0.05

Type I and Type II errors

- Imagine that we rejected a hypothesis and it happens to be true. This kind of error is called **Type I and the probability is described by α**.
- If we accepted a hypothesis which is not true, we say we made **Type II** error **and the probability is described by β.**
- In both cases we failed and error in judgement was made, so we'd like to have these values low.



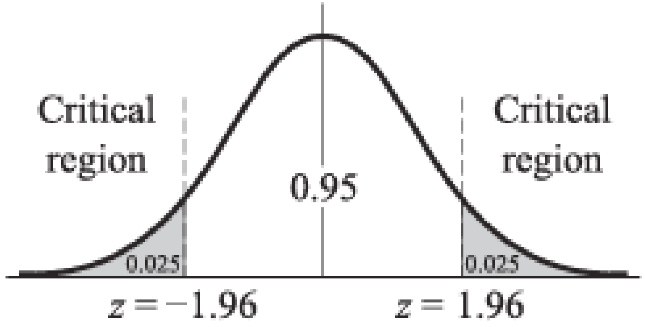
Truth	*Decision made*	
	Ho not rejected	Ho rejected
Ho is true	Correct (1-α)	Type 1 error (α)
Ho is false	Type II error (β)	Correct (1-β)

Tests with normal distribution

- Let’s assume that we are considering a statistics S that is approximately normal with the mean and standard deviation μ_S and σ_S. First thing first – standardisation:

$$\boxed{Z_S = \frac{S - \mu_S}{\sigma_S}}$$

- For the time being, we assume that we reject the hypothesis if our statistics is either too small or too large:



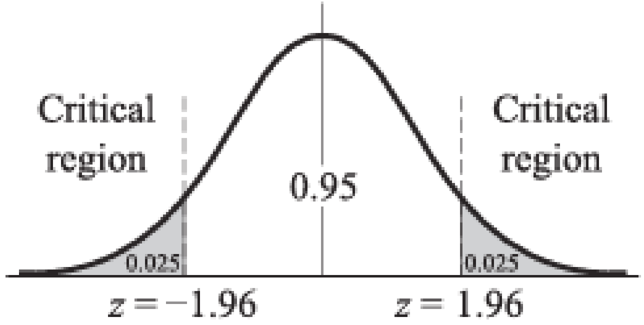
Confidence Level	99.73%	99%	98%	96%	95.45%	95%	90%	80%	68.27%	50%
z_c	3.00	2.58	2.33	2.05	2.00	1.96	1.645	1.28	1.00	0.6745

Tests with normal distribution

- **So, how to reject the test**? If for a sample, its z-score corresponding to the value of S is outside the 95% interval, we could conclude that this is **unlikely event** that would occur with probability of 5%, if our null hypothesis were true
- In other words, we would say, that the z-score for this sample differed significantly from our expectations based on the H_0 and it should be rejected!

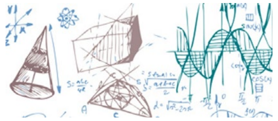
✓ The grey area represents the **level of significance** of the test, i.e., the probability of **Type I error**

Or, we say, that the hypothesis was rejected at 5% level of significance

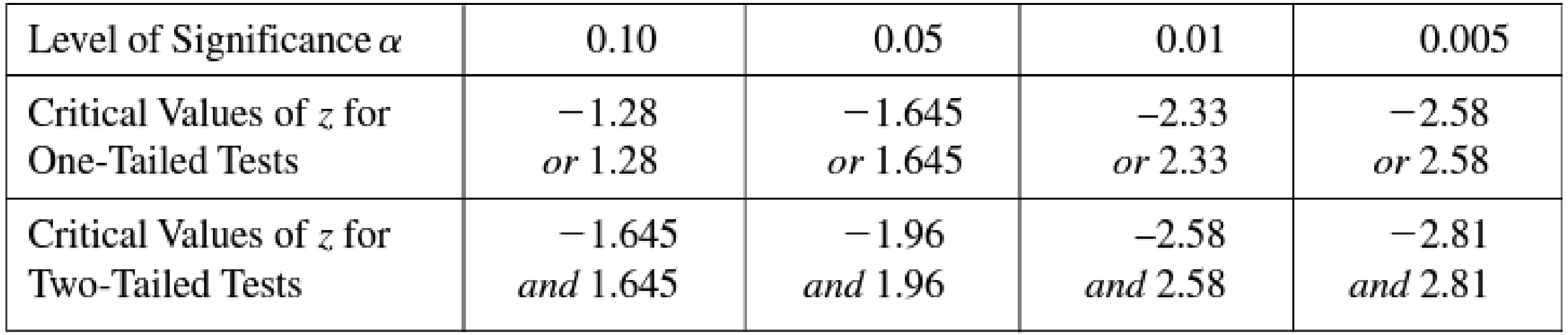


- All z-scores outside -1.96 and 1.96 are called **critical region** of the rejection of the null hypothesis (the region of significance)
- Not too hard to notice that all z-scores inside this interval will be called the **region of acceptance** of the hypothesis or the region of non-significance

The decision rule

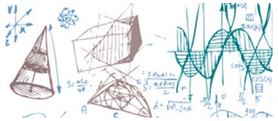


- **Reject the null hypothesis** at α level of significance if the z-score of a statistic **S falls outside** the range $\pm z_S$ (e.g., for $\alpha = 0.05$, $\pm z_S = \pm 1.96$). We also say, that the observed sample statistics is significant at level α
- We accept (or do nothing!) otherwise
  - Note! You could choose an arbitrary value for the level of significance
  - Note! We can also use one-tailed tests (the one discussed above is called two-tailed test). The difference: check if a technological process A **is better or worse** than B, check if a technological process A **is better** than B



Level of Significance α	0.10	0.05	0.01	0.005
Critical Values of z for One-Tailed Tests	−1.28 *or* 1.28	−1.645 *or* 1.645	–2.33 *or* 2.33	−2.58 *or* 2.58
Critical Values of z for Two-Tailed Tests	−1.645 *and* 1.645	−1.96 *and* 1.96	–2.58 *and* 2.58	−2.81 *and* 2.81

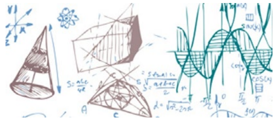
Test statistics



null hyphotesis H_0	$\mu = \mu_0$ (σ known)	$\mu = \mu_0$	$\sigma = \sigma_0^2$
Test statistic T	$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$	$\frac{\bar{X} - \mu}{S/\sqrt{n}}$	$\frac{(n-1)S^2}{\sigma_0}$
Distribution of T under H_0	$N(0,1)$	$t(\nu = n - 1)$	$\chi^2(\nu = n - 1)$

Pop. prop. p $\quad p = p_0 \quad z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} \quad n\hat{p} \geq 10,\ n(1 - \hat{p}) \geq 10$

Tests for means



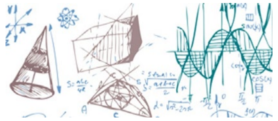
- ***Ex. 1*** Imagine somebody working for a Casino that is tasked with checking that new batch of tossing coins is fair. Say, we just draw one coin.
- We could start with deciding on the decision rule, say, we want to perform **100 tosses and obtain a number of heads for which the prob. is „high”. Let’s pick this to be over 95%.** We hypothesising an initial range to be from 40 to 60 heads. Let’s check the figures.

$$P(40 \leq X_{\mathcal{B}} \leq 60) = C_{100}^{40}\left(\frac{1}{2}\right)^{40}\left(\frac{1}{2}\right)^{60} + \cdots + C_{100}^{60}\left(\frac{1}{2}\right)^{60}\left(\frac{1}{2}\right)^{40}$$

- This calculations are cumbersome…, however we could use the normal distribution approximation:

$$\mu = np = 100 \cdot \frac{1}{2} = 50, \sigma = \sqrt{npq} = 5, \mu \gg \sigma$$

Tests for means – example cont.



- For categorical numbers we get: $40 \equiv 39.5, 60 \equiv 60.5$, we do that so the results are nice numbers (mind the standardisation transformation):

$$z_{39.5} = \frac{39.5 - 50}{5} = -2.1, \qquad z_{60.5} = \frac{60.5 - 50}{5} = 2.1$$

- So, we have our critical points: $z_{39.5} = -2.1, z_{60.5} = 2.1$

$$P(40 \leq X_{\mathcal{B}} \leq 60) = P(-2.1 \leq X_{\mathcal{N}} \leq 2.1) = 0.9642$$

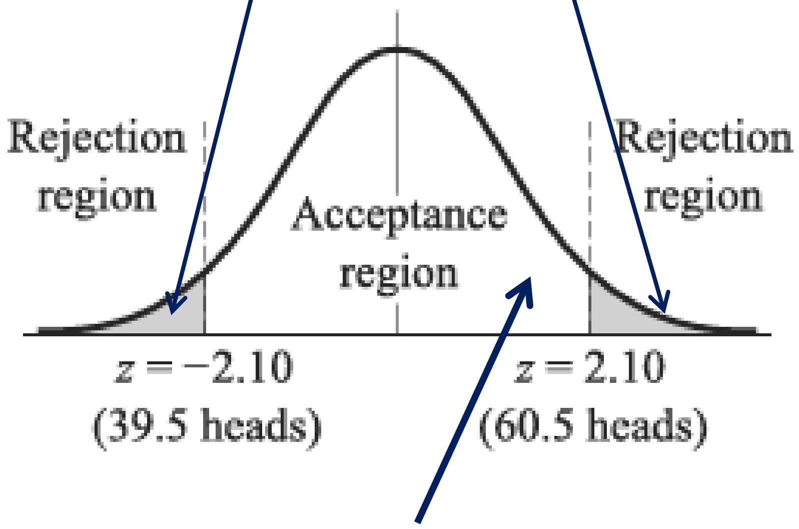
- This looks reasonable, the probability is high enough. We can formulate the decision rule as:
  - ✓ **accept the hypothesis** (H_0) if the number of heads (or tails) observed in a data sample of 100 tosses is between 40 and 60.
  - ✓ If we observe a different number of heads – **reject the hypothesis** that the coin is fair

Tests for means – example cont.

- Note, that the Type I error (rejecting the H_0 when it is correct) can occur with the probability of $(1-0.9642)=\mathbf{0.0358}$
- The decision rule can be showed using a plot:

Type I error → Level of significance

$\mathbf{0.0358/2}$

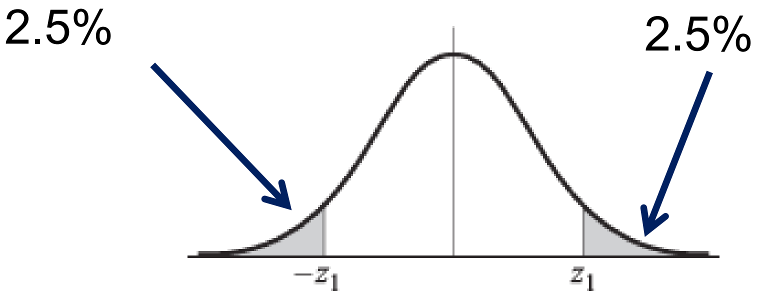


Accept H_0 if the result is here

Tests for means – example cont.

- **If the observed result (given a data sample) would land within the shaded area, we say: the z-score differs significantly from the expected value (taking into account pure statistical chances)**. This is equivalent to accept the alternative hypothesis (however, there is a catch… Very often we chose not to do that, and just reject the H_0)
- Note, that we chosen very unusual significance level $\boldsymbol{\alpha = 3.58}$ %**.** As mentioned before – it does not matter! This value is up to you!
- Also, consider this: what if the true value of $p = 0.7$? It would be perfectly possible to observe 60 heads – so, we would accept the wrong hypothesis… (**Type II error**)

> ➢ Just for training, design a decision rule when the data sample is 64 tosses? Assume that the level of significance $\alpha = 0.05$

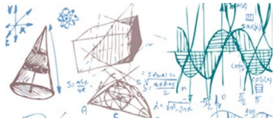


$$P(0 \geq X \geq z_+) = 0.5 \ - 0.025$$

$$z_+ = 1.96, z_- = -1.96$$

$$\mu = np = 64 \cdot \frac{1}{2} = 32, \sigma = \sqrt{npq} = 4$$

Tests for means



- We need to translate the z scores into heads count

$$Z = \frac{X - \mu}{\sigma} \rightarrow \pm 1.96 = \frac{X - 32}{4}$$

$$X_{1.96} = 39.84, \equiv \mathbf{39}, \qquad X_{-1.96} \quad = 24.16 \equiv \mathbf{24}$$

we have to accept the hypothesis (H_0) if the number of heads (or tails) observed in a data sample of 64 tosses is between 24 and 39 @ 5% SL.

- Note, that this calculations go exactly the same way as if we want to estimate the $C.I.$ representing 95% probability!

- Ok, how we should deal with Type II error? Remember?

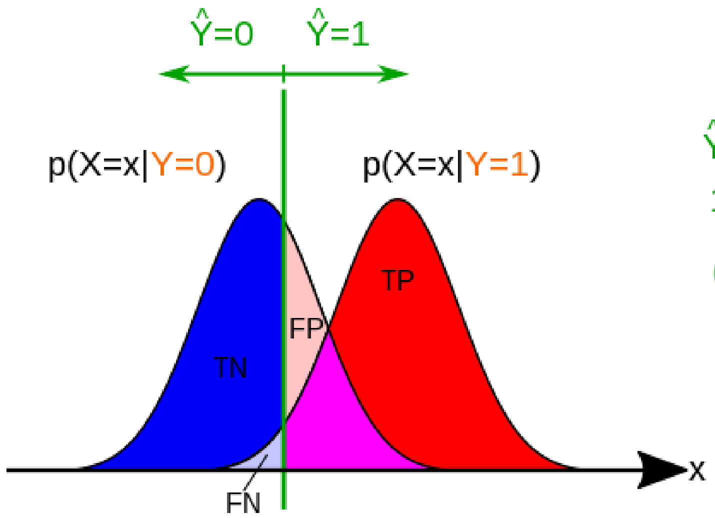
  - If we accepted a hypothesis which is not true, we say we made **Type II** error **and the probability is described by $\boldsymbol{\beta}$.**

ML spoiler

Let’s see the problem with Type I and Type II error – we cannot have them both low: efford to reduce one type increase the other.

alternative hyphotesis H_A

null hyphotesis H_0



$\hat{Y}$ \ Y	1	0
1	TP	FP
0	FN	TN

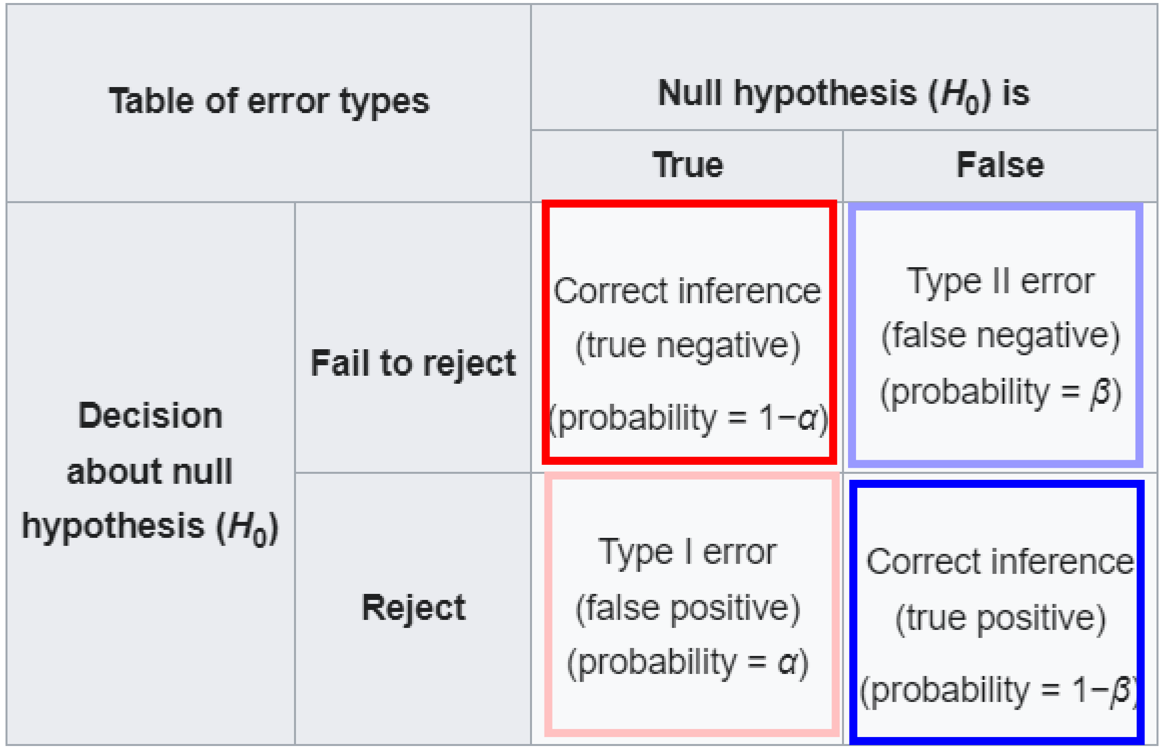


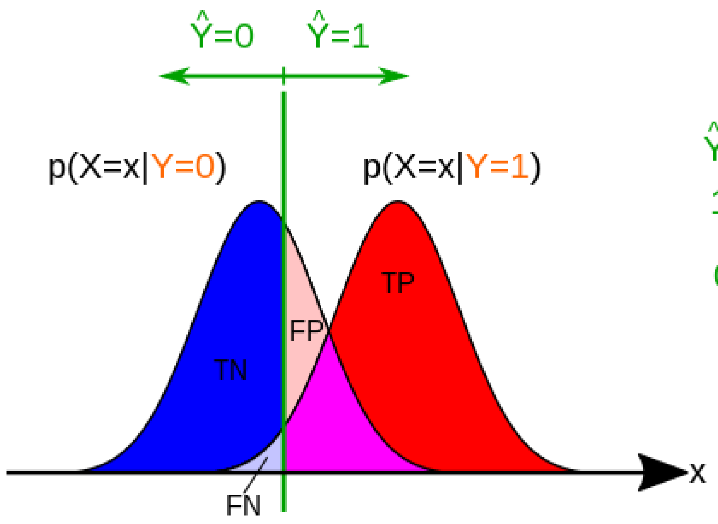
Table of error types		Null hypothesis (H_0) is	
		True	False
Decision about null hypothesis (H_0)	Fail to reject	Correct inference (true negative) (probability = 1−α)	Type II error (false negative) (probability = β)
	Reject	Type I error (false positive) (probability = α)	Correct inference (true positive) (probability = 1−β)

ML spoiler

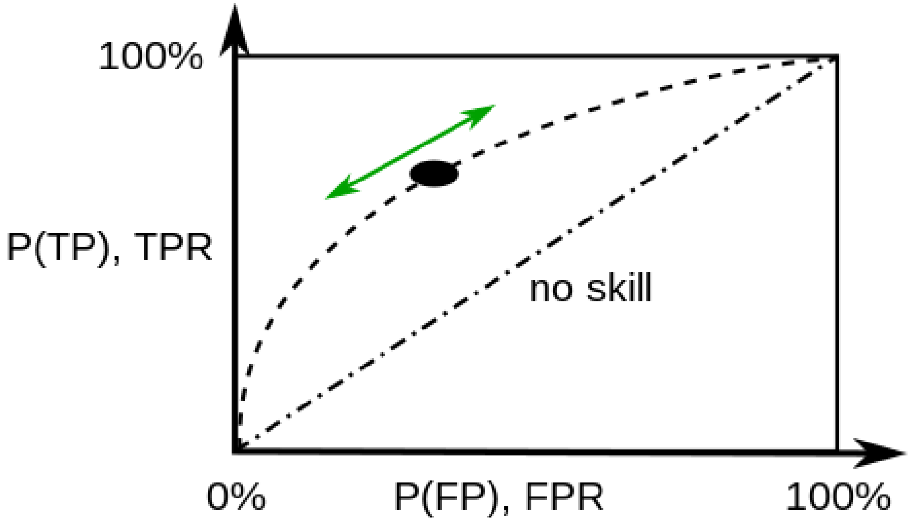
Let’s shift the critical value – has out „classification” improved?

alternative hyphotesis H_A

null hyphotesis H_0



$\hat{Y}$ \ Y	1	0
1	TP	FP
0	FN	TN

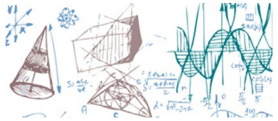


The power of the test depends on:

a) alpha ?

b) beta ?

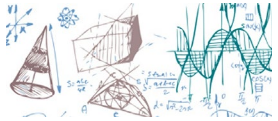
Type II error – comming back



$$X_{1.96} = 39.84, \equiv \mathbf{39}, \qquad X_{-1.96} = 24.16 \equiv \mathbf{24}$$

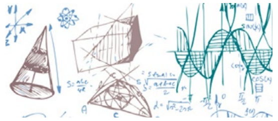
- If we accepted a hypothesis which is not true, we say we made **Type II** error **and the probability is described by $\boldsymbol{\beta}$.**
- So, how we should deal with Type II error? Well, one way or another we need to see the result and decide if we want to re-word the decision rule (this may be tricky). For instance if we see the number of heads to be **23**, it is so close… We could say, we do not reject the zero hypothesis if we find the result between 22 and 42.
- This is not nice and nasty way of doing things, though. **We should never ever change the decision rule after seeing data** – blinded analysis
- We have additional tools to enforce our decision – **p-value technique**

p-value



- **This technique is very powerful** and is of great interest in practical data analyses
- *Say, we have a problem where two hypotheses were formulated H_0 and H_1. Say, the zero hypo makes attempt at assertion that a given parameter has some specified value, the alternative hypos can de defined as follow*:
- Its value is indeed greater than stated (**right-tailed test**)
- The parameter is less than stated (**left-tailed test**)
- It is different (greater than or less than) than stated (**two-tailed test**)
- We can use the following definition: the p-value (probability value) is the probability of obtaining the value of the test statistic at least as extreme as the one calculated using data sample, assuming that the null hypothesis is correct
- In other words, p-value estimates how well the observed data support the null hypothesis (if it is true). It measures how compatible are your data with the H_0: **large values makes your null look good, small suggest rejection**

p-value



- Let’s see some action! Say, we have a R.V. following the normal distribution with the standard deviation $\sigma = 3$. Now, H_0 states that the mean value $\mu = 12$. Next, we drawn a sample, n=36, and got $\bar{x} = 12.95$, we choose „standard” test statistics to be:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{12.95 - 12}{0.5} = 1.9$$

- The p-value will depend on the alternative hypothesis!!

  ✓ H_1: $\mu > 12$ in this cast the p-value is defined as **the probability that for a random sample of size $n = 36$ we would observe a sample mean of 12.95 or higher if the true mean were 12**: $P(Z \geq 1.9) = 0.029$. This represents 3 chances in 100 that $\bar{x} = 12.95$ if $\mu = 12$

p-value

- Let’s see some action! Say, we have a R.V. following the normal distribution with the standard deviation $\sigma = 3$. Now, H_0 states that the mean value $\mu = 12$. Next, we drawn a sample, n=36, and got $\bar{x} = 12.95$, we choose „standard” test statistics to be:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{12.95 - 12}{0.5} = 1.9$$

- The p-value will depend on the alternative hypothesis!!

✓ $H_1\colon \mu > 12$

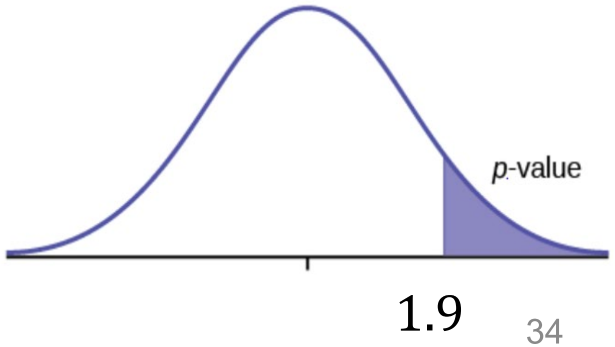
in this cast the p-value is defined as:

the probability that for a random sample of size $n = 36$ we would observe a sample mean of 12.95 or more,

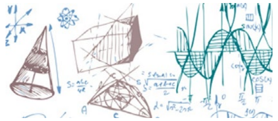
if the true mean were 12:

$P(Z > 1.9) = 0.03.$

This represents 3 chances in 100 that $\bar{X} = 12.95$ if $\mu = 12$



p-value



- Let’s see some action! Say, we have a R.V. following the normal distribution with the standard deviation $\sigma = 3$. Now, H_0 states that the mean value $\mu = 12$. Next, we drawn a sample, n=36, and got $\bar{x} = 12.95$, we choose „standard” test statistics to be:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{12.95 - 12}{0.5} = 1.9$$

- The p-value will depend on the alternative hypothesis!!

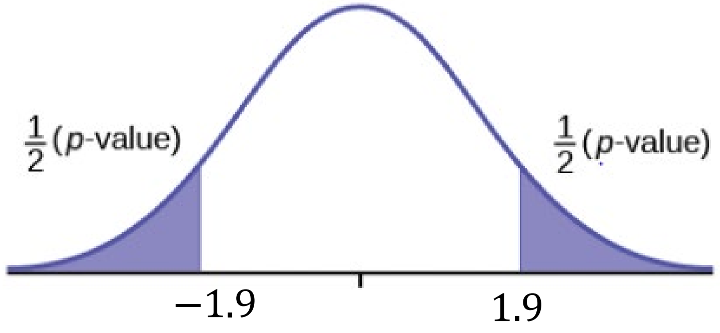
✓ $H_1\colon\ \mu \neq 12$

in this cast the p-value is defined as:

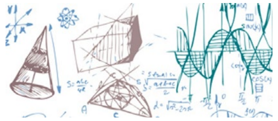
- **the probability that for a random sample of size $n = 36$ we would observe a sample mean of 0.95 or more units (standard deviation) away from 12:**

$$P(Z \geq 1.9) + P(Z \leq -1.9) = 0.057.$$

- This represents 6 chances in 100 that $|\bar{x} - 12| \geq 0.95$ if $\mu = 12$



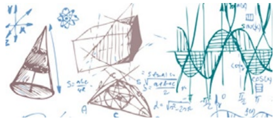
p-value



Discussion:

- First of all, the p-value does not provide a tool for rejecting or keeping the null hypothesis on its own! We always need an alternative hypo against which we are going to make a judgement. The same value of test statistics based on the same data sample may lead to completely different conclusions
- In the first case we have **small p-value**, which suggests we **should reject the null** in favour of the given alternative
- The second one **sports large p-value** and suggests strongly **not to reject the null** in favour of the alternative (less than 12)
- Final example again **shows low p-value** and suggests that the alternative hypothesis should be considered (**note that this one is not so strong as the first one**)

p-value



- **Ex. 3** Imagine, a psychic wants to confirm his abilities for extrasensory perception (ESP). He was asked to guess the colour (yellow and blue) of a card chosen from a deck of 50 cards by somebody else in other room. The test subject does not know how many yellow or blue cards there are in the deck. Say, he identified correctly 32 cards. Is he really the psychic? Assume level of significance to be $\alpha = 0.05$
- Let p be the probability of the individual stating the colour of a card correctly, we formulate the following hypothesis
  - ✓ $H_0 : p = 0.5$ – he is just randomly guessing
  - ✓ $H_1 : p > 0.5$ – he indeed has ESP abilities
- We choose the one-tail test – we are going for high scores!

p-value

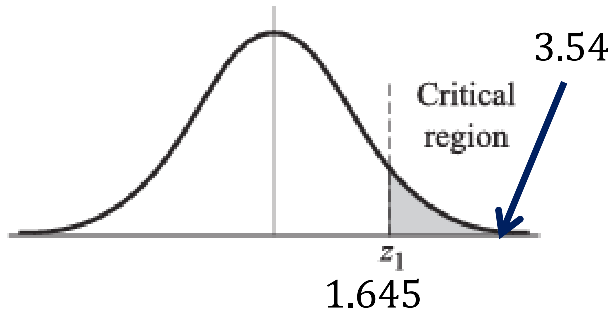
- ✓ $H_0: p = 0.5$ – he is just randomly guessing
- ✓ $H_1: p > 0.5$ – he indeed has ESP abilities

- We choose the one-tail test – we are going for high scores!
- For the null hypo we have:

$$\mu = np = 50 \cdot 0.5 = 25, \sigma = \sqrt{npq} = 3.54$$

- For one-tailed test we have:

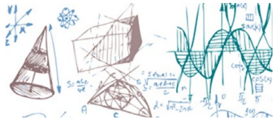
$$P(0 \geq Z \geq z_1) = 0.45, z_1 = 1.645$$



so it is rather unlikely that he identified the cards by guessing only….

but how much unlikely?

p-value



- Approximating the binomial distribution by the normal one, we find that the result of 32 has the following z-score

$$\frac{X-\mu}{\sigma}=\frac{32-25}{3.54}=1.98$$

- We should conclude, that there is something extraordinary in this result
- Please, repeat the calculations for the significance level $\alpha = 0.01$
- Now, let's calculate the p-value: what is the probability that the 32 cards will be identified correctly using just random guessing?

$$P(Z \geq 1.98) \approx 2\%$$

- So, the chances of concluding that the test subject does not have EPS abilities would be like **2 in 100** – this supports rejection of the null hypo.