

# **Introduction to statistics**

## **(Day 3)**

# Recap



# Recap

- When can't study a population, we select a representative sample
- Categorical variables are described with absolute and relative frequencies, numerical variables are described with measures of central tendency (mode, median, mean) and dispersion (range, IQR, standard deviation)
- Parameters (calculated on the population) vs statistics (calculated on the sample)

# Recap

- Multiple phenomena and statistical distributions are normally distributed, and the Normal distribution describes both the probability of an observation and its proportion in the population
- We use statistics to estimate parameters (point estimates), with interval estimates (confidence intervals) estimating their uncertainty
- 95% confidence intervals tell us the the true value has 95% probability of being inside the given range

# Making decision with data



# Learning objectives

- Make and test hypotheses
- Interpret P values
- Understand Type I and II errors
- Understand the power of a study



Copyrighted material

Spiegelhalter, D., *The Art of Statistics: Learning From Data*, Pelican, 2019

# Making hypotheses

“ A hypothesis can be defined as a proposed explanation for a phenomenon. It is not the absolute truth, but a provisional, working assumption, perhaps best thought of as a potential suspect in a criminal case.

”

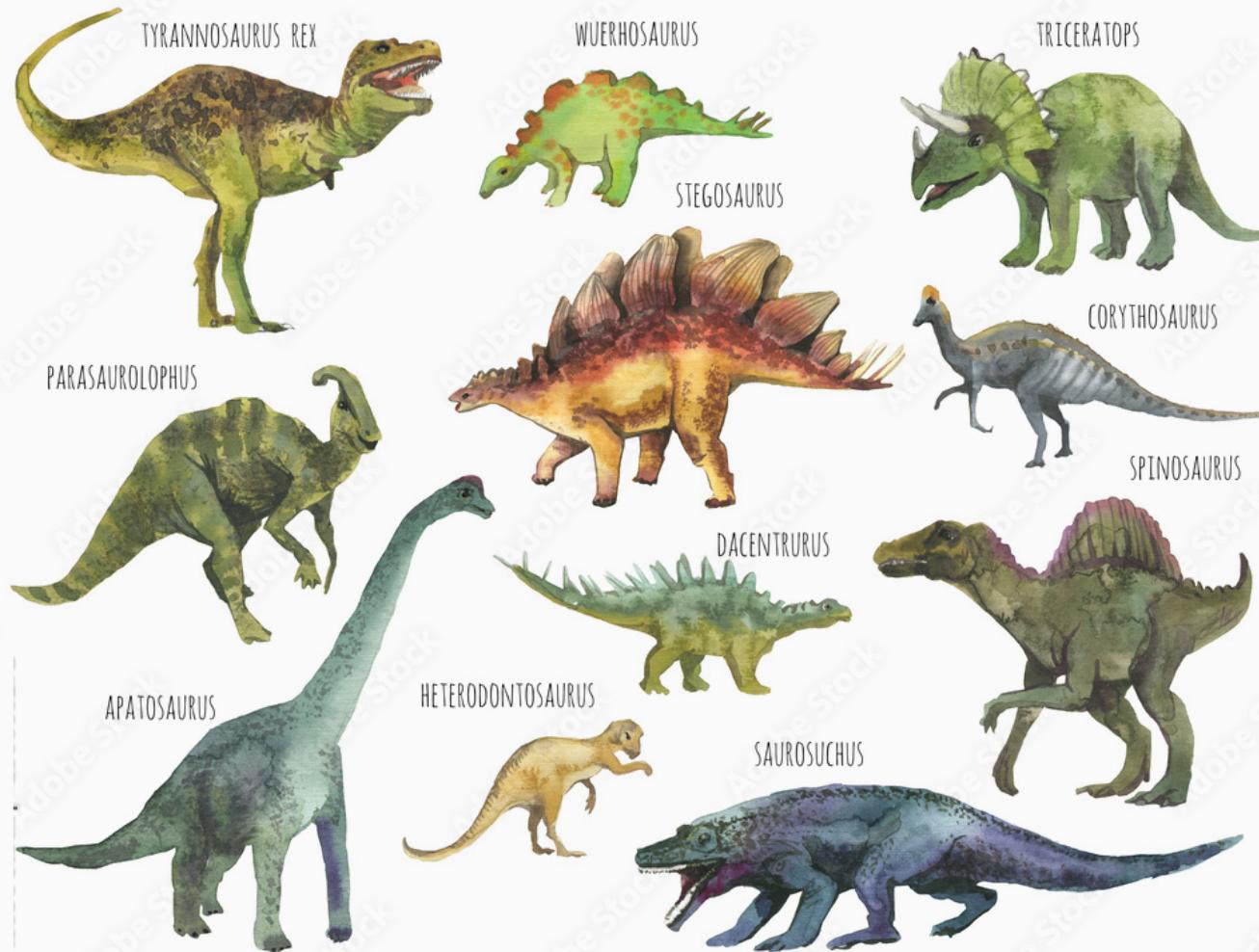
# Making hypotheses

- The outcomes in the intervention and the control group are different
- The proportion of an event in the intervention and control group is different

# The falsification principle and the null hypothesis

- The outcomes in the intervention and the control group are ~~different~~ **the same**
- The proportion of an event in the intervention and control group is ~~different~~ **the same**

# The Falsification Principle



# The Falsification Principle

DINOSAUR EVOLUTION

## A Jurassic ornithischian dinosaur from Siberia with both feathers and scales

Pascal Godefroit,<sup>1\*</sup> Sofia M. Sinitsa,<sup>2</sup> Danielle Dhouailly,<sup>3</sup> Yuri L. Bolotsky,<sup>4</sup>  
Alexander V. Sizov,<sup>5</sup> Maria E. McNamara,<sup>6,7</sup> Michael J. Benton,<sup>7</sup> Paul Spagna<sup>1</sup>

# Exercise #1

**Objective** To determine whether intravenous dexamethasone increases the number of ventilator-free days among patients with COVID-19-associated ARDS.

**Design, Setting, and Participants** Multicenter, randomized, open-label, clinical trial conducted in 41 intensive care units (ICUs) in Brazil. Patients with COVID-19 and moderate to severe ARDS, according to the Berlin definition, were enrolled from April 17 to June 23, 2020. Final follow-up was completed on July 21, 2020. The trial was stopped early following publication of a related study before reaching the planned sample size of 350 patients.

- ? Which is the null hypothesis of this study?
- a) Dexamethasone plus standard care is **more effective** than standard care alone
  - b) Dexamethasone plus standard care is **less effective** than standard care alone
  - c) Dexamethasone plus standard care is **as effective** as standard care alone
  - d) Dexamethasone plus standard care is not **as effective** as standard care alone

# Exercise #1 -- Solution

**Objective** To determine whether intravenous dexamethasone increases the number of ventilator-free days among patients with COVID-19-associated ARDS.

**Design, Setting, and Participants** Multicenter, randomized, open-label, clinical trial conducted in 41 intensive care units (ICUs) in Brazil. Patients with COVID-19 and moderate to severe ARDS, according to the Berlin definition, were enrolled from April 17 to June 23, 2020. Final follow-up was completed on July 21, 2020. The trial was stopped early following publication of a related study before reaching the planned sample size of 350 patients.

- ? Which is the null hypothesis of this study?
  - a) Dexamethasone plus standard care is **more effective** than standard care alone
  - b) Dexamethasone plus standard care is **less effective** than standard care alone
  - c) Dexamethasone plus standard care is **as effective** as standard care alone
  - d) Dexamethasone plus standard care is not **as effective** as standard care alone

## Exercise #2

- ? If one **doesn't** reject the null hypothesis it means that...
- a) the null hypothesis is true
  - b) the null hypothesis is false
  - c) the observations are compatible with the null hypothesis
  - d) the observations aren't compatible with the null hypothesis
  - e) it depends on the research question

01:00

## Exercise #2 -- Solution

- ? If one **doesn't** reject the null hypothesis it means that...
- a) the null hypothesis is true
  - b) the null hypothesis is false
  - c) the observations are compatible with the null hypothesis
  - d) the observations aren't compatible with the null hypothesis
  - e) it depends on the research question

# Exercise #3

**Objective** To determine whether intravenous dexamethasone increases the number of ventilator-free days among patients with COVID-19-associated ARDS.

**Design, Setting, and Participants** Multicenter, randomized, open-label, clinical trial conducted in 41 intensive care units (ICUs) in Brazil. Patients with COVID-19 and moderate to severe ARDS, according to the Berlin definition, were enrolled from April 17 to June 23, 2020. Final follow-up was completed on July 21, 2020. The trial was stopped early following publication of a related study before reaching the planned sample size of 350 patients.

? How do you define the null hypothesis in this study?

- a)  $\mu_i - \mu_c = 0$
- b)  $\mu_i - \mu_c \neq 0$
- c)  $\bar{x}_i - \bar{x}_c = 0$
- d)  $\bar{x}_i - \bar{x}_c \neq 0$

# Exercise #3 -- Solution

**Objective** To determine whether intravenous dexamethasone increases the number of ventilator-free days among patients with COVID-19-associated ARDS.

**Design, Setting, and Participants** Multicenter, randomized, open-label, clinical trial conducted in 41 intensive care units (ICUs) in Brazil. Patients with COVID-19 and moderate to severe ARDS, according to the Berlin definition, were enrolled from April 17 to June 23, 2020. Final follow-up was completed on July 21, 2020. The trial was stopped early following publication of a related study before reaching the planned sample size of 350 patients.

? How do you define the null hypothesis in this study?

- a)  $\mu_i - \mu_c = 0$  
- b)  $\mu_i - \mu_c \neq 0$
- c)  $\bar{x}_i - \bar{x}_c = 0$
- d)  $\bar{x}_i - \bar{x}_c \neq 0$

# Making hypotheses

- 📌 Dexamethasone plus standard care is **as effective as** standard care

$$n_i = 151, \bar{x}_i = 6.6, s_i = 10.0$$

$$n_c = 148, \bar{x}_c = 4.0, s_c = 8.7$$

$$\mu_i - \mu_c = 0$$

→ Null hypothesis ( $\mathcal{H}_0$ )

# Making hypotheses

- 📌 Dexamethasone plus standard care is **as effective as** standard care

$$n_i = 151, \bar{x}_i = 6.6, s_i = 10.0$$

$$n_c = 148, \bar{x}_c = 4.0, s_c = 8.7$$

$$\mu_i - \mu_c = 0$$

→ Null hypothesis ( $\mathcal{H}_0$ )

$$\mu_i - \mu_c \neq 0$$

→ Alternative hypothesis ( $\mathcal{H}_1/\mathcal{H}_A$ )

# Testing hypotheses

- 📌 Dexamethasone plus standard care is **as effective as** standard care

**Interventions** Twenty mg of dexamethasone intravenously daily for 5 days, 10 mg of dexamethasone daily for 5 days or until ICU discharge, plus standard care ( $n=151$ ) or standard care alone ( $n=148$ ).

**Results** A total of 299 patients (mean [SD] age, 61 [14] years; 37% women) were enrolled and all completed follow-up. Patients randomized to the dexamethasone group had a mean 6.6 ventilator-free days (95% CI, 5.0-8.2) during the first 28 days vs 4.0 ventilator-free days (95% CI, 2.9-5.4) in the standard care group

$$n_i = 151, \quad \bar{x}_i = 6.6, \quad s_i = 10.0$$

$$n_c = 148, \quad \bar{x}_c = 4.0, \quad s_c = 8.7$$

# Testing hypotheses

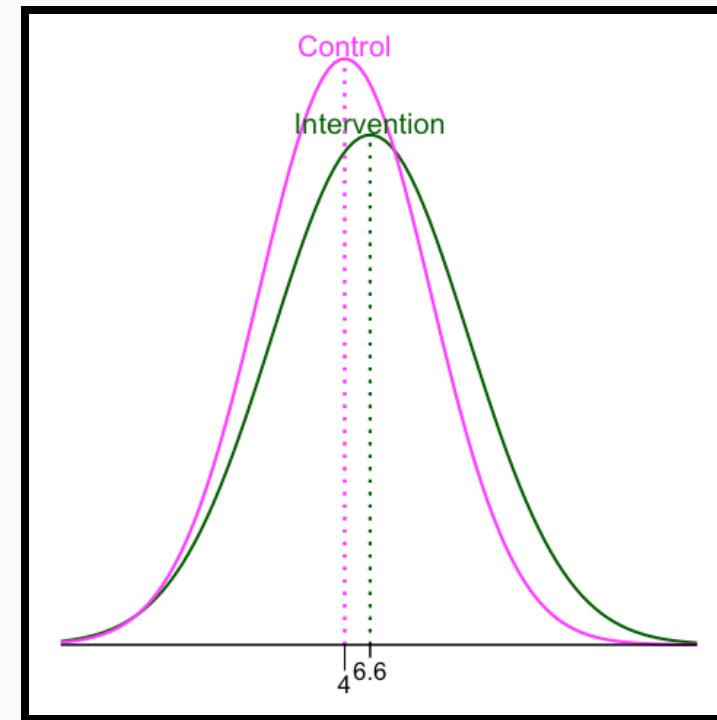
- 📌 Dexamethasone plus standard care is **as effective as** standard care

$$n_i = 151, \bar{x}_i = 6.6, s_i = 10.0$$

$$n_c = 148, \bar{x}_c = 4.0, s_c = 8.7$$

$$\mu_i - \mu_c = 0 \quad \leftarrow$$

$$\bar{x}_i - \bar{x}_c = 6.6 - 4.0 = 2.6$$



# Testing hypotheses

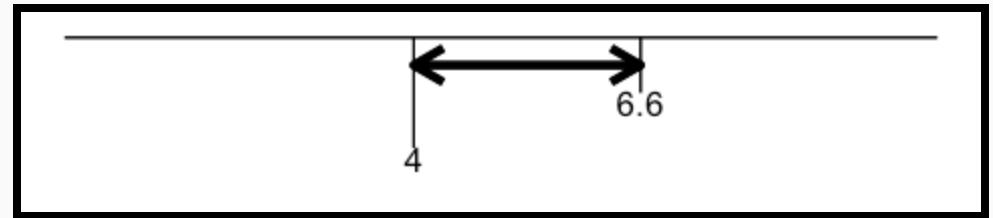
- 📌 Dexamethasone plus standard care is **as effective as** standard care

$$n_i = 151, \bar{x}_i = 6.6, s_i = 10.0$$

$$n_c = 148, \bar{x}_c = 4.0, s_c = 8.7$$

$$\mu_i - \mu_c = 0 \quad \leftarrow$$

$$\bar{x}_i - \bar{x}_c = 6.6 - 4.0 = 2.6$$



- ? What is the probability of observing a difference of 2.6 days if  $\mu_d - \mu_m = 0$ ?

# Let's take a step back

1. The Normal distribution is defined by its mean and standard deviation and corresponds to a probability distribution  
→ Area  $Z \equiv$  probability  $\mathcal{P}$
2. Sampling distributions (including the difference of means) show a Normal distribution (CLT)

# Let's take a step back

1. The Normal distribution is defined by its mean and standard deviation and corresponds to a probability distribution  
→ Area  $Z \equiv$  probability  $\mathcal{P}$
2. Sampling distributions (including the difference of means) show a Normal distribution (CLT)

For the difference of means:

$$\mathcal{N} = \left( \mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right) \text{ with } \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \rightarrow \text{standard error}$$

# Testing hypotheses

- 📌 Dexamethasone plus standard care is **as effective as** standard care alone

$$n_i = 151, \bar{x}_i = 6.6, s_i = 10.0$$

$$n_c = 148, \bar{x}_c = 4.0, s_c = 8.7$$

$$\mu_c - \mu_i = 0 \quad \leftarrow$$

$$\bar{x}_c - \bar{x}_i = 6.6 - 4.0 = 2.6$$

$$\mathcal{N} = \left( \mu_c - \mu_i, \frac{\sigma_c^2}{n_c} + \frac{\sigma_i^2}{n_i} \right) \rightarrow \mu_c - \mu_i = 0 \text{ and } \hat{SE} = \sqrt{\frac{s_c^2}{n_c} + \frac{s_i^2}{n_i}} = 1.08$$

# Testing hypotheses

- 📌 Dexamethasone plus standard care is **as effective as** standard care

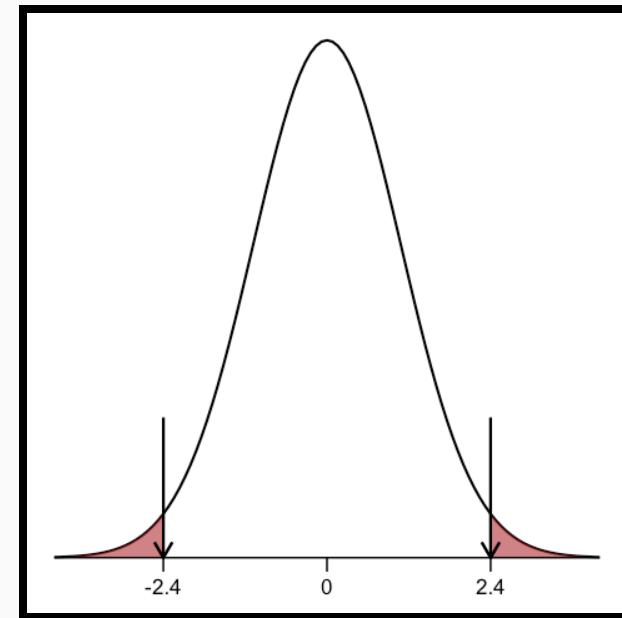
$$\mu_c - \mu_i = 0$$

$$\hat{SE} = 1.08$$

$$\bar{x}_c - \bar{x}_i = 6.6 - 4.0 = 2.6$$

- ❓ What is the probability of observing a difference of 2.6 days if  $\mu_c - \mu_i = 0$ ?

$$z = \frac{(\bar{x}_c - \bar{x}_i) - (\mu_c - \mu_i)}{\hat{SE}} = \frac{2.6 - 0}{1.08} = 2.4$$



# Testing hypotheses

- 📌 Dexamethasone plus standard care is **as effective as** standard care

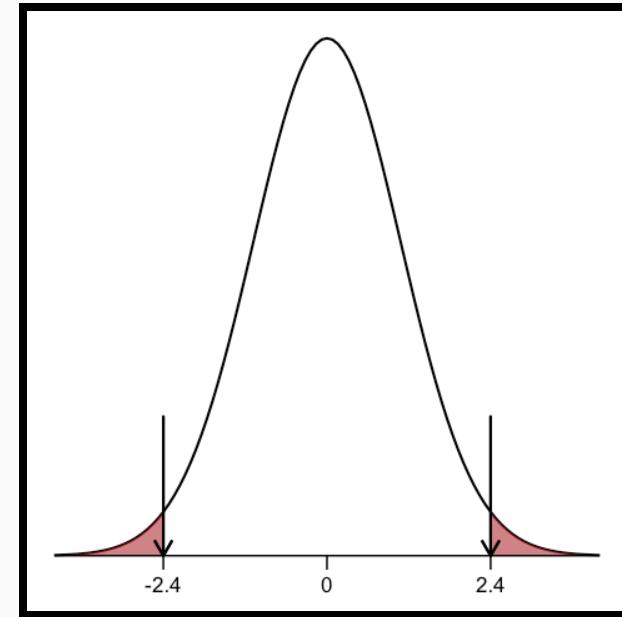
$$\mu_c - \mu_i = 0$$

$$\hat{SE} = 1.08$$

$$\bar{x}_c - \bar{x}_i = 6.6 - 4.0 = 2.6$$

- ❓ What is the probability of observing a difference of 2.6 days if  $\mu_c - \mu_i = 0$ ?

$$z = \frac{(\bar{x}_c - \bar{x}_i) - (\mu_c - \mu_i)}{\hat{SE}} = \frac{2.6 - 0}{1.08} = 2.4 \quad \rightarrow \quad \mathcal{P} = 2 \times 0.0082 = 0.0164$$



# P-value

- 🎯 The P-value measures the discrepancy between the data and the null hypothesis  $\mathcal{H}_0$  and correspond to the probability of observing such an extreme value, if  $\mathcal{H}_0$  was true

# P-value

- 🎯 The P-value measures the discrepancy between the data and the null hypothesis  $\mathcal{H}_0$  and correspond to the probability of observing such an extreme value, if  $\mathcal{H}_0$  was true

P-value = 0.5 → 50% → 1 sample out of 2

P-value = 0.1 → 10% → 1 sample out of 10

P-value = 0.05 → 5% → 1 sample out of 20

P-value = 0.01 → 1% → 1 sample out of 100

P-value = 0.005 → 0.5% → 1 sample out of 200

# P-value e statistical significance

- 🎯 The P-value measures the discrepancy between the data and the null hypothesis  $\mathcal{H}_0$  and correspond to the probability of observing such an extreme value, if  $\mathcal{H}_0$  was true

If the P-value is less than some pre-specified level  $\alpha$ , we consider the observed difference as statistically significant

$$\alpha = 0.05 \text{ or } 0.01$$

# **Hypothesis testing, one step at a time**

# Hypothesis testing, one step at a time

1. Define a null hypothesis ( $\mathcal{H}_0$ )

Dexamethasone plus standard care is **as effective as** standard care

$$\mathcal{H}_0 : \mu_c - \mu_i = 0$$

# Hypothesis testing, one step at a time

1. Define a null hypothesis ( $\mathcal{H}_0$ )
2. Choose a test statistic that estimates something that, if extreme enough, would lead one to doubt  $\mathcal{H}_0$

*t*-test<sup>(\*)</sup> for differences in mean

We are using the *t*-test for differences in mean and not the *z*-test because we don't know the standard deviation in the population (and are using the sample's standard deviation instead).

# Hypothesis testing, one step at a time

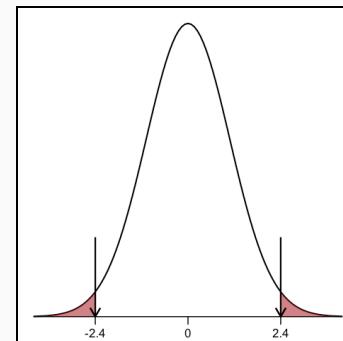
1. Define a null hypothesis ( $\mathcal{H}_0$ )
2. Choose a test statistic that estimates something that, if extreme enough, would lead one to doubt  $\mathcal{H}_0$
3. Generate the sampling distribution of the chosen test statistic, assuming  $\mathcal{H}_0$  to be true

$$\mathcal{N} = (\mu_c - \mu_i, \hat{SE}), \text{ with } \mu_c - \mu_i = 0 \text{ and } \hat{SE} = \sqrt{\frac{s_c^2}{n_c} + \frac{s_i^2}{n_i}}$$

# Hypothesis testing, one step at a time

1. Define a null hypothesis ( $\mathcal{H}_0$ )
2. Choose a test statistic that estimates something that, if extreme enough, would lead one to doubt  $\mathcal{H}_0$
3. Generate the sampling distribution of the chosen test statistic, assuming  $\mathcal{H}_0$  to be true
4. Check whether the observed statistic lies in the tails of this distribution, and calculate a probability (P-value) for this event

$$\mathcal{P} = 2 \times 0.0082 = 0.0164$$



# Hypothesis testing, one step at a time

1. Define a null hypothesis ( $\mathcal{H}_0$ )
2. Choose a test statistic that estimates something that, if extreme enough, would lead one to doubt  $\mathcal{H}_0$
3. Generate the sampling distribution of the chosen test statistic, assuming  $\mathcal{H}_0$  to be true
4. Check whether the observed statistic lies in the tails of this distribution, and calculate a probability (P-value) for this event
5. Declare the result statistically significant if the P-value is below some critical threshold  $\alpha$

$$\mathcal{P} = 2 \times 0.0082 = 0.0164 < \alpha = 0.05 \quad \rightarrow \quad \text{one rejects } \mathcal{H}_0$$

## Exercize #4

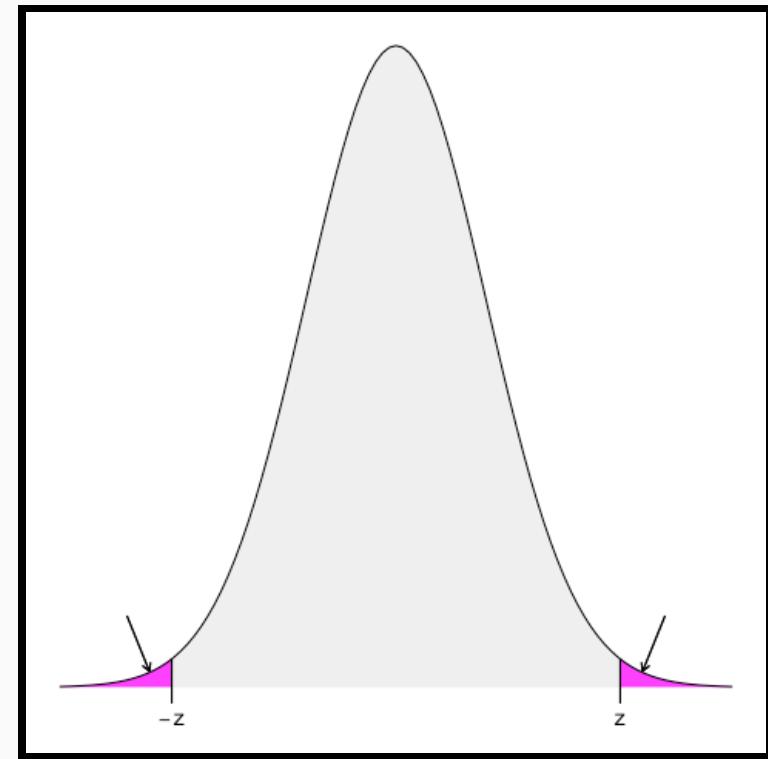
- ? In a randomised control trial, the P-value for one of the outcomes is 0.48. With an  $\alpha$  level of 5%, are there statistically significant differences in the outcome between the two arms of the trial?
- a) Yes, because the P value is lower than the  $\alpha$  level
  - b) Yes, because the P value is greater than the  $\alpha$  level
  - c) No, because the P value is lower than the  $\alpha$  level
  - d) No, because the P value is greater than the  $\alpha$  level

## Exercize #4 -- Solution

- ? In a randomised control trial, the P-value for one of the outcomes is 0.48. With an  $\alpha$  level of 5%, are there statistically significant differences in the outcome between the two arms of the trial?
- a) Yes, because the P value is lower than the  $\alpha$  level
  - b) Yes, because the P value is greater than the  $\alpha$  level
  - c) No, because the P value is lower than the  $\alpha$  level
  - d) No, because the P value is greater than the  $\alpha$  level 

# One- and two-tailed tests

- 🎯  $\mathcal{H}_1: \mu_i - \mu_c \neq 0$   
 $\mathcal{H}_0: \mu_i - \mu_c = 0$   
→ two-tailed test



# One- and two-tailed tests



$$\mathcal{H}_1: \mu_i - \mu_c \neq 0$$

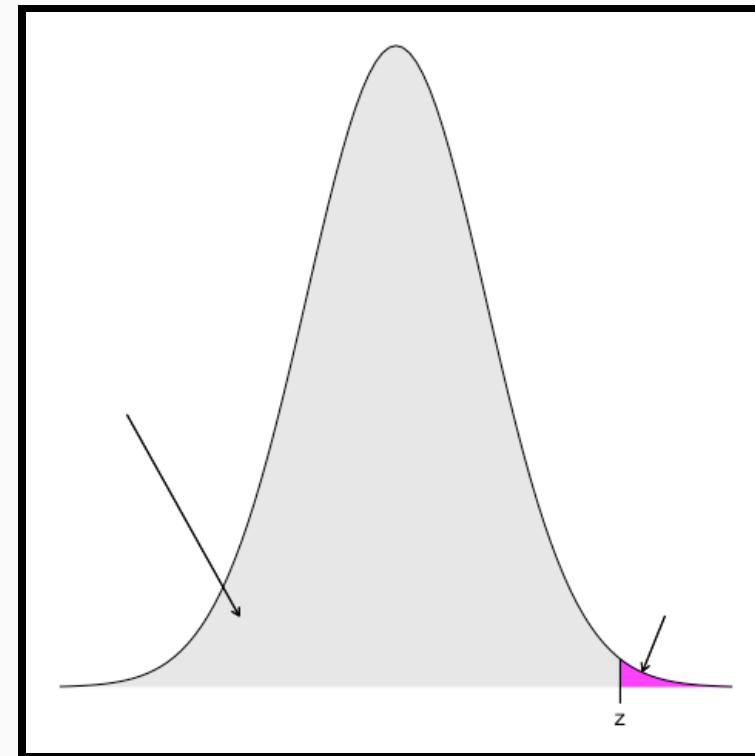
$$\mathcal{H}_0: \mu_i - \mu_c = 0$$

→ two-tailed test

$$\mathcal{H}_1: \mu_i - \mu_c < 0$$

$$\mathcal{H}_0: \mu_i - \mu_c \geq 0$$

→ one-tailed test



# One- and two-tailed tests



$$\mathcal{H}_1: \mu_i - \mu_c \neq 0$$

$$\mathcal{H}_0: \mu_i - \mu_c = 0$$

→ two-tailed test

$$\mathcal{H}_1: \mu_i - \mu_c < 0$$

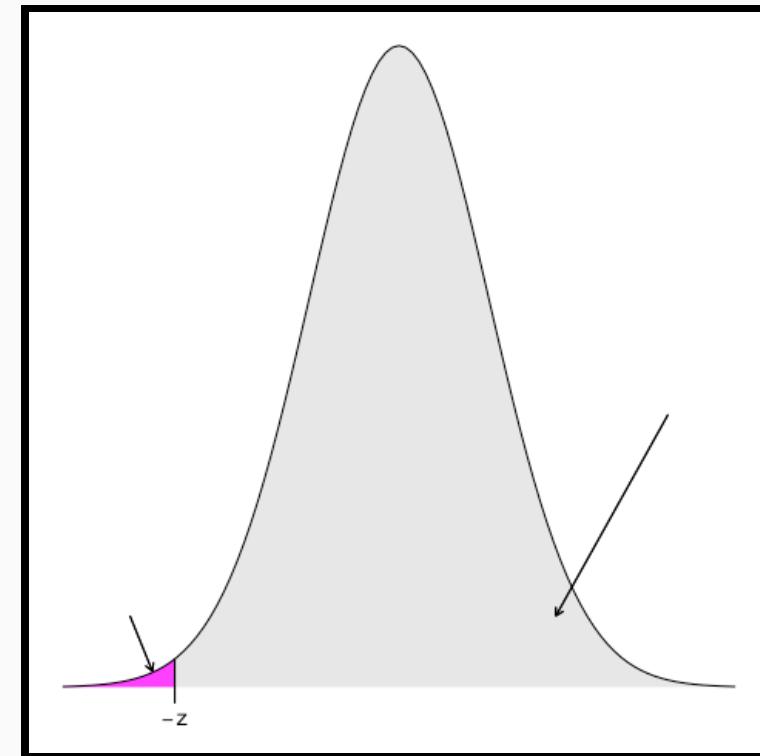
$$\mathcal{H}_0: \mu_i - \mu_c \geq 0$$

or

$$\mathcal{H}_1: \mu_i - \mu_c > 0$$

$$\mathcal{H}_0: \mu_i - \mu_c \leq 0$$

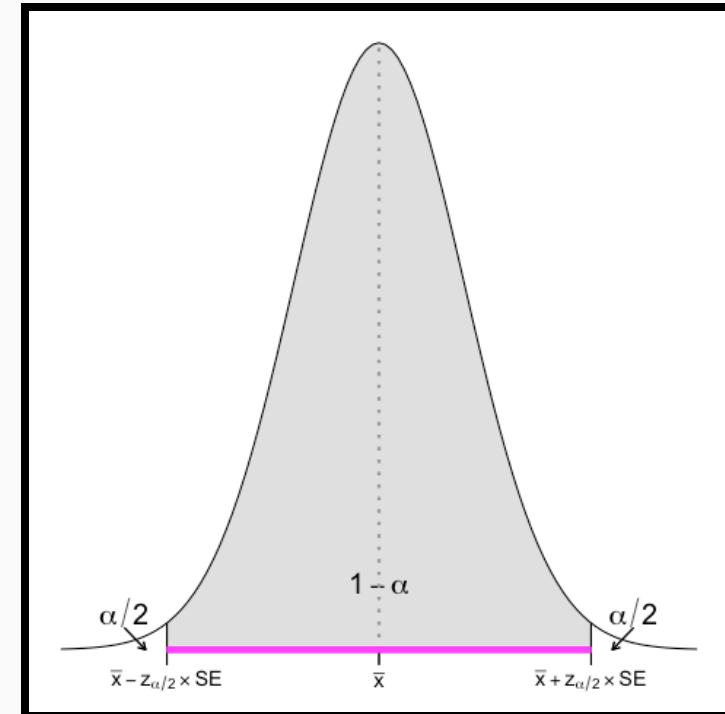
→ one-tailed test



# Hypothesis testing & confidence intervals

- 🎯 A 95% confidence interval is the set of null hypotheses that are not rejected with  $\alpha = 0.05$

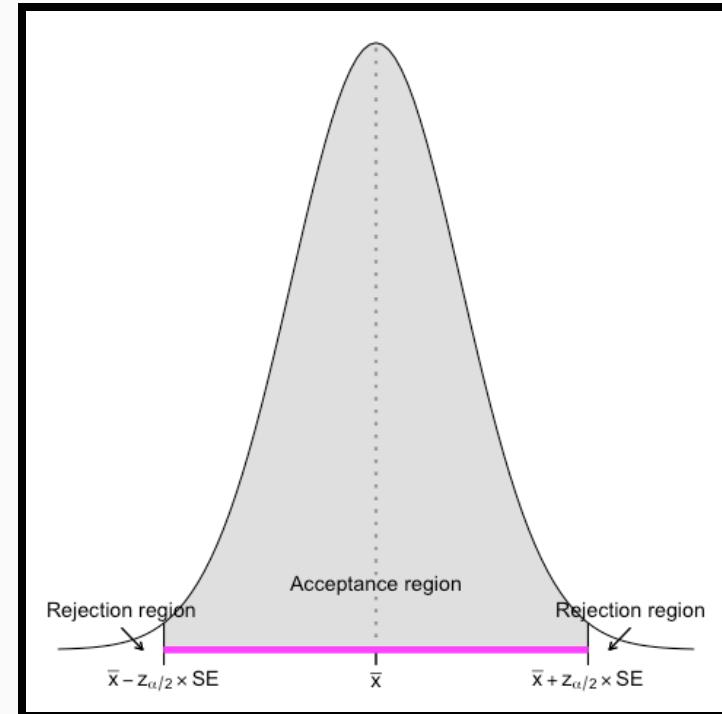
Confidence Level	$\alpha$	$\alpha/2$	$z_{\alpha/2}$
95%	5%	2.5%	1.96



# Hypothesis testing & confidence intervals

- 🎯 A 95% confidence interval is the set of null hypotheses that are not rejected with  $\alpha = 0.05$

In a two-sided test, P-value  $< 0.05$  if the 95% confidence interval does not include the null hypothesis (usually  $0$ ).



# Exercise #5

- Is Zidovudine better than placebo to reduce mother-infant HIV transmission?

From April 1991 through December 20, 1993, the cutoff date for the first interim analysis of efficacy, 477 pregnant women were enrolled; during the study period, 409 gave birth to 415 live-born infants. HIV-infection status was known for 363 births (180 in the zidovudine group and 183 in the placebo group). Thirteen infants in the zidovudine group and 40 in the placebo group were HIV-infected.

$$n_i = 180, \quad m_i = 13, \quad p_i = \frac{m_i}{n_i} = \frac{13}{180} = 0.07$$

$$n_c = 183, \quad m_c = 40, \quad p_c = \frac{m_c}{n_c} = \frac{40}{183} = 0.22$$

$$\mathcal{N} = (\pi_i - \pi_c, \frac{\pi_i \times (1-\pi_i)}{n_i} + \frac{\pi_c \times (1-\pi_c)}{n_c})$$

05:00

# Exercise 5 -- Solution

- ? Is Zidovudine better than placebo to reduce mother-infant HIV transmission?

$$n_i = 180, \quad m_i = 13, \quad p_i = \frac{m_i}{n_i} = \frac{13}{180} = 0.07$$

$$n_c = 183, \quad m_c = 40, \quad p_c = \frac{m_c}{n_c} = \frac{40}{183} = 0.22$$

$$\mathcal{N} = (\pi_i - \pi_c, \frac{\pi_i \times (1 - \pi_i)}{n_i} + \frac{\pi_c \times (1 - \pi_c)}{n_c})$$

1. Define a null hypothesis ( $\mathcal{H}_0$ )

Zidovudine is **as effective as** placebo to reduce mother-infant HIV transmission

$$\rightarrow \mathcal{H}_0 : \pi_i - \pi_c = 0$$

# Exercise 5 -- Solution

- ? Is Zidovudine better than placebo to reduce mother-infant HIV transmission?

$$n_i = 180, \quad m_i = 13, \quad p_i = \frac{m_i}{n_i} = \frac{13}{180} = 0.07$$

$$n_c = 183, \quad m_c = 40, \quad p_c = \frac{m_c}{n_c} = \frac{40}{183} = 0.22$$

$$\mathcal{N} = (\pi_i - \pi_c, \frac{\pi_i \times (1 - \pi_i)}{n_i} + \frac{\pi_c \times (1 - \pi_c)}{n_c})$$

2. Choose a test statistic that estimates something that, if extreme enough, would lead one to doubt  $\mathcal{H}_0$   
→  $t$ -test<sup>(\*)</sup> for differences in proportion

We are using the  $t$ -test for differences in proportion and not the  $z$ -test because we don't know the standard deviation in the population (and are using the sample's standard deviation instead).

# Exercise 5 -- Solution

- ? Is Zidovudine better than placebo to reduce mother-infant HIV transmission?

$$n_i = 180, \quad m_i = 13, \quad p_i = \frac{m_i}{n_i} = \frac{13}{180} = 0.07$$

$$n_c = 183, \quad m_c = 40, \quad p_c = \frac{m_c}{n_c} = \frac{40}{183} = 0.22$$

$$\mathcal{N} = (\pi_i - \pi_c, \frac{\pi_i \times (1 - \pi_i)}{n_i} + \frac{\pi_c \times (1 - \pi_c)}{n_c})$$

3. Generate the sampling distribution of the chosen test statistic,  
assuming  $\mathcal{H}_0$  to be true

$$\begin{aligned}\mathcal{N} &= (\pi_i - \pi_c, \text{SE}), \text{ with } \pi_i - \pi_c = 0 \text{ and } \hat{\text{SE}} = \sqrt{\frac{\pi_i \times (1 - \pi_i)}{n_i} + \frac{\pi_c \times (1 - \pi_c)}{n_c}} \\ \hat{\text{SE}} &= \sqrt{\frac{0.15 \times (1 - 0.15)}{180} + \frac{0.15 \times (1 - 0.15)}{183}} = 0.037 \quad \rightarrow \quad \mathcal{N} = (0, 0.037),\end{aligned}$$

# Exercise 5 -- Solution

- ? Is Zidovudine better than placebo to reduce mother-infant HIV transmission?

$$n_i = 180, \quad m_i = 13, \quad p_i = \frac{m_i}{n_i} = \frac{13}{180} = 0.07$$

$$n_c = 183, \quad m_c = 40, \quad p_c = \frac{m_c}{n_c} = \frac{40}{183} = 0.22$$

4. Check whether the observed statistic lies in the tails of this distribution, and calculate a probability (P-value) for this event

$$\mathcal{N} = (0, 0.037),$$

$$z = \frac{(\bar{p}_s - \bar{p}_c) - (\pi_c - \pi_i)}{\hat{SE}} = \frac{(0.07 - 0.22) - 0}{0.037} = \frac{-0.15}{0.037} = 4.06$$

$$\mathcal{P}(|(\bar{p}_s - \bar{p}_c)| \geq 4.06) = 2 \times 0.00005 = 0.0001 = 1 \times 10^{-4}$$

# Exercise 5 -- Solution

- ? Is Zidovudine better than placebo to reduce mother-infant HIV transmission?

$$n_i = 180, \quad m_i = 13, \quad p_i = \frac{m_i}{n_i} = \frac{13}{180} = 0.07$$

$$n_c = 183, \quad m_c = 40, \quad p_c = \frac{m_c}{n_c} = \frac{40}{183} = 0.22$$

5. Declare the result statistically significant if the P-value is below some critical threshold  $\alpha$

$$\mathcal{P} = 2 \times 0.00005 = 0.0001 = 1 \times 10^{-4} < \alpha = 0.05 \quad \rightarrow \quad \text{one rejects } \mathcal{H}_0$$

# Pearson's $\chi^2$ test

- ? Is Zidovudine better than placebo to reduce mother-infant HIV transmission?

$$n_i = 180, \quad m_i = 13, \quad p_i = \frac{m_i}{n_i} = \frac{13}{180} = 0.07$$

$$n_c = 183, \quad m_c = 40, \quad p_c = \frac{m_c}{n_c} = \frac{40}{183} = 0.22$$

Let's use another test to compare differences in proportion!

# Pearson's $\chi^2$ test

- ? Is Zidovudine better than placebo to reduce mother-infant HIV transmission?

$$n_i = 180, \quad m_i = 13, \quad p_i = \frac{m_i}{n_i} = \frac{13}{180} = 0.07$$

$$n_c = 183, \quad m_c = 40, \quad p_c = \frac{m_c}{n_c} = \frac{40}{183} = 0.22$$

1. Define a null hypothesis ( $\mathcal{H}_0$ )

Zidovudine is **as effective as** placebo to reduce mother-infant HIV transmission

$$\rightarrow \mathcal{H}_0 : \pi_i - \pi_c = 0$$

# Pearson's $\chi^2$ test

- ? Is Zidovudine better than placebo to reduce mother-infant HIV transmission?

$$n_i = 180, \quad m_i = 13, \quad p_i = \frac{m_i}{n_i} = \frac{13}{180} = 0.07$$

$$n_c = 183, \quad m_c = 40, \quad p_c = \frac{m_c}{n_c} = \frac{40}{183} = 0.22$$

2. Choose a test statistic that estimates something that, if extreme enough, would lead one to doubt  $\mathcal{H}_0$

→ Pearson's  $\chi^2$  test for categorical data

# Pearson's $\chi^2$ test

- ? Is Zidovudine better than placebo to reduce mother-infant HIV transmission?

$$n_i = 180, \quad m_i = 13, \quad p_i = \frac{m_i}{n_i} = \frac{13}{180} = 0.07$$

$$n_c = 183, \quad m_c = 40, \quad p_c = \frac{m_c}{n_c} = \frac{40}{183} = 0.22$$

3. Generate the sampling distribution of the chosen test statistic, assuming  $\mathcal{H}_0$  to be true

# Pearson's $\chi^2$ test

- ? Is Zidovudine better than placebo to reduce mother-infant HIV transmission?

$$n_i = 180, \quad m_i = 13, \quad p_i = \frac{m_i}{n_i} = \frac{13}{180} = 0.07$$

$$n_c = 183, \quad m_c = 40, \quad p_c = \frac{m_c}{n_c} = \frac{40}{183} = 0.22$$

3. Generate the sampling distribution of the chosen test statistic, assuming  $\mathcal{H}_0$  to be true

- ? Let's fill this contingency table

Treatment/Infected	Yes	No	Total
Zidovudine			
Placebo			
Total			

# Pearson's $\chi^2$ test

- ? Is Zidovudine better than placebo to reduce mother-infant HIV transmission?

$$n_i = 180, \quad m_i = 13, \quad p_i = \frac{m_i}{n_i} = \frac{13}{180} = 0.07$$

$$n_c = 183, \quad m_c = 40, \quad p_c = \frac{m_c}{n_c} = \frac{40}{183} = 0.22$$

3. Generate the sampling distribution of the chosen test statistic, assuming  $\mathcal{H}_0$  to be true

Observed values

Treatment/Infected	Yes	No	Total
Zidovudine	13	167	180
Placebo	40	143	183
Total	53	310	363

# Pearson's $\chi^2$ test

- ? Is Zidovudine better than placebo to reduce mother-infant HIV transmission?

$$\Pi = \frac{tot_{infected}}{total} = \frac{53}{363} = 0.146$$

Observed values

Treatment/Infected	Yes	No	Total
Zidovudine	13	167	180
Placebo	40	143	183
Total	53	310	363

# Pearson's $\chi^2$ test

- ? Is Zidovudine better than placebo to reduce mother-infant HIV transmission?

$$\Pi = \frac{tot_{infected}}{total} = \frac{53}{363} = 0.146$$

Observed values

Treatment/Infected	Yes	No	Total
Zidovudine	13	167	180
Placebo	40	143	183
Total	53	310	363

Expected values

Treatment/Infected	Yes	No	Total
Zidovudine	$180 * 0.146$		180
Placebo	$183 * 0.146$		183
Total	53	310	363

# Pearson's $\chi^2$ test

- ? Is Zidovudine better than placebo to reduce mother-infant HIV transmission?

$$\Pi = \frac{tot_{infected}}{total} = \frac{53}{363} = 0.146$$

Observed values

Treatment/Infected	Yes	No	Total
Zidovudine	13	167	180
Placebo	40	143	183
Total	53	310	363

Expected values

Treatment/Infected	Yes	No	Total
Zidovudine	26.28		180
Placebo	26.72		183
Total	53	310	363

# Pearson's $\chi^2$ test

- ? Is Zidovudine better than placebo to reduce mother-infant HIV transmission?

$$\Pi = \frac{tot_{infected}}{total} = \frac{53}{363} = 0.146$$

Observed values

Treatment/Infected	Yes	No	Total
Zidovudine	13	167	180
Placebo	40	143	183
Total	53	310	363

Expected values

Treatment/Infected	Yes	No	Total
Zidovudine	26.28	153.72	180
Placebo	26.72	156.28	183
Total	53	310	363

# Pearson's $\chi^2$ test

- ? Is Zidovudine better than placebo to reduce mother-infant HIV transmission?

$$\Pi = \frac{tot_{infected}}{total} = \frac{53}{363} = 0.146$$

Observed values

Treatment/Infected	Yes	No	Total
Zidovudine	13	167	180
Placebo	40	143	183
Total	53	310	363

Expected values

Treatment/Infected	Yes	No	Total
Zidovudine	26.28	153.72	180
Placebo	26.72	156.28	183
Total	53	310	363

$$\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected} = \frac{(13 - 26.28)^2}{26.28} + \frac{(167 - 153.72)^2}{153.72} + \frac{(40 - 26.72)^2}{26.72} + \frac{(143 - 156.28)^2}{156.28} = 15.57$$

# Pearson's $\chi^2$ test

? Is Zidovudine better than placebo to reduce mother-infant HIV transmission?

$$\Pi = \frac{tot_{infected}}{total} = \frac{53}{363} = 0.146$$

Observed values

Treatment/Infected	Yes	No	Total
Zidovudine	13	167	180
Placebo	40	143	183
Total	53	310	363

Expected values

Treatment/Infected	Yes	No	Total
Zidovudine	26.28	153.72	180
Placebo	26.72	156.28	183
Total	53	310	363

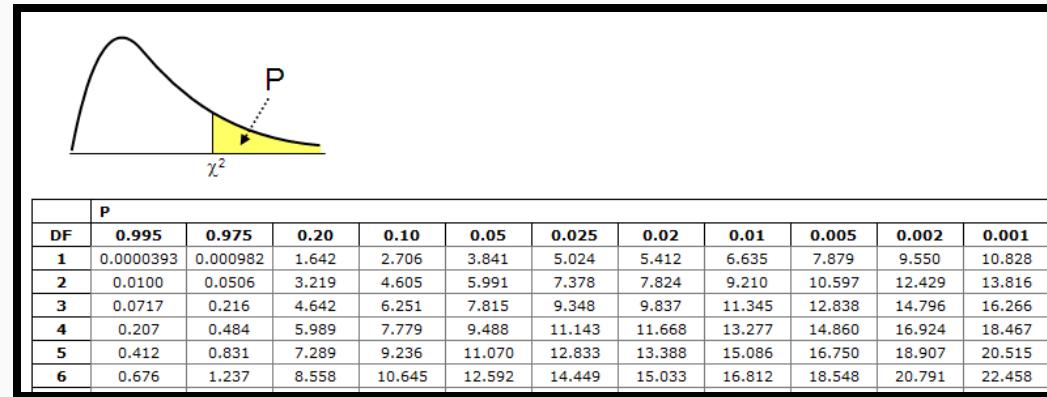
$$\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected} = \frac{(13 - 26.28)^2}{26.28} + \frac{(167 - 153.72)^2}{153.72} + \frac{(40 - 26.72)^2}{26.72} + \frac{(143 - 156.28)^2}{156.28} = 15.57$$

$$df = (n_{\text{righe}} - 1) \times (n_{\text{colonne}} - 1) = 1$$

# Pearson's $\chi^2$ test

? Is Zidovudine better than placebo to reduce mother-infant HIV transmission?

$$\Pi = \frac{tot_{infected}}{total} = \frac{53}{363} = 0.146$$



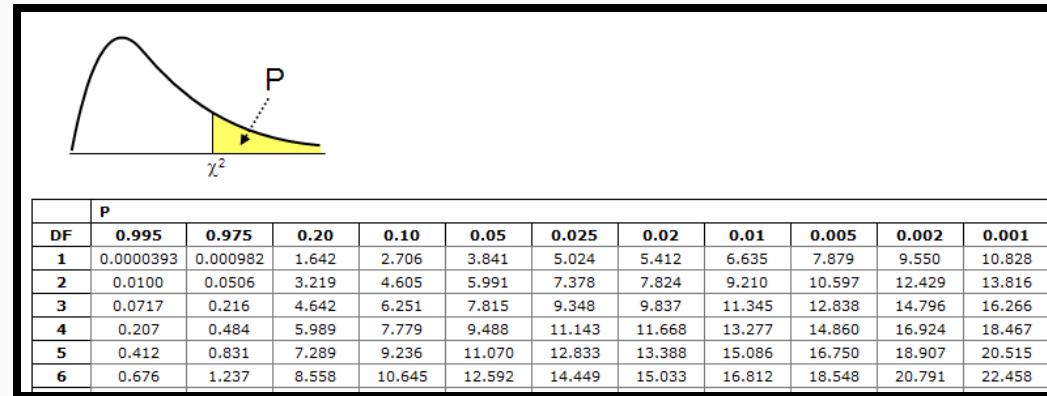
4. Check whether the observed statistic lies in the tails of this distribution, and calculate a probability (P-value) for this event

$$\chi^2 = 15.57 \quad df = 1 \quad \rightarrow \quad P < 0.001 = 7.9 \times 10^{-5}$$

# Pearson's $\chi^2$ test

? Is Zidovudine better than placebo to reduce mother-infant HIV transmission?

$$\Pi = \frac{tot_{infected}}{total} = \frac{53}{363} = 0.146$$



4. Declare the result statistically significant if the P-value is below some critical threshold  $\alpha$

$$\chi^2 = 15.57 \quad df = 1 \quad \rightarrow \quad P < 0.001 = 7.9 \times 10^{-5} < \alpha = 0.05 \rightarrow \text{reject } \mathcal{H}_0$$

# Pearson's $\chi^2$ test

- 📌 Does education level influences physical activity frequency?

Observed values

	No Exercise	Sporadic Exercise	Regular Exercise	Total
Primary education				
Secondary education				
Bachelor/Master				
Doctorate				
Total				

# Pearson's $\chi^2$ test

- 📌 Does education level influences physical activity frequency?

Expected values

	No Exercise	Sporadic Exercise	Regular Exercise	Total
Primary education	$\frac{\Sigma \text{Row}_1 \times \Sigma \text{Column}_1}{\text{Total}}$	$\frac{\Sigma \text{Row}_1 \times \Sigma \text{Column}_2}{\text{Total}}$	$\frac{\Sigma \text{Row}_1 \times \Sigma \text{Column}_3}{\text{Total}}$	$\Sigma \text{Row}_1$
Secondary education	$\frac{\Sigma \text{Row}_2 \times \Sigma \text{Column}_1}{\text{Total}}$	...	...	$\Sigma \text{Row}_2$
Bachelor/Master	$\frac{\Sigma \text{Row}_3 \times \Sigma \text{Column}_1}{\text{Total}}$	...	...	$\Sigma \text{Row}_3$
Doctorate	$\frac{\Sigma \text{Row}_4 \times \Sigma \text{Column}_1}{\text{Total}}$	...	...	$\Sigma \text{Row}_4$
Total	$\Sigma \text{Column}_1$	$\Sigma \text{Column}_2$	$\Sigma \text{Column}_3$	Total

?       $df = ?$

# Pearson's $\chi^2$ test

- 📌 Does education level influences physical activity frequency?

Expected values

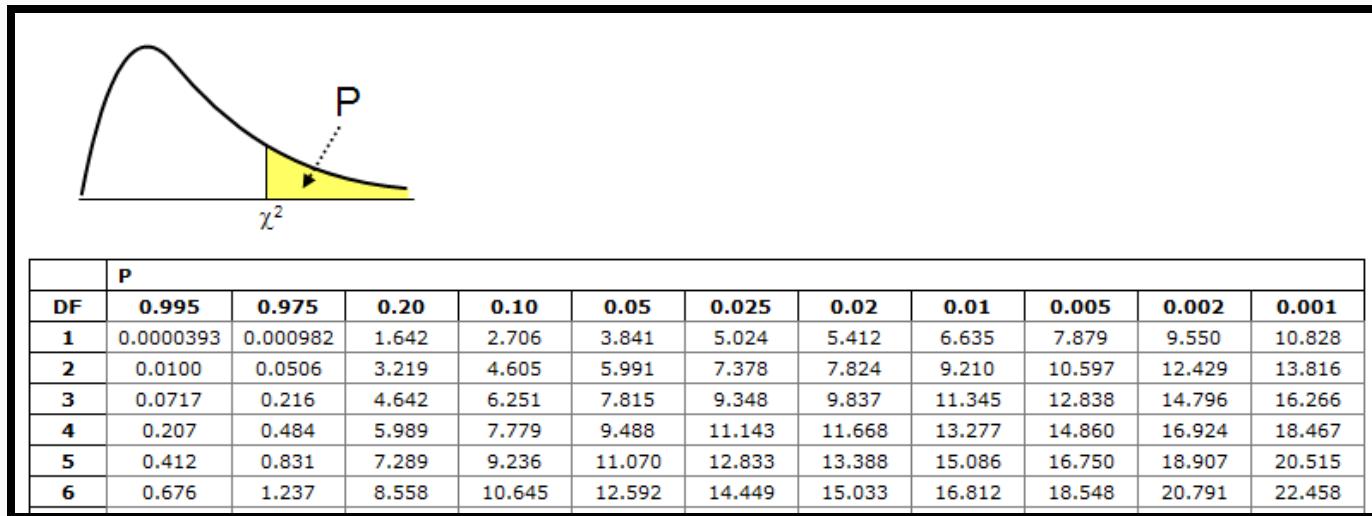
	No Exercise	Sporadic Exercise	Regular Exercise	Total
Primary education	$\frac{\Sigma \text{Row}_1 \times \Sigma \text{Column}_1}{\text{Total}}$	$\frac{\Sigma \text{Row}_1 \times \Sigma \text{Column}_2}{\text{Total}}$	$\frac{\Sigma \text{Row}_1 \times \Sigma \text{Column}_3}{\text{Total}}$	$\Sigma \text{Row}_1$
Secondary education	$\frac{\Sigma \text{Row}_2 \times \Sigma \text{Column}_1}{\text{Total}}$	...	...	$\Sigma \text{Row}_2$
Bachelor/Master	$\frac{\Sigma \text{Row}_3 \times \Sigma \text{Column}_1}{\text{Total}}$	...	...	$\Sigma \text{Row}_3$
Doctorate	$\frac{\Sigma \text{Row}_4 \times \Sigma \text{Column}_1}{\text{Total}}$	...	...	$\Sigma \text{Row}_4$
Total	$\Sigma \text{Column}_1$	$\Sigma \text{Column}_2$	$\Sigma \text{Column}_3$	Total

?  $\text{df} = (n_{\text{row}} - 1) \times (n_{\text{column}} - 1) = (4 - 1) \times (3 - 1) = 3 \times 2 = 6$

# Exercise #6

- Does the area of practice influences drinking habits of Italian healthcare workers?

Out of 279, 230, and 130 healthcare professionals working in medicine, surgery, and other wards, 122, 107, and 51 were non-drinkers, respectively.



# Exercise #6 -- Solution

- ? Does the area of practice influences drinking habits of Italian healthcare workers?

Out of 279, 230, and 130 healthcare professionals working in medicine, surgery, and other wards, 122, 107, and 51 were non-drinkers, respectively.

$\mathcal{H}_0$ : The area of practice **does not** influence drinking habits

# Exercise #6 -- Solution

- ? Does the area of practice influences drinking habits of Italian healthcare workers?

Out of 279, 230, and 130 healthcare professionals working in medicine, surgery, and other wards, 122, 107, and 51 were non-drinkers, respectively.

$\mathcal{H}_0$ : The area of practice **does not** influence drinking habits

Observed values

Ward/Drinking	No	Yes	Total
Medicine	122	157	279
Surgery	107	123	230
Others	51	79	130
Total	280	359	639

# Exercise #6 -- Solution

- ? Does the area of practice influences drinking habits of Italian healthcare workers?

$\mathcal{H}_0$ : The area of practice **does not** influence drinking habits

Observed values

Ward/Drinking	No	Yes	Total
Medicine	122	157	279
Surgery	107	123	230
Others	51	79	130
Total	280	359	639

# Exercise #6 -- Solution

- ? Does the area of practice influences drinking habits of Italian healthcare workers?

$H_0$ : The area of practice **does not** influence drinking habits

Observed values

Ward/Drinking	No	Yes	Total
Medicine	122	157	279
Surgery	107	123	230
Others	51	79	130
Total	280	359	639

Expected values

Ward/Drinking	No	Yes	Total
Medicine	$\frac{279 \times 280}{639}$	$\frac{279 \times 359}{639}$	279
Surgery	$\frac{230 \times 280}{639}$	$\frac{230 \times 359}{639}$	230
Others	$\frac{130 \times 280}{639}$	$\frac{130 \times 359}{639}$	130
Total	280	359	639

# Exercise #6 -- Solution

- Does the area of practice influences drinking habits of Italian healthcare workers?

$\mathcal{H}_0$ : The area of practice **does not** influence drinking habits

Observed values

Ward/Drinking	No	Yes	Total
Medicine	122	157	279
Surgery	107	123	230
Others	51	79	130
Total	280	359	639

Expected values

Ward/Drinking	No	Yes	Total
Medicine	122.25	156.75	279
Surgery	100.78	129.22	230
Others	56.96	73.04	130
Total	280	359	639

$$\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected}$$

# Exercise #6 -- Solution

- ? Does the area of practice influences drinking habits of Italian healthcare workers?

Observed values

Ward/Drinking	No	Yes	Total
Medicine	122	157	279
Surgery	107	123	230
Others	51	79	130
Total	280	359	639

Expected values

Ward/Drinking	No	Yes	Total
Medicine	122.25	156.75	279
Surgery	100.78	129.22	230
Others	56.96	73.04	130
Total	280	359	639

$$\chi^2 = \frac{(122-122.25)^2}{122.25} + \frac{(107-100.78)^2}{100.78} + \frac{(51-56.96)^2}{56.96} + \frac{(157-156.75)^2}{156.75} + \frac{(123-129.22)^2}{129.22} + \frac{(79-73.04)^2}{73.04} = 1.17$$

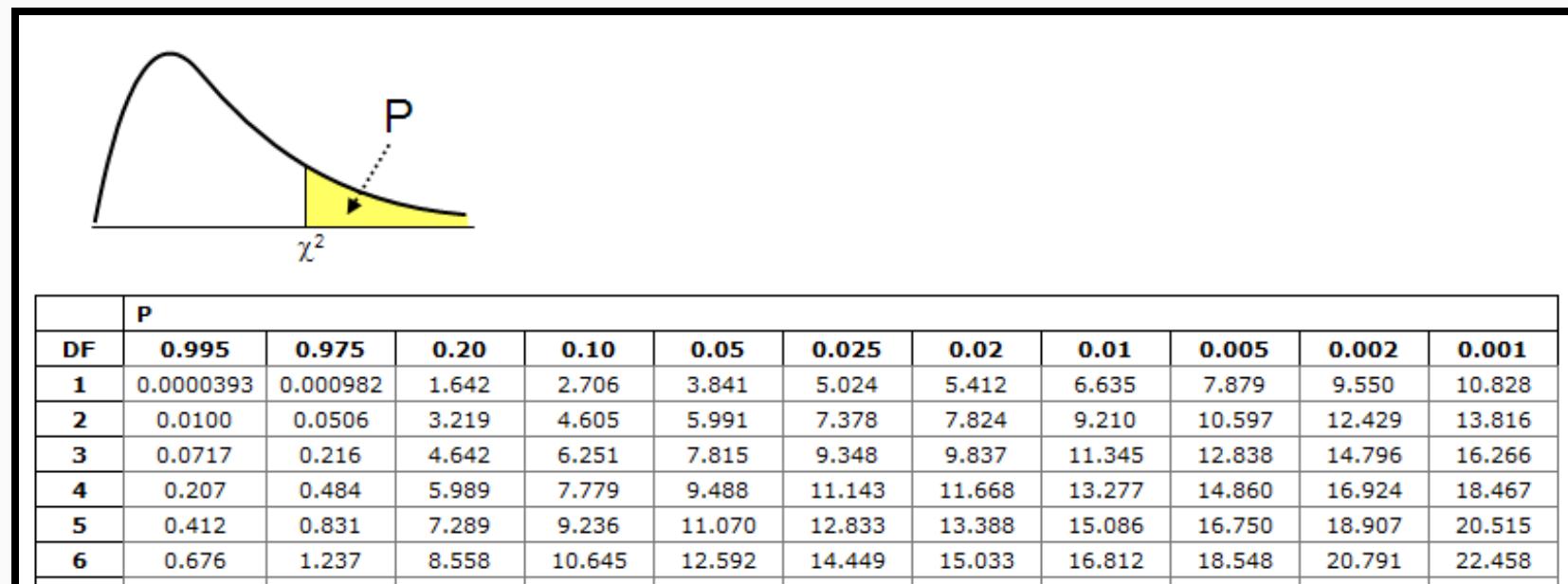
$$df = (n_{\text{row}} - 1) \times (n_{\text{column}} - 1) = 2$$

# Exercise #6 -- Solution

- Does the area of practice influences drinking habits of Italian healthcare workers?

$\mathcal{H}_0$ : The area of practice **does not** influence drinking habits

$$\chi^2 = 1.17, \quad df = 2$$

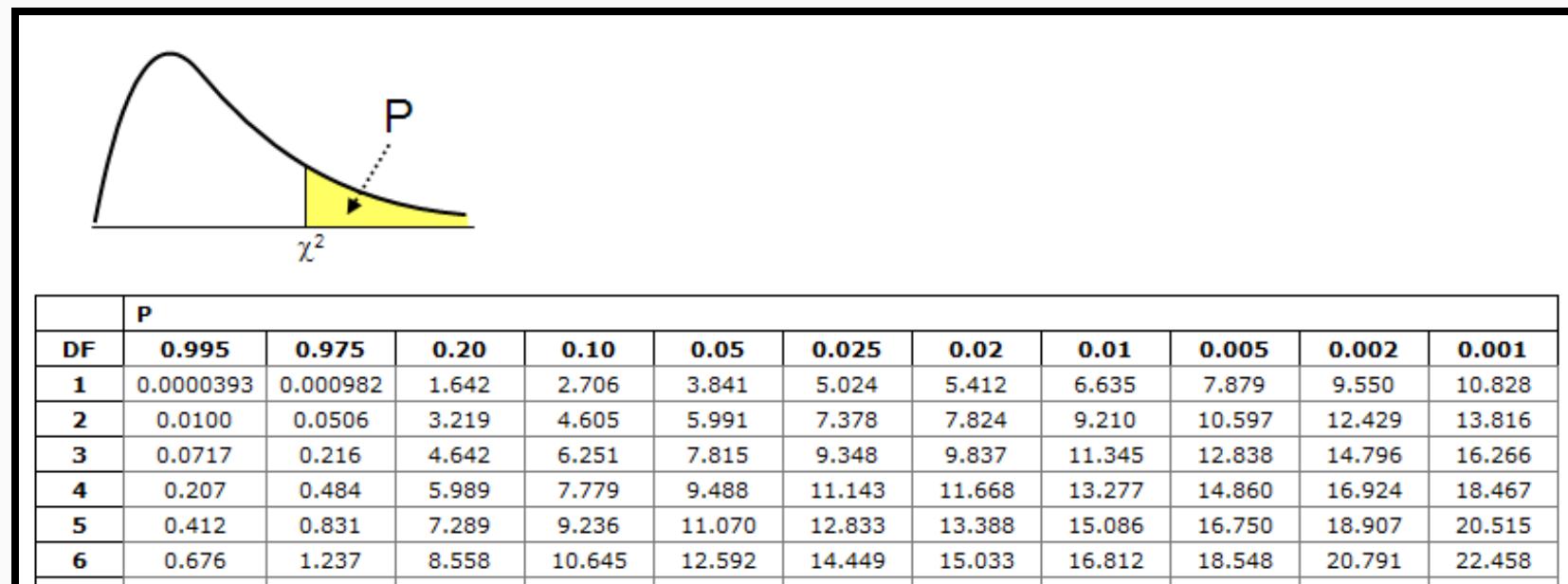


# Exercise #6 -- Solution

- Does the area of practice influences drinking habits of Italian healthcare workers?

$\mathcal{H}_0$ : The area of practice **does not** influence drinking habits

$$\chi^2 = 1.17, \quad df = 2 \quad \rightarrow \quad P > 0.20 = 0.41 > \alpha = 0.05 \rightarrow \text{does not reject } \mathcal{H}_0$$



# Pearson's $\chi^2$ test -- Yates' correction


$$\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected}$$



$$\chi^2 = \sum \frac{(|Observed - Expected| - 0.5)^2}{Expected}$$

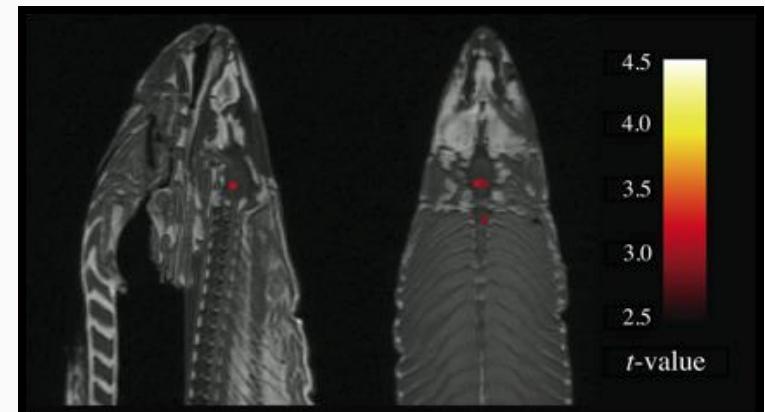
# Multiple testing comparisons

- 📌 A research group showed to a single subject (\*) a series of pics of humans expressing different emotions while carrying out brain imaging (fMRI). They identified 16 brain regions showing a significant response at  $P < 0.001$ .

# Multiple testing comparisons

- 📌 A research group showed to a single subject (\*) a series of pics of humans expressing different emotions while carrying out brain imaging (fMRI). They identified 16 brain regions showing a significant response at  $P < 0.001$ .

(\*) Atlantic salmon, '*not alive at the time of scanning*'



# Multiple testing comparisons

- 📌  $\alpha = 0.05 \rightarrow 5\%$  chance one rejects  $\mathcal{H}_0$  when is true  
 $\mathcal{P} = 1 - 0.95 = 0.05$

# Multiple testing comparisons

- 📌  $\alpha = 0.05 \rightarrow 5\% \text{ chance of rejecting } \mathcal{H}_0 \text{ when is true}$   
 $\mathcal{P} = 1 - 0.95 = 0.05$

with 2 tests, the chance of getting at least 1 significant ( $P < 0.05$ ) is:  
 $\mathcal{P} = 1 - 0.95 \times 0.95 = 1 - 0.95^2 = 0.0975 \rightarrow \approx 10\%$

with 3 tests, the chance of getting at least 1 significant is:  
 $\mathcal{P} = 1 - 0.95^3 = 0.145 \rightarrow \approx 14\%$

with 10 tests, the chance of getting at least 1 significant is:  
 $\mathcal{P} = 1 - 0.95^{10} = 0.40 \rightarrow \approx 40\%$

# Multiple testing correction

- 🎯 When one carries out multiple testing comparisons, they should ask for a smaller  $\alpha$

$$\textbf{Bonferroni-correction: } \alpha = \frac{0.05}{N_{\text{test}}}$$

with 10 tests, the chance of getting at least 1 significant ( $P < \frac{0.05}{10}$ ):

$$\mathcal{P} = 1 - 0.995^{10} = 0.049 \rightarrow \approx 5\%$$

# Multiple testing correction

- 🎯 When one carries out multiple testing comparisons, they should ask for a smaller  $\alpha$
- When one carries out multiple testing comparisons, they should fix the expected proportion of "discoveries" that are false

**False discovery rate (FDR, Benjamini-Hochberg procedure):**

1. Sort test results from the smallest to the largest P-value
2. For a given  $\alpha$ , find the largest  $k$  such that  $\mathcal{P}(k) \leq \frac{k}{m}\alpha$
3. Reject the null hypothesis for  $i = 1, \dots, k$

# Errors in decision making



$p < \alpha \rightarrow$  reject  $\mathcal{H}_0$

$p \geq \alpha \rightarrow$  does not reject  $\mathcal{H}_0$

$\alpha = 0.05 \rightarrow$  5% chance of rejecting  $\mathcal{H}_0$  when is true

# Errors in decision making



$p < \alpha \rightarrow$  reject  $\mathcal{H}_0$

$p \geq \alpha \rightarrow$  does not reject  $\mathcal{H}_0$

$\alpha = 0.05 \rightarrow$  5% chance of rejecting  $\mathcal{H}_0$  when is true

$\mathcal{H}_0$ is		Not rejected	Rejected
True			
False			

# Errors in decision making



$p < \alpha \rightarrow$  reject  $\mathcal{H}_0$

$p \geq \alpha \rightarrow$  does not reject  $\mathcal{H}_0$

$\alpha = 0.05 \rightarrow$  5% chance of rejecting  $\mathcal{H}_0$  when is true

$\mathcal{H}_0$ is		Not rejected	Rejected
True			False positive
False	False negative		

# Errors in decision making

Suspect is Absolved		Convicted
Innocent		One convicts an innocent
Guilty	One absolve an offender	

# Errors in decision making



$p < \alpha \rightarrow$  reject  $\mathcal{H}_0$

$p \geq \alpha \rightarrow$  does not reject  $\mathcal{H}_0$

$\alpha = 0.05 \rightarrow$  5% chance of rejecting  $\mathcal{H}_0$  when is true

$\mathcal{H}_0$ is		Not rejected	Rejected
True		Type I error ( $\alpha$ )	
False	Type II error ( $\beta$ )		

## Exercise #7

- There was a shepherd boy who repeatedly cried wolf when there was no wolf. Yet, each time, villagers went to help him. Then, the wolf arrived, but, when the boy cried wolf, no villager helped.

First we see an example of a...

- a) Type I error, then Type II error
- b) Type II error, then Type I error
- c) Null error, then alternative error
- d) None of the above

## Exercise #7 -- Solution

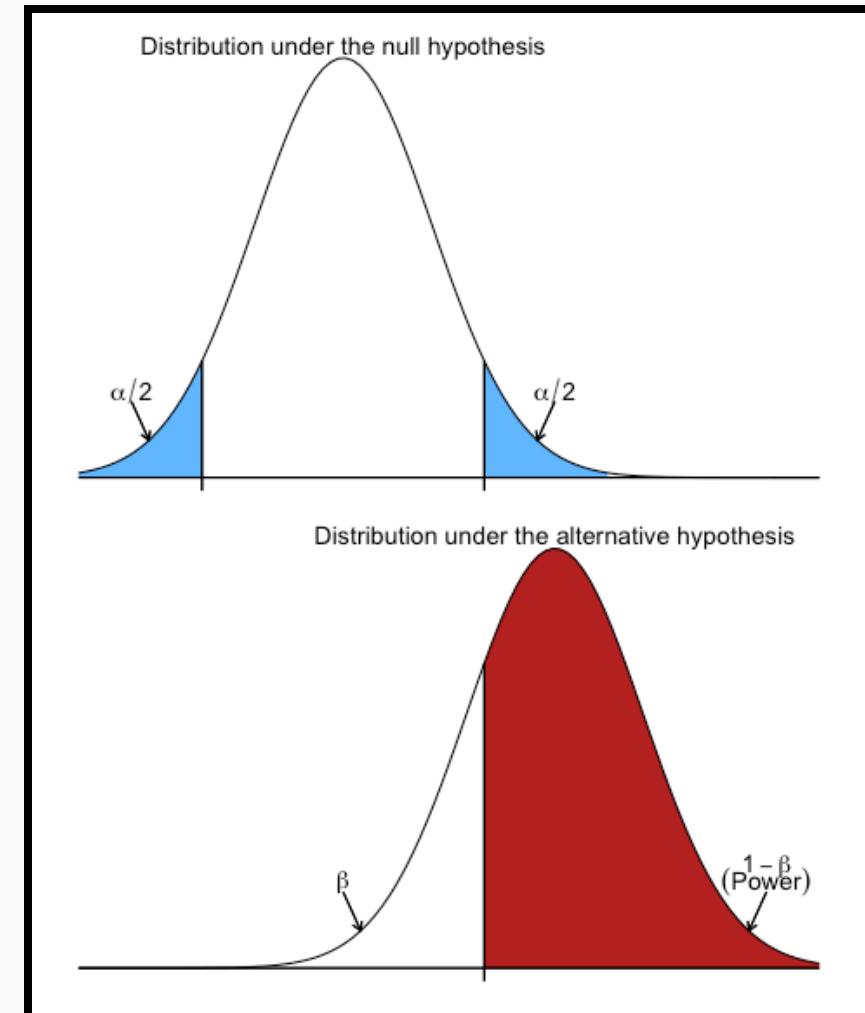
- ? There was a shepherd boy who repeatedly cried wolf when there was no wolf. Yet, each time, villagers went to help him. Then, the wolf arrived, but, when the boy cried wolf, no villager helped.

First we see an example of a...

- a) Type I error, then Type II error 
- b) Type II error, then Type I error
- c) Null error, then alternative error
- d) None of the above

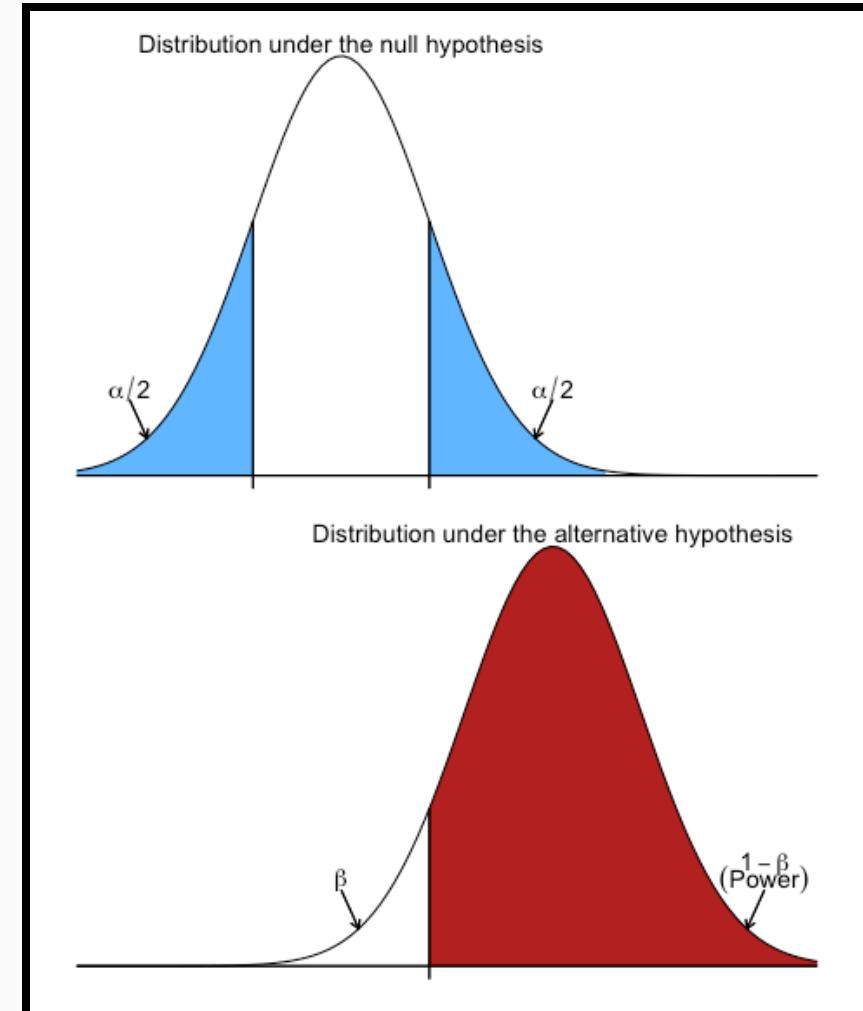
# The power of a study

$\mathcal{H}_0$ is Not rejected		Rejected
True		$\alpha$
False	$\beta$	$1 - \beta$ Statistical power



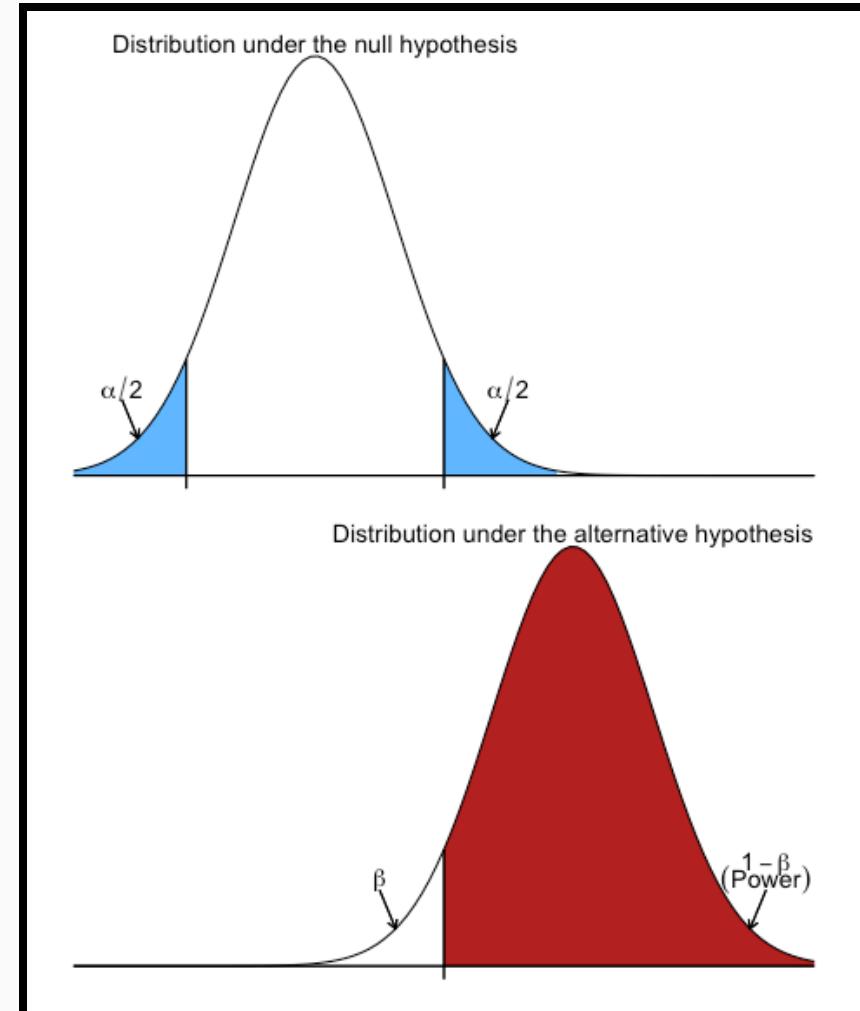
# The power of a study

- 🎯 The power is increased by:
  - larger  $\alpha$



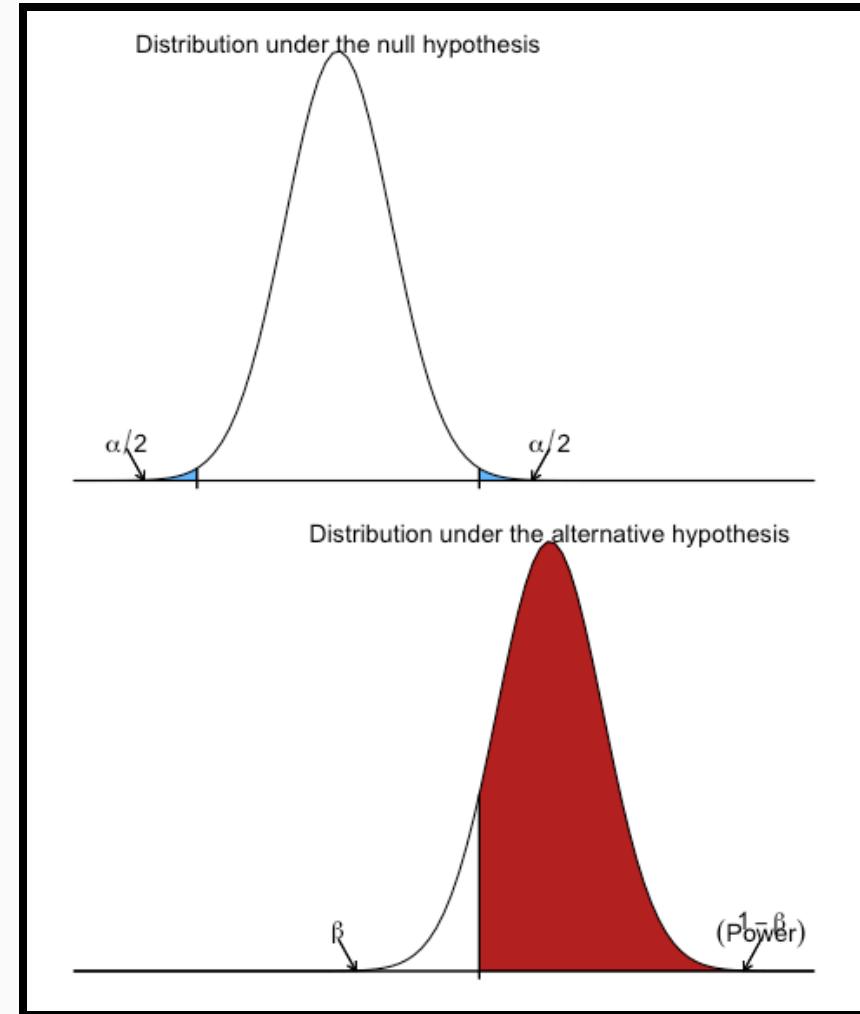
# The power of a study

- 🎯 The power is increased by:
  - larger  $\alpha$
  - larger  $\mu_i - \mu_c$



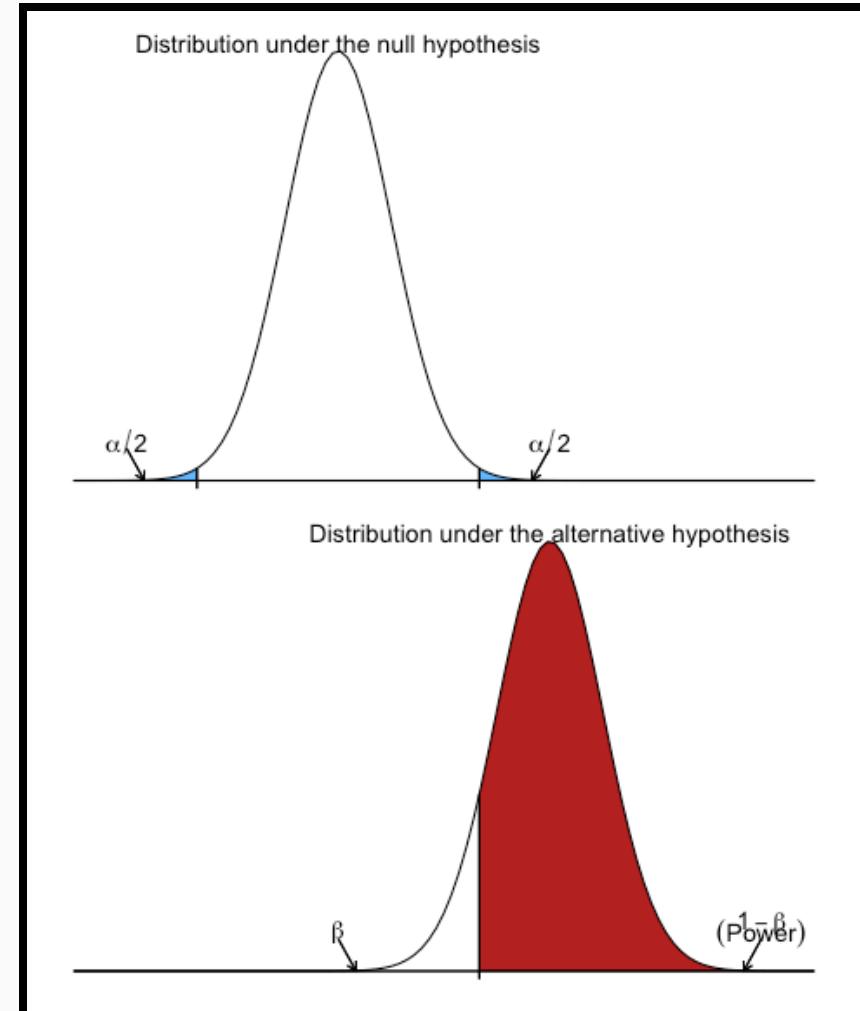
# The power of a study

- 🎯 The power is increased by:
  - larger  $\alpha$
  - larger  $\mu_i - \mu_c$
  - smaller  $\sigma^2$



# The power of a study

- 🎯 The power is increased by:
  - larger  $\alpha$
  - larger  $\mu_i - \mu_c$
  - smaller  $\sigma^2$
  - larger sample size  $n$



## Exercise #8

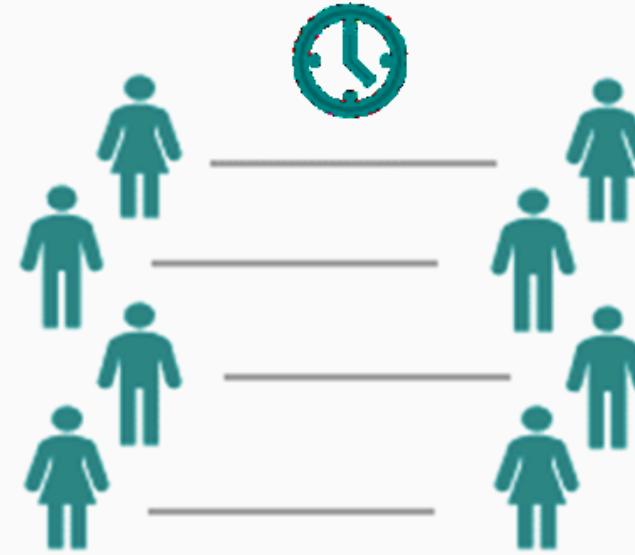
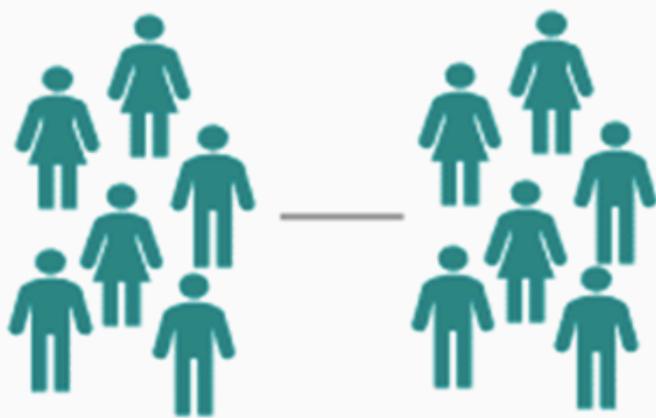
- ? If one'd like to increase the power of their study, which factor(s) could modify?
- a) the level of significance  $\alpha$
  - b) the difference  $\mu_i - \mu_c$
  - c) the samples'  $\sigma^2$
  - d) the samples' size  $n$

01:00

## Exercise #8 -- Solution

- ? If one'd like to increase the power of their study, which factor(s) could modify?
  - a) the level of significance  $\alpha$  
  - b) the difference  $\mu_i - \mu_c$
  - c) the samples'  $\sigma^2$
  - d) the samples' size  $n$  

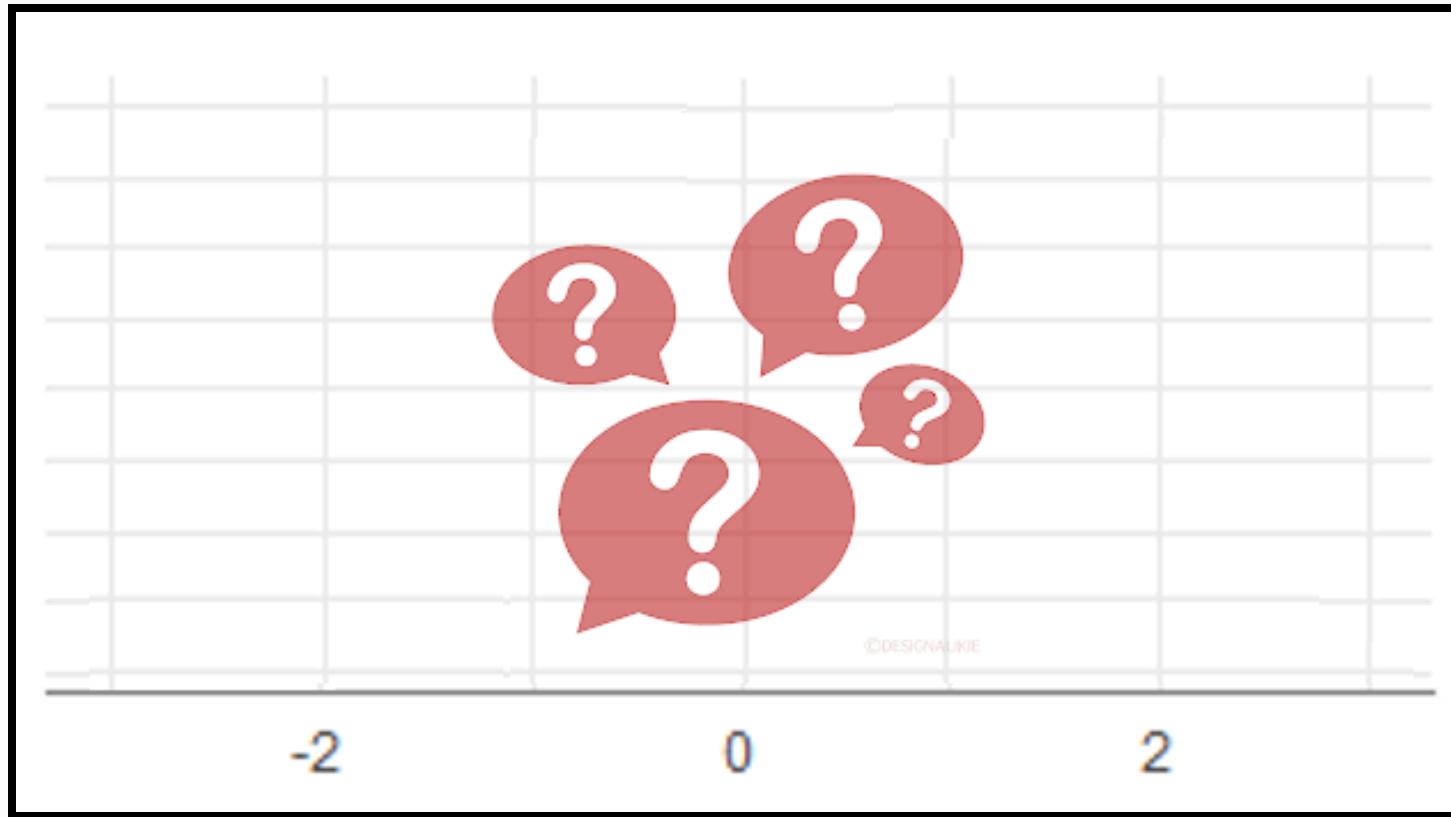
# Independent and paired samples



# Independent and paired samples



# Non-parametric tests

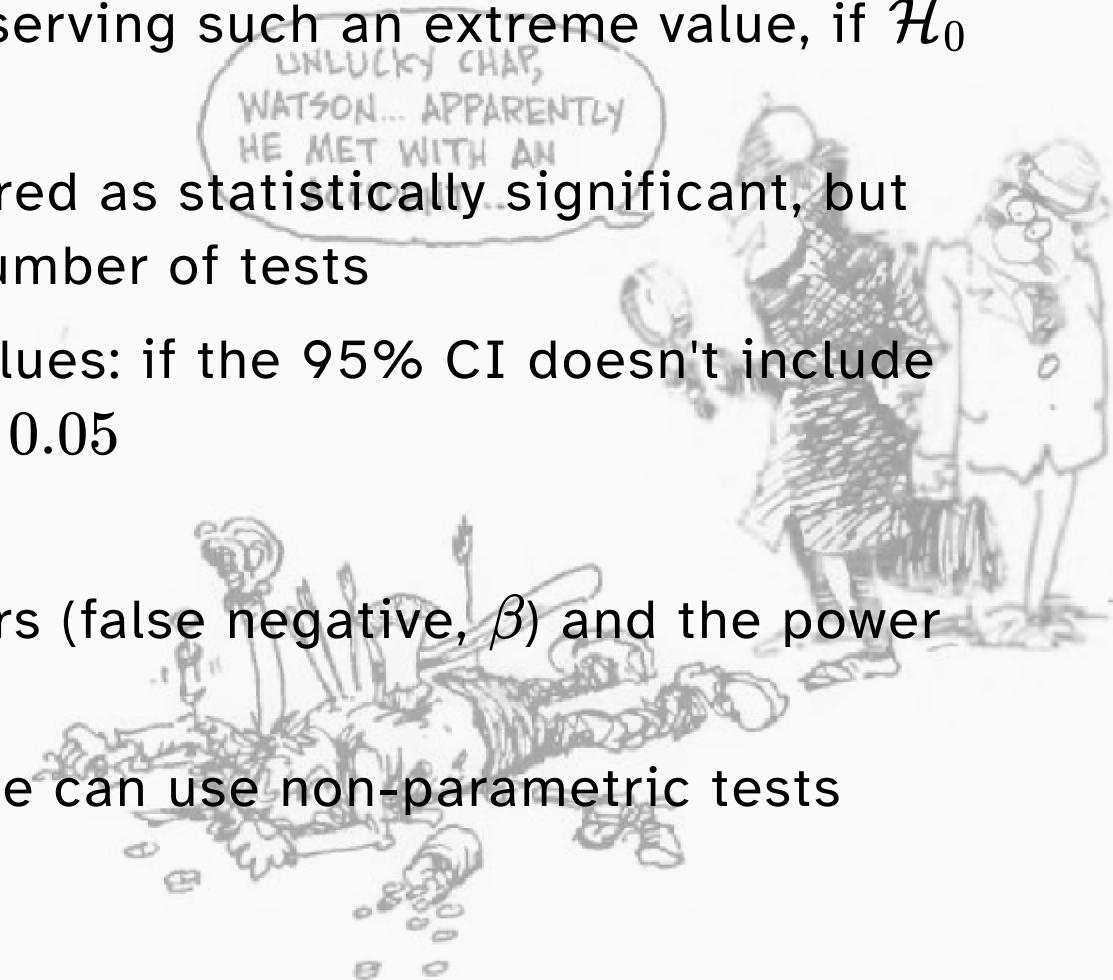


# Non-parametric tests

Sample	Data type	$\mathcal{H}_0$	Non-parametric test
Independent	Numerical	$\mu_1 = \mu_2$	Mann-Whitney's test
Paired	Numerical	$\mu_1 = \mu_2$	Wilcoxon's test
Independent	Categorical	$\pi_1 = \pi_2$	Fisher's test
Paired	Categorical	$\pi_1 = \pi_2$	McNemar's test

# Summary

- The P-value measures the discrepancy between the data and the null hypothesis  $\mathcal{H}_0$ , and correspond to the probability of observing such an extreme value, if  $\mathcal{H}_0$  was true
- Historically,  $P < 0.05$  or  $< 0.01$  are considered as statistically significant, but these  $\alpha$  levels should be corrected by the number of tests
- There is a relationship between CI and P-values: if the 95% CI doesn't include the null hypothesis, one can reject it at  $\alpha = 0.05$
- Type I errors (false positive) depend on  $\alpha$
- There is a relationship between type II errors (false negative,  $\beta$ ) and the power of a study
- When data have non-Normal distribution, one can use non-parametric tests



# Wrap up



# The PARACHUTE trial

## RESEARCH

### Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial

Robert W Yeh,<sup>1</sup> Linda R Valsdottir,<sup>1</sup> Michael W Yeh,<sup>2</sup> Changyu Shen,<sup>1</sup> Daniel B Kramer,<sup>1</sup> Jordan B Strom,<sup>1</sup> Eric A Secemsky,<sup>1</sup> Joanne L Healy,<sup>1</sup> Robert M Domeier,<sup>3</sup> Dhruv S Kazi,<sup>1</sup> Brahmajee K Nallamothu<sup>4</sup> On behalf of the PARACHUTE Investigators

#### WHAT IS ALREADY KNOWN ON THIS TOPIC

Parachutes are routinely used to prevent death or major traumatic injury among individuals jumping from aircraft, but their efficacy is based primarily on biological plausibility and expert opinion  
No randomized controlled trials of parachute use have yet been attempted, presumably owing to a lack of equipoise

#### WHAT THIS STUDY ADDS

This randomized trial of parachute use found no reduction in death or major injury compared with individuals jumping from aircraft with an empty backpack  
Lack of enrolment of individuals at high risk could have influenced the results of the trial

# The PARACHUTE trial



# Closing remarks

“ *To consult the statistician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of.* ”

R. Fisher

# Thank you

