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

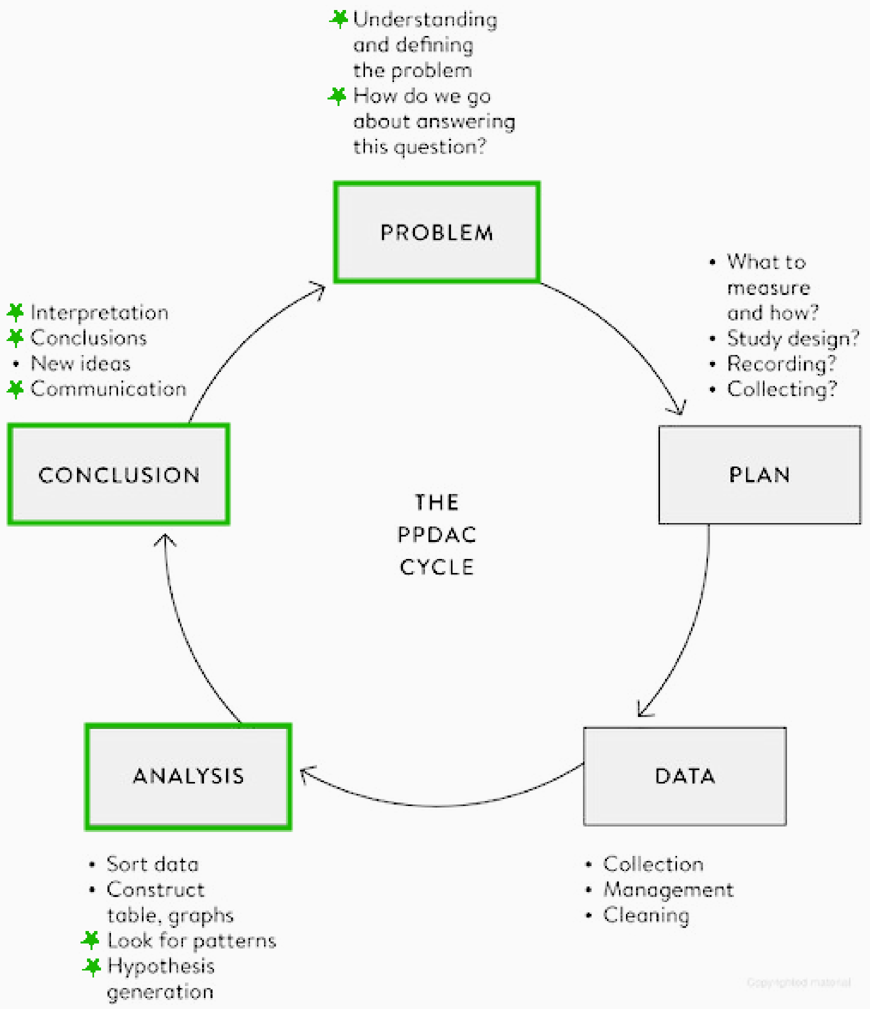
- Multiple phenomena and statistical distributions are normally distributed, and the Normal distribution describes both the probability of an observation and its proportion in the population

- 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



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.* ”

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

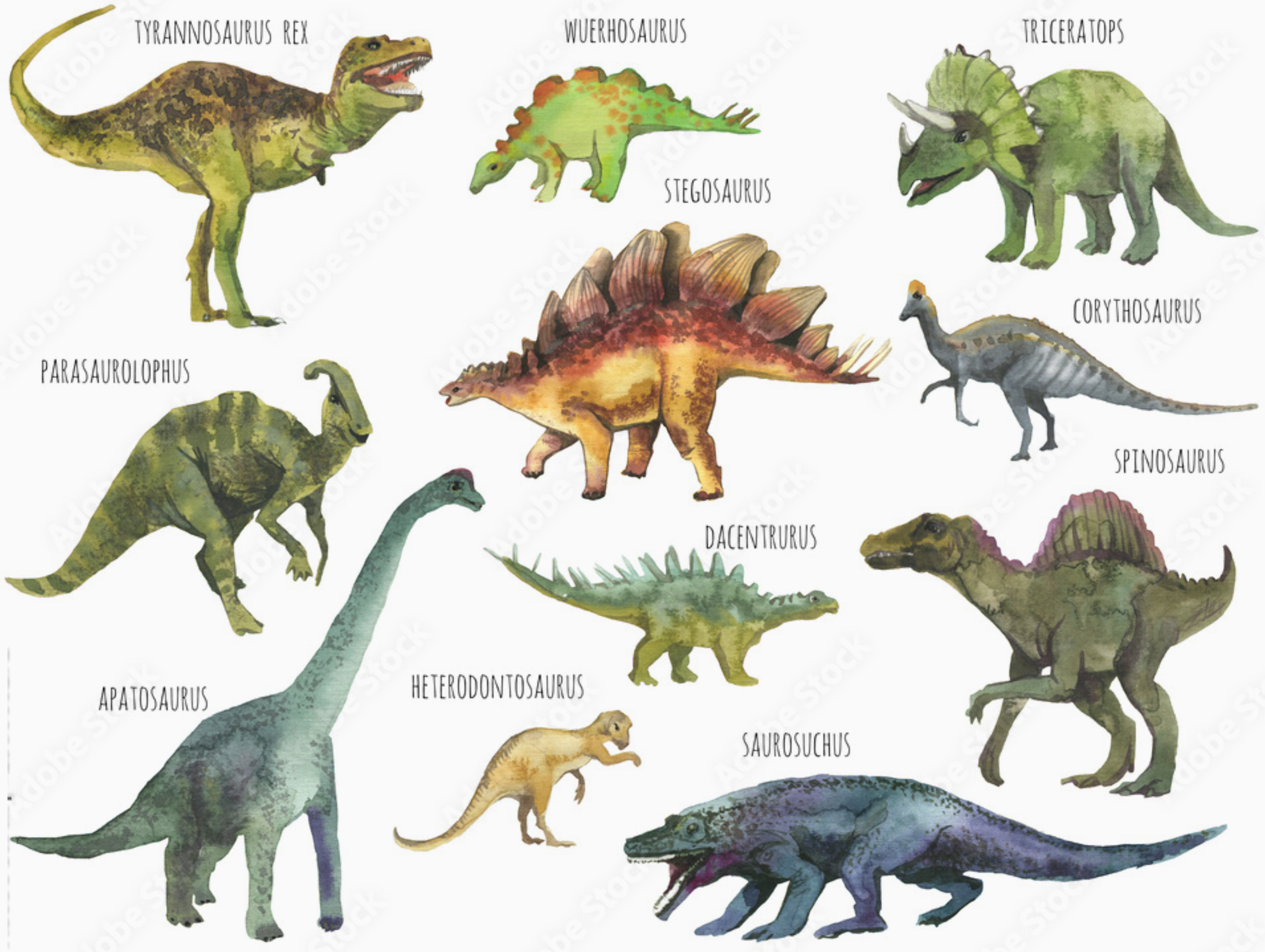
Making hypotheses

- 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,[1*] Sofia M. Sinitsa,[2] Danielle Dhouailly,[3] Yuri L. Bolotsky,[4] Alexander V. Sizov,[5] Maria E. McNamara,[6,7] Michael J. Benton,[7] Paul Spagna[1]

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 a **as effective** as standard care alone

Tomazini, B.M., *et al.*, Effect of dexamethasone on days alive and ventilator-free in patients with moderate or severe acute respiratory distress syndrome and COVID-19: the CoDEX randomized clinical trial.", JAMA, 2020, doi:10.1001/jama.2020.17021

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

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) $\mu_{\mathrm{i}} - \mu_{\mathrm{c}} = 0$
b) $\mu_{\mathrm{i}} - \mu_{\mathrm{c}} \neq 0$
c) $\bar{x}_{\mathrm{i}} - \bar{x}_{\mathrm{c}} = 0$
d) $\bar{x}_{\mathrm{i}} - \bar{x}_{\mathrm{c}} \neq 0$

Making hypotheses

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

$$n_{\mathrm{i}} = 151, \bar{x}_{\mathrm{i}} = 6.6, s_{\mathrm{i}} = 10.0$$
$$n_{\mathrm{c}} = 148, \bar{x}_{\mathrm{c}} = 4.0, s_{\mathrm{c}} = 8.7$$

$$\mu_{\mathrm{i}} - \mu_{\mathrm{c}} = 0$$
$$\rightarrow \text{Null hypothesis } (\mathcal{H}_0)$$

Making hypotheses

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

$$n_{\mathrm{i}} = 151, \bar{x}_{\mathrm{i}} = 6.6, s_{\mathrm{i}} = 10.0$$
$$n_{\mathrm{c}} = 148, \bar{x}_{\mathrm{c}} = 4.0, s_{\mathrm{c}} = 8.7$$

$$\mu_{\mathrm{i}} - \mu_{\mathrm{c}} = 0$$
$$\rightarrow \text{Null hypothesis } (\mathcal{H}_0)$$

$$\mu_{\mathrm{i}} - \mu_{\mathrm{c}} \neq 0$$
$$\rightarrow \text{Alternative hypothesis } (\mathcal{H}_1/\mathcal{H}_A)$$

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

$$n_{\mathrm{i}} = 151, \quad \bar{x}_{\mathrm{i}} = 6.6, \quad s_{\mathrm{i}} = 10.0$$
$$n_{\mathrm{c}} = 148, \quad \bar{x}_{\mathrm{c}} = 4.0, \quad s_{\mathrm{c}} = 8.7$$

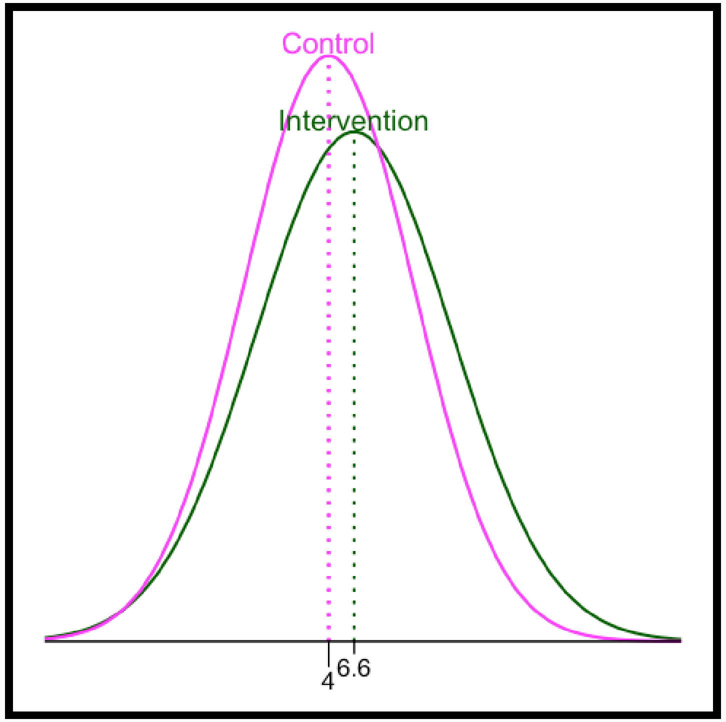
Tomazini B.M., *et al.*, *Effect of Dexamethasone on Days Alive and Ventilator-Free in Patients With Moderate or Severe Acute Respiratory Distress Syndrome and COVID-19*, JAMA, 2020

Testing hypotheses

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

$$n_{\mathrm{i}} = 151, \bar{x}_{\mathrm{i}} = 6.6, s_{\mathrm{i}} = 10.0$$
$$n_{\mathrm{c}} = 148, \bar{x}_{\mathrm{c}} = 4.0, s_{\mathrm{c}} = 8.7$$

$$\mu_{\mathrm{i}} - \mu_{\mathrm{c}} = 0 \quad \leftarrow$$
$$\bar{x}_i - \bar{x}_c = 6.6 - 4.0 = 2.6$$

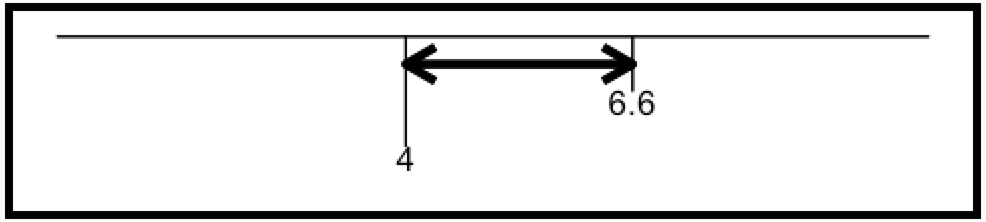


Testing hypotheses

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

$$n_{\mathrm{i}} = 151, \bar{x}_{\mathrm{i}} = 6.6, s_{\mathrm{i}} = 10.0$$
$$n_{\mathrm{c}} = 148, \bar{x}_{\mathrm{c}} = 4.0, s_{\mathrm{c}} = 8.7$$

$$\mu_{\mathrm{i}} - \mu_{\mathrm{c}} = 0 \qquad \leftarrow$$
$$\bar{x}_i - \bar{x}_c = 6.6 - 4.0 = 2.6$$



❓ What is the probability of observing a difference of 2.6 days if $\mu_{\mathrm{d}} - \mu_{\mathrm{m}} = 0$?

Let's take a step back

1. The Normal distribution is defined by its mean and standard deviation and corresponds to a probability distribution

   $\rightarrow$ 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
   $\rightarrow$ Area $Z \equiv$ probability $\mathcal{P}$
2. Sampling distributions (including the difference of means) show a Normal distribution (CLT)

For the difference of means:

$\mathcal{N} = (\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$ with

$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \rightarrow$ standard error

Testing hypotheses

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

$$n_{\mathrm{i}} = 151, \bar{x}_{\mathrm{i}} = 6.6, s_{\mathrm{i}} = 10.0$$
$$n_{\mathrm{c}} = 148, \bar{x}_{\mathrm{c}} = 4.0, s_{\mathrm{c}} = 8.7$$

$$\mu_{\mathrm{c}} - \mu_{\mathrm{i}} = 0 \qquad \leftarrow$$
$$\bar{x}_{\mathrm{c}} - \bar{x}_{\mathrm{i}} = 6.6 - 4.0 = 2.6$$

$$\mathcal{N} = \left(\mu_{\mathrm{c}} - \mu_{\mathrm{i}}, \frac{\sigma_c^2}{n_c} + \frac{\sigma_i^2}{n_i}\right) \rightarrow \mu_{\mathrm{c}} - \mu_{\mathrm{i}} = 0 \text{ and } \hat{\mathrm{SE}} = \sqrt{\frac{s_{\mathrm{c}}^2}{n_{\mathrm{c}}} + \frac{s_{\mathrm{i}}^2}{n_{\mathrm{i}}}} = 1.08$$

Testing hypotheses

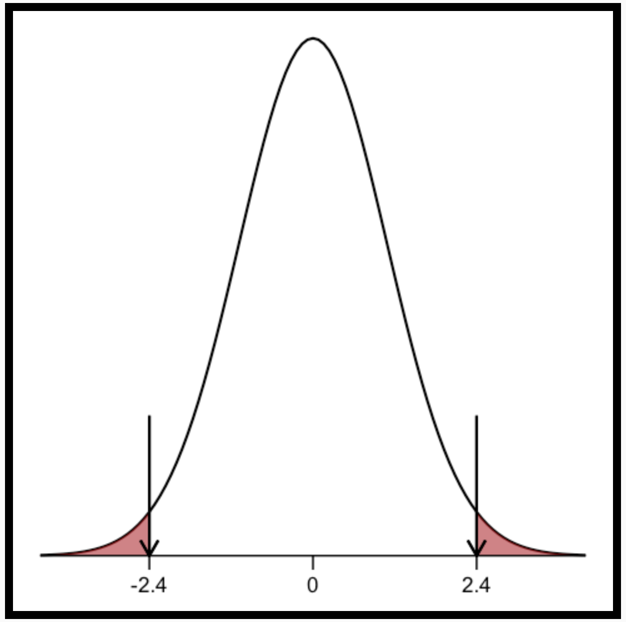
📌 Dexamethasone plus standard care is **as effective as** standard care

$$\mu_c - \mu_i = 0$$
$$\hat{\text{SE}} = 1.08$$

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

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

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

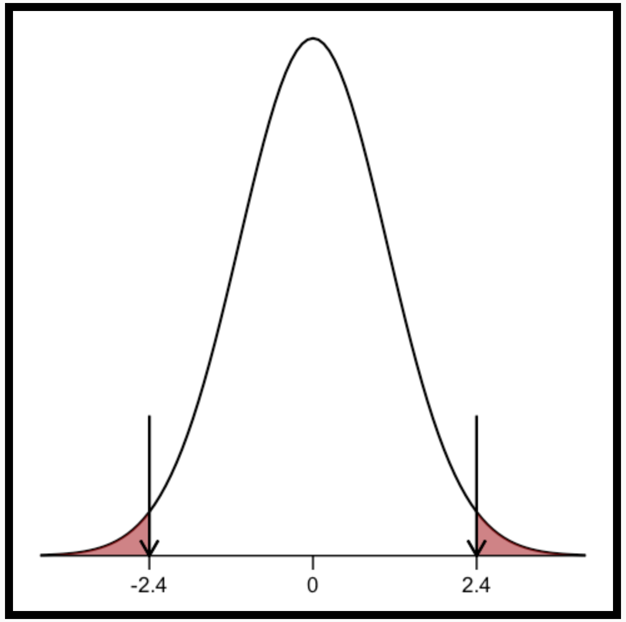


Testing hypotheses

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

$$\mu_{\mathrm{c}} - \mu_{\mathrm{i}} = 0$$
$$\hat{\mathrm{SE}} = 1.08$$

$$\bar{x}_{\mathrm{c}} - \bar{x}_{\mathrm{i}} = 6.6 - 4.0 = 2.6$$



❓ What is the probability of observing a difference of 2.6 days if $\mu_{\mathrm{c}} - \mu_{\mathrm{i}} = 0$?

$$z = \frac{(\bar{x}_{\mathrm{c}} - \bar{x}_{\mathrm{i}}) - (\mu_{\mathrm{c}} - \mu_{\mathrm{i}})}{\hat{SE}} = \frac{2.6 - 0}{1.08} = 2.4 \quad \rightarrow \quad \mathcal{P} = 2 \times 0.0082 = 0.0164$$

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

$$
\begin{aligned}
&\text{P-value} = 0.5 \rightarrow 50\% \rightarrow 1 \text{ sample out of } 2\\
&\text{P-value} = 0.1 \rightarrow 10\% \rightarrow 1 \text{ sample out of } 10\\
&\text{P-value} = 0.05 \rightarrow 5\% \rightarrow 1 \text{ sample out of } 20\\
&\text{P-value} = 0.01 \rightarrow 1\% \rightarrow 1 \text{ sample out of } 100\\
&\text{P-value} = 0.005 \rightarrow 0.5\% \rightarrow 1 \text{ sample out of } 200
\end{aligned}
$$

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 α, 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

$$\mathcal{H}_0 : \mu_{\mathrm{c}} - \mu_{\mathrm{i}} = 0$$

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$^{(*)}$ 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

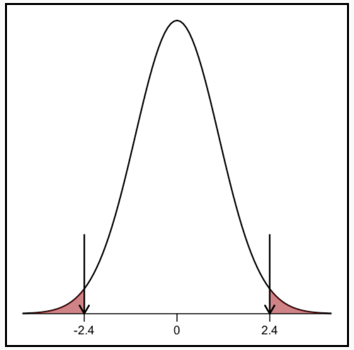
1. Define a null hypothesis $(\mathcal{H}_0)$
2. Choose a test statistic that estimates something that, if extreme enough, would lead one to doubt $\mathcal{H}_0$
3. Generate the sampling distribution of the chosen test statistic, assuming $\mathcal{H}_0$ to be true

$$\mathcal{N} = (\mu_\mathrm{c} - \mu_\mathrm{i}, \mathrm{SE}), \text{ with } \mu_\mathrm{c} - \mu_\mathrm{i} = 0 \text{ and } \hat{\mathrm{SE}} = \sqrt{\frac{s_\mathrm{c}^2}{n_\mathrm{c}} + \frac{s_\mathrm{i}^2}{n_\mathrm{i}}}$$

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 α

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

Exercise #4

? In a randomised control trial, the P-value for one of the outcomes is 0.48. With an α level of 5%, are there statistically significant differences in the outcome between the two arms of the trial?

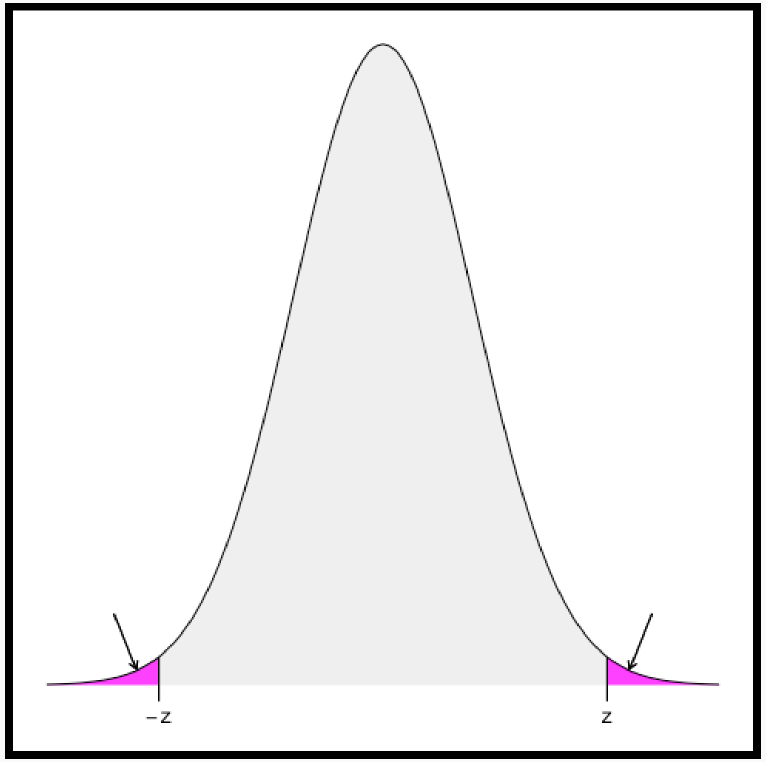
a) Yes, because the P value is lower than the α level
b) Yes, because the P value is greater than the α level
c) No, because the P value is lower than the α level
d) No, because the P value is greater than the α level

One- and two-tailed tests

🎯 $\mathcal{H}_1$: $\mu_\mathrm{i} - \mu_\mathrm{c} \neq 0$

$\mathcal{H}_0$: $\mu_\mathrm{i} - \mu_\mathrm{c} = 0$

$\rightarrow$ two-tailed test



One- and two-tailed tests

🎯 $\mathcal{H}_1$: $\mu_\mathrm{i} - \mu_\mathrm{c} \neq 0$

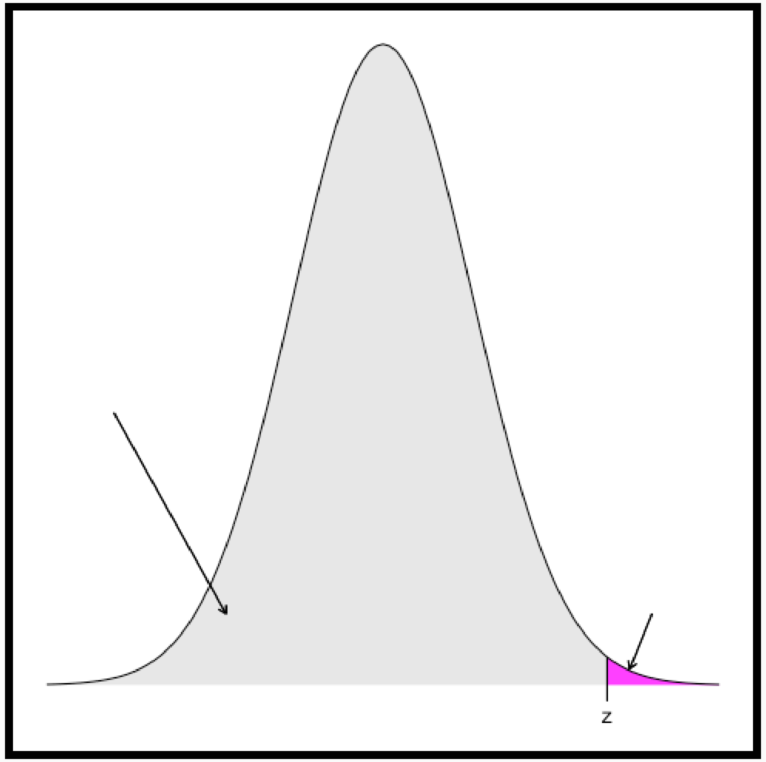
$\mathcal{H}_0$: $\mu_\mathrm{i} - \mu_\mathrm{c} = 0$

$\rightarrow$ two-tailed test

$\mathcal{H}_1$: $\mu_\mathrm{i} - \mu_\mathrm{c} < 0$

$\mathcal{H}_0$: $\mu_\mathrm{i} - \mu_\mathrm{c} \geq 0$

$\rightarrow$ one-tailed test



One- and two-tailed tests

🎯 $\mathcal{H}_1$: $\mu_\mathrm{i} - \mu_\mathrm{c} \neq 0$

$\mathcal{H}_0$: $\mu_\mathrm{i} - \mu_\mathrm{c} = 0$

$\rightarrow$ two-tailed test

$\mathcal{H}_1$: $\mu_\mathrm{i} - \mu_\mathrm{c} < 0$

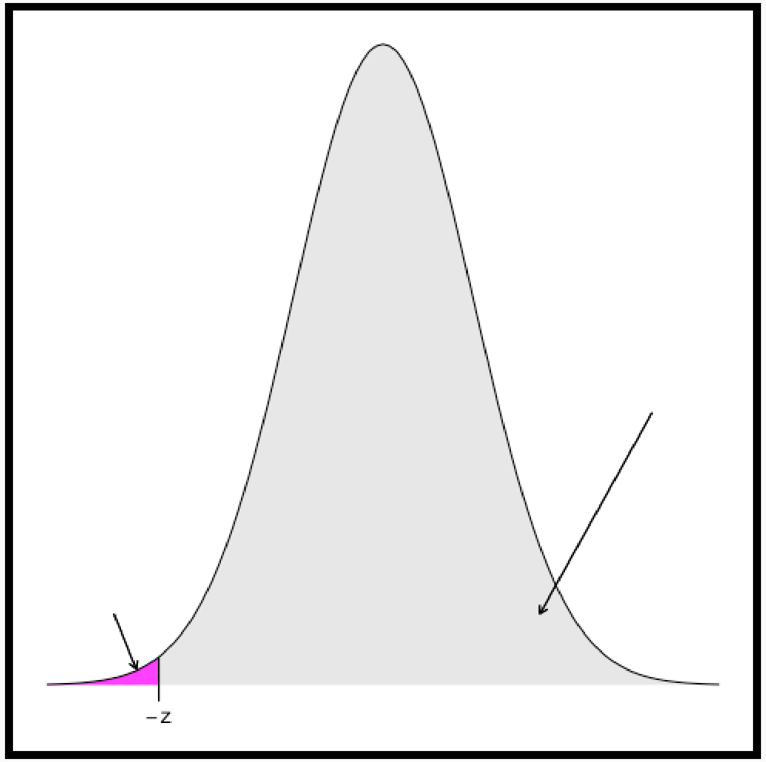
$\mathcal{H}_0$: $\mu_\mathrm{i} - \mu_\mathrm{c} \geq 0$

or

$\mathcal{H}_1$: $\mu_\mathrm{i} - \mu_\mathrm{c} > 0$

$\mathcal{H}_0$: $\mu_\mathrm{i} - \mu_\mathrm{c} \leq 0$

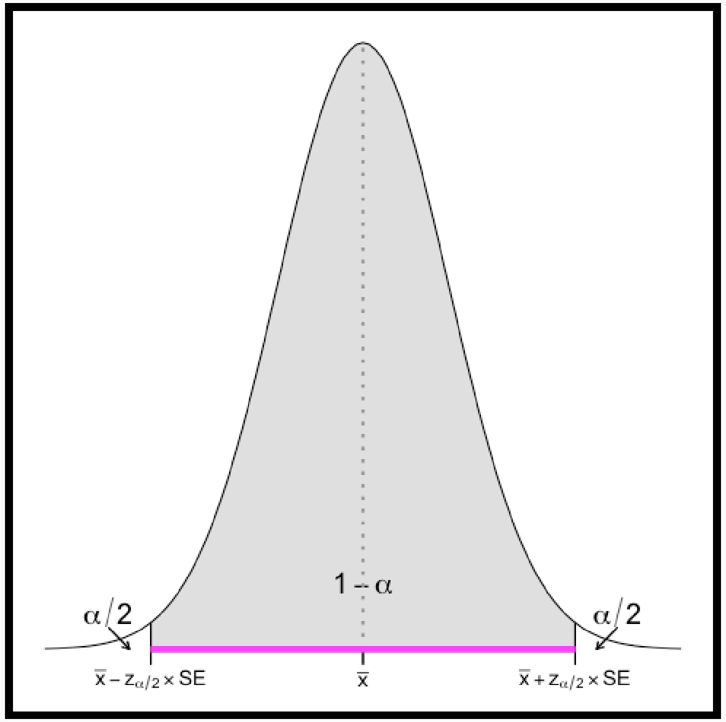
$\rightarrow$ one-tailed test



Hypothesis testing & confidence intervals

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

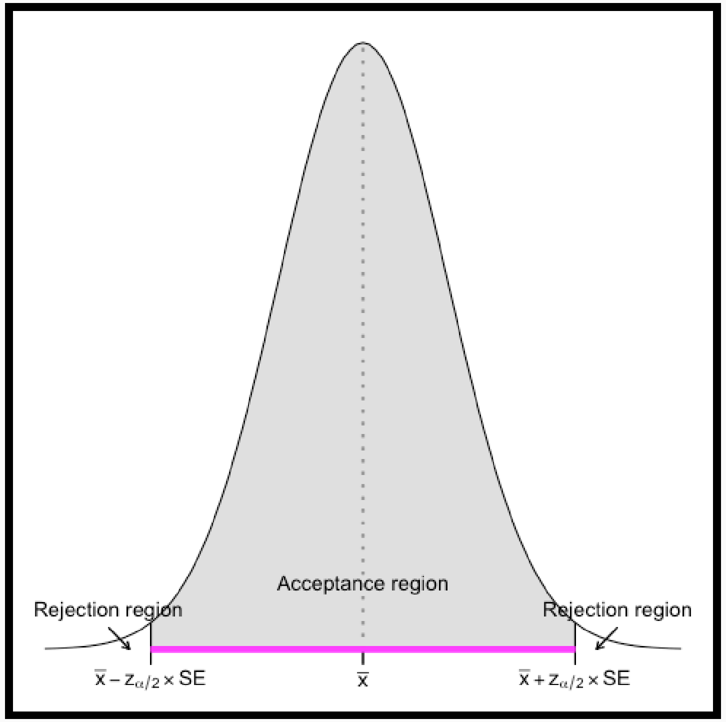
Confidence Level	α	$\alpha/2$	$z_{\alpha/2}$
95%	5%	2.5%	1.96



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, $\text{P-value} < 0.05$ 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.

$$n_\mathrm{i} = 180, \quad m_\mathrm{i} = 13, \quad p_\mathrm{i} = \frac{m_\mathrm{i}}{n_\mathrm{i}} = \frac{13}{180} = 0.07$$
$$n_\mathrm{c} = 183, \quad m_\mathrm{c} = 40, \quad p_\mathrm{c} = \frac{m_\mathrm{c}}{n_\mathrm{c}} = \frac{40}{183} = 0.22$$

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

Connor, E.M. *et al.*, *Reduction of Maternal-Infant Transmission of Human Immunodeficiency Virus Type 1 with Zidovudine Treatment*, NEJM, 1994

Pearson's χ^2 test

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

$$n_\mathrm{i} = 180, \quad m_\mathrm{i} = 13, \quad p_\mathrm{i} = \frac{m_\mathrm{i}}{n_\mathrm{i}} = \frac{13}{180} = 0.07$$
$$n_\mathrm{c} = 183, \quad m_\mathrm{c} = 40, \quad p_\mathrm{c} = \frac{m_\mathrm{c}}{n_\mathrm{c}} = \frac{40}{183} = 0.22$$

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

Pearson's χ^2 test

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

$$n_{\text{i}} = 180, \quad m_{\text{i}} = 13, \quad p_{\text{i}} = \frac{m_{\text{i}}}{n_{\text{i}}} = \frac{13}{180} = 0.07$$
$$n_{\text{c}} = 183, \quad m_{\text{c}} = 40, \quad p_{\text{c}} = \frac{m_{\text{c}}}{n_{\text{c}}} = \frac{40}{183} = 0.22$$

1. Define a null hypothesis $(\mathcal{H}_0)$
   Zidovudine is **as effective as** placebo to reduce mother-infant HIV transmission
   $\rightarrow \quad \mathcal{H}_0 : \pi_{\text{i}} - \pi_{\text{c}} = 0$

Pearson's χ^2 test

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

$$n_{\mathrm{i}} = 180, \quad m_{\mathrm{i}} = 13, \quad p_{\mathrm{i}} = \frac{m_{\mathrm{i}}}{n_{\mathrm{i}}} = \frac{13}{180} = 0.07$$
$$n_{\mathrm{c}} = 183, \quad m_{\mathrm{c}} = 40, \quad p_{\mathrm{c}} = \frac{m_{\mathrm{c}}}{n_{\mathrm{c}}} = \frac{40}{183} = 0.22$$

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

   $\rightarrow$ Pearson's χ^2 test for categorical data

Pearson's χ^2 test

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

$$
\begin{aligned}
n_\mathrm{i} = 180, \quad m_\mathrm{i} = 13, \quad p_\mathrm{i} = \tfrac{m_\mathrm{i}}{n_\mathrm{i}} = \tfrac{13}{180} = 0.07 \\
n_\mathrm{c} = 183, \quad m_\mathrm{c} = 40, \quad p_\mathrm{c} = \tfrac{m_\mathrm{c}}{n_\mathrm{c}} = \tfrac{40}{183} = 0.22
\end{aligned}
$$

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

Pearson's χ^2 test

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

$$n_\mathrm{i} = 180, \quad m_\mathrm{i} = 13, \quad p_\mathrm{i} = \frac{m_\mathrm{i}}{n_\mathrm{i}} = \frac{13}{180} = 0.07$$
$$n_\mathrm{c} = 183, \quad m_\mathrm{c} = 40, \quad p_\mathrm{c} = \frac{m_\mathrm{c}}{n_\mathrm{c}} = \frac{40}{183} = 0.22$$

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

? Let's fill ths contingency table

Treatment/Infected	Yes	No	Total
Zidovudine			
Placebo			
Total			

Pearson's χ^2 test

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

$$n_\mathrm{i} = 180, \quad m_\mathrm{i} = 13, \quad p_\mathrm{i} = \frac{m_\mathrm{i}}{n_\mathrm{i}} = \frac{13}{180} = 0.07$$
$$n_\mathrm{c} = 183, \quad m_\mathrm{c} = 40, \quad p_\mathrm{c} = \frac{m_\mathrm{c}}{n_\mathrm{c}} = \frac{40}{183} = 0.22$$

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 χ^2 test

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

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

Observed values

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

Pearson's χ^2 test

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

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

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 χ^2 test

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

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

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 χ^2 test

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

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

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 χ^2 test

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

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

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 χ^2 test

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

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

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

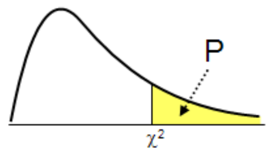
$$\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected} = \frac{(13-26.28)^2}{26.28} + \frac{(167-153.72)^2}{153.72} + \frac{(40-26.72)^2}{26.72} + \frac{(143-156.28)^2}{156.28} = 15.57$$

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

Pearson's χ^2 test

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

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



	P										
DF	0.995	0.975	0.20	0.10	0.05	0.025	0.02	0.01	0.005	0.002	0.001
1	0.0000393	0.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458

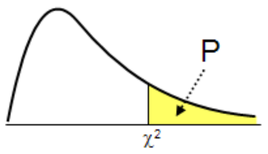
4. Check whether the observed statistic lies in the tails of this distribution, and calculate a probability (P-value) for this event

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

Pearson's χ^2 test

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

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



	P										
DF	0.995	0.975	0.20	0.10	0.05	0.025	0.02	0.01	0.005	0.002	0.001
1	0.0000393	0.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458

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

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

Pearson's χ^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 χ^2 test

📌 Does education level influences physical activity frequency?

Expected values

	No Exercise	Sporadic Exercise	Regular Exercise	Total
Primary education	$\frac{\Sigma\text{Row}_1 \times \Sigma\text{Column}_1}{\text{Total}}$	$\frac{\Sigma\text{Row}_1 \times \Sigma\text{Column}_2}{\text{Total}}$	$\frac{\Sigma\text{Row}_1 \times \Sigma\text{Column}_3}{\text{Total}}$	ΣRow_1
Secondary education	$\frac{\Sigma\text{Row}_2 \times \Sigma\text{Column}_1}{\text{Total}}$	...	...	ΣRow_2
Bachelor/Master	$\frac{\Sigma\text{Row}_3 \times \Sigma\text{Column}_1}{\text{Total}}$	...	...	ΣRow_3
Doctorate	$\frac{\Sigma\text{Row}_4 \times \Sigma\text{Column}_1}{\text{Total}}$	...	...	ΣRow_4
Total	ΣColumn_1	ΣColumn_2	ΣColumn_3	Total

❓ $\text{df} = \ ?$

Pearson's χ^2 test

📌 Does education level influences physical activity frequency?

Expected values

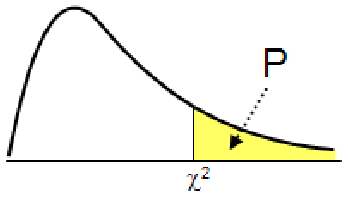
	No Exercise	Sporadic Exercise	Regular Exercise	Total
Primary education	$\frac{\Sigma\text{Row}_1 \times \Sigma\text{Column}_1}{\text{Total}}$	$\frac{\Sigma\text{Row}_1 \times \Sigma\text{Column}_2}{\text{Total}}$	$\frac{\Sigma\text{Row}_1 \times \Sigma\text{Column}_3}{\text{Total}}$	ΣRow_1
Secondary education	$\frac{\Sigma\text{Row}_2 \times \Sigma\text{Column}_1}{\text{Total}}$	...	...	ΣRow_2
Bachelor/Master	$\frac{\Sigma\text{Row}_3 \times \Sigma\text{Column}_1}{\text{Total}}$	...	...	ΣRow_3
Doctorate	$\frac{\Sigma\text{Row}_4 \times \Sigma\text{Column}_1}{\text{Total}}$	...	...	ΣRow_4
Total	ΣColumn_1	ΣColumn_2	ΣColumn_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.



	P										
DF	**0.995**	**0.975**	**0.20**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.002**	**0.001**
1	0.0000393	0.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458

Albano, L. *et al.*, *Alcohol consumption in a sample of Italian healthcare workers: A cross-sectional study*, Archives of Environmental & Occupational Health, 2020

Pearson's χ^2 test -- Yates' correction

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

$$\downarrow$$

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

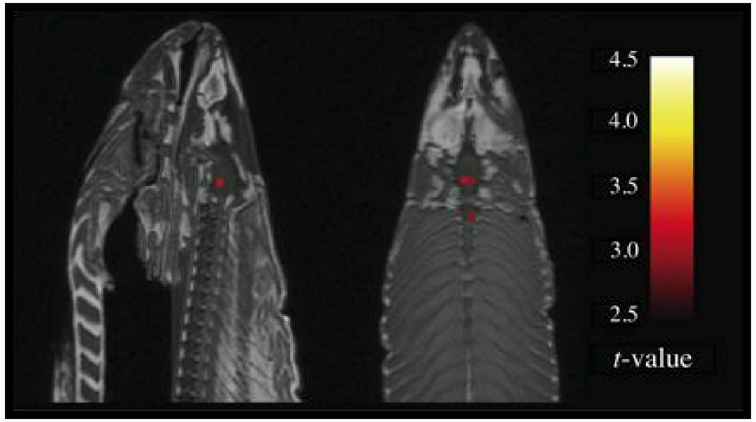
Multiple testing comparisons

📌 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 $\mathrm{P} < 0.001$.

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 $\mathrm{P} < 0.001$.

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



Bennett, C. M., Miller M.B., and Wolford G.L.,. *Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction.* Neuroimage 47.Suppl 1 (2009) doi:10.1016/S1053-8119(09)71202-9

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

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

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

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

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

Multiple testing correction

🎯 When one carries out multiple testing comparisons, they should ask for a smaller α

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

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

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

Multiple testing correction

🎯 When one carries out multiple testing comparisons, they should ask for a smaller α
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 α, find the largest k such that $\mathcal{P}(k) \leq \frac{k}{m}\alpha$
3. Reject the null hypothesis for $i = 1, \ldots, k$

Errors in decision making

🎯 $p < \alpha \rightarrow \text{reject } \mathcal{H}_0$
$p \geq \alpha \rightarrow \text{does not reject } \mathcal{H}_0$
$\alpha = 0.05 \rightarrow 5\%$ chance of rejecting $\mathcal{H}_0$ when is true

Errors in decision making

🎯 $p < \alpha \rightarrow \text{reject } \mathcal{H}_0$
$p \geq \alpha \rightarrow \text{does not reject } \mathcal{H}_0$
$\alpha = 0.05 \rightarrow 5\%$ chance of rejecting $\mathcal{H}_0$ when is true

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

Errors in decision making

🎯 $p < \alpha \to \text{reject } \mathcal{H}_0$
$p \geq \alpha \to \text{does not reject } \mathcal{H}_0$
$\alpha = 0.05 \to 5\%$ 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

🎯 $p < \alpha \rightarrow \text{reject } \mathcal{H}_0$
$p \geq \alpha \rightarrow \text{does not reject } \mathcal{H}_0$
$\alpha = 0.05 \rightarrow 5\%$ chance of rejecting $\mathcal{H}_0$ when is true

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

Exercise #7

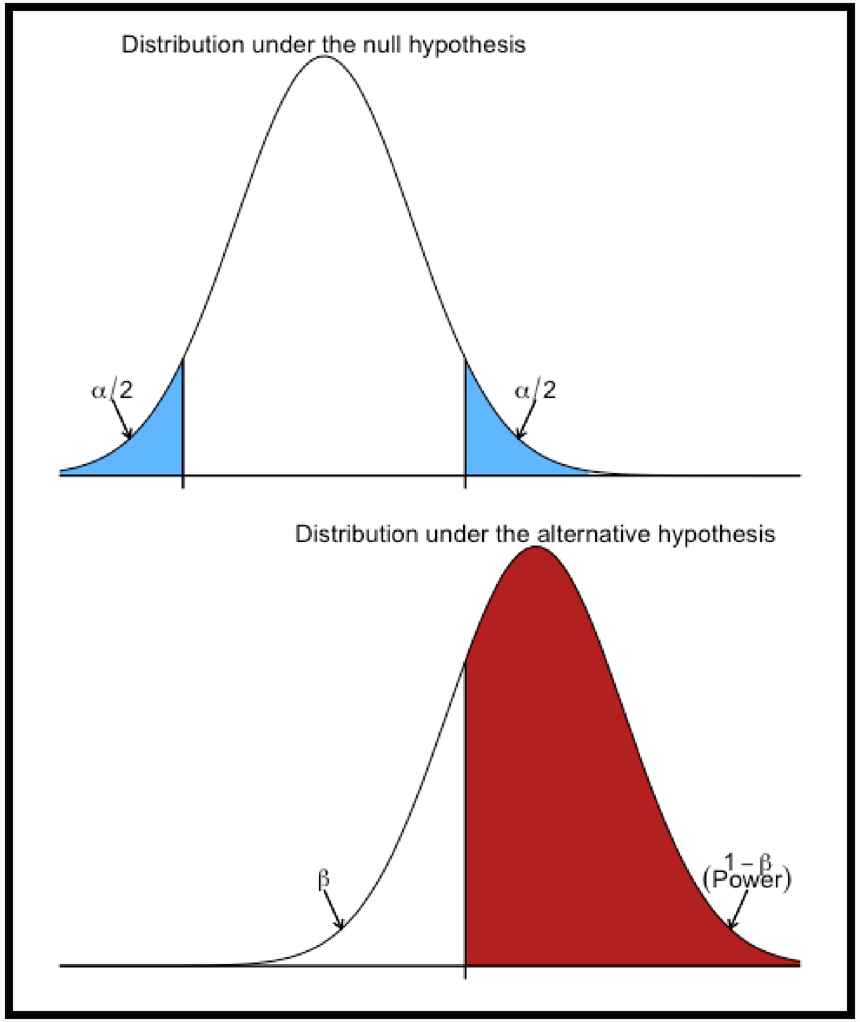
? There was a shepherd boy who repeatedly cried wolf when there was no wolf. Yet, each time, villagers went to help him. Then, the wolf arrived, but, when the boy cried wolf, no villager helped.

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

The power of a study

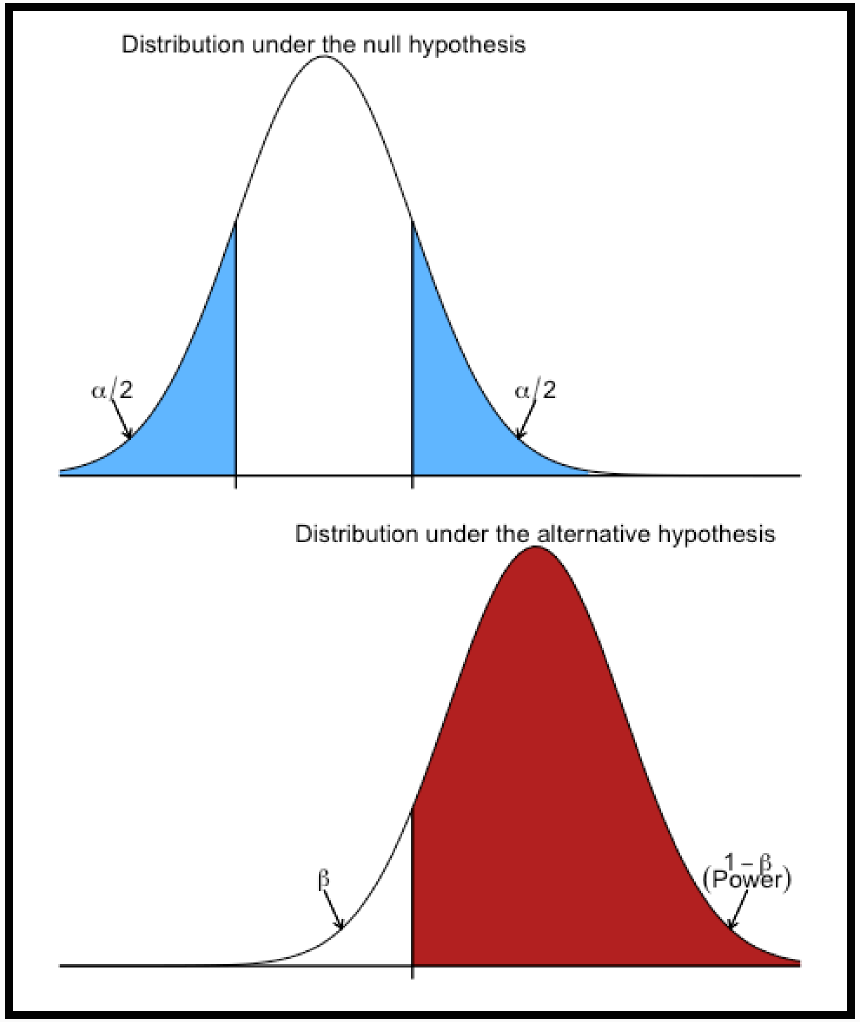
$\mathcal{H}_0$ is	Not rejected	Rejected
True		α
False	β	$1-\beta$ Statistical power



The power of a study

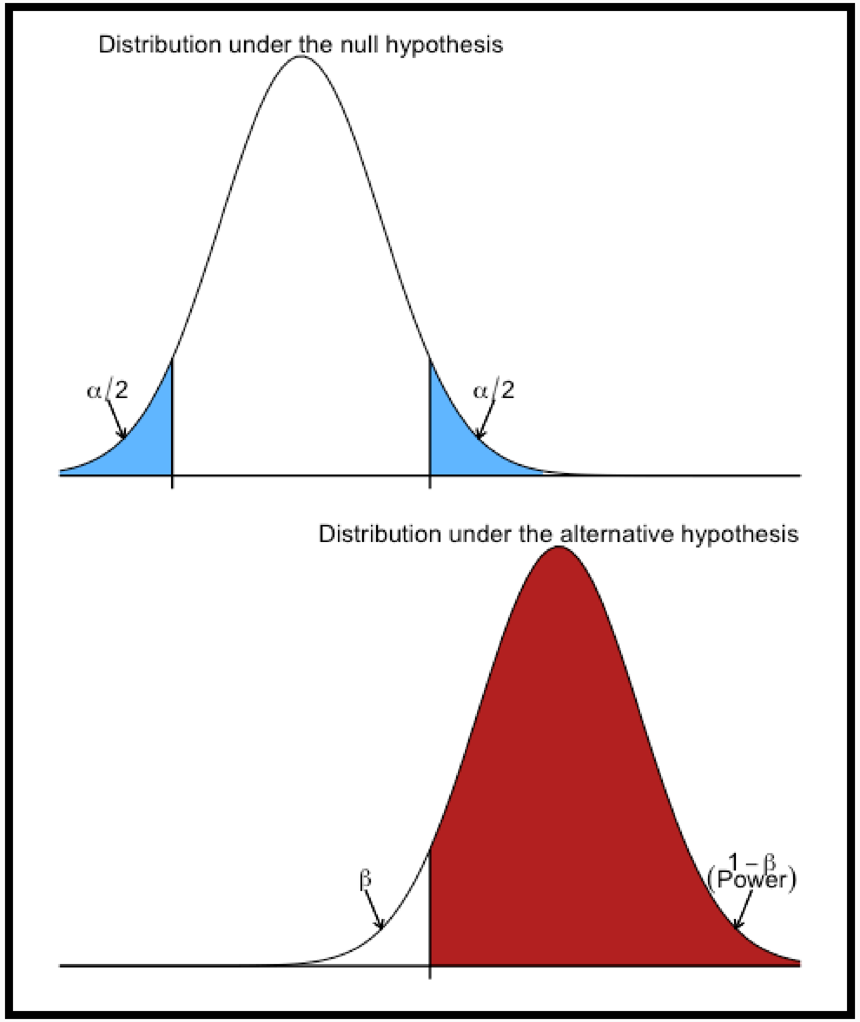
🎯 The power is increased by:

- larger α



The power of a study

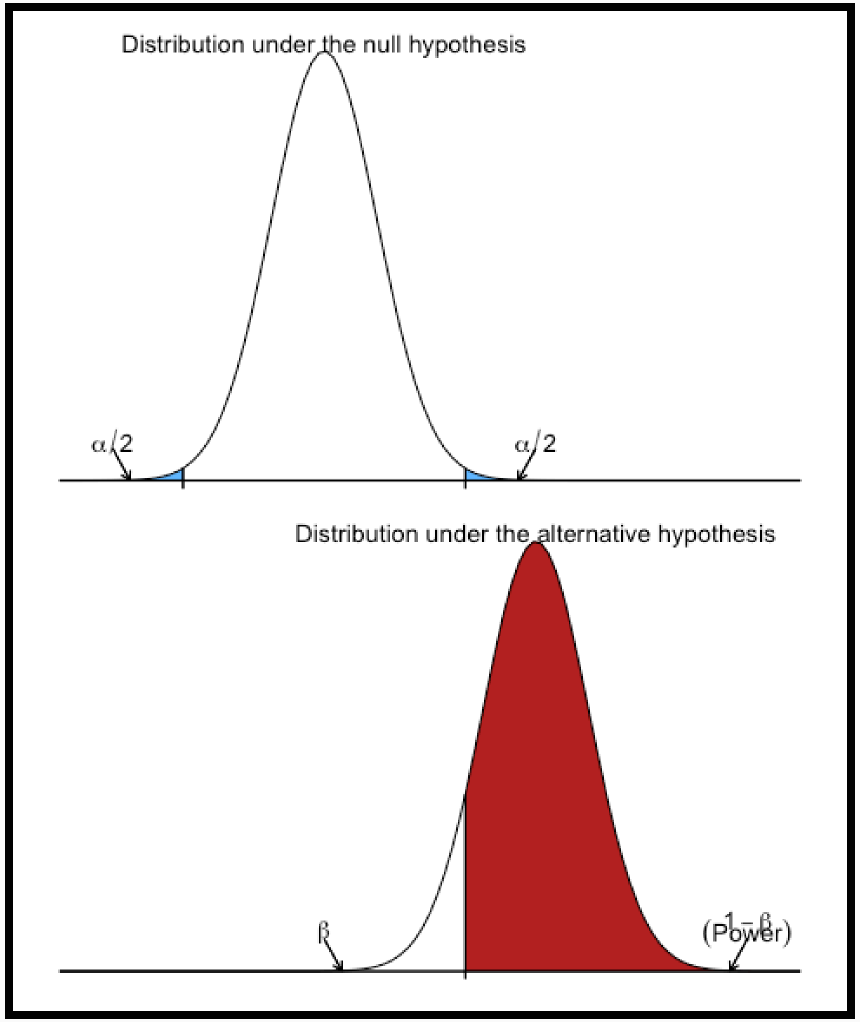
🎯 The power is increased by:
- larger α
- larger $\mu_i - \mu_c$



The power of a study

🎯 The power is increased by:

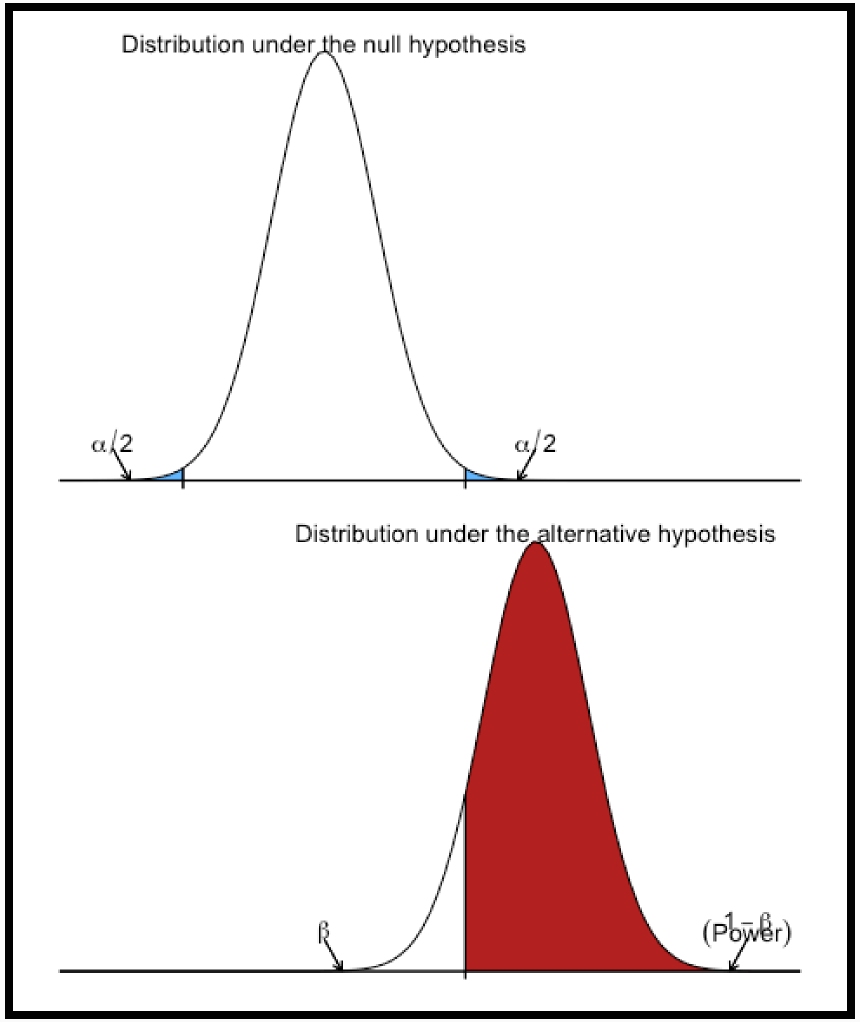
- larger α
- larger $\mu_i - \mu_c$
- smaller σ^2



The power of a study

🎯 The power is increased by:

- larger α
- larger $\mu_i - \mu_c$
- smaller σ^2
- larger sample size n

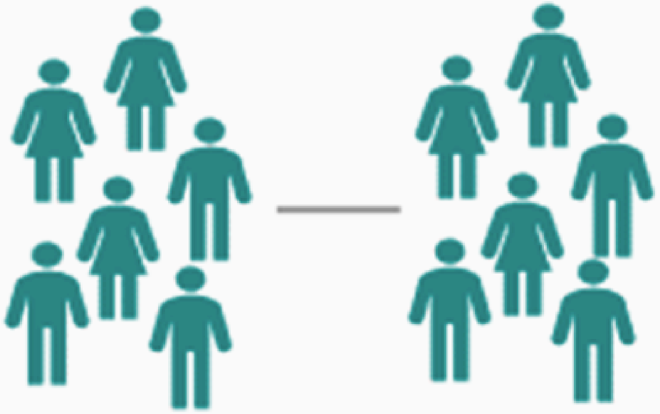


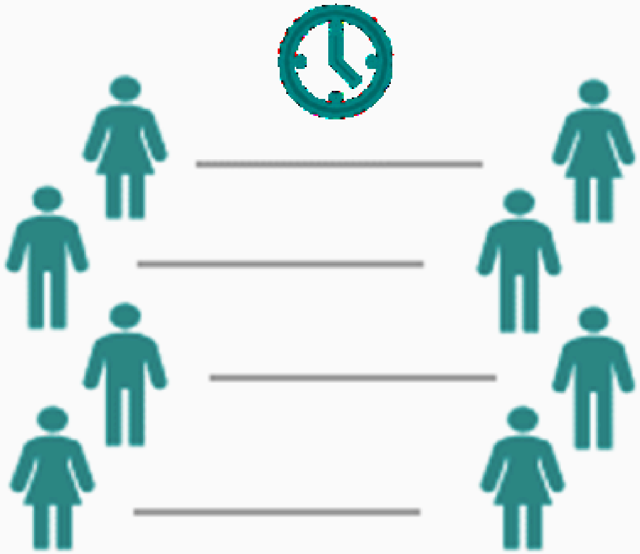
Exercise #8

? If one'd like to increase the power of their study, which factor(s) could modify?

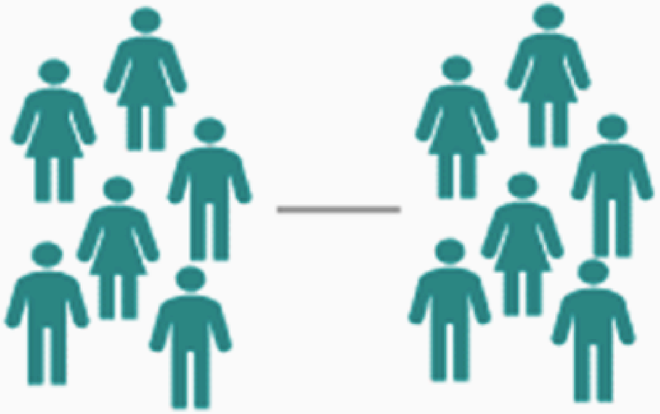
a) the level of significance α
b) the difference $\mu_i - \mu_c$
c) the samples' σ^2
d) the samples' size n

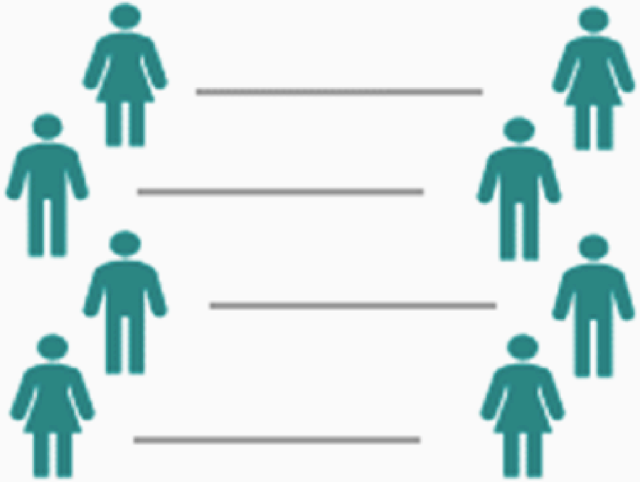
Independent and paired samples



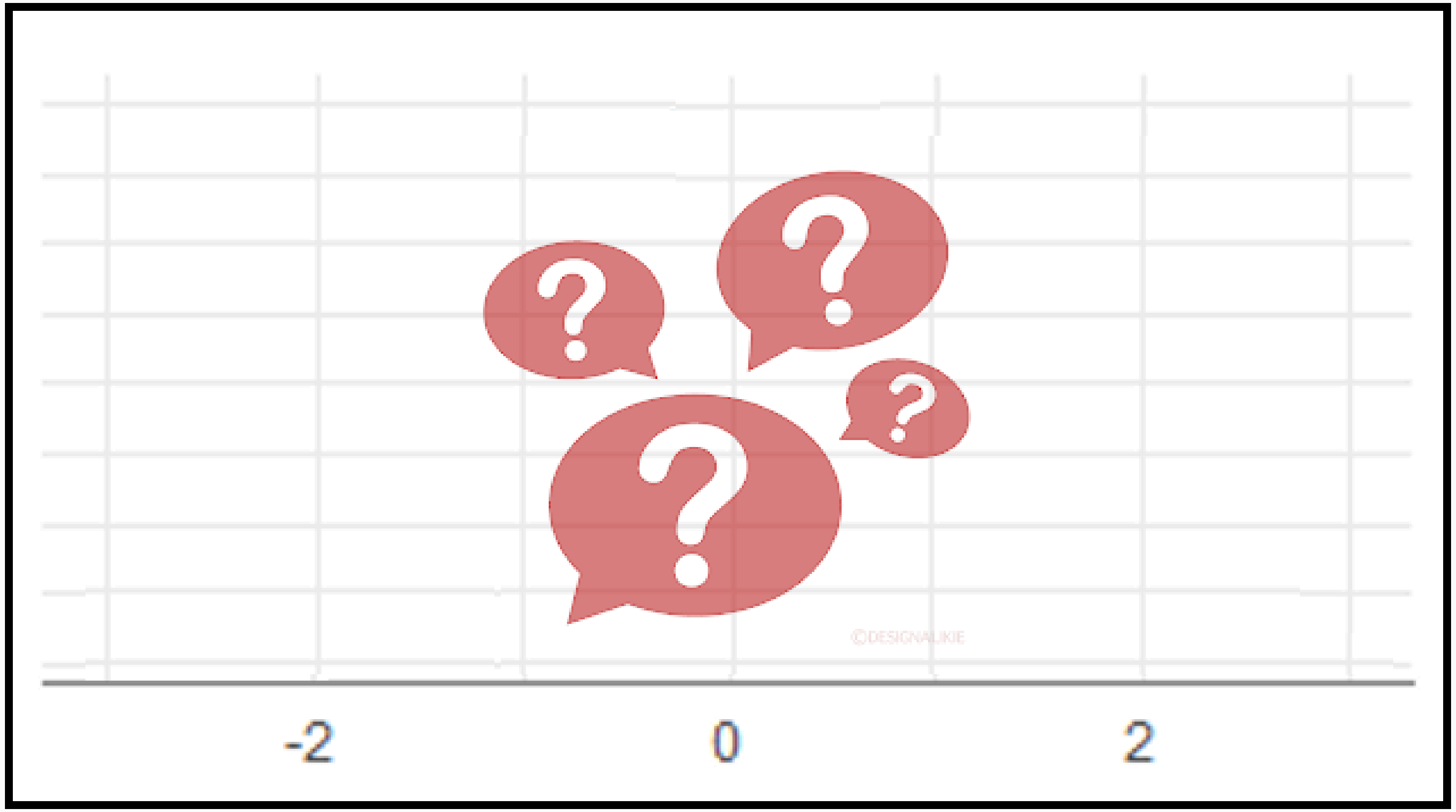


Independent and paired samples





Non-parametric tests

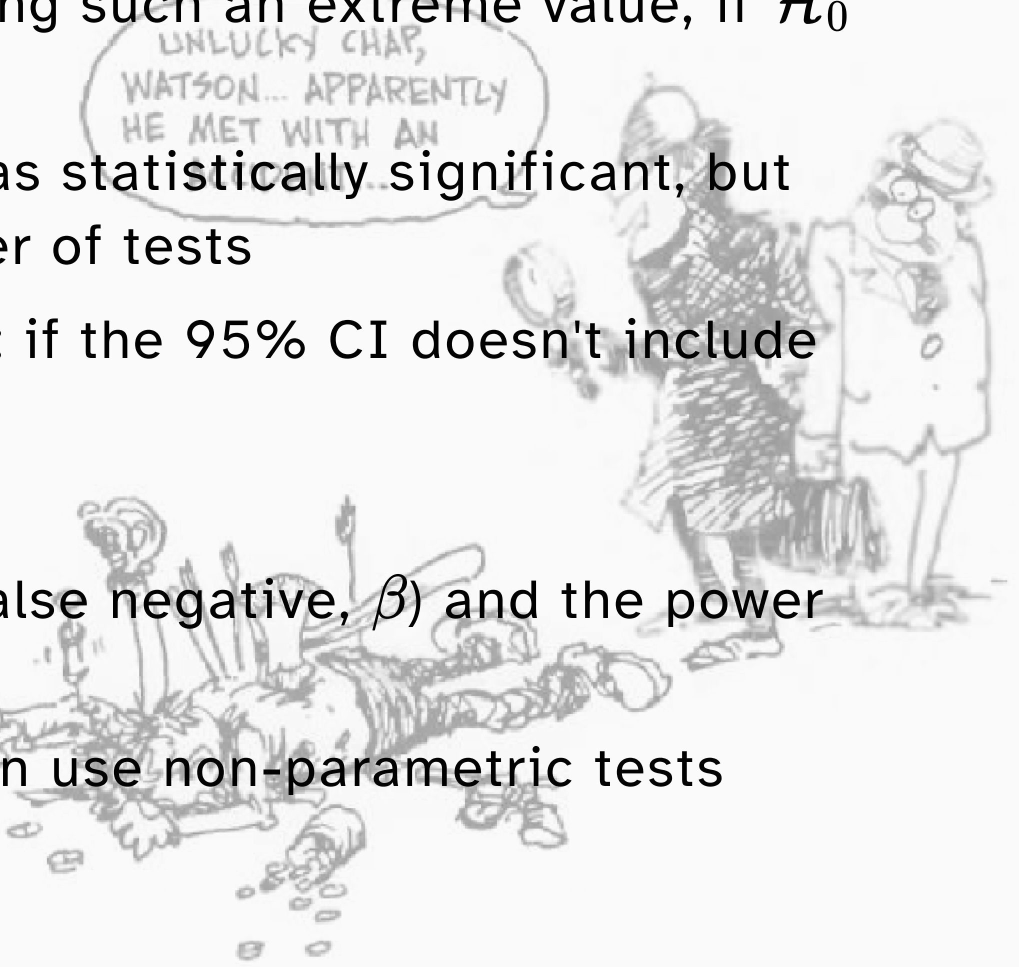


Non-parametric tests

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 α 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 α
- There is a relationship between type II errors (false negative, β) 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,[1] Linda R Valsdottir,[1] Michael W Yeh,[2] Changyu Shen,[1] Daniel B Kramer,[1] Jordan B Strom,[1] Eric A Secemsky,[1] Joanne L Healy,[1] Robert M Domeier,[3] Dhruv S Kazi,[1] Brahmajee K Nallamothu[4] 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