Statistical Reasoning

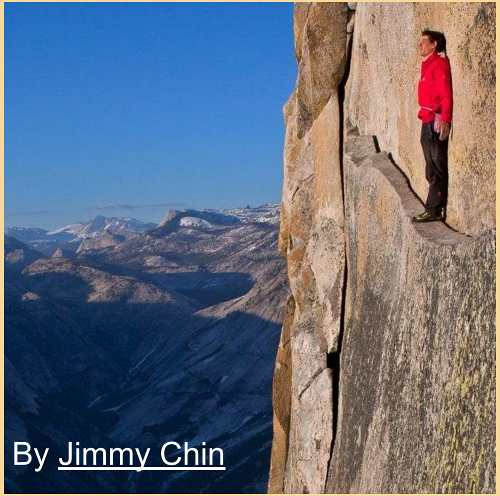
"The most important questions of life are, for the most part, really only problems of probability."
— Pierre-Simon, Marquis de Laplace

👨‍🏫 Statistical Reasoning Lecture #1
Alexander Savi, 2025
📚 Mehmetoglu & Mittner Ch. (1, 2, 3), 7

Curled by Danielle Navarro (jasmines package, diy) 🎨

Does the unprescribed use of Ritalin (yes, no) lead to mental illnesses, such as sleeplessness and psychological dependence (yes, no)?

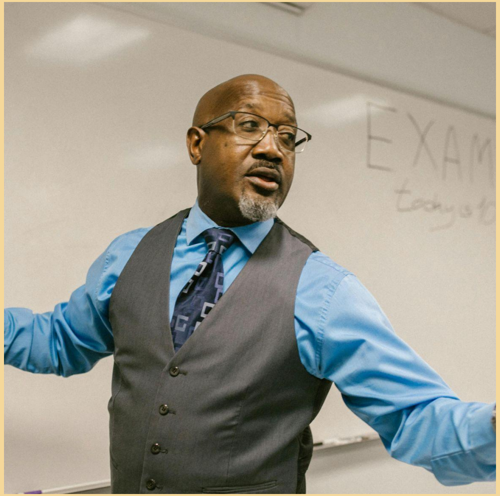
By Bhekisisa.org

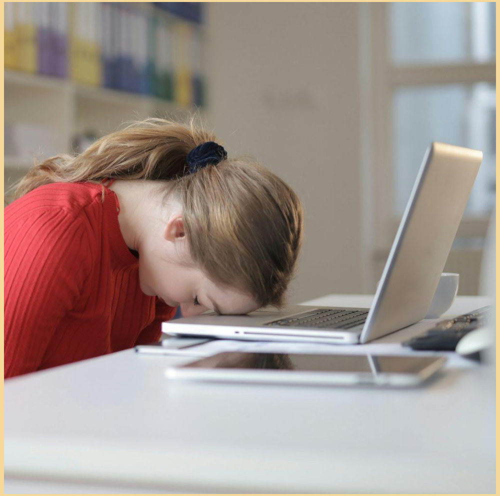


By Jimmy Chin

How does the amount of daily physical activity (measured in hours) predict the likelihood of experiencing anxiety symptoms (yes, no)?

Is there a difference in average test scores among students from different teaching methods (traditional, online, hybrid)?





What is the relationship between hours of sleep per night and cognitive function (measured by memory recall)?

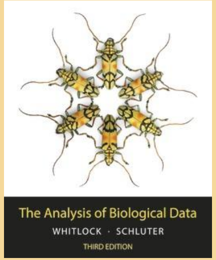
What are you supposed to know already?

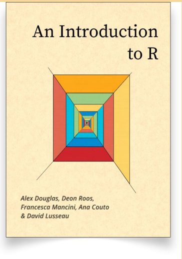
UNIVERSITEIT VAN AMSTERDAM

CANVAS BY INSTRUCTURE

RStudio®

SOWISO

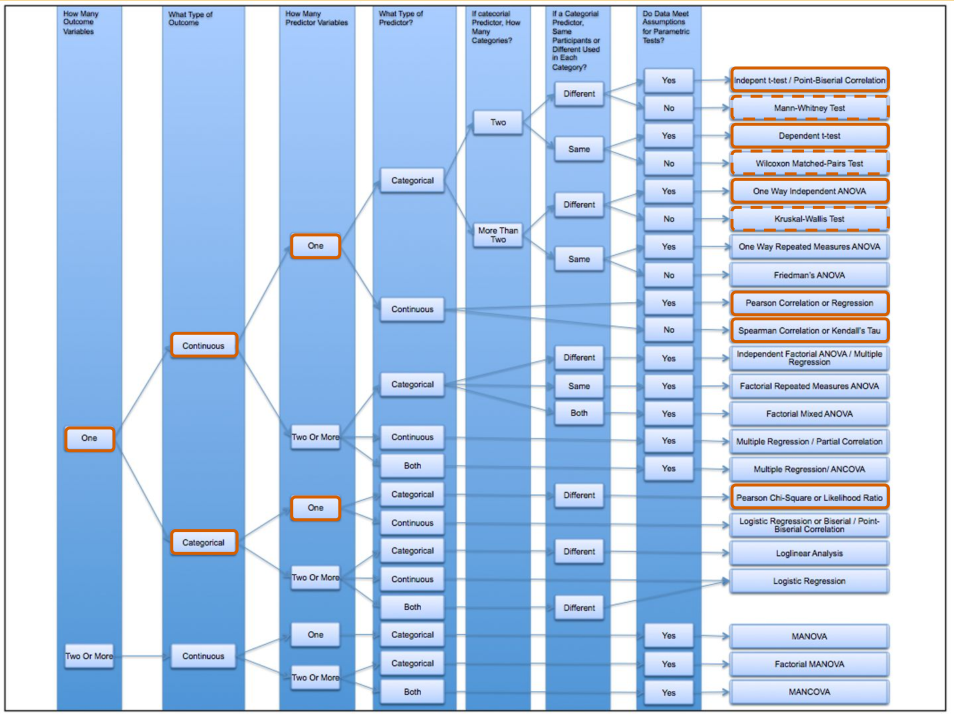




Methoden van Onderzoek en Statistiek

Samples & sampling distributions
NHST
P values
Confidence intervals
Probabilities
Bayes theorem
Test choice & research design
Data visualization

What are you supposed to know already?



Methoden van Onderzoek en Statistiek

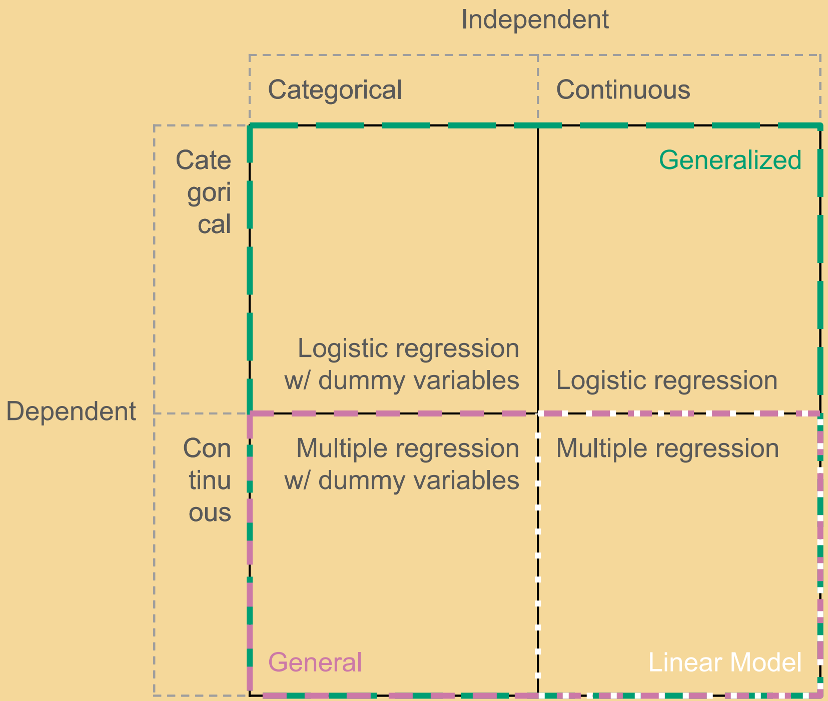
Binomial test
Chi-squared test
Student's *t*-test
One-way ANOVA
Correlation
Simple linear regression
Some nonparametric alternatives

How do we go from here?

		Independent	
		Categorical	Continuous
Dependent	Categorical	Chi-squared test	
	Continuous	Student's *t* test One-way ANOVA Factorial ANOVA	Simple regression Multiple regression

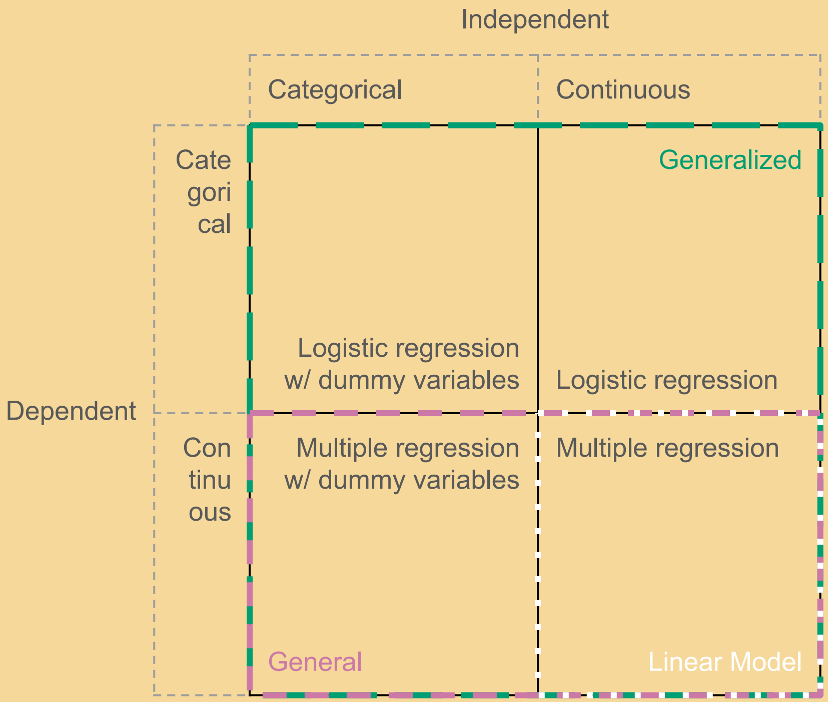
- ❑ From one predictor to multiple predictors

How do we go from here?



- ❏ From one predictor to multiple predictors
- ❏ From a toolbox to a multitool (“Swiss Army knife”)

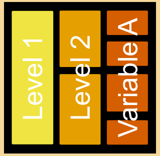
How do we go from here?



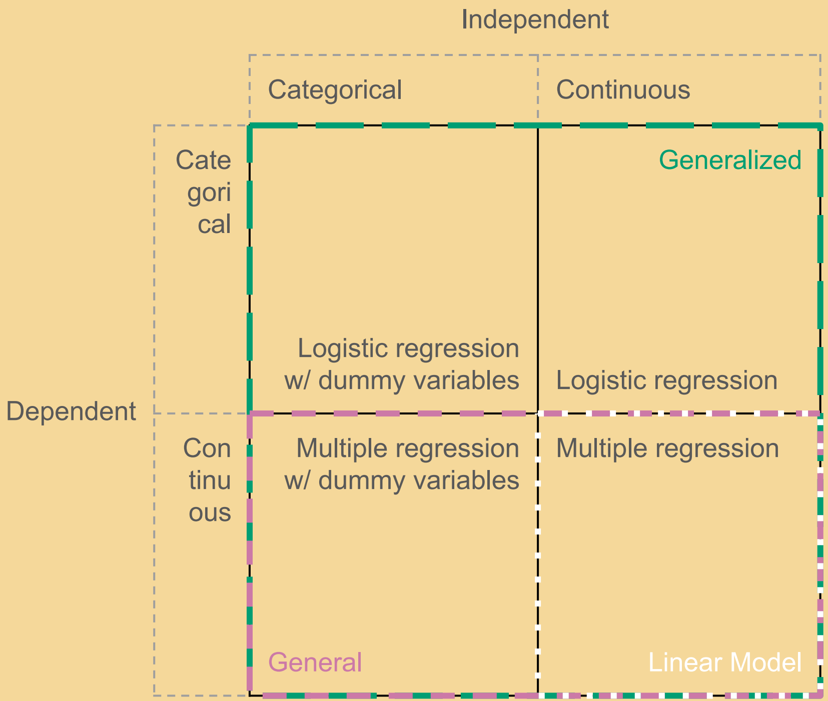
- ❏ From one predictor to multiple predictors
- ❏ From a toolbox to a multitool
- ❏ Hierarchical data analysis

Independent data

Dependent data



How do we go from here?



- ❑ From one predictor to multiple predictors
- ❑ From a toolbox to a multitool
- ❑ Hierarchical data analysis
- ❑ Bayesian inference (vs. frequentist inference)

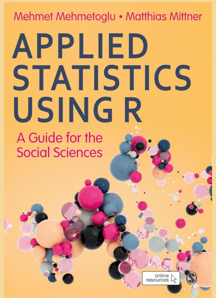
How do we go from here?

UNIVERSITEIT VAN AMSTERDAM

CANVAS BY INSTRUCTURE

RStudio

SOWISO



💡 Created with R package [ggfx]

Knowledge and skills in statistical reasoning and inference.

- Conceptual understanding of statistical inference (frequentist, Bayesian and non-parametric statistics)
- Match statistical test to research design (and vice versa)
- Perform statistical tests with R
- Interpret test results from R

Foundation for follow-up courses (e.g., *Experimentation Year 2*) and independent learning.

Why should you (not) trust me?

Me

- ❏ Assistant Professor in Psychological Methods
- ❏ Not a statistician (but well-connected)

Book

- ❏ Written by experts (see “About the Authors”)
- ❏ Recently written (pros and cons)
- ❏ Supported by evidence (see “Bibliography”)
- ❏ Well-reviewed by experts (see reviews)

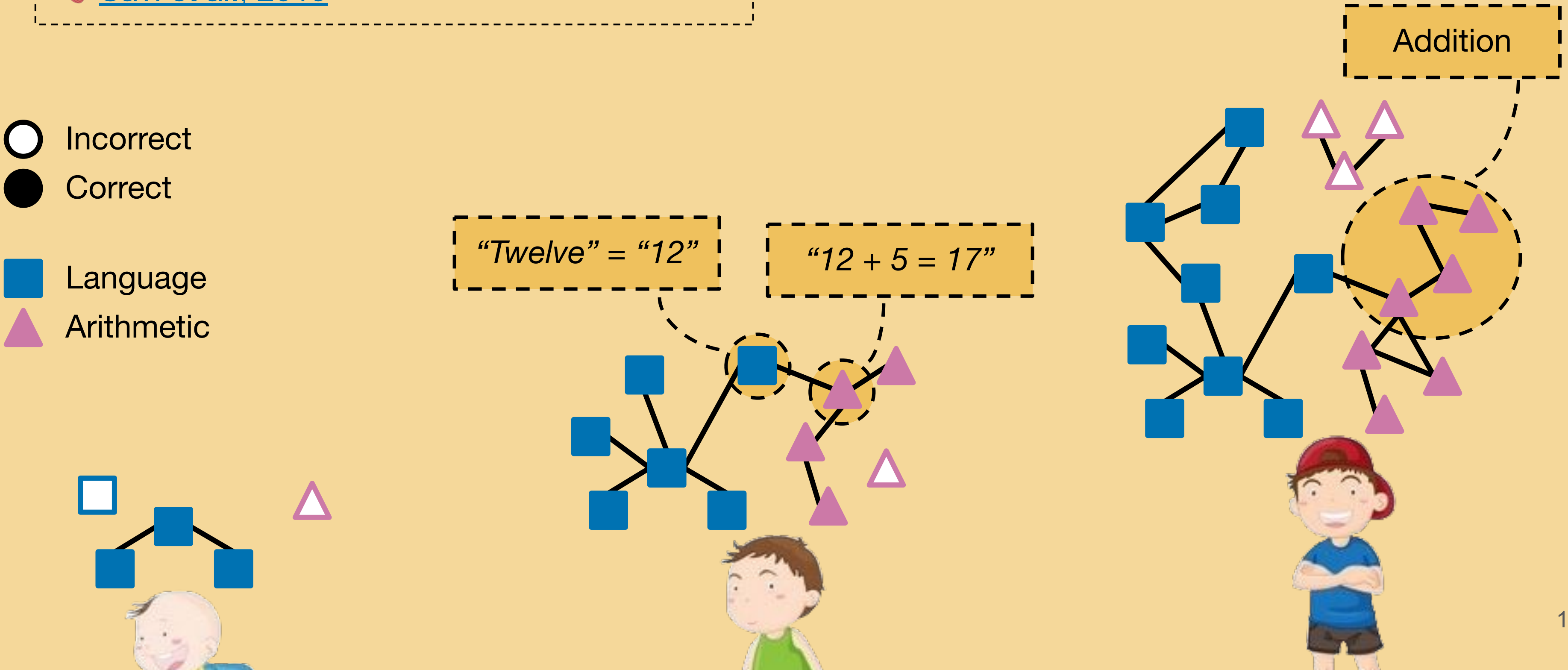
Purpose

- ❏ See “How can you benefit…”
- ❏ No hidden agenda

By Frederick Burr Opper (1894)

The wiring of intelligence

🧠 [Savi et al., 2019]



How can you benefit the most from this course?

- Learn key abilities for future projects and your thesis
  - What are the type of questions I'll be able to answer?
  - What does it take to formulate a statistically sound answer to that question?
- Expand your critical thinking skills
  - How does knowledge come about?
  - How can I judge knowledge using my statistical reasoning skills?
- Understand some of the basics of AI
- *Anything we should add?*

Make it relevant. What is most relevant for you?

Today

Topics

Statistical reasoning with GLM
| 🔁 Simple linear regression
| Simulation superpower
| 🔁 Frequentist inference
| Course overview & organization
Multiple linear regression
Dummy-variable regression
Logistic regression
Multilevel and longitudinal analysis
Bayesian statistics

Learning goals

Revisit simple linear regression, which is at the core of this course.

Get acquainted with statistical simulations and see how they can help elucidate statistical concepts.

Revisit some of the concepts of MvO1.

Simple linear regression

Curled by Danielle Navarro (jasmines package, diy) 🎨

IJskoud de Beste

Joop

Can I use the weather forecast to predict the amount of ice cream to make?

What is the relationship between temperature and ice cream sales?

Me

temperature - - - - - - - - - sales

Photo by IJskoud de Beste

📺 [JOOP! reality show]

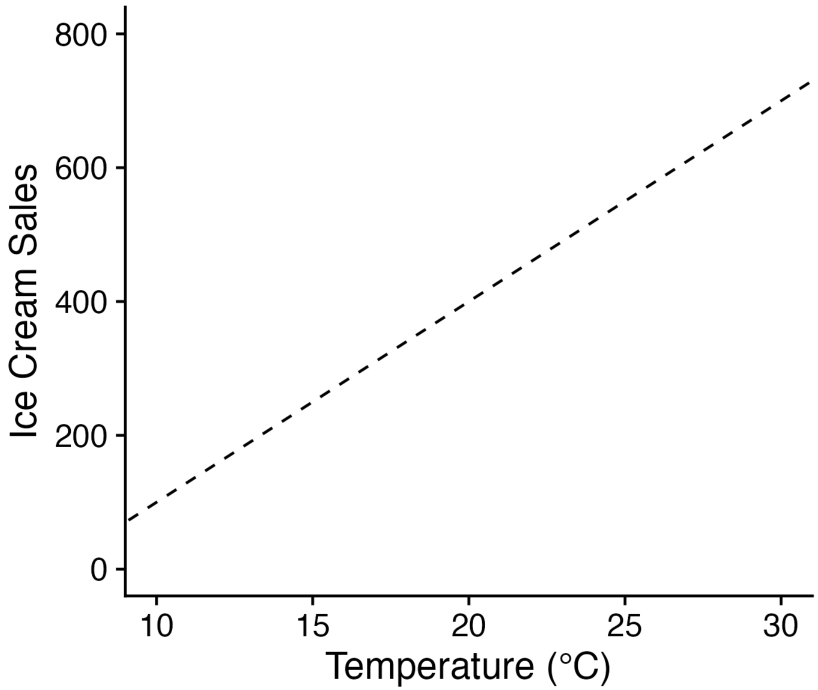
Conceptual model | Relationships between variables

independent	dependent
predictor	outcome
explanatory	response
exogenous	endogenous
[feature]()	target
x	y
continuous	continuous

Dependent \ Independent	Categorical	Continuous
Categorical		
Continuous		Simple regression

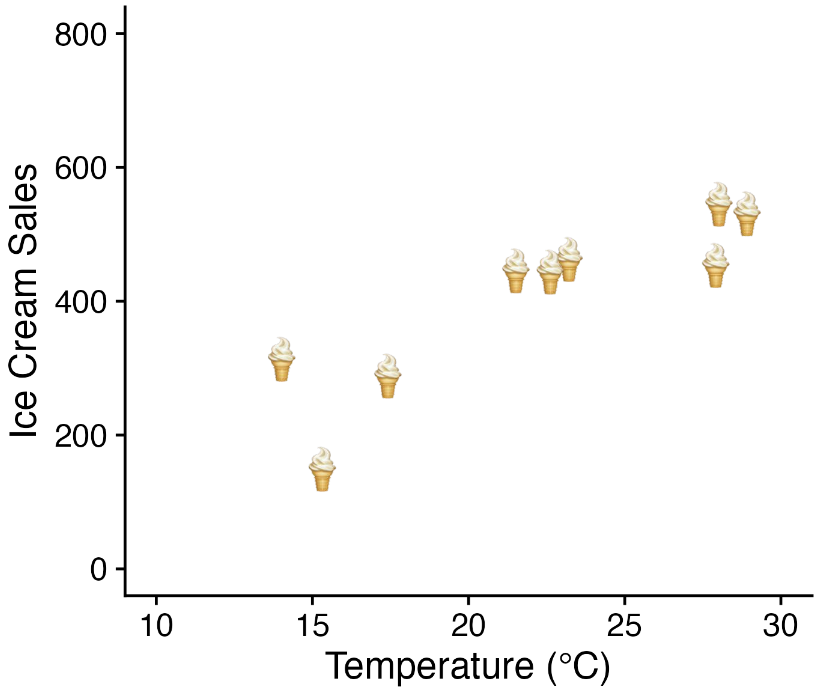
📖 7.1

Functional form | Linear relationships between variables



Sales = -200 + 30 Temperature

Data collection

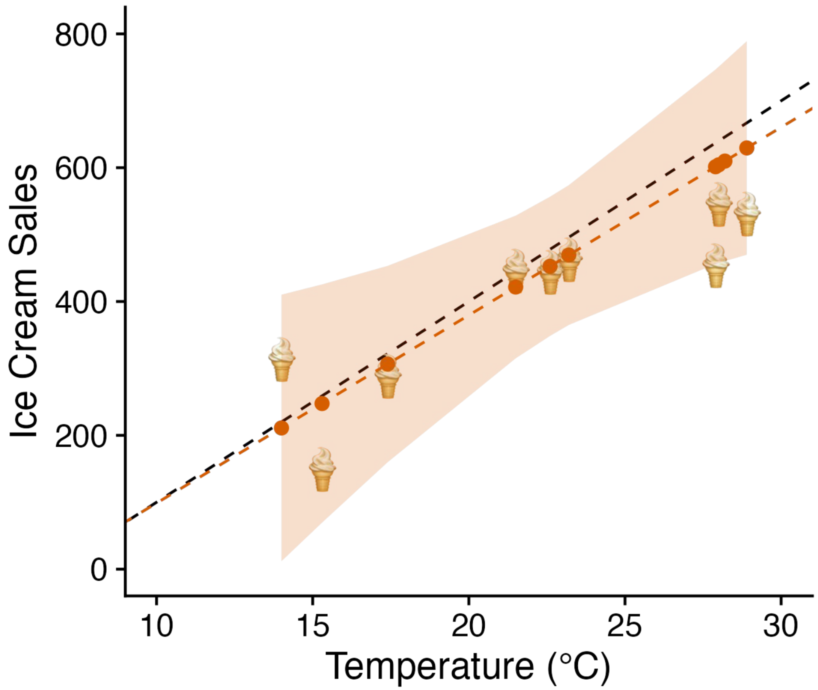



```
ice_cream_data <- tibble::tribble(
  ~temperature, ~ice_cream_sales,
  27.9,          452,
  15.3,          148,
  17.4,          287,
  21.5,          444,
  28.2,          935,
  14,            312,
  28,            544,
  28.9,          529,
  23.2,          461,
  22.6,          442)
```

$$\text{ice\_cream\_sales} = \alpha + \beta_1(\text{temperature}) + \epsilon$$

```
mod <- ice_cream_sales ~ temperature  # y ~ x
```

Modeling | Linear regression



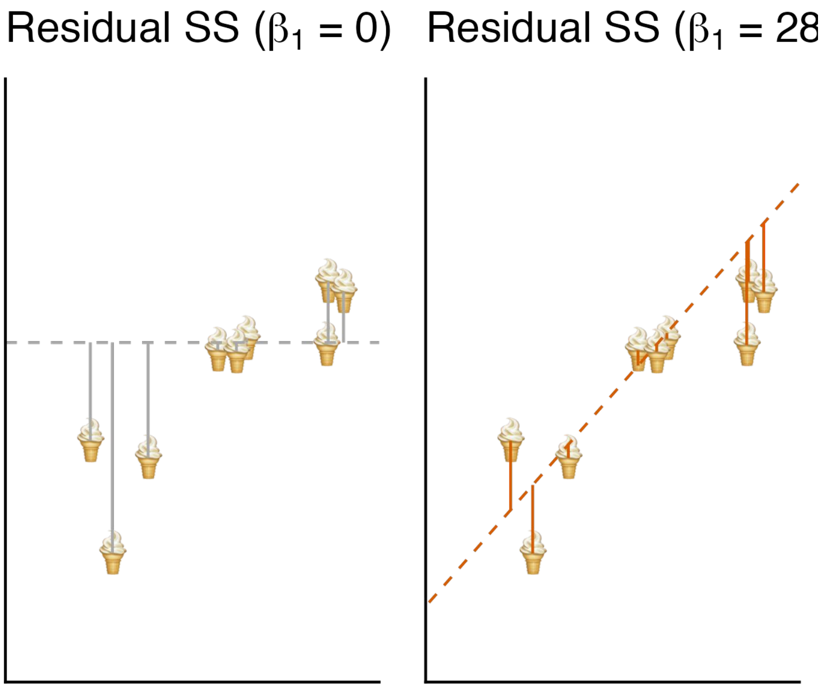

```
fit <- lm(mod, data = ice_cream_data)
summary(fit); confint(fit, level = .95)
```

```
Call:
lm(formula = model, data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-149.46  -89.73  -15.06   14.62  325.12 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -182.199    197.447  -0.923   0.3831
temperature   28.088      8.468   3.317   0.0106 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 142.7 on 8 degrees of freedom
Multiple R-squared:  0.579,   Adjusted R-squared:  0.5264 
F-statistic:    11 on 1 and 8 DF,  p-value: 0.01059
```

For every °C increase, sales increase on average with 9 to 48 ice creams.

$$\widehat{\text{ice\_cream\_sales}} = -182.2 + 28.09(\text{temperature})$$

Modeling | Ordinary least squares



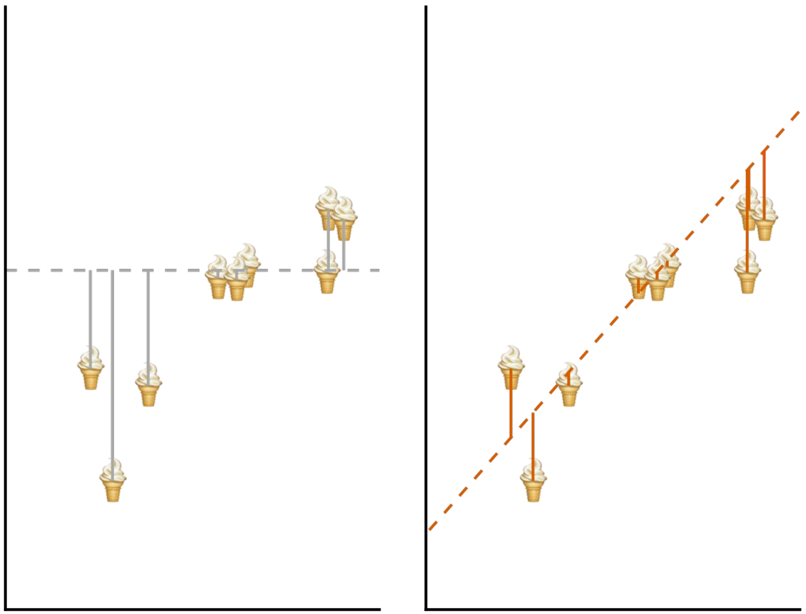
Objective: minimize residual sums of squares (RSS)

Baseline RSS $Sales_i = \beta_o + e_i$: 387052.4
Model RSS $Sales_i = \beta_o + \beta_1\ Temperature_i + e_i$: 162946

```
mod_baseline <- ice_cream_sales ~ 1
fit_baseline <- lm(mod_baseline, data = data)
summary(fit_baseline)

deviance(fit)
deviance(fit_baseline)
```

Evaluation | How well does the model describe the data?

Residual SS (β_1 = 0) Residual SS (β_1 = 28



On average, how far off is my regression estimate from the observed values?

Standard deviation of the residuals, or root mean squared error (RMSE): √*(RSS / (n - K)*

Model RMSE: 142.717 ice creams off

Is this good or bad?

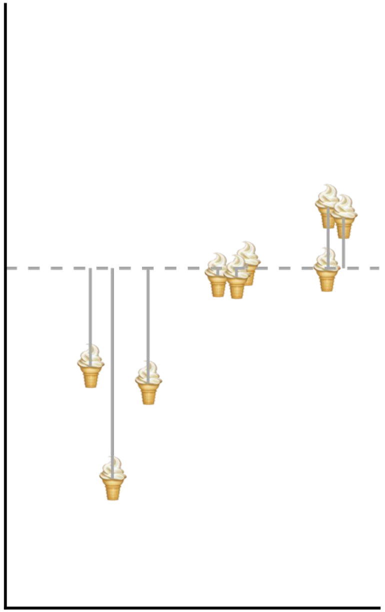
Baseline RMSE: 207.379

Still, is it good or bad? Prediction!

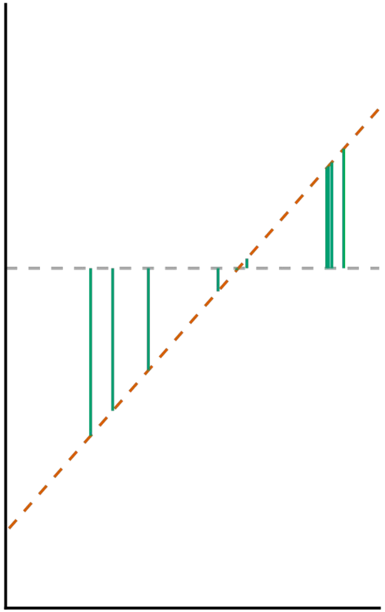
```
summary(fit)$sigma
summary(fit_baseline)$sigma
```

Evaluation | How well does the model describe the data?

Total SS



Explained SS



What is the proportion of variance shared by the data and the model?

Coefficient of determination (R^2), or *'explained' variance: ESS / TSS*

Model R^2: 0.579

58% of variation in sales is shared with the model (i.e., temperature)

R^2 = 1 (model fits data perfectly)
R^2 = 0 (model doesn't do better than mean)

```
summary_fit$r.squared
```

Evaluation | Is the relationship statistically significant?

Global significance (full model): *F* test

H_0: $R^2 = 0$
H_1: $R^2 = 0.579$ (fitted R^2)

See next lectures.

Local significance (one coefficient): *t* value

H_0: $\beta_1 = 0$
H_1: $\beta_1 = 28.088$ (fitted β_1)

See simulation superpower.

💡Compute *t*-statistic for β_1 (same procedure as for the mean): $t = (28.088 - 0) / 8.468 = 3.317$

Evaluation | What is the size of the effect?

Global effect size (full model)

R^2 (correlation family, variance explained)

- *$R^2 \geq .02$ weak; .13 moderate; .26 substantial; Cohen (1988)*

Cohen's f^2 (correlation family, variance explained)

- *$f^2 = R^2 / (1 - R^2)$ = explained / unexplained*
- *$f^2 \geq .02$ small; .15 medium; .35 large; Cohen (1988)*

Local effect size (one coefficient)

Cohen's f^2 (correlation family, variance explained)

- *$f^2 = (R^2_{AB} - R^2_A) / (1 - R^2_{AB})$*

📖 not in book

Evaluation | Sign, size, significance, smelt

💡 For every °C increase, sales increase on average with 9 to 48 ice creams.

💡 On average, the model estimate is 142.717 ice creams off from the observed values.

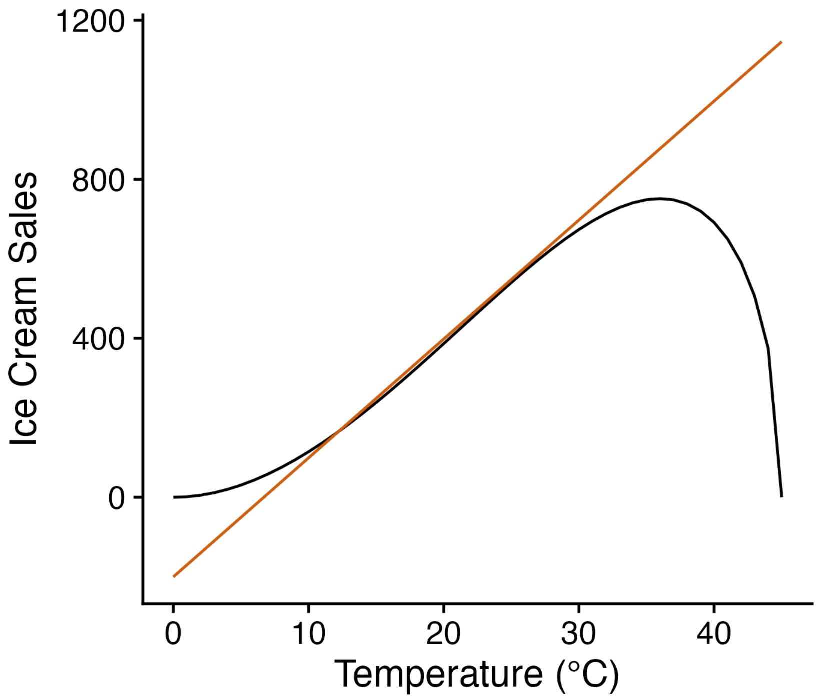
💡 58% of variation in sales can be attributed to the model; F(1, 8) = 11.003, p = .011 (alpha = .05).

💡 The slope of temperature differed significantly from zero; t(8) = 3.50, p = .011 (two-tailed, alpha = .05).

💡 Effect size f^2 is 1.375, which is considered large (Cohen, 1988).

Prediction

- ❏ Linearity assumption
- ❏ Evaluate model on new data



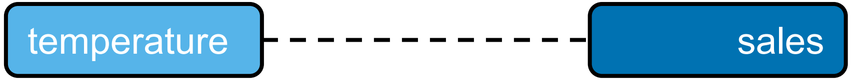

```
new_data <- tibble(temperature = c(-10, 0, 10, 20,
30, 100))
predict(fit, newdata = new_data)
```

📖 7.2.4

Model thinking

What is the relationship between temperature and ice cream sales?

Conceptual model



Linear regression equation

$$Sales_i = \beta_o + \beta_1 \, Temperature_i + e$$

R formula

```
mod <- ice_cream_sales ~ temperature
```

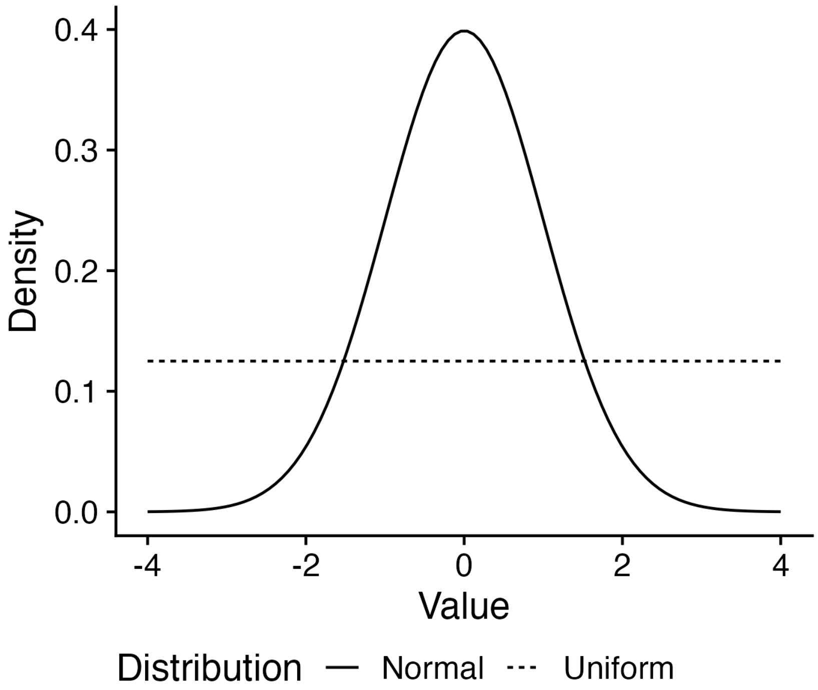
Simulation Superpower

🪄

Curled by Danielle Navarro (jasmines package, diy) 🎨

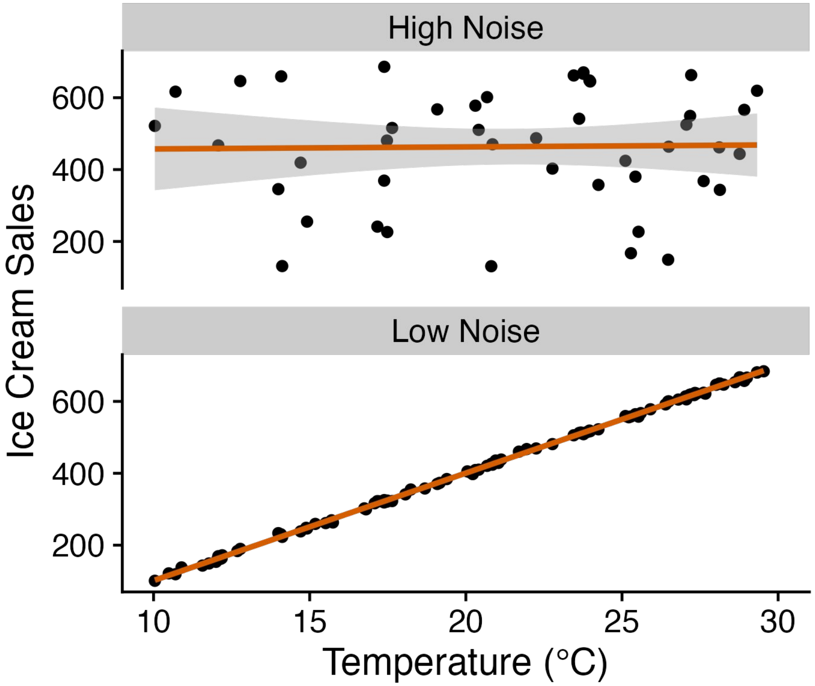
Simulation #1 | Ice cream data

```
set.seed(0)

n <- 10
temperature <- runif(n = n, min = 10, max
= 30)

intercept <- -200
slope <- 30
noise_sd <- 120
ice_cream_sales <- intercept + slope *
temperature + rnorm(n = n, mean = 0, sd =
noise_sd)

data <- tibble(temperature,
ice_cream_sales)
```



Simulation #2 | High noise, low noise

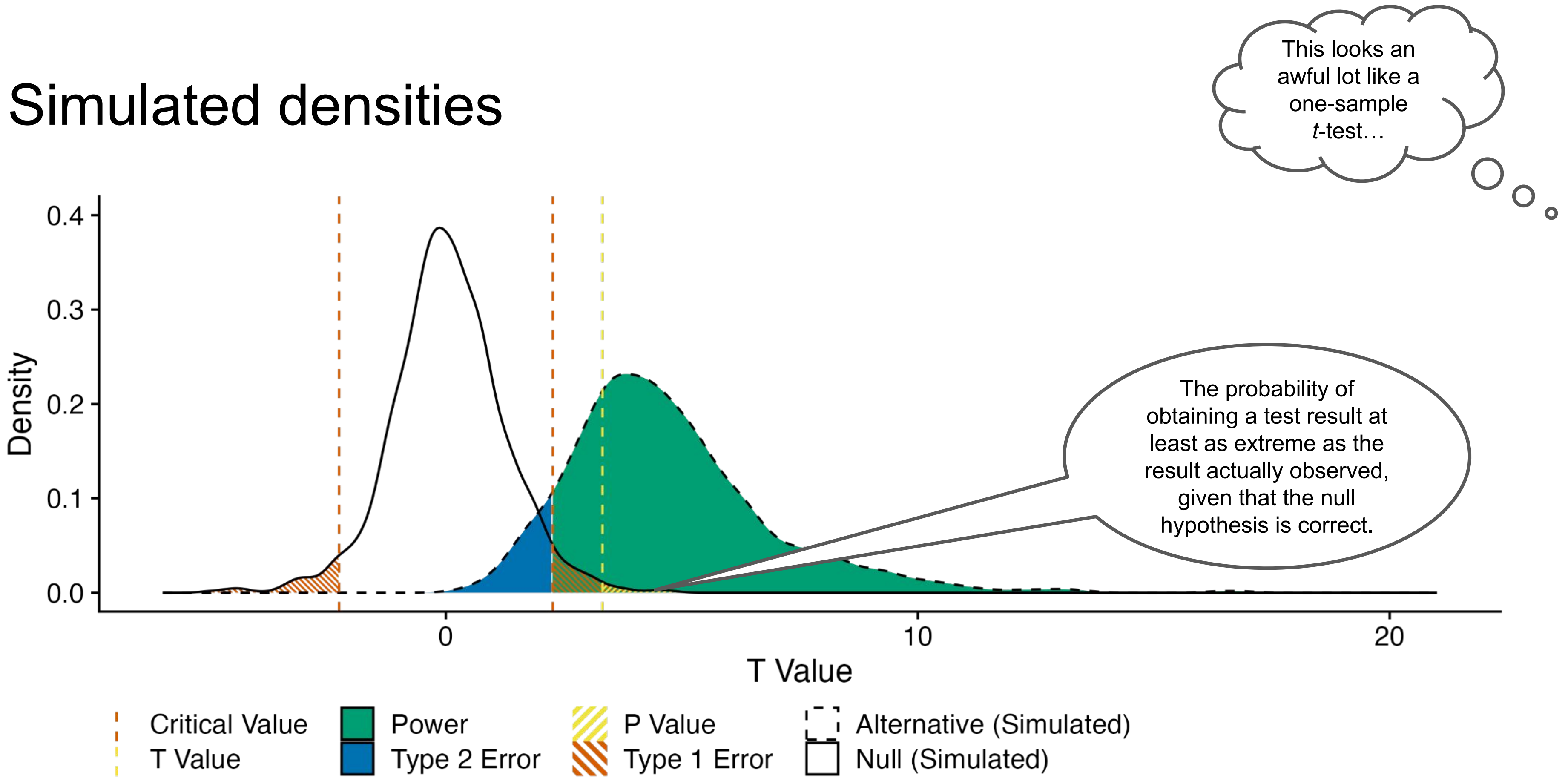
```
noise_sd_low <- 5
ice_cream_sales_sd_low <- intercept +
slope * temperature + rnorm(n = n, mean =
0, sd = noise_sd_high)

noise_sd_high <- 500
ice_cream_sales_sd_high <- intercept +
slope * temperature + rnorm(n = n, mean =
0, sd = noise_sd_high)
```



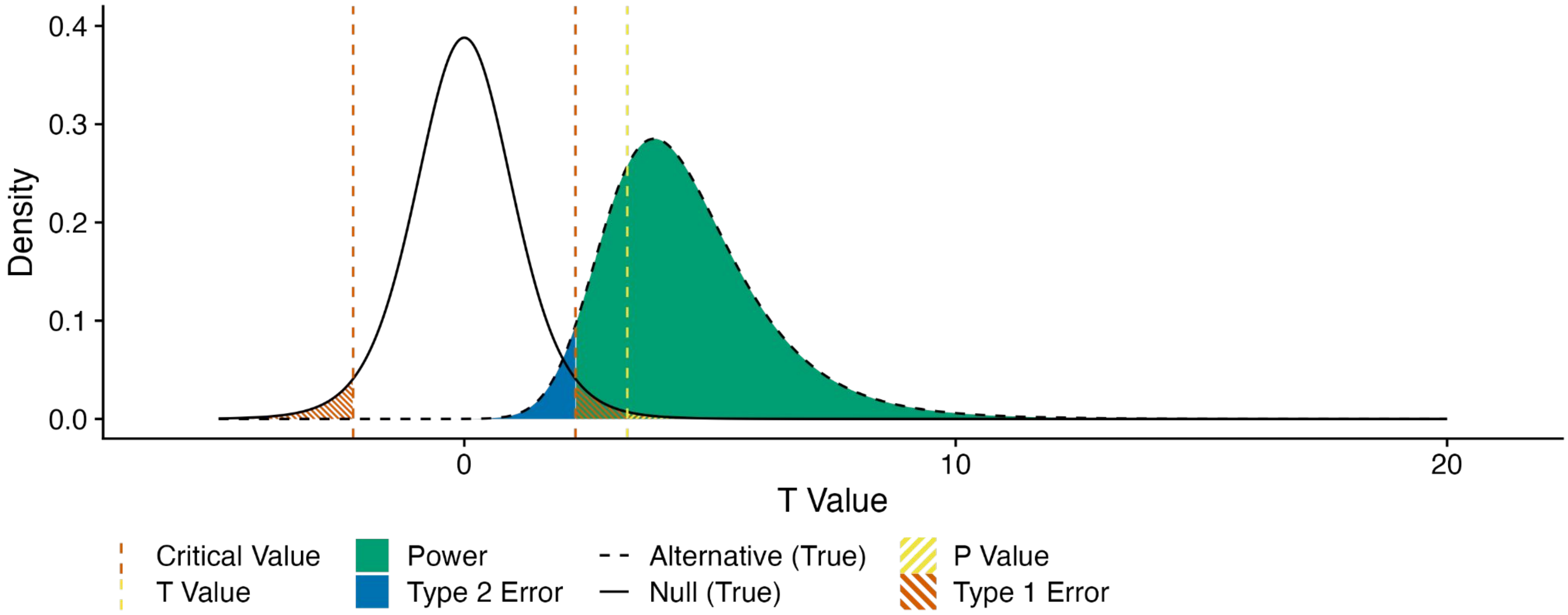
Simulation #3 | Frequentist inference (NHST)

```
n_sim <- 1000
slope_null <- 0
slope_alt <- slope

null_t_stats <- numeric(n_sim)
for (i in 1:n_sim) {
  obs_temperature <- runif(n = n, min = 10, max = 30)
  ice_cream_sales <- intercept + slope_null * obs_temperature
  noise <- rnorm(n = n, mean = 0, sd = noise_sd)

  data <- tibble(
    temperature = obs_temperature,
    sales = ice_cream_sales + noise)

  fit <- lm(sales ~ temperature, data = data)
  null_t_stats[i] <- summary(fit)$coefficients[2, "t value"]
}
```

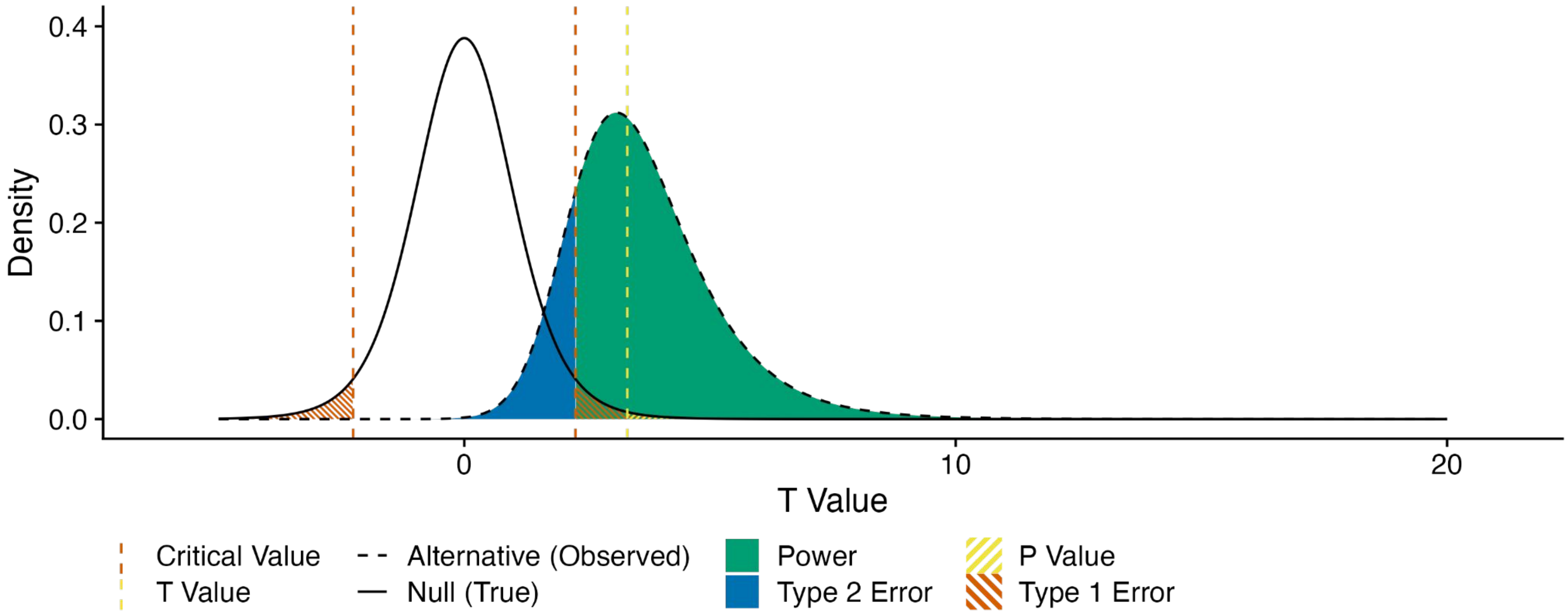
Simulated densities



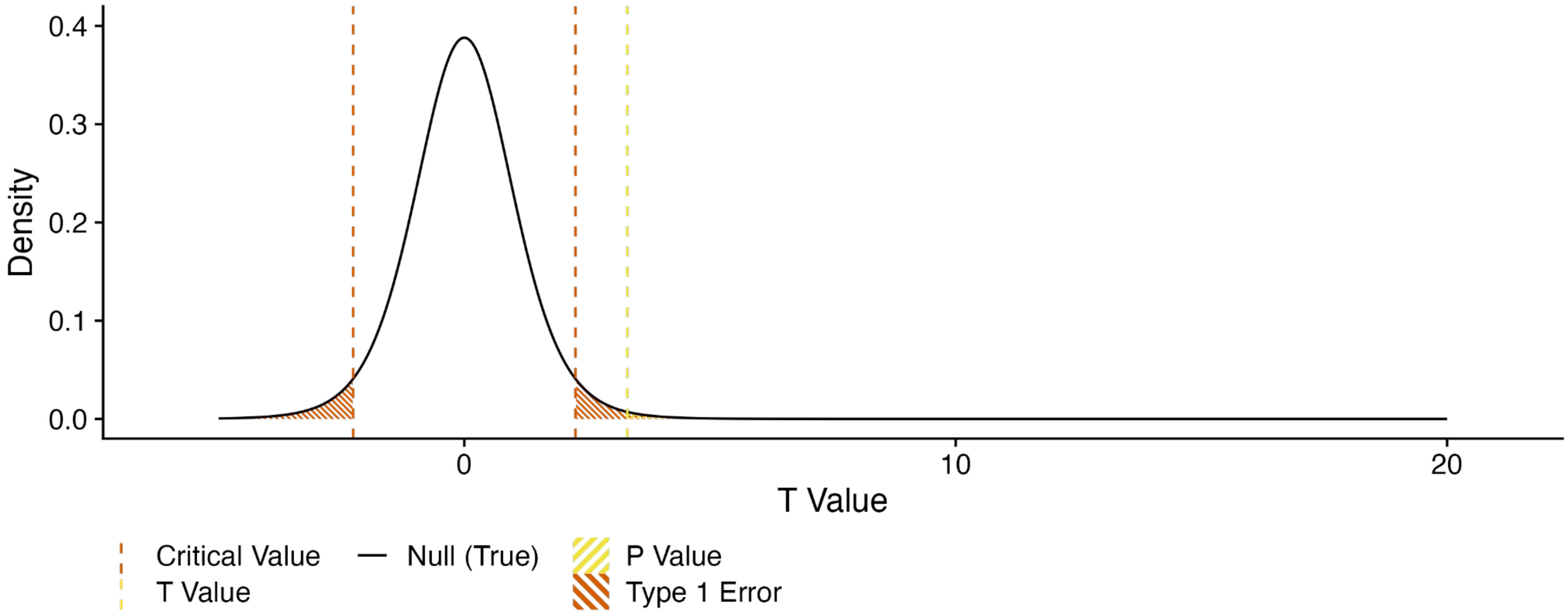
True densities



Observed densities

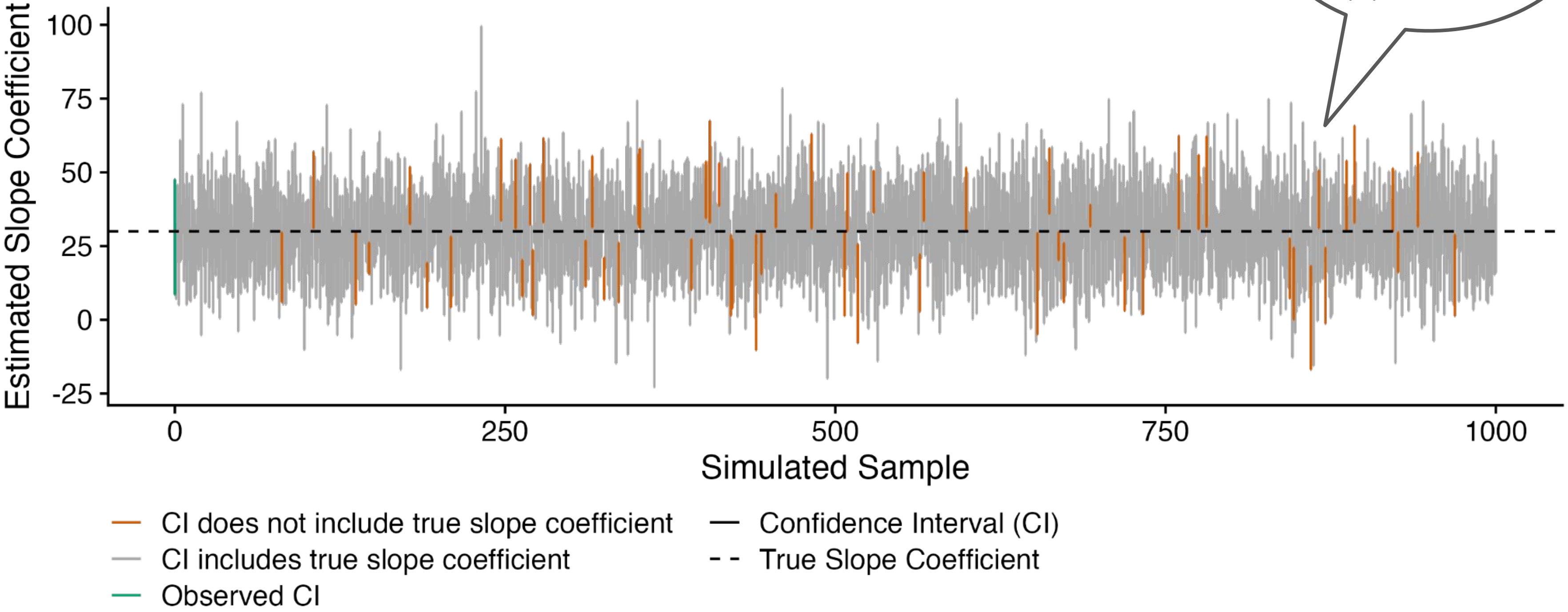


NHST in practice



Confidence interval

When repeatedly sampling from the target population, 95% of the calculated CIs contain the actual population value.



Introduction to R

Curled by Danielle Navarro (jasmines package, diy) 🎨

Best practices

- ❑ Create a *new project* for this course
- ❑ Use scripts and save regularly
  - ❑ .R
  - ❑ .Rmd (R Markdown)
  - ❑ .qmd (Quarto)
- ❑ Use *comments* excessively
- ❑ Clear history
- ❑ Understand pipes
  - ❑ function(function(x))
  - ❑ function(x) |> function()
  - ❑ function(x) %>% function()

Course

Overview & organization

Curled by Danielle Navarro (jasmines package, diy) 🎨

Book

Applied Statistics Using R

- ❏ Builds on Chapter 18 of Whitlock & Schluter (where you left last year)
- ❏ Includes R examples for both frequentist and Bayesian analyses
- ❏ Not only for social sciences
- ❏ Check the online resources
- ❏ Don't pay more than 40 euro
- ❏ Rent or buy eBook: VitalSource (PayPal/creditcard only)
- ❏ Use it as a reference during the the remainder of your study

Active reading

- Use skimming for some chapters, so you can scan them when there's something you really need to know.
- Use intensive reading (or 'study') for others (check online resources for executing R code).

Prepare the literature before the start of the week.

🔮 What will this course add?

Week	Topic	Literature	
		Skimming (R Programming)	**Intensive reading (Statistical Inference)**
1	Statistical reasoning with GLM	*Revisit Chapter 18 from Whitlock & Schluter.* Preface and Chapter 1, 2, 3.	Preface sections “How To Use This Book” and “Regression Approach to ANOVA”. Chapter 7.
2	Multiple linear regression	Chapter 4, 5.	Chapter 8, 10.1, 10.2, 10.4, 10.7.
3	Dummy-variable regression	Chapter 6.	Chapter 9, 10.3, 10.5, 10.6.
4	Logistic regression		Chapter 11.
5	Multilevel and longitudinal analysis		Chapter 12.
6	Bayesian statistics		Chapter 15.
7	Surprise		

Weekly assignments

Online consultation (drop-in):
every Wednesday 11:00–12:00

Assignment sufficient, if at least 80% correct

0	10	20	30	40	50	60	70	80	90	100

% correct response

effort counts!

Pass weekly assignment requisite, if at least 5 sufficient

0	1	2	3	4	5	6	7

sufficient assignments

Catch-up insufficient assignments after exam

Exam or Retake	Inspection	Final exam grades	Catch-up week[1,2]

[1]if exam/retake grade is sufficient
[2]you must contact Alexander & Jonas to participate

I. Attempt problem independently
   - Consult official resources (literature, lectures)
   - This is what will be examined

II. Consult fellow students or teachers
   - Collaborate and discuss with peers
   - Attend lecture & ask questions
   - Join the weekly drop-in consultation
   - Active engagement will increase your success

III. Use UvA AI chat* responsibly
   - Only if I & II are satisfied
   - Treat as a coach, not as a solver
   - Verify and reflect, beware of plagiarism
   - Prediction ≠ intelligence

* Available this academic year

Module organization*

(*course manual is leading)

Team

- Alexander Savi (lecturer, coordinator)
- Jonas van Nijnatten (assignments, co-coordinator)
- Sytske Schep and Tim Vervenne (teaching assistants)

Contact

- General: during lectures and through psychobiologiejaar2-science@uva.nl.
- Statistical reasoning: during lectures.
- Weekly assignments: during walk-in consultation hours.

Assignments (formative)

- Weekly, sufficient/insufficient, deadline every Sunday 23:55 (correct answers available after deadline)
- Pass with 5/7 sufficient assignments
- 3 attempts per assignment, 2 prior checks per attempt
- ~~Exemption for recidivists (if previously sufficient)~~ brand new topics!
- Use decimal *points* (.)

Exam (summative)

- 80% of final grade (⅝ SR, ⅜ PhS)
- SR digital open book (PhS *not* open book)

Cooling Down

Curled by Danielle Navarro (jasmines package, diy) 🎨

Takeaways

Illustration by [Zach Weinersmith]

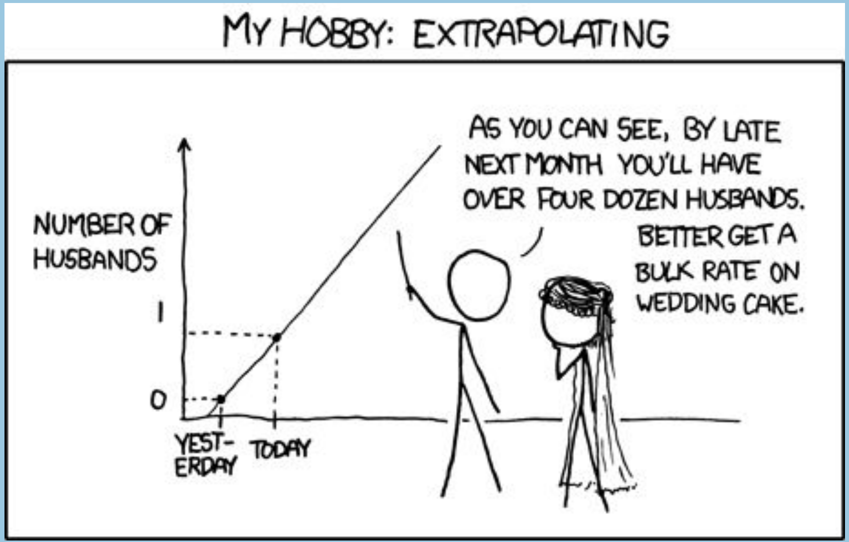


Illustration by [Randall Munroe] ([wtf])

Takeaways

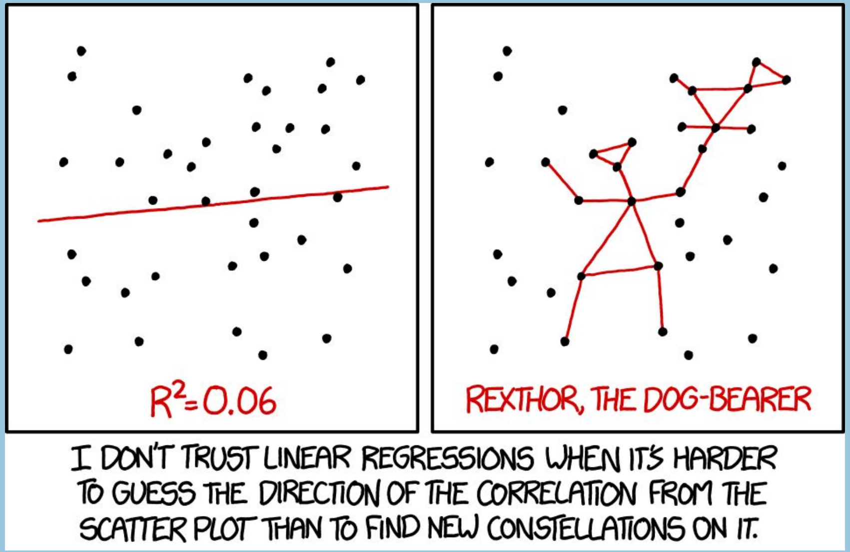


Illustration by Randall Munroe (wtf)

🥡 Takeaways

Illustration by Timo Elliott

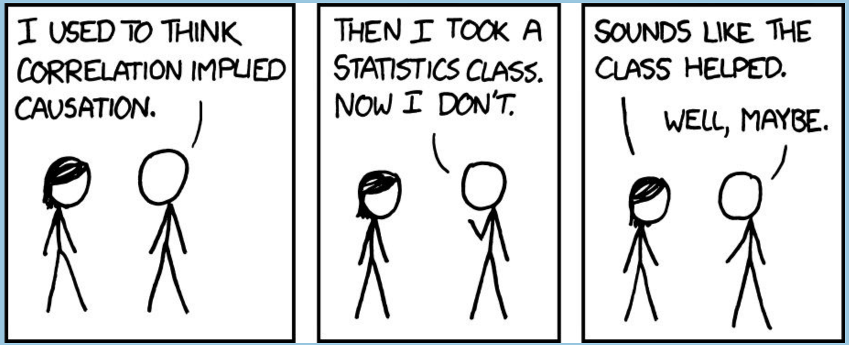


Illustration by [Randall Munroe](#) ([wtf](#))

⚠️ Don’t look here!

Show the distribution of *p*-values when the null-hypothesis is true (e.g., two samples come from the same population).

Additional challenge I: what happens to the distribution if the null hypothesis is false?

Additional challenge II: argue why the distribution in the comic would be problematic.

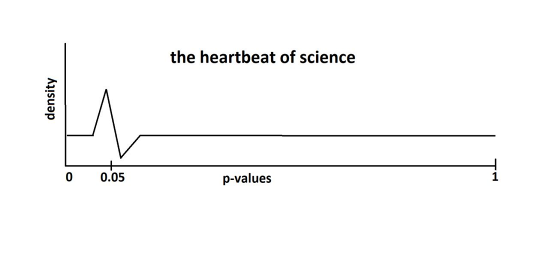


Illustration by František Bartoš

Hints (select and copy/paste the invisible text below to reveal it)

0.

1.

2.

3.

✍️ Colophon

Slides

alexandersavi.nl/teaching/

License

Statistical Reasoning by Alexander Savi is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License. An Open Educational Resource. Approved for Free Cultural Works.