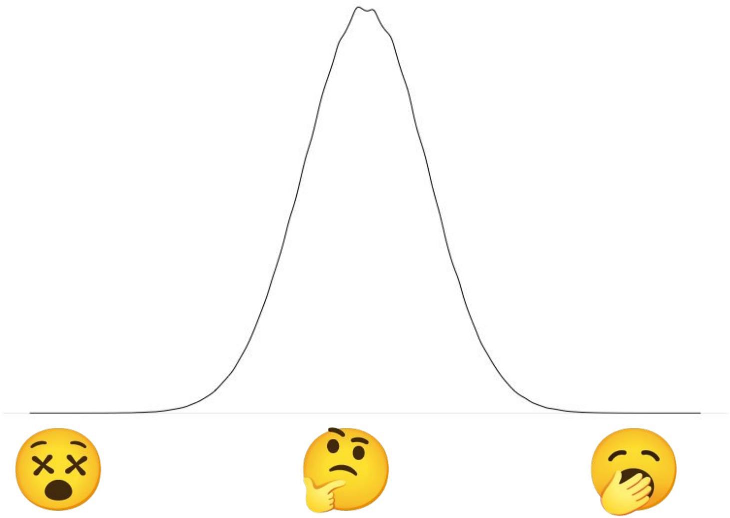
Philosophy of Science and Statistical Reasoning

Bayesian Inference

📢 But first, …



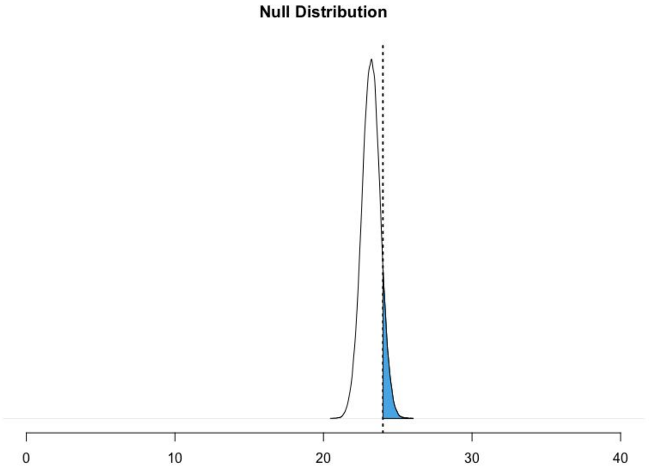
Student Puzzlement Scale

📰 News

🍺 Pub quiz

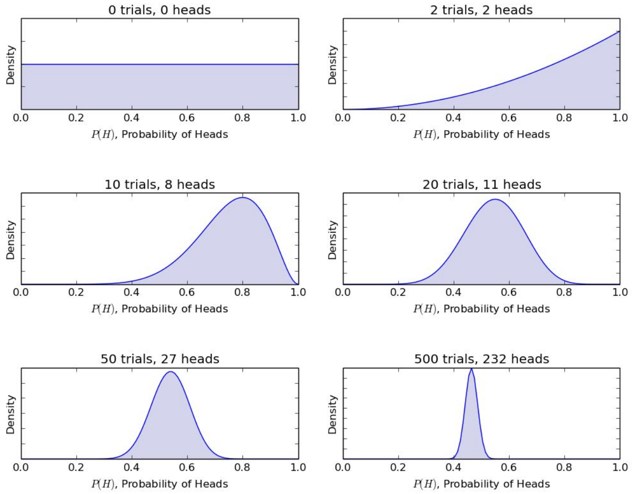
🔁 Previously, on statistical reasoning

🪙 Frequentist inference



p(D | H_0)

🪙 Bayesian inference



p(H | D)

🔮 What will we learn today?

Topics

Statistical reasoning
Empirical cycle
Probability distributions
Frequentist inference
Sample / sampling distribution
Central limit theorem
Normal distribution
P-value
Type I/II errors
Effect size
Confidence interval
Power
Test statistics
Linear regression
t-Test
Moderation
F-distribution
Nonparametric inference
ANOVA
Bayesian inference

Questions

What is Bayesian inference? What is Bayes theorem?

How can Bayesian inference be used for hypothesis testing?

How can Bayesian inference be used for parameter estimation?

🎬 Course retrospect

Bayes theorem

D: 📋 positive ADHD test result
$H_{\square}$: ADHD

💭 What's the probability of $H_{\square}$?

- High?
- Low?
- Why?

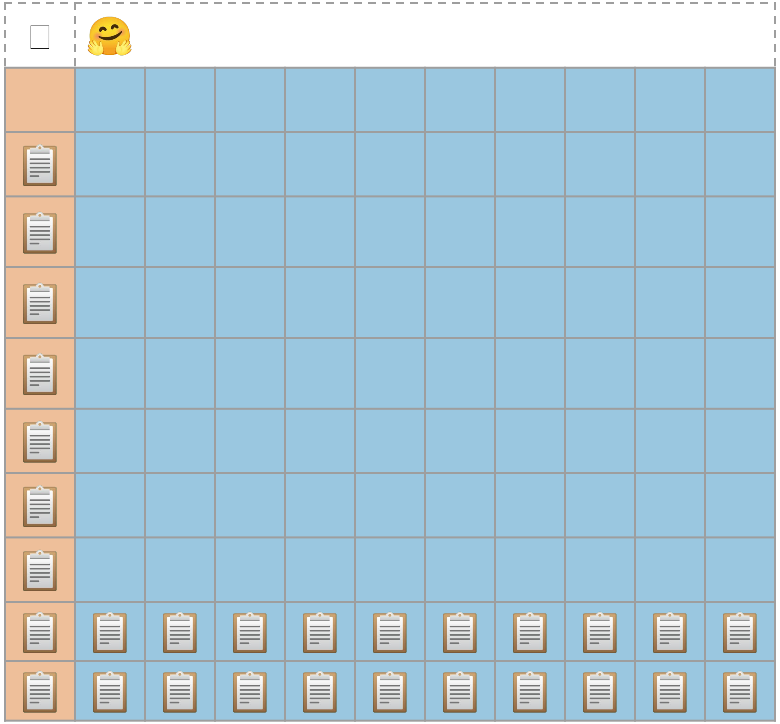
Bayes theorem

D: 📋 positive ADHD test result

$H_{□}$: ADHD

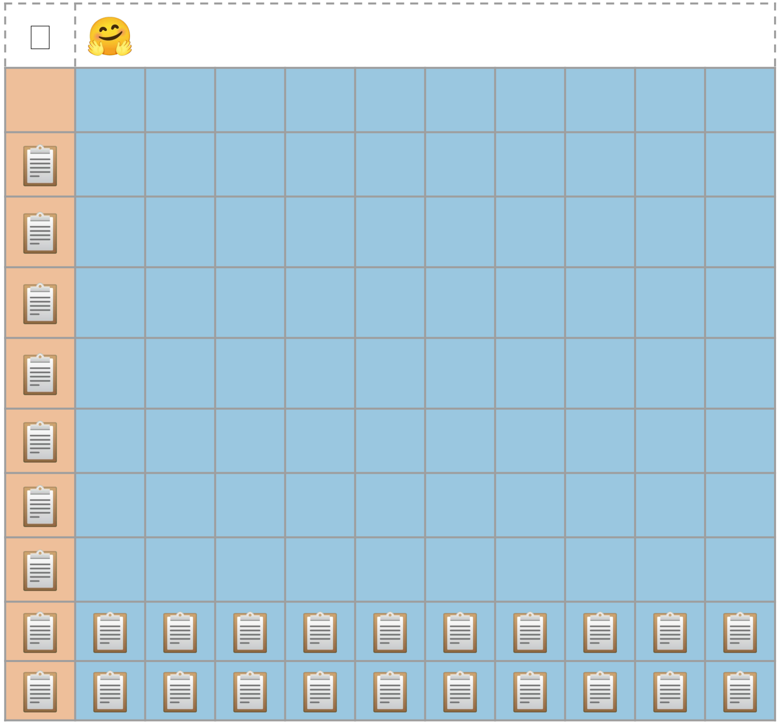
💭 What's the probability of $H_{□}$?

- High?
- Low?
- Why?



Base rate

> " It is the proportion of individuals in a population who have a certain characteristic or trait. — [Wikipedia]

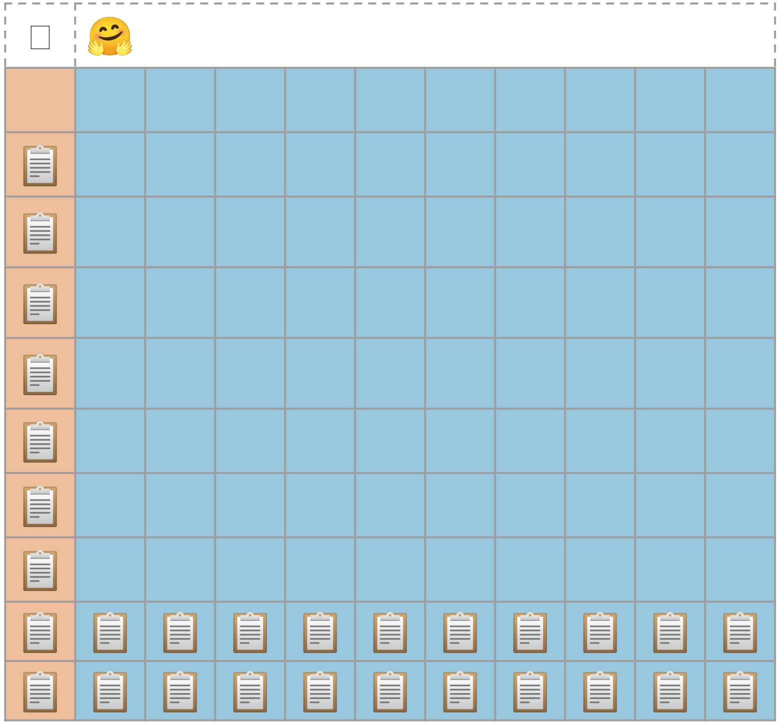


Base rate fallacy

> “ [A] type of fallacy in which people tend to ignore the base rate (e.g., general prevalence) in favor of the individuating information (i.e., information pertaining only to a specific case). — Wikipedia

💭 Out of the number of people who test positive, how many have ADHD? (9/29 ≈ .3)

💭 Keith Devlin on base rates (Edge.org).



Base rate fallacy

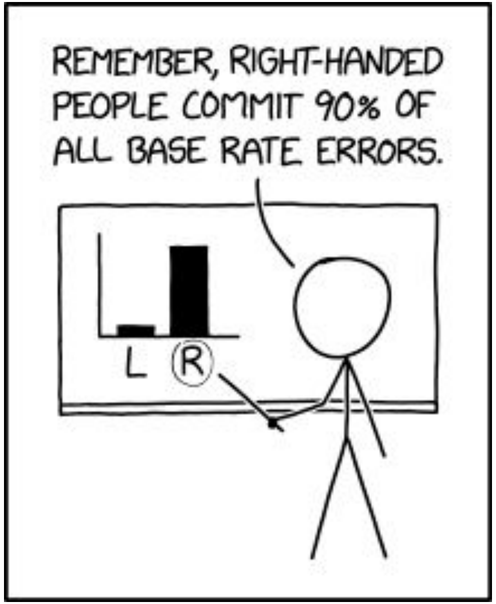
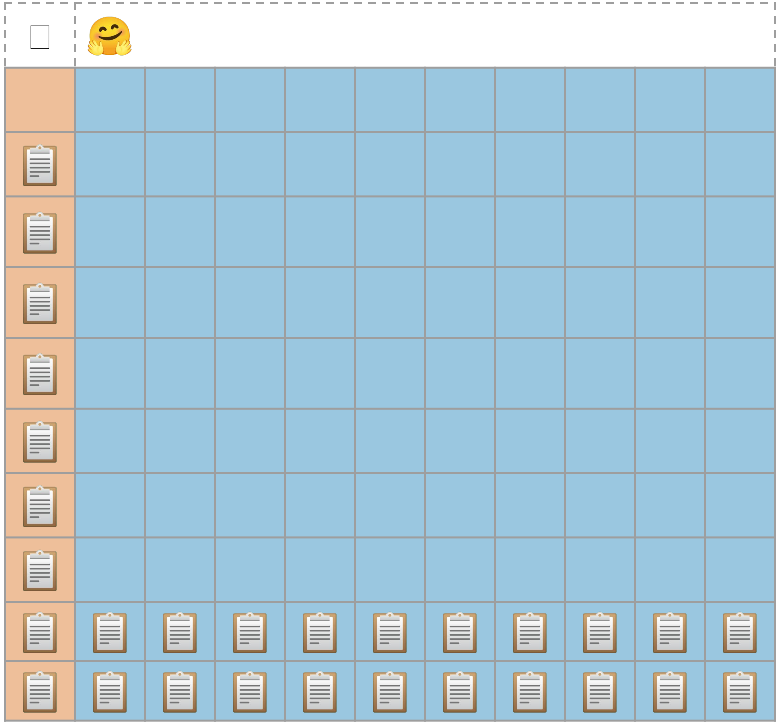
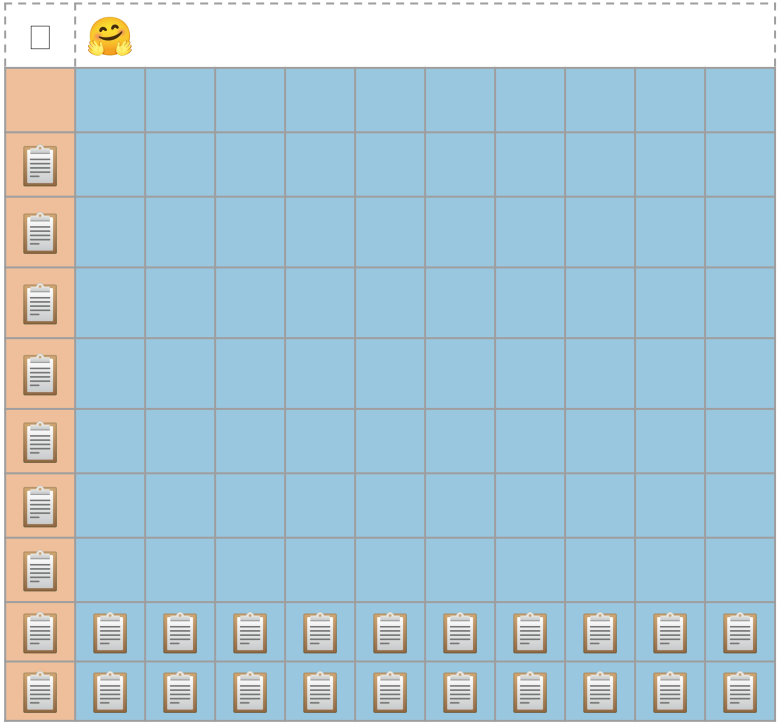


Illustration by [Randall Munroe](#) ([wtf](#))



Bayes theorem

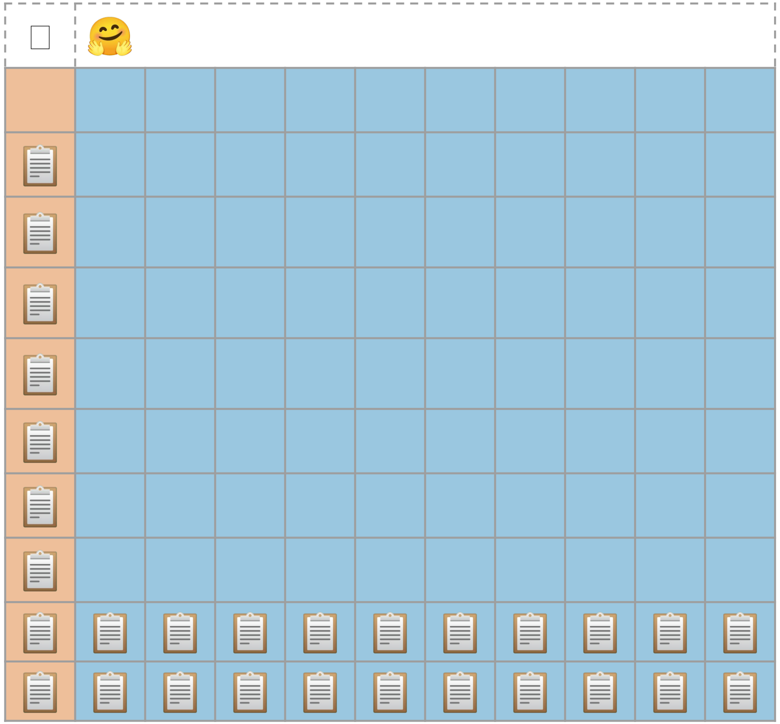
p(H | D) = (p(H) × p(D | H)) / p(D)



Prior

p(H | D) = (**p(H)** × p(D | H)) / p(D)

p(H) = ■ / (■ + ■) = 10 / (10 + 100) = .09



The probability of ADHD.

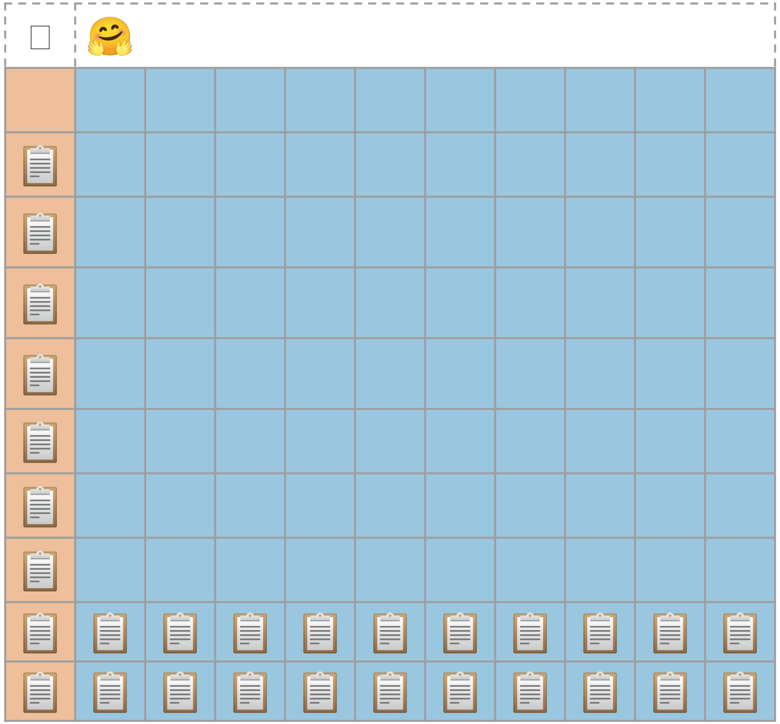
Likelihood

p(H | D) = (p(H) × **p(D | H)**) / p(D)

p(H) = / (+) = 10 / (10 + 100) = .09

p(D | H) = / = 9 / 10 = .9

The probability of a positive test result, given ADHD.



Prior × likelihood

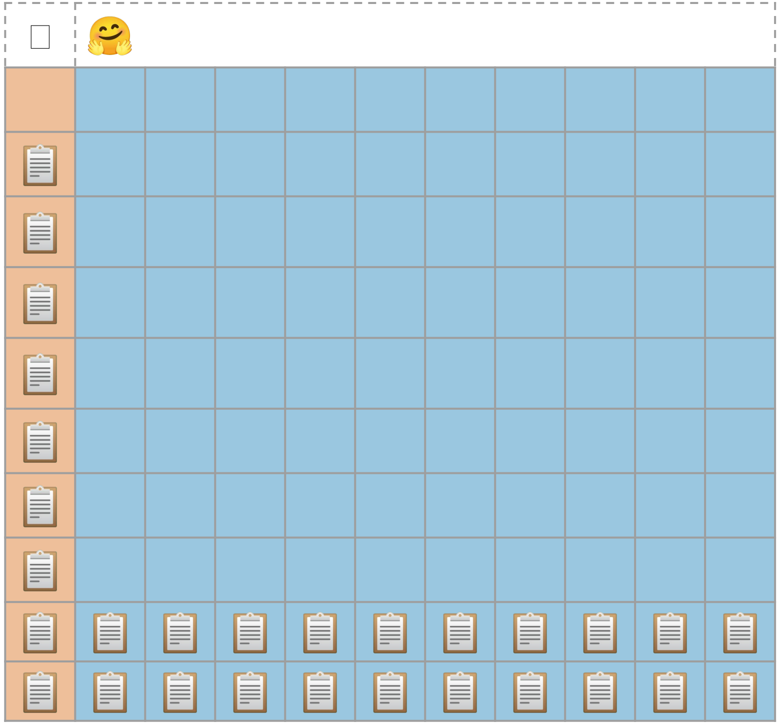
p(H | D) = (**p(H) × p(D | H)**) / p(D)

p(H) = / (+) = 10 / (10 + 100) = .09

p(D | H) = / = 9 / 10 = .9

p(H) × p(D | H) = / (+) = .09 × .9 = .08

The probability of a positive test result *and* ADHD.



Marginal likelihood

p(H | D) = (p(H) × p(D | H)) / **p(D)**

p(H) = 🟧 / (🟧 + 🟦) = 10 / (10 + 100) = .09

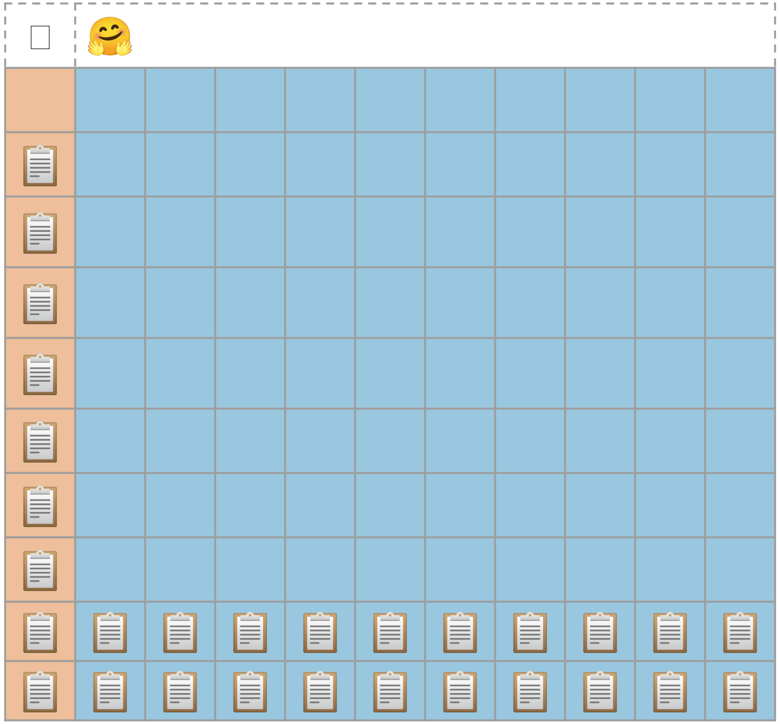
p(D | H) = 📋 / 🟧 = 9 / 10 = .9

p(H) × p(D | H) = 📋 / (🟧 + 🟦) = .09 × .9 = .08

p(D) = (p(H) × p(D | H)) + (p(¬H) × p(D | ¬H)) =

.08 + .18 = .26

The probability of a positive test result.



Posterior

p(H | D) = (p(H) × p(D | H)) / p(D) = .08 / .26 ≈ .3

p(H) = 🟧 / (🟧 + 🟦) = 10 / (10 + 100) = .09

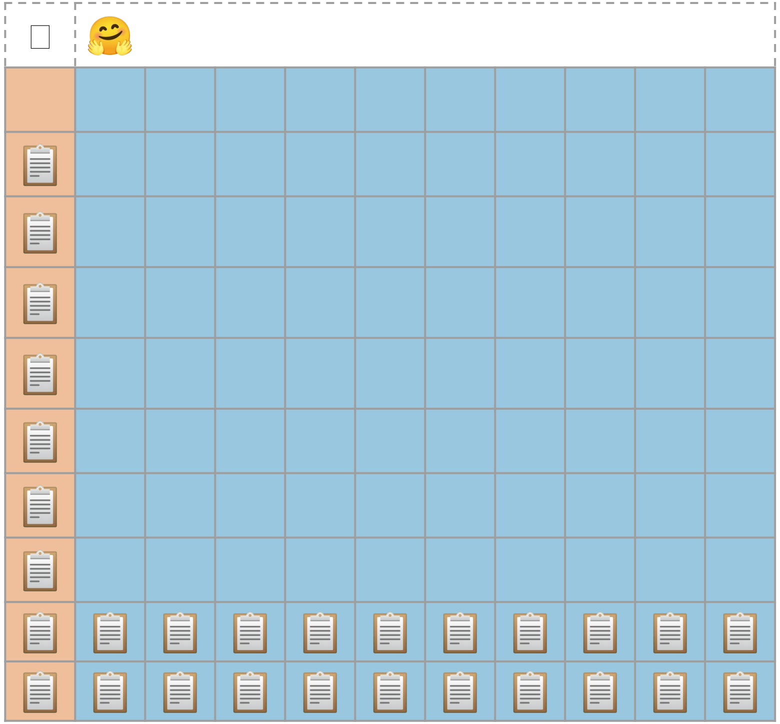
p(D | H) = 📋 / 🟧 = 9 / 10 = .9

p(H) × p(D | H) = 📋 / (🟧 + 🟦) = .09 × .9 = .08

p(D) = (p(H) × p(D | H)) + (p(¬H) × p(D | ¬H)) =

.08 + .18 = .26

The probability of someone with ADHD, given a positive test result.



Posterior

$p(H \mid D)$ $= (p(H) \times p(D \mid H)) / p(D) = .08 / .26 \approx .3$

$p(H) =$ 🟧 / (🟧 + 🟦) $= 10 / (10 + 100) = .09$

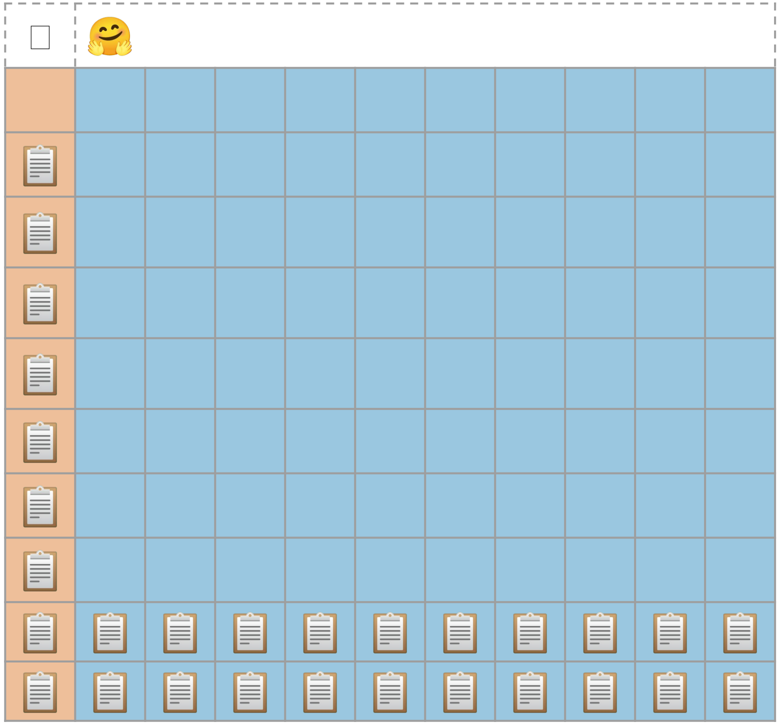
$p(D \mid H) =$ 📋 / 🟧 $= 9 / 10 = .9$

$p(H) \times p(D \mid H) =$ 📋 / (🟧 + 🟦) $= .09 \times .9 = .08$

$p(D) = (p(H) \times p(D \mid H)) + (p(\neg H) \times p(D \mid \neg H)) =$

$.08 + .18 = .26$

🔁 Out of the number of people who test positive, how many have ADHD? (9/29 ≈ .3)

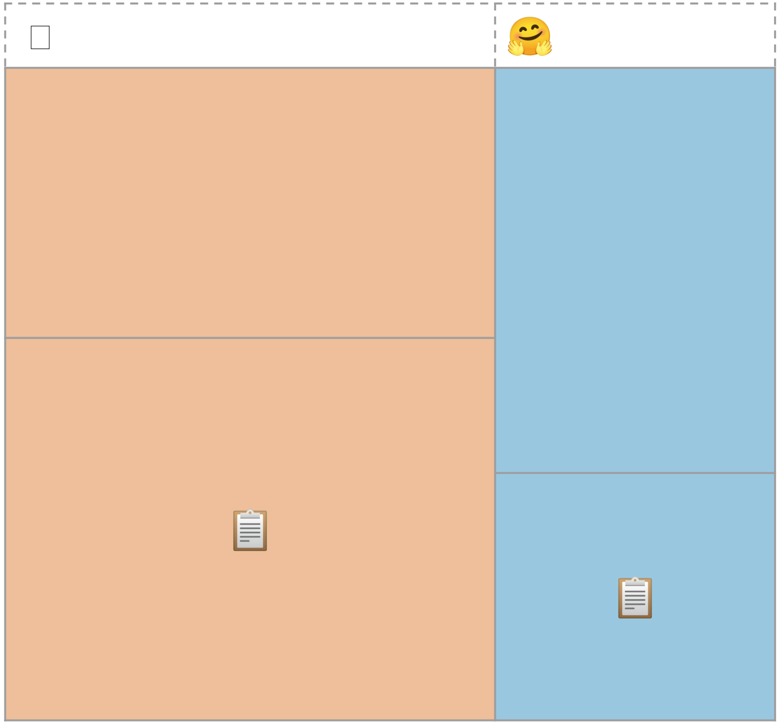


Bayes theorem

$p(H_{\text{□}} \mid D) = p(H_{\text{□}}) \times p(D \mid H_{\text{□}}) \,/\, p(D) \approx .3$
$p(H_{\text{🤗}} \mid D) = p(H_{\text{🤗}}) \times p(D \mid H_{\text{🤗}}) \,/\, p(D) \approx .7$
(‘alternative’ hypothesis)

💡 3Blue1Brown on Bayes theorem and the product rule (prior × likelihood).
🛠️ Bayes theorem visualization by Seeing Theory.

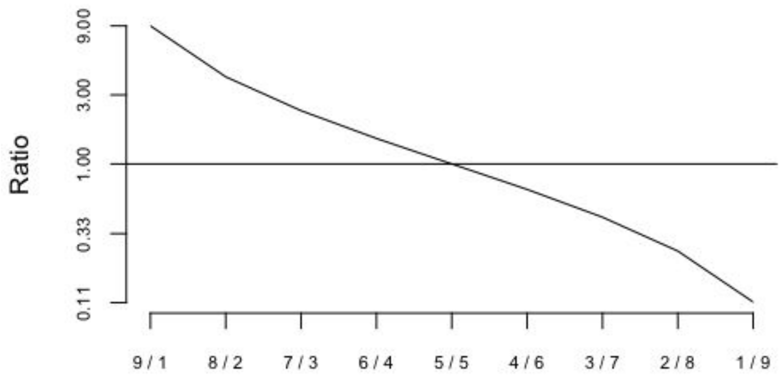
💭 3Blue1Brown on improving Bayes theorem.



Bayesian hypothesis testing

> The Bayes factor is a ratio of two competing statistical models represented by their evidence, and is used to quantify the support for one model over the other.
> — Wikipedia

🔁 Remember the *F*-ratio?



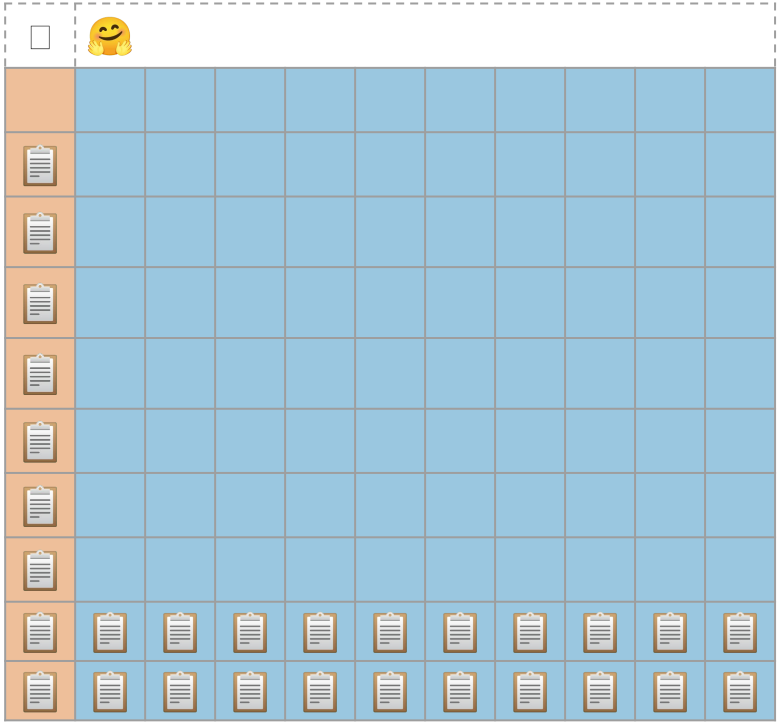
Bayes factor

K = p(D | H□) / p(D | H🤗) = ?

p(H | D) = (p(H) × **p(D | H)**) / p(D)
p(D | H□) = 📋 / ■ = 9 / 10 = .9
p(D | H🤗) = 📋 / ■ = 20 / 100 = .2

K = p(D | H□) / p(D | H🤗) = 4.5

- Continuous degree of evidence (vs. all-or-none)
- Monitor evidence during data collection
- Evidence of absence (data support a null effect) and absence of evidence (data are not informative)



Bayes factor interpretation

K	dHart	bits	Strength of evidence
$< 10^0$	< 0	< 0	Negative (supports M_2)
10^0 to $10^{1/2}$	0 to 5	0 to 1.6	Barely worth mentioning
$10^{1/2}$ to 10^1	5 to 10	1.6 to 3.3	Substantial
10^1 to $10^{3/2}$	10 to 15	3.3 to 5.0	Strong
$10^{3/2}$ to 10^2	15 to 20	5.0 to 6.6	Very strong
$> 10^2$	> 20	> 6.6	Decisive

— Wikipedia

$\log_{10} K$	K	Strength of evidence
0 to 1/2	1 to 3.2	Not worth more than a bare mention
1/2 to 1	3.2 to 10	Substantial
1 to 2	10 to 100	Strong
> 2	> 100	Decisive

— Wikipedia

💭 Discussion on Andrew Gelman's blog.
Mindless (Bayesian) statistics 🔒.

Bayes factor interp

K	dHart	bits	Str
$< 10^0$	< 0	< 0	Ne
10^0 to $10^{1/2}$	0 to 5	0 to 1.6	Ba
$10^{1/2}$ to 10^1	5 to 10	1.6 to 3.3	
10^1 to $10^{3/2}$	10 to 15	3.3 to 5.0	
$10^{3/2}$ to 10^2	15 to 20	5.0 to 6.6	
$> 10^2$	> 20	> 6.6	

Strength of evidence
t worth more than a bare mention
Substantial
Strong
Decisive

— Wikipedia

n Andrew Gelman's blog.
sian) statistics 🔒.

Illustration by Viktor Beekman, Eric-Jan Wagenmakers

☕

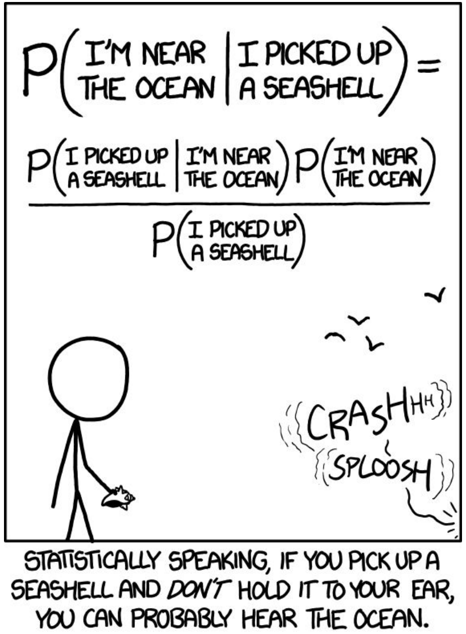


Illustration by Randall Munroe ([wtf](wtf))

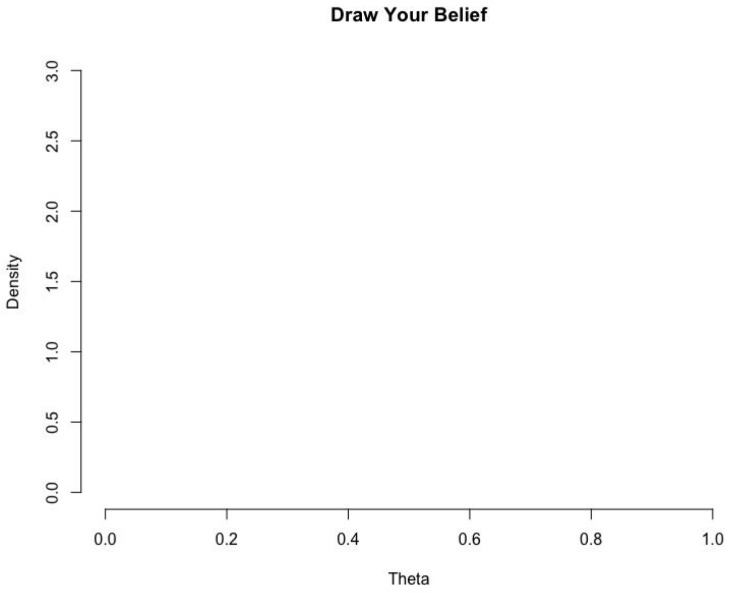
15:00

🎲 Bayesian parameter estimation

🪙: theta = p(heads)

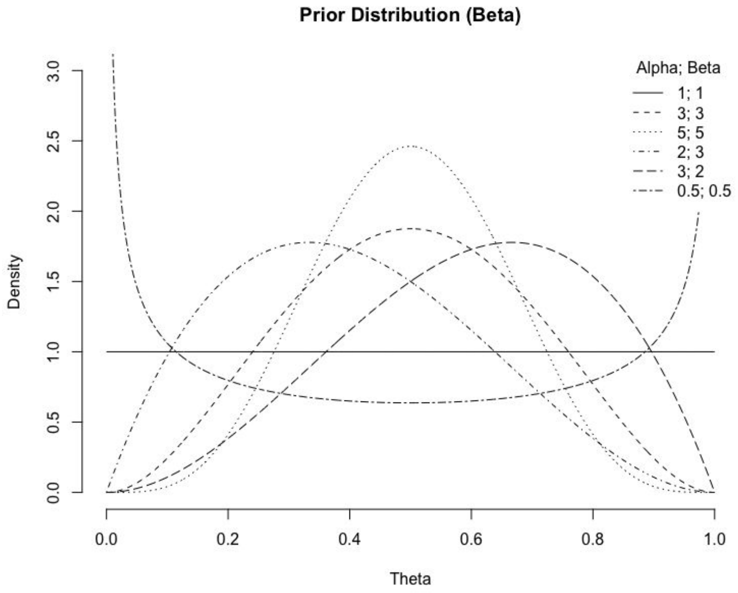
"Which values of theta *are good estimates (H), given my data (D)?"*

p(H | D) = (p(H) × p(D | H)) / p(D)



💭 How to get fair results from a biased coin?

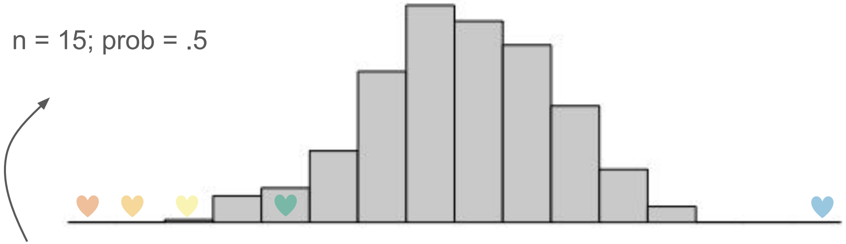
Prior distribution p(H)

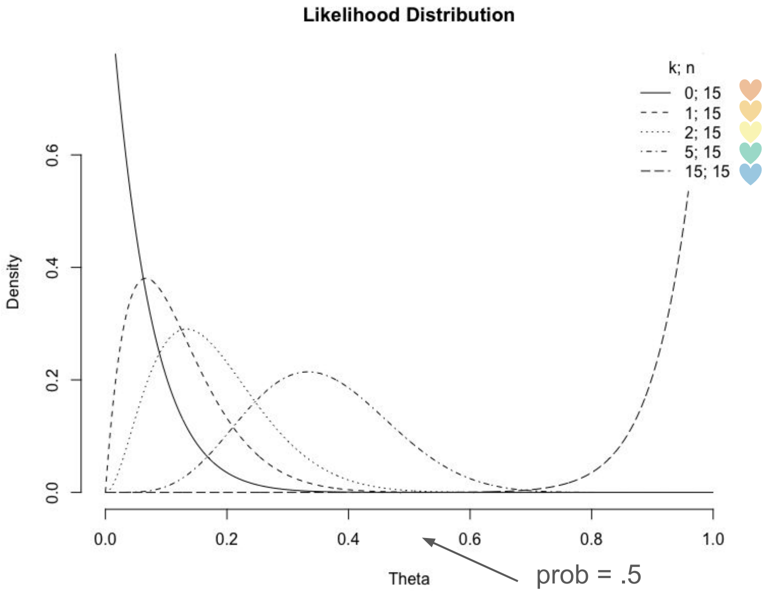


💭 Beta distribution.

- Prior belief/information
- Uninformative/informative
- Weakly/strongly informative
- Skeptical
- Point-valued

Likelihood distribution p(D | H)





🧑‍💻 dbinom(x = k, size = n, prob = seq(0, 1, .01))

🛠️ Web simulations by [Seeing Theory](#).

Posterior distribution p(H | D)

p(H | D) = (p(H) × p(D | H)) / p(D)

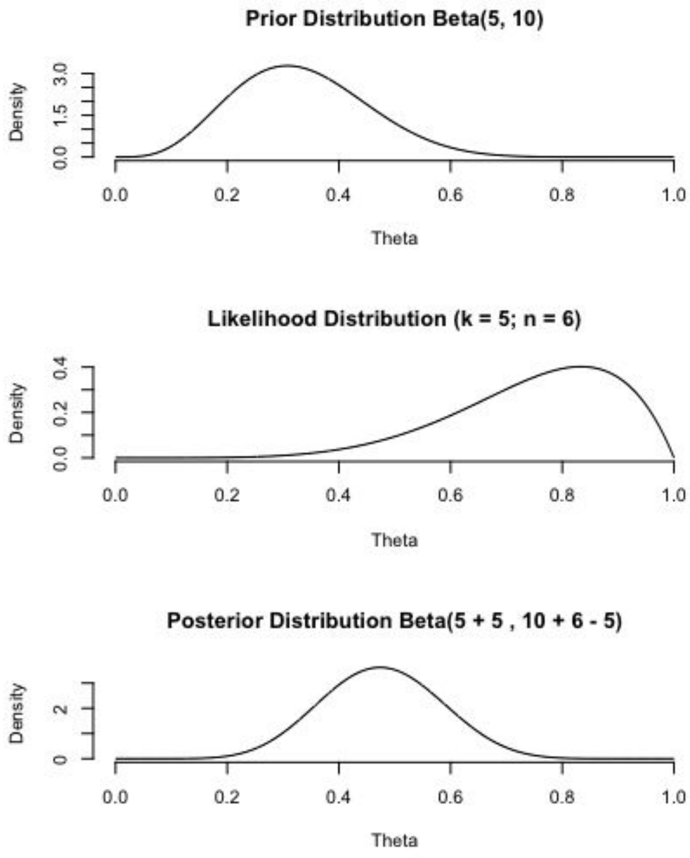
p(D) (marginal likelihood distribution): “integral of doom” 💀

But, beta distribution is *conjugate prior* of binomial distribution:

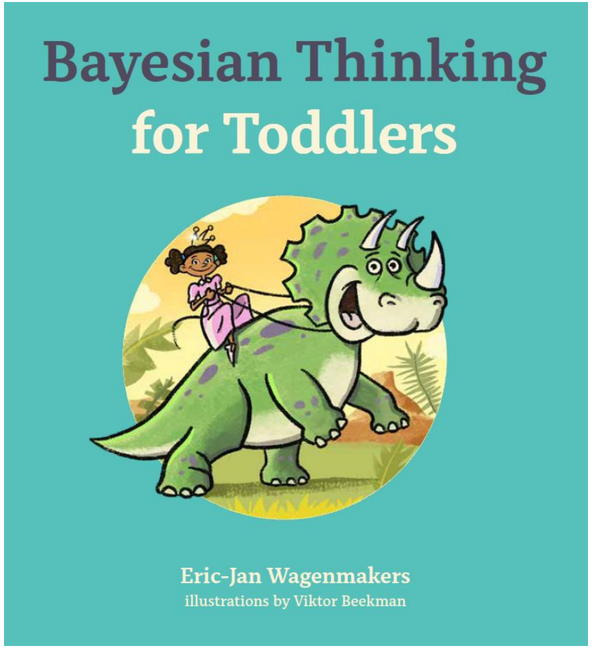
For a prior theta~Beta(a,b) and data *k* and *N*, the posterior is theta~Beta(a+k,b+N−k).

Repeat! Iterative!

🛠️ Web simulations by Kristoffer Magnusson and Seeing Theory.



💩 Bayesian thinking for toddlers



(grab a free pdf)

A brief introduction to Bayesian inference by Johnny van Doorn.

Bayesian and Frequentist inference side-by-side (Shiny app by John Kruschke, video tutorial by Eero Liski).

JASP

How to use JASP: Lots of resources.

Cooling down

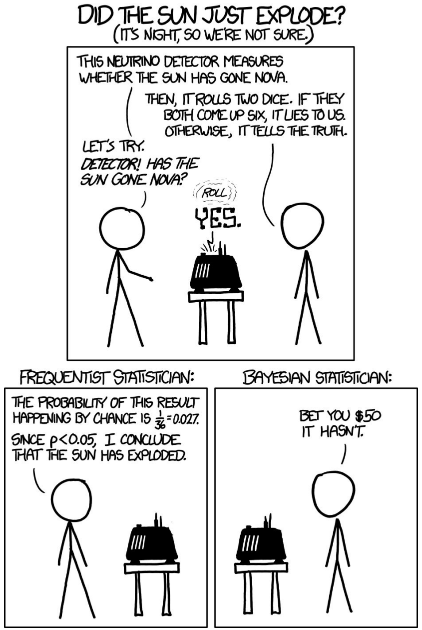


Illustration by Randall Munroe (wtf / controversy)

🔁 What did we learn?

🔮 What have we learned?

Topics

Statistical reasoning
Empirical cycle
Probability distributions
Frequentist inference
Sample / sampling distribution
Central limit theorem
Normal distribution
P-value
Type I/II errors
Effect size
Confidence interval
Power
Test statistics
Linear regression
t-Test
Moderation
F-distribution
Nonparametric inference
ANOVA
Bayesian inference

Illustration by Jennifer Cheuk

🧑‍💻 Take-home assignments

📆 Weekly assignment

💯 Exam

SR

- Open and closed questions
- Practical and theoretical questions
- R
- JASP output (no access to software)
- No follow-up questions (like in weekly assignments)
- Majority of questions

PoS

- *Ask Steven, previously essay questions*

🎬 Fin

'How can we use statistical inference to learn about the world?'

- Expand foundation for follow-up courses and independent learning.
- Expand knowledge and skills in statistics and statistical reasoning, develop statistical intuition, avoid common pitfalls and fallacies, and so on.

📅 Weekly assignments 🧑‍💻🛠️
🧑‍🏫 Lectures 💡💭
🧑‍🏫 Lecture slides 🗃️
☐ Evaluations 👎 👍

👋

Courses

- Improve Your Statistical Inferences / Questions (Daniel Lakens)
- Learning Statistics with R (Danielle Navarro)

Blogs

- Statistical Modeling, Causal Inference, and Social Science (Andrew Gelman et al.) (recommendations)
- Statistical Thinking (Frank Harrell)
- Bayesian Spectacles (Eric-Jan Wagenmakers et al.)
- Error Statistics Philosophy / PhilStatWars (Deborah Mayo)
- Simply Statistics (Jeff Leek, Roger Peng, Rafa Irizarry)
- The 20% Statistician (Daniel Lakens)
- Data Colada (Uri Simonsohn, Leif Nelson, Joe Simmons)
- Doing Bayesian Data Analysis (John Kruschke)
- R-Bloggers

Podcasts

- Nullius in Verba (Smriti Mehta, Daniel Lakens)

Colophon

Slides
alexandersavi.nl/teaching/

License