

An Introduction To Bayesian Statistics



Simon Thornewill von Essen

Data Analyst, Goodgame Studios

@sthornewillve



How do we estimate the probability?

- **Classical:** By considering equal outcomes
- **Frequentist:** Relative Frequency over time
- **Bayesian:** By updating our beliefs for each obs.

Coin Toss: Classical Est.



Dice: Classical Est.



Classical Stats

- Requirements
 - All Outcomes are known
 - Outcomes are assumed to be equally likely
- Advantages
 - Fast Estimation
 - Easy to understand
- Disadvantages
 - High Bias
 - Outcomes must be known
 - Cannot create sophisticated (high variance) models

How do we estimate the probability?

- ~~Classical~~
- Frequentist
- Bayesian

Thermometer Calibration: Frequentist Est.

- Calibrating Thermometer to show accurate values
- Follows a Normal Distribution

Frequentist Approach:

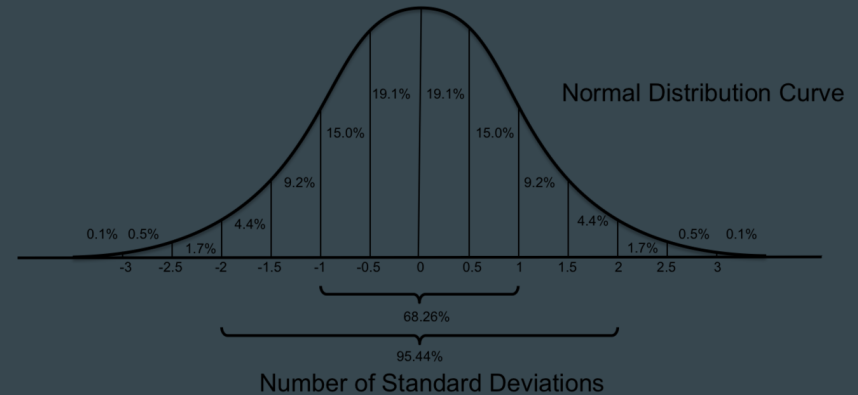
Take many readings and use the expectation value (mean) to find value over time.



Thermometer Calibration: Frequentist Est.

Confidence Interval:

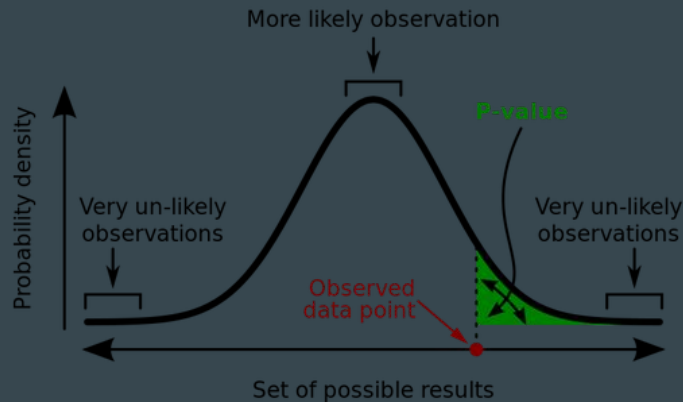
- From sample mean and standard deviation, calculate an interval
- “Interval that contains the true parameter some percent of the time”



Probability of Rain: Frequentist Est.

P-value:

- Probability of data given a parameter
- “The probability that outcome is due to random chance given that there is no difference between experimental groups”
- $P(X | \mu)$



Thermometer Calibration: Test

- 1. Mean thermometer temp is higher than assumed param,
P-value = 0.001 (highly significant),

Does this mean that the probability of mean thermometer temp is 0.999? ❌

- 2. 95% Confidence interval is [98°C, 102 °C] and mean = 100°C,

Does this mean that 100°C will fall inside this interval 95% of the time? ❌

Thermometer Calibration: Test

- 1. Mean thermometer temp is higher than assumed param,
P-value = 0.001 (highly significant),

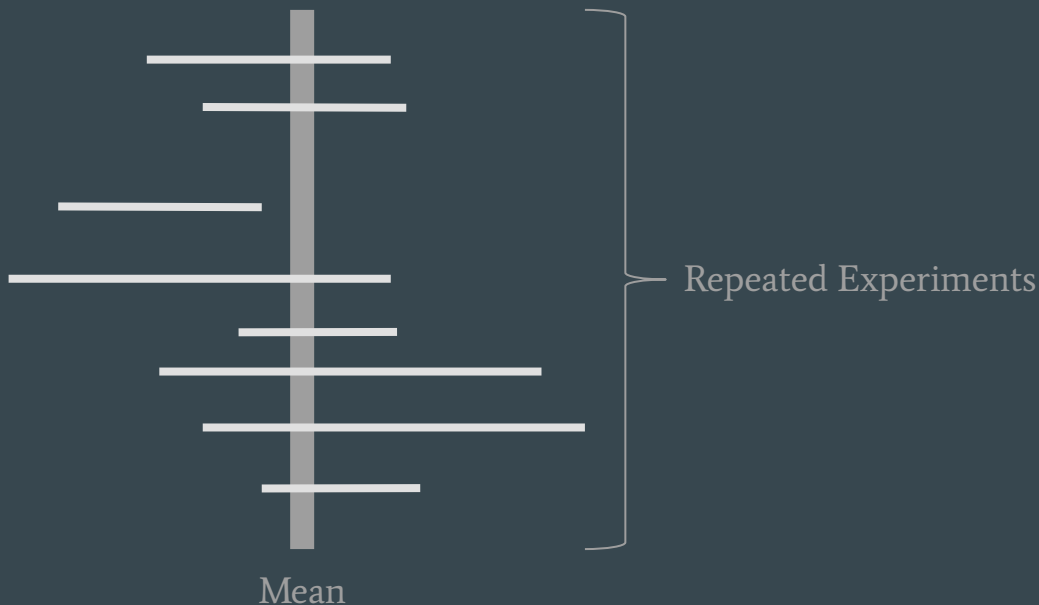
Probability of getting this result given no difference in experimental groups is 0.001

- 2. 95% Confidence interval is [98°C, 102 °C] and mean = 100°C,

Interval will contain the parameter 95% of the time

Thermometer Calibration: Test Learnings

i Frequentism expects that parameters exist and are fixed, the probabilities are the likelihood of our data given these expectations



Thermometer Calibration: Test Learnings



Child doesn't move, but you will only take a picture of them 95% of the time

Frequentist Stats

- Requirements
 - Possibility to perform experiments indefinitely
 - Parameters are assumed to be specific values
 - Able to estimate params given enough experiments
- Advantages
 - Works well for simulations
 - “Objective”
- Disadvantages
 - Requires large sample size
 - Does not allow for integration of domain knowledge
 - P-values and confidence intervals are unintuitive
 - Difficult to communicate

Frequentist Stats Disav. Cont.

What if?

- Amount of data you have is limited? ✓
- You have relevant and applicable prior information ✓
- “Infinite” experiments are not possible? (Cost, feasibility) ✓
- Stakeholders have a hard time understanding frequentist logic? ✓
- Children never stay still and assuming they don't is blasphemy ✓

How do we estimate the probability?

- ~~Classical~~
- ~~Frequentist~~
- Bayesian

Bayes Theorem

How

Foreshadow MCMC

Conclusion and “Call to Action”

Find Slides on Github

<https://cutt.ly/zGqux9>



A screenshot of the GitHub profile page for Simon Thornevill von Essen. The profile includes a profile picture, a bio, and a list of pinned repositories. The pinned repositories are: 'Udacity-DataScience-Nanodegree', 'Udacity-DataAnalyst-Nanodegree', 'Pet-Project---Bodybuilding-WFPB-Diet', 'Pet-Project---Tygem-Fuseki-Web-Scraper-using-Python', 'Udemy_LazyProgrammer_Courses', and 'Hamburg.DS.Meetup.Bayesian-Stats_Intro'. The 'Hamburg.DS.Meetup.Bayesian-Stats_Intro' repository is highlighted with a yellow background. Below the pinned repositories, there is a section for '178 contributions in the last year' with a calendar grid showing contributions from July to July. The grid shows a pattern of green squares indicating contributions, with a higher density in the latter half of the year. The grid is labeled with days of the week (Mon, Wed, Fri) and months (Jul, Aug, Sep, Oct, Nov, Dec, Jan, Feb, Mar, Apr, May, Jun, Jul). A link to 'Learn how we count contributions.' is provided at the bottom left of the grid, and a 'More' link is at the bottom right.

Find me on Social Media!



@sthornewillve

