STRINGFEST ANALYTICS

LEARNING STATISTICS IN EXCEL

George Mount

Data Analyst & Educator at Stringfest Analytics

George works as an independent analyst and data analytics educator with the goal to help clients manage their data so they think more creatively. He serves as a technical expert and lead curriculum developer for Thinkful's data analytics program and is the instructor of the DataCamp course "Survey and Measure Development in R."

George blogs about data, innovation, and career development at georgejmount.com. He holds a master's degree in information systems with a certificate of achievement in quantitative methods from Case Western Reserve University

OBJECTIVES FOR TODAY

- Normality explained
- Normality and statistical inference
- Sampling and the margin of error
- Frequentist versus Bayesian probability
- Next steps for learning statistics

PREREQUISITES

- Familiarity with descriptive statistics (mean/median/mode, standard deviation/variance)
- Experience with intermediate Excel functions (`MATCH()`, nested `IF()` statements, etc.)
- Ability to insert and modify Excel charts

FOLLOWING ALONG

- Each section is a sub-folder

- Demos = follow along with me
  - Refresh your memory with the demo notes

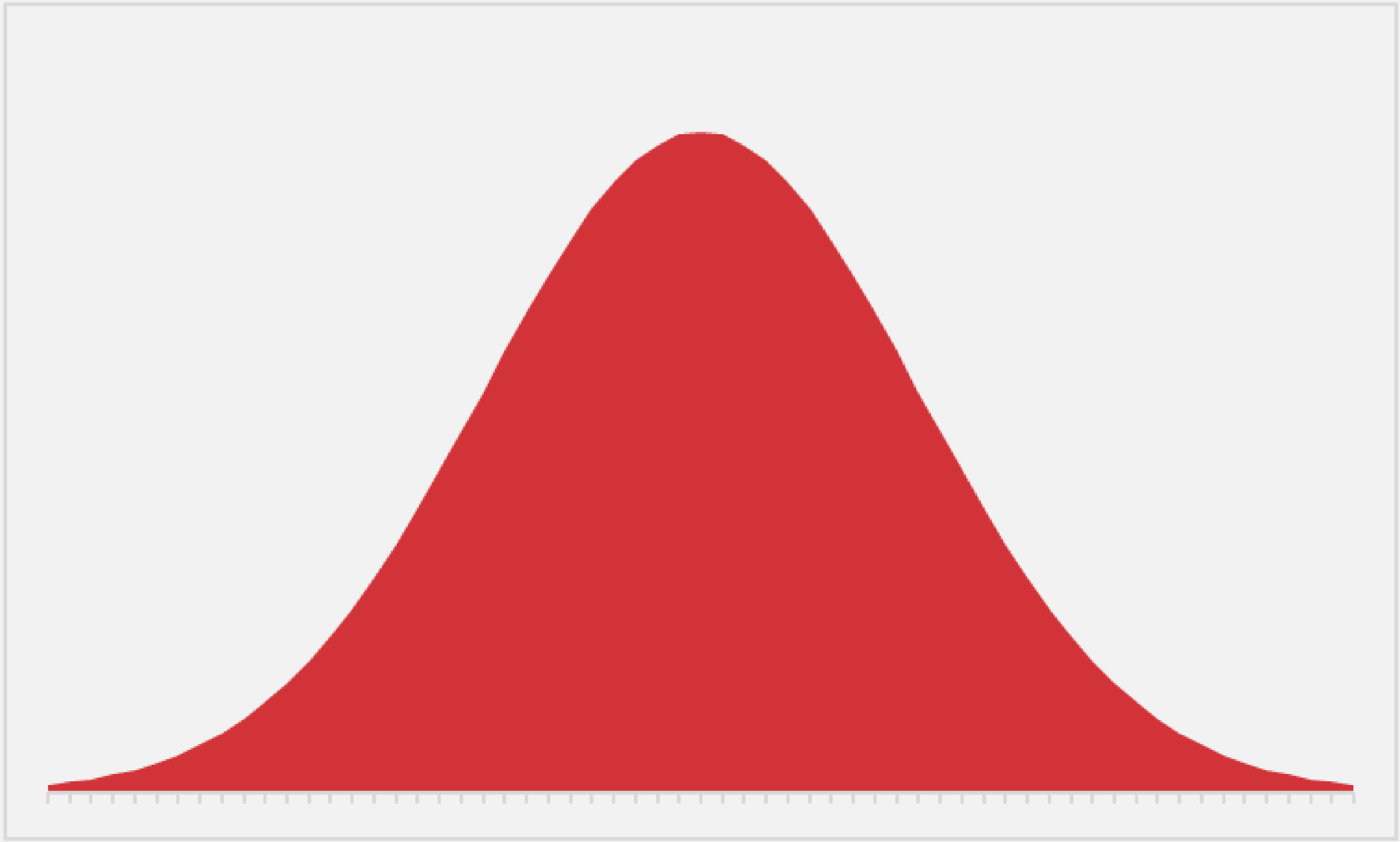
WHY WOULD WE DO THIS IN EXCEL?

"You get to look at the data every step of the way, building confidence while learning the tricks of the trade."

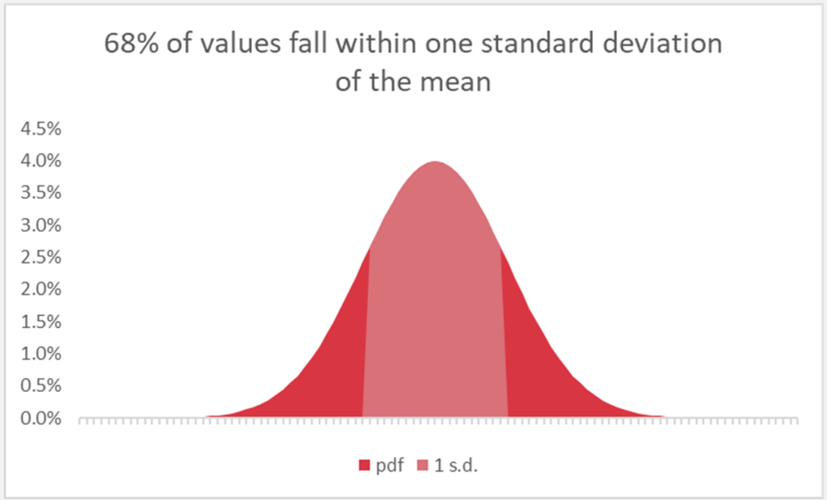
-- John Foreman

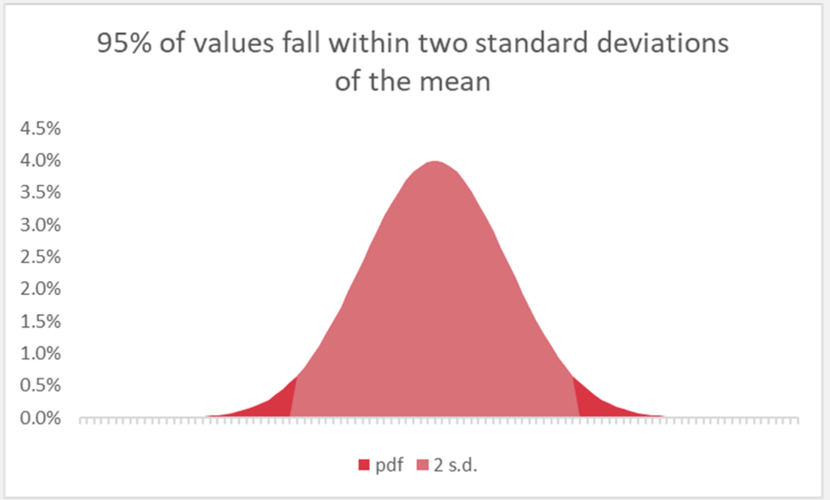
1. NORMALITY EXPLAINED

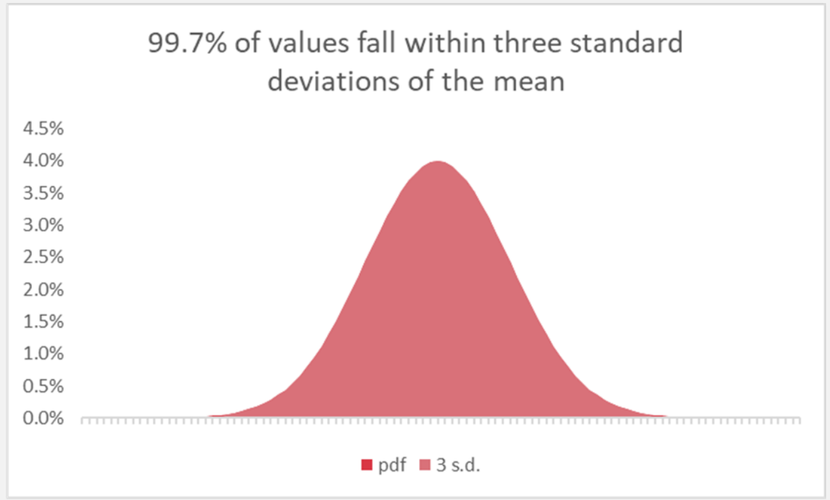
EVER WANTED TO JUST BE NORMAL? STATISTICS IS YOUR CHANCE.



WHAT DOES IT MEAN TO BE NORMAL?
“Empirical rule”







DEMO

- Demo: `empirical-rule.xlsx`
  - *Probability density function* : tells us what percent of values we expect to find within a given interval of a distribution
    - e.g. We would expect to find about 2% of values ranging between 34 and 42 for a normally-distributed variable with a mean of 50 and standard deviation of 10

QUESTIONS?

2. NORMALITY AND STATISTICAL INFERENCE

STATROULETTE

- A roulette wheel returns values between 0 and 36.
- Let’s simulate a game of roulette in Excel

DEMO

- `central-limit.xlsx`

  First:

  - Simulate 500 rounds of a roulette spin.
  - Plot the resulting frequency distribution as a histogram.

  Then:

  - Simulate a roulette spin 100 times.
  - Take the average spin.
  - Do this for 500 trials.
  - Plot the distribution of trial means as a histogram.

MAGIC… OR STATISTICS?

- Central limit theorem: the sampling distribution of the mean of any independent, random variable will be normal or nearly normal, *if the sample size is large enough.*

How “large enough” is large enough?

- N = 30? 60? 100?
- It depends on how “normal” your sample is

DEMO

- `large-numbers.xlsx`
- A roulette wheel returns values between 0 and 36.
- What is the average roulette spin given more and more spins?

MAGIC… OR STATISTICS?

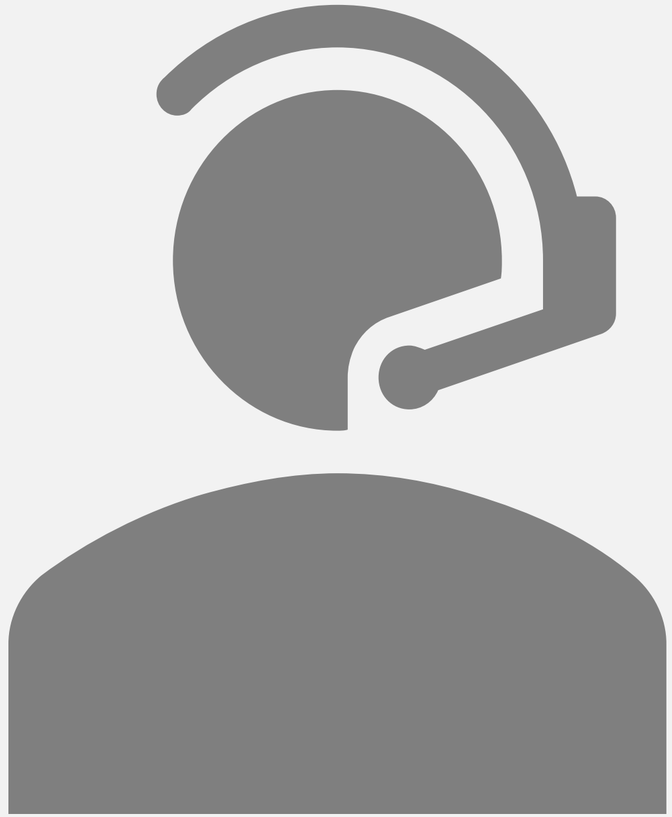
- Law of large numbers: the average of results obtained from trials become closer to the expected value as more trials are performed

QUESTIONS?

3. SAMPLING AND THE MARGIN OF ERROR

DEMO

- File: `margin-of-error.xlsx`
- Pollsters often report a “margin of error of +/- 2-3%.”
  - *What does that mean?*



QUESTIONS?

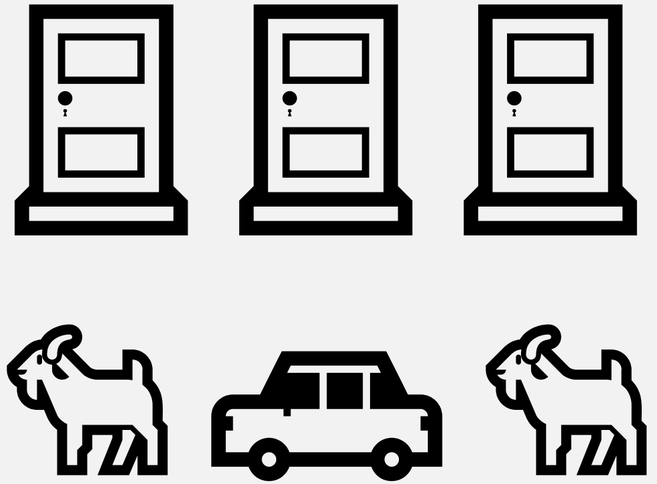
4. FREQUENTIST AND BAYESIAN PROBABILITY

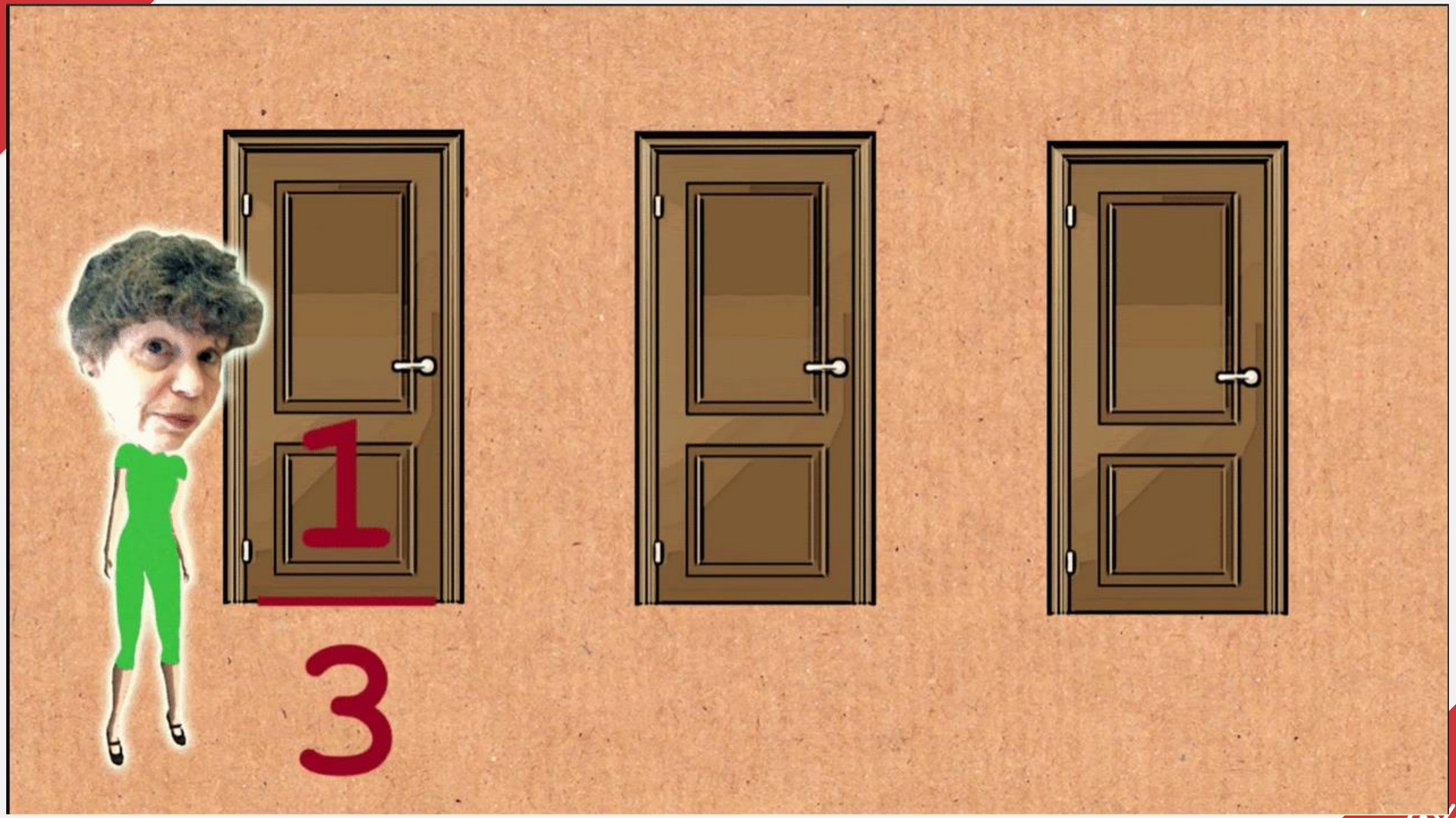
Statistical concept	Frequentist
Variables	Random and deterministic
Parameters	Fixed, unknown constants
Estimators	Should be good when averaged across many trials

LET'S MAKE A DEAL
1
2
3 A B C
PICK A DOOR!
YES...I PLAYED
LET'S MAKE A DEAL!!

DEMO

- `monty-hall.xlsx`
- Three doors: one car, two goats
- You pick a door
- Monty opens another door: it has a goat
- Do you stick to your door, or switch doors? Does it matter?





https://www.youtube.com/watch?v=4Lb-6rxZxx0

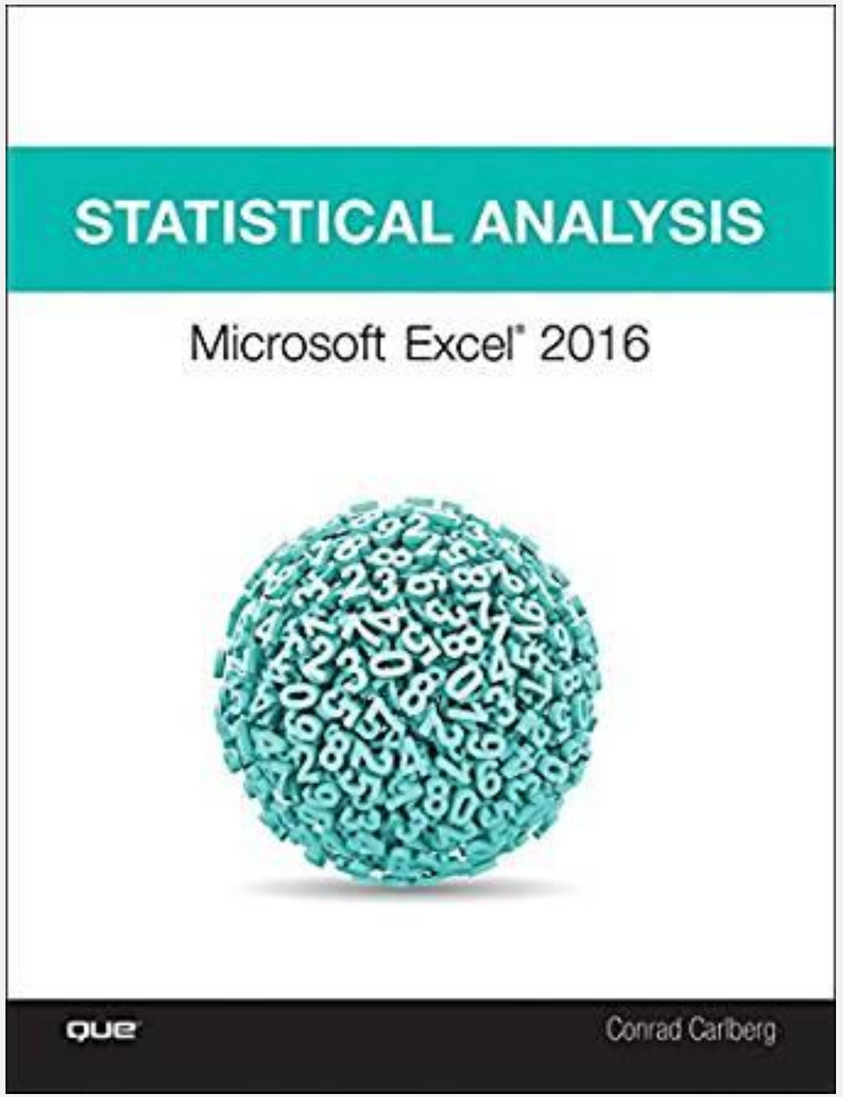
Statistical concept	Frequentist	Bayesian
Variables	Random and deterministic	Everything is random
Parameters	Fixed, unknown constants	Subjective belief
Estimators	Should be good when averaged across many trials	Should be good for the available data

QUESTIONS?

4. CONCLUSION

Statistical Analysis: Microsoft Excel 2016, by Conrad Carlberg

- On O'Reilly Learning at https://learning.oreilly.com/library/view/statistical-analysis-microsoft/9780134840437/

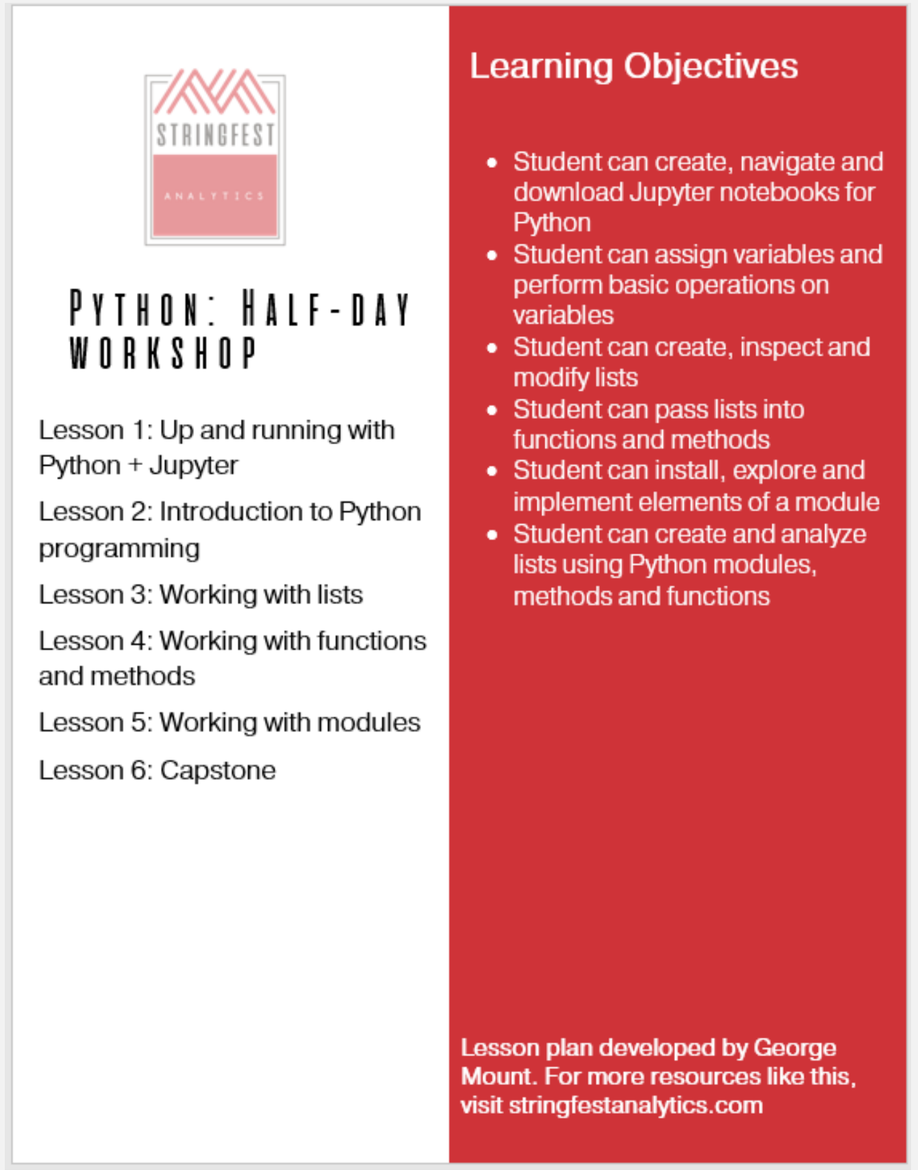


Data Smart: Using Data Science to Transform Information into Insight, **by John Foreman**

- On O'Reilly Learning at https://learning.oreilly.com/library/view/data-smart-using/9781118661468/

Stringfest Analytics Resource Library

- stringfestanalytics.com



STRINGFEST ANALYTICS

LET'S TALK

LINKEDIN

linkedin.com/in/gjmount

EMAIL ADDRESS

george@stringfestanalytics.com

WEBSITE

stringfestanalytics.com

GITHUB

github.com/summerofgeorge

QUESTIONS?