Hypothesis Testing

Data Science Immersive

How can we know things?

?

How can we know things?

- Even more fundamental question:

What does it mean to know something?

Quick Philosophy

"All I know is that I know nothing"

"I think therefore I am"

"A wise man proportions his belief to the evidence"

Principles of Empiricism

- Knowledge is *a posteriori*

- Proportion beliefs to evidence

- *Tabula Rasa*

- Ok to use knowledge from experience to synthesize new knowledge

Hypothesis formulation

- Any scientific study begins with a hypothesis, then a conclusion is eventually made about this hypothesis.

- Following *tabula rasa*, if we cannot prove something, we assume that it's untrue by default.

H_0 - Null Hypothesis H_A - Alternative Hypothesis

Proportioning Beliefs with the Evidence

- Error
  - I: False positive rate
  - II: False negative rate

- P-value
  - We make an observation. We are able to calculate (with statistics!) a degree of uncertainty.

- Choosing the right error rate
  - Alpha, α
  - Sigma, σ
  - Depends on field of study, $0.2 \leq \alpha \leq 0.00001$
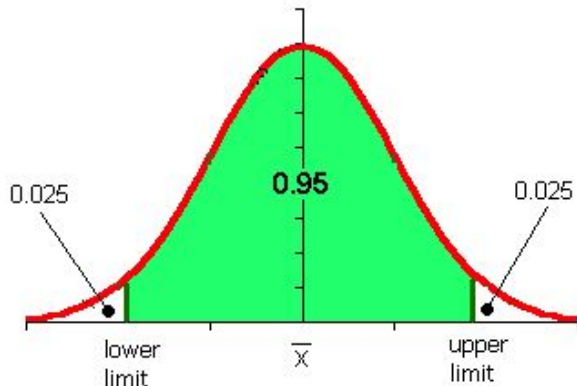 - 0.05 is the most common by far.

Example
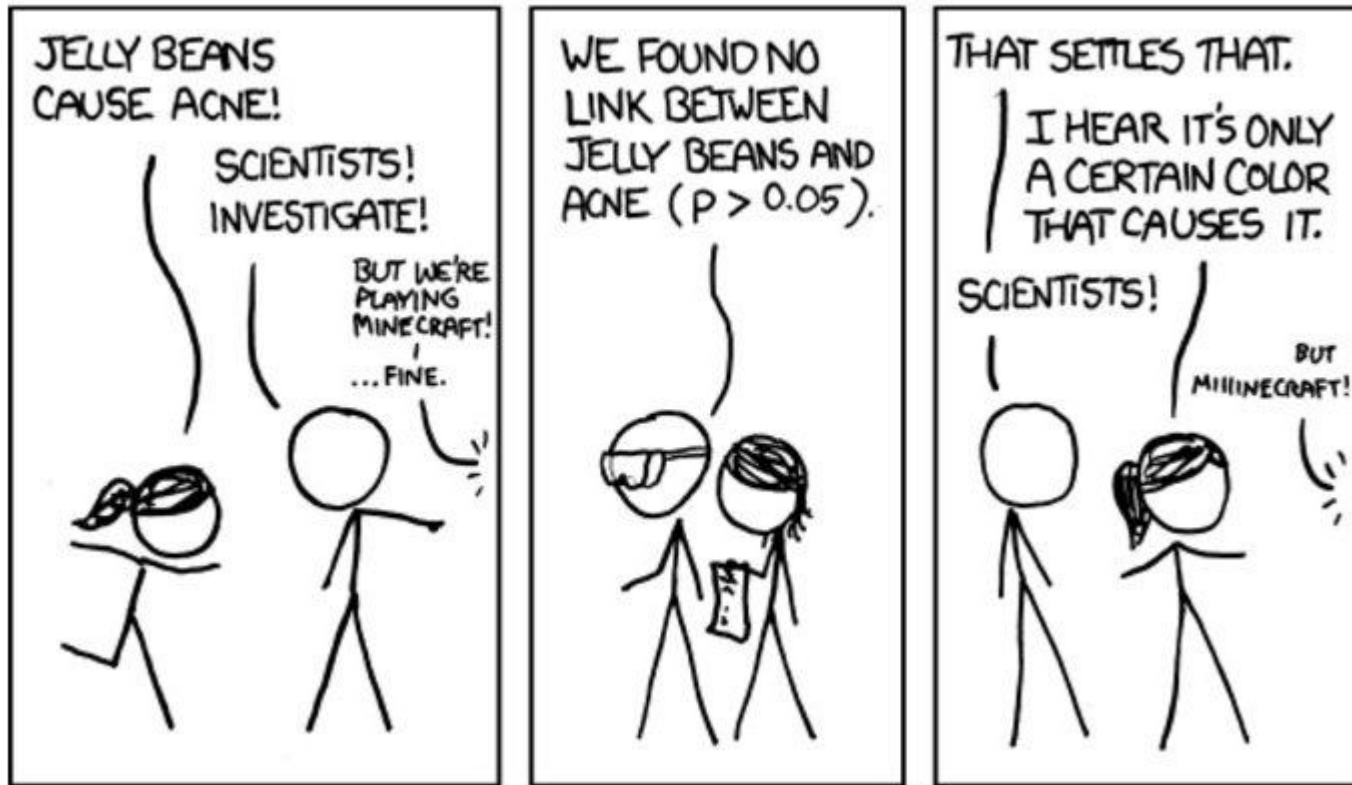
Revisiting Confidence Intervals

Our level of confidence that if we obtained a sample of equal size through the same process, our sample would contain the population mean.

IT IS NOT: The % chance the population mean lies within our sample interval. (Many people will say this!)

Any guesses what this measure will be???
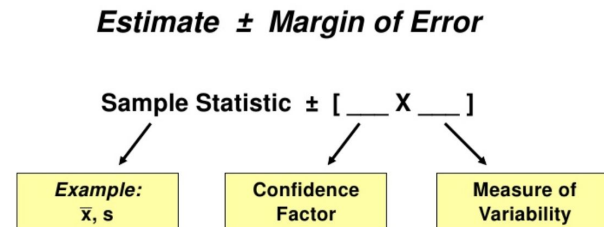

Estimate ± Margin of Error

Sample Statistic ± [___ X ___]

| Example: x̄, s | Confidence Factor | Measure of Variability |

Revisiting Confidence Intervals

Assuming a 95% confidence interval….

Estimate ± Margin of Error

Sample Statistic ± [___ X ___]

| *Example:* $\bar{x}$, s | Confidence Factor | Measure of Variability |

If we know population variance

$$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

If we do not know population variance

$$\bar{x} \pm 1.96 \frac{s}{\sqrt{n}}$$

If we have a small sample size (generally n < 100)

$$\bar{x} \pm t_{(\alpha/2, n-1)} \frac{s}{\sqrt{n}}$$

What would $t_{alpha/2}$ be if we had a sample size of 25?

Conclusion

- Established a framework for knowledge that is practical and sensible for the scientific process.

- Provided some meaning behind "error" and "confidence"

- Next up - doing some hypothesis testing with our first statistical tests, z and t tests