Mathematics and Problem Solving

Lecture 11.2

Hypothesis Testing

Starting with a question

- Often you’ll come to data with a question:
  - “Do players prefer character A or character B?”

Hypotheses

- To answer this with statistics, you need to turn it into **two** hypotheses:
  - **Null** Hypothesis
  - **Alternate** Hypothesis

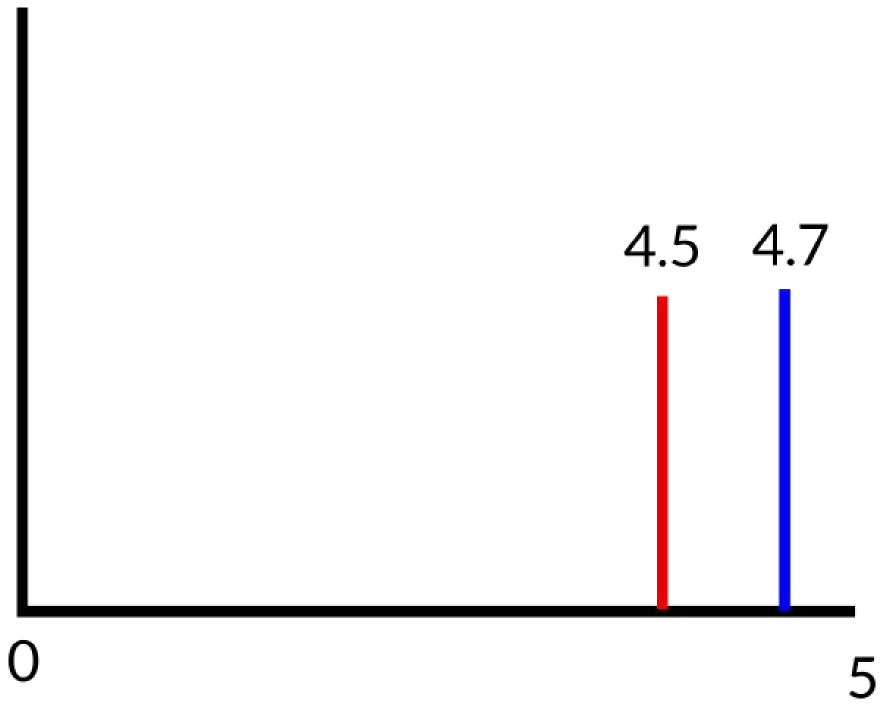
Null Hypothesis

- The **null hypothesis** says that
  - whatever results you got were the result of chance
  - i.e. nothing interesting is happening
- e.g. “Players like both characters more or less the same”

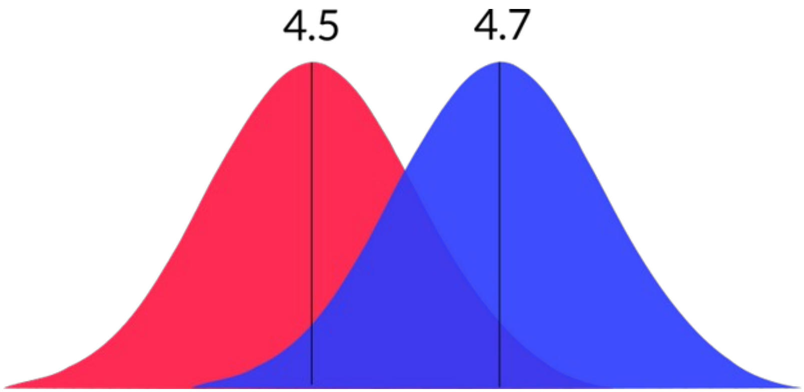
Alternate Hypothesis

- Your **alternate hypothesis** is what you’re trying to prove
  - e.g. “Players either prefer Character A or they prefer character B”
  - This is the hypothesis that we’re going to test

- Now, imagine we find the mean score for characters A and B to be 4.5 and 4.7
- Can we confirm our hypothesis?
  - No, because we only **sampled** our population, the difference might be chance
- Inferential statistics tell you how likely it is that the effect you observe is the result of chance.



- Instead, imagine we have the distributions of the results.
  - Now can we confirm our hypothesis?



Statistics

- Key values
  - Test statistic (varies)
  - p
  - α
  - Effect size (if possible)

Statistical Significance

- p = probability the effect observed was chance
  - $p < \alpha$
    - statistically significant
  - $p > \alpha$
    - not statistically significant

Alpha

- **Alpha** (α) is usually set at 0.05
- Reduced for multiple testing

Effect Size

- **How big** an effect is
  - e.g. Cohen's *d*
    - $d = 0.2$ is small
    - $d = 0.5$ is medium
    - $d = 0.8$ is large
- Important for interpreting results

Type I error

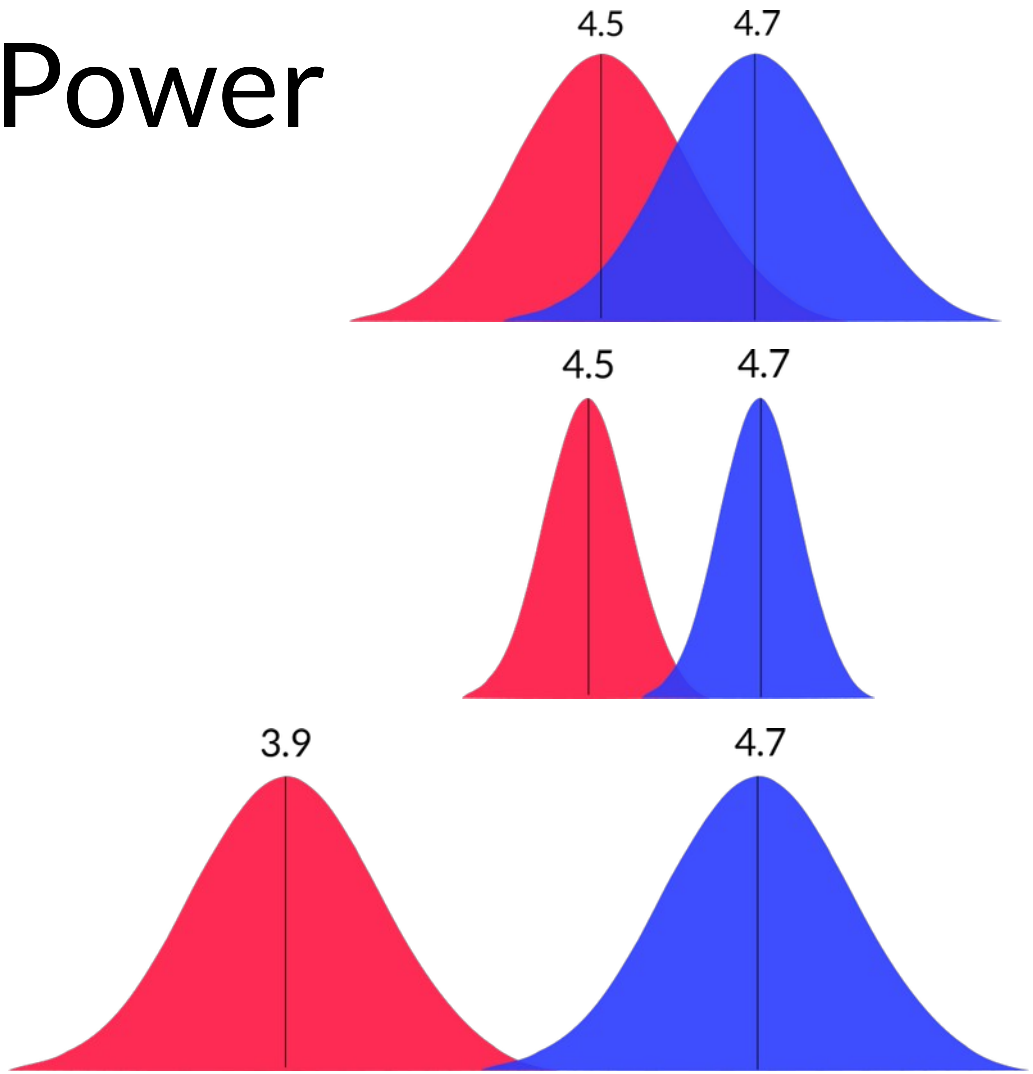
- Falsely **rejecting** the **null hypothesis**
  - we **claim an effect** when there is **no effect**
- If $\alpha = 0.05$, we have a 5% chance of making a Type I error.

Type II error

- Falsely **accepting** the **null hypothesis**
  - **there is an effect**, but you **don’t detect it**
- The probability of making a type II error is called **beta** (β)
  - This is related to your statistical **power**. (= 1-β)
  - A good **power** is 0.8 (β = 0.2)

Power

- Related to
  - sample size
  - effect size
  - spread of data



In closing

- Inferential statistics = **lots of numbers** about your data
- It's **frighteningly easy** to make mistakes
  - If that is in science that people rely on, that's a **big problem**