Hypothesis Testing

And Causal Inference

Probability and Statistics for Data Science

Carlos Fernandez-Granda

These slides are based on the book Probability and Statistics for Data Science by Carlos Fernandez-Granda, available for purchase here. A free preprint, videos, code, slides and solutions to exercises are available at https://www.ps4ds.net

Goal

Discuss relationship between hypothesis testing and causal inference

Antetokounmpo's free throws

Conjecture: Free throw percentage is higher at home than away

Null hypothesis: Percentage is the same

Test statistic:

$$\frac{\text{Made at home}}{\text{Attempted at home}} - \frac{\text{Made away}}{\text{Attempted away}}$$

Significance level: $\alpha := 0.05$

P value: $0.011 < \alpha$

Antetokounmpo's free throws

P value: $0.011 < \alpha$

What does this mean?

Free-throw outcome: $\tilde{y}$

Fans taunting: $\tilde{t}$

We're pretty sure that

$$p_{\tilde{y} \mid \tilde{t}}(\text{made} \mid \text{no taunt}) > p_{\tilde{y} \mid \tilde{t}}(\text{made} \mid \text{taunt})$$

Does this mean taunting causes free-throw % to decrease? No!

Could be due to confounding factors

Evaluating NBA players

Goal: Evaluate impact of a player on team performance

Statistic: Difference of mean point differential with/without player

$$t_{\text{data}} := m_{\text{with}} - m_{\text{without}}$$

Permutation test with Bonferroni's correction

What does this mean?

	Mean point diff.	P value	Mins per game
L. James (CLE)	16.7	$< 10^{-7}$	36.6
B. Caboclo (TOR)	16.4	$< 10^{-7}$	4.6
N. Mirotic (CHI)	10.3	$3 \cdot 10^{-7}$	23.1
C. Anthony (NY)	8.1	$5 \cdot 10^{-7}$	36.3
Ricky Rubio (MIN)	7.6	$7 \cdot 10^{-7}$	31.4
James Jones (MIA)	8.2	$6 \cdot 10^{-6}$	7.8
Brandon Rush (GS)	6.7	$6 \cdot 10^{-6}$	12.6
Joel Embiid (PHI)	8.7	$2 \cdot 10^{-5}$	28.7

We're pretty sure that

Conditional mean with $>$ Conditional mean without

Does this mean the player causes the increase?

No!

	Mean point diff.	P value	Mins per game
L. James (CLE)	16.7	$< 10^{-7}$	36.6
B. Caboclo (TOR)	16.4	$< 10^{-7}$	4.6
N. Mirotic (CHI)	10.3	$3 \cdot 10^{-7}$	23.1
C. Anthony (NY)	8.1	$5 \cdot 10^{-7}$	36.3
Ricky Rubio (MIN)	7.6	$7 \cdot 10^{-7}$	31.4
James Jones (MIA)	8.2	$6 \cdot 10^{-6}$	7.8
Brandon Rush (GS)	6.7	$6 \cdot 10^{-6}$	12.6
Joel Embiid (PHI)	8.7	$2 \cdot 10^{-5}$	28.7

Played 24 games over 4 years (missing 200)

Were Raptors winning *because* Caboclo was playing?

Caboclo was playing because Raptors were winning

Causal inference

To identify causal effect, outcome and treatment must be independent

How can we achieve this? Randomizing the treatment

COVID-19 vaccine

43,448 patients randomly divided into

- Treatment group of 21,720 patients: 8 cases (0.037%)
- Control group of 21,728 patients: 162 (0.746%)

Two-sample z test

Null hypothesis: All data are i.i.d. Bernoulli with parameter θ_{null}

Test statistic: Infection rate without vaccine - Infection rate with

$$\text{pv}(t_{\text{data}}) = \mathrm{P}\left(\tilde{t}_{\text{null}} \geq t_{\text{data}}\right) < 10^{-23}$$

Causal inference vs hypothesis testing

Causal inference and hypothesis testing have complementary roles

Is there a difference between control and treatment groups?

Yes, the difference is statistically significant by the hypothesis test

Is the difference due to a causal effect?

Yes, because the trial is randomized

A/B testing

Goal: Comparing two options (A/B) when designing a product

- Users are randomly assigned to each option, so that differences reveal causal effect
- Hypothesis testing is applied to determine whether differences are statistically significant

Obama's presidential campaign

Options: Image or video on website

Metric: Sign-up rate

- Images: 14,016 out of 155,280
- Videos: 10,337 out of 155,102

Two-sample z test

Null hypothesis: All data are i.i.d. Bernoulli with parameter θ_{null}

Test statistic: |sign-up rate images − sign-up rate videos|

$$\text{pv}(t_{\text{data}}) = \text{P}\left(\tilde{t}_{\text{null}} \geq t_{\text{data}}\right) < 10^{-80}$$

Definitely statistically significant!

Difference reveals causal effect thanks to randomization

Does this imply practical significance? Not necessarily!

What have we learned

Statistical significance does not imply a causal effect

Randomization and hypothesis testing have complementary roles