Stat 88: Probability & Mathematical Statistics in Data Science

Lecture 29 : 4/5/2021

Section 9.2

A/B Testing

A/B testing

So far : hypothesis testing : One sample

- Data 8, section 12.3, randomized controlled trial to see if botulinum toxin could help manage chronic pain
- 31 patients → 15 in treatment group, 16 in control group. 2 patients in the control group reported pain relief and 9 in the treatment group.

- A/B testing is a (relatively recent) term used to describe hypothesis tests which involve comparing the distributions of two random samples. (Earlier we had *one* sample and made a hypothesis about its distribution.)
- In particular, we can conduct an A/B test for hypothesis tests involving results of randomized controlled trials.

Fisher's exact test

31 people took part in study

SRS out of 31

- Control group: 16 patients, 2 reported relief
- Treatment group: 15 patients, 9 reported relief

31 patients

- H_0: The treatment has no effect (there would have been 11 patients reporting pain relief no matter what, and it just so happens that 9 of them were in the treatment group)
- H_A = H_1: The treatment has an effect

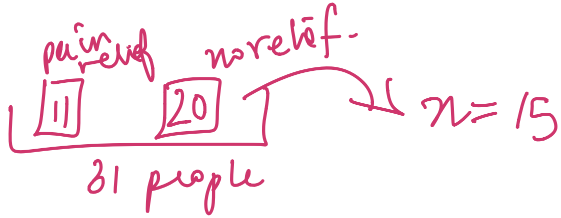
Test statistic: X = # of treated patients who had relief.

SRS of 15 people out of 31 assigned to the treatment gp.

If H_0 is true, the treatment has no effect & the popn. of 31 people has 11 successes, 20 failures

Observed value of X = 9

Example continued



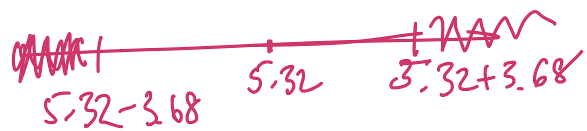
$$X \sim HG(31, 11, 15)$$

$$\mathbb{E}(X \mid H_0) = 15 \cdot \frac{11}{31} \approx 5.32$$

$T = |X - 5.32|$, observed value of $T = 9 - 5.32$
$= 3.68$

$$\text{P-value} = P(T \geq 3.68)$$

$$= P(|X - 5.32| \geq 3.68)$$



$$= P(X \leq 5.32 - 3.68) + P(X \geq 5.32 + 3.68)$$

$$= \underbrace{P(X \leq 1.64)}_{P(X \leq 1)} + P(X \geq 9)$$

$$P(X \leq 1) + P(X \geq 9) \approx 0.00915.$$

Example: The Lady Tasting Tea

- The first person to describe this sort of hypothesis test was the famous British statistician Ronald Fisher. In his book *The Design of Experiments*, he describes a tea party in which a lady of his acquaintance claimed that she could tell from tasting a cup of tea if the milk had been poured first or the tea.
- Fisher immediately set up an experiment in which she was given multiple cups of tea and asked to identify which of them had had the tea poured first. She tasted 8 cups of tea, of which 4 had the tea poured first, and identified 3 of them correctly. Does this data support her claim?

Actually poured first

Lady says	Tea	Milk	
Tea first	3	1	
Milk first	1	3	
	4	4	8 N

H_0: The lady cannot tell & guesses at random.

H_1: She can tell when tea was poured first.

X = # of correct guesses

N = 8, n = 4, G = 4, obs. value of X = 3

P-value: $\underline{P(X \geq 3)} \approx 0.243$

Example: Gender bias?

- Rosen and Jerdee conducted several experiments using male bank supervisors (this was in 1974) who were given a personnel file and asked to decide whether to promote or hold the file. 24 were randomly assigned to a file labeled as that of a male employee and 24 to a female.
- 21 of the 24 males were promoted, and 14 of the females. Is there evidence of gender bias?

	M	F	
promoted	21	14	35
not promoted	3	10	13
	24	24	48

H_0: No gender bias

H_1: There is a gender bias.

Let X = # of successes among males. $X \sim HG(48, 35, 24)$

Observed value = 21, $E(X|H_0) = 24 \cdot \frac{35}{48} = 17.5$

Our test statistic will see how far the observed value is from the expected value.

$T = |X - 17.5|$ Obs. value of $T = |21 - 17.5|$

p-value $= P(T \geq 3.5) = P(|X - 17.5| \geq 3.5)$

$$\underbrace{P(X \leq 17.5 - 3.5)}_{P(X \leq 14)} + P(X \geq 17.5 + 3.5)$$

$$\underbrace{P(X \leq 14)}_{\text{min value of } X \text{ is } 11} + P(X \geq 21)$$

$P(X \leq 14) + P(X \geq 21) \approx 0.0489$.

Reject the null.