Statistical Inference

Foundations for Inference

Behnam Bahrak
Spring 2020

Experiment on Bank Managers

		Promotion		
		Promoted	Not Promoted	total
Gender	Male	21	3	24
	Female	14	10	24
	total	35	13	48

% of males promoted = 21/24 ≈ 88%
% of females promoted = 14/24 ≈ 58%

Two Competing Claims

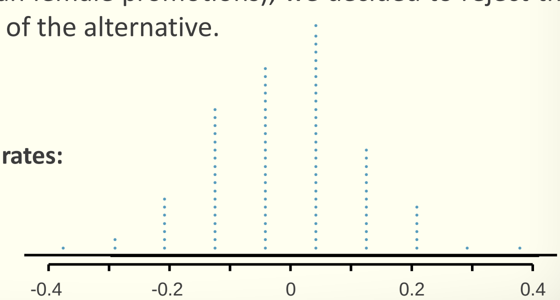
1. "There is nothing going on."
Promotion and gender are independent, no gender discrimination, observed difference in proportions is simply due to chance. → Null hypothesis

2. "There is something going on."
Promotion and gender are dependent, there is gender discrimination, observed difference in proportions is not due to chance. → Alternative hypothesis

Simulation-based Inference

➢ Since it was quite unlikely to obtain results like the actual data or something more extreme in the simulations (male promotions being 30% or more higher than female promotions), we decided to reject the null hypothesis in favor of the alternative.

Difference in promotion rates:



Recap: Hypothesis Testing Framework

- We start with a null hypothesis (H_0) that represents the status quo.
- We also have an alternative hypothesis (H_A) that represents our research question, i.e. what we're testing for.
- We conduct a hypothesis test under the assumption that the null hypothesis is true, either via simulation or theoretical methods — methods that rely on the CLT.
- If the test results suggest that the data do not provide convincing evidence for the alternative hypothesis, we stick with the null hypothesis. If they do, then we reject the null hypothesis in favor of the alternative.

Hypotheses

null - H_0	Often either a skeptical perspective or a claim to be tested	=
alternative - H_A	Represents an alternative claim under consideration and is often represented by a range of possible parameter values.	$<, >, \neq$

- The skeptic will not abandon the H_0 unless the evidence in favor of the H_A is so strong that she rejects H_0 in favor of H_A.

Hypothesis Testing via CI

- Earlier we calculated a 95% confidence interval for the average textbook costs that college students have paid in to be (270, 370). Based on this confidence interval, do these data support the hypothesis that college students on average have paid more than \$300 for textbooks.

H_0: $\mu = 300$ College students have paid \$300 for textbooks, on average.

H_A: $\mu > 300$ College students have paid more than \$300 for textbooks, on average.

always about pop. parameters, never about sample statistics

Hypothesis Testing via CI

H_0: $\mu = 300$ College students have paid \$300 for textbooks, on average.

H_A: $\mu > 300$ College students have paid more than \$300 for textbooks, on average.

- Since the null value is included in the interval, we do not reject the null hypothesis in favor of the alternative.
- This is a quick-and-dirty approach for hypothesis testing. However it doesn't tell us the likelihood of certain outcomes under the null hypothesis, i.e. the p-value, based on which we can make a decision on the hypotheses.

p-value

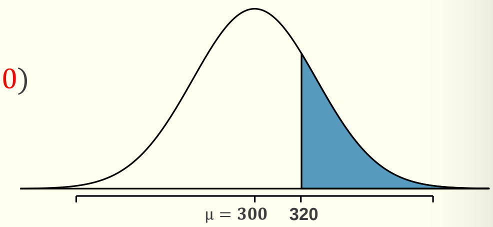
- p-value = P(observed or more extreme outcome | H_0 true)

For previous example:

p-value = $P(\bar{x} > 320 | H_0: \mu = 300)$

$s = 174\,, n = 50 \;\; \Rightarrow \;\; SE = 24.6$

$\bar{x} \sim N(\mu = 300\,, SE = 24.6)$



test statistics: $Z = \frac{320-300}{24.6} = 0.81$

p-value = $\mathrm{P}(Z > 0.81) = 0.209$

Decision based on the p-value

- We used the test statistic to calculate the p-value, the probability of observing data at least as favorable to the alternative hypothesis as our current data set, if the null hypothesis was true.

- If the p-value is low (lower than the significance level, α, which is usually 5%) we say that it would be very unlikely to observe the data if the null hypothesis were true, and hence reject H_0.

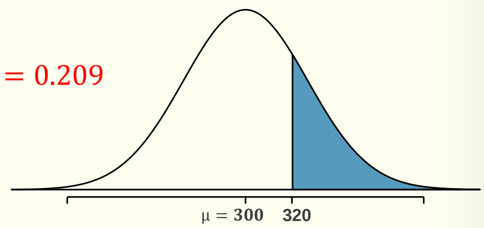
- If the p-value is high (higher than α) we say that it is likely to observe the data even if the null hypothesis were true, and hence do not reject H_0.

Example

p-value = $P(\bar{x} > 320 | \mu = 300) = 0.209$

significance level: $\alpha = 0.05$

p-value = $0.209 > 0.05$

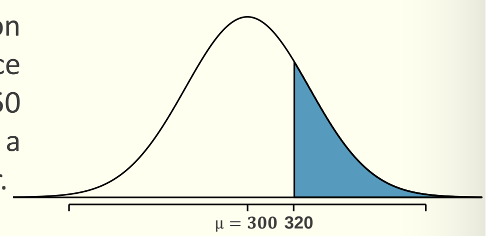


Since p-value is high, we do not reject H_0.

Interpreting the p-value

- If in fact college students have paid \$300 for textbooks on average, there is a 21% chance that a random sample of 50 college students would yield a sample mean of \$320 or higher.

p-value = $0.209 \approx 0.21$



- This is a pretty high probability, so we think that a sample mean of \$320 or more is likely to happen simply by chance.

احتمال اینکه μ جامعه واقعاً این بوده و دیدن این نمونه شانسی بوده چند درصد بوده؟

Making a Decision

- Since p-value is high (higher than 5%) we fail to reject H_0.
- The data do not provide convincing evidence that college students have paid more than $300 for textbooks on average.
- The difference between the null value of $300 textbooks cost and the observed sample mean of $320 is due to chance or sampling variability.

Two-sided Tests

- Often instead of looking for a divergence from the null in a specific direction, we might be interested in divergence in any direction.
- We call such hypothesis tests two-sided (or two-tailed).
- The definition of a p-value is the same regardless of doing a one or two-sided test, however the calculation is slightly different since we need to consider “at least as extreme as the observed outcome” in both directions.

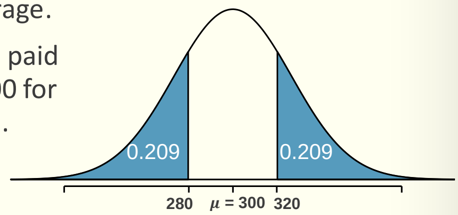
H_0 : $\mu >$ یا $\mu <$ $\Longleftrightarrow$ آزمون یک طرفه

H_1 : $\mu \neq ..$ $\Longleftrightarrow$ " دو "

Example

$H_0: \mu = 300$ College students have paid $300 for textbooks, on average.

$H_A: \mu \neq 300$ College students have paid more or less than $300 for textbooks, on average.



p-value = $P(\bar{x} > 320 \text{ or } \bar{x} < 280 | H_0: \mu = 300)$

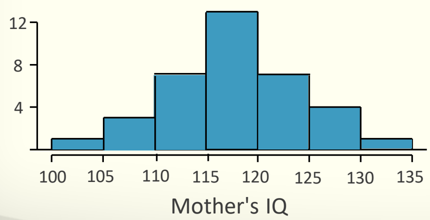
p-value = $P(Z > 0.81) + P(Z < -0.81) = 0.209 + 0.209 = 0.418$

Recap: Hypothesis testing for a single mean

1. Set the hypotheses: H_0: $\mu = null\ value$
   H_A: $\mu < or > or \neq null\ value$
2. Calculate the point estimate: $\bar{x}$
3. Check conditions:
   - ***Independence:*** Sampled observations must be independent (random sample/ assignment & if sampling without replacement, $n <$ 10% of population)
   - ***Sample size/skew:*** $n \geq 30$, larger if the population distribution is very skewed.
4. Draw sampling distribution, shade p-value, calculate test statistic:
   $$Z = \frac{\bar{x} - \mu}{SE} \quad , \qquad SE = \frac{s}{\sqrt{n}}$$
5. Make a decision, and interpret it in context of the research question:
   - If p-value $< \alpha$, reject H_0; the data provide convincing evidence for H_A.
   - If p-value $> \alpha$, fail to reject H_0; the data *do not* provide convincing evidence for H_A.

Example 1

➢ Researchers investigating characteristics of gifted children collected data from schools in a large city on a random sample of thirty-six children who were identified as gifted children soon after they reached the age of four. In this study, along with variables on the children, the researchers also collected data on their mothers' IQ scores. The histogram shows the distribution of these data, and also provided are some sample statistics.



n	36
min	101
mean	118.2
sd	6.5
max	131

Example 1

➢ Perform a hypothesis test to evaluate if these data provide convincing evidence of a difference between the average IQ score of mothers of gifted children and the average IQ score for the population at large, which is 100. Use a significance level of 0.01.

1. **Set the hypotheses**

$\mu =$ average IQ score of mothers of gifted children

$H_0: \mu = 100$ $\quad$ $H_A: \mu \neq 100$

2. **Calculate the point estimate**

$\bar{x} = 118.2$

3. **Check conditions**

1. random & 36 < 10% of all gifted children → independence
2. $n > 30$ & sample not skewed → nearly normal sampling distribution

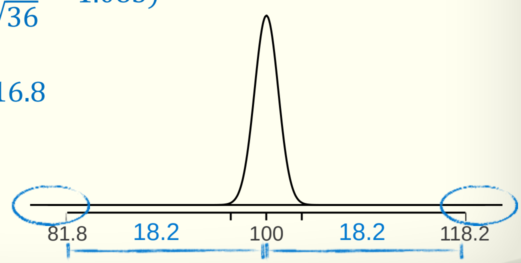
Example 1

4. Draw sampling distribution, shade p-value, calculate test statistic

$\bar{x} \sim N(\mu = 100\,, SE = \frac{s}{\sqrt{n}} = \frac{6.5}{\sqrt{36}} \approx 1.083)$

test statistic: $Z = \frac{118.2-100}{1.083} = 16.8$

p-value ≈ 0



Example 1

5. Make a decision, and interpret it in context of the research question

p-value is very low → strong evidence against the null

We reject the null hypothesis and conclude that the data provide convincing evidence of a difference between the average IQ score of mothers of gifted children and the average IQ score for the population at large.

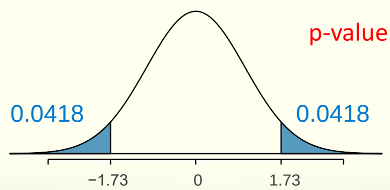
Example 2

➢ A statistics student interested in sleep habits of domestic cats took a random sample of 144 cats and monitored their sleep. The cats slept an average of 16 hours/day. According to online resources domestic dogs sleep, on average, 14 hours/day. We want to find out if these data provide convincing evidence of different sleeping habits for domestic cats and dogs with respect to how much they sleep. The test statistic is 1.73.

$\bar{x} = 16$

H_0: $\mu = 14$

H_A: $\mu \neq 14$



p-value = $2 \times 0.0418 = 0.0836$

Example 2

➢ What is the interpretation of this p-value in context of these data?

p-value = P(observed or more extreme outcome | H_0 true)

= P(obtaining a random sample of 144 cats that sleep 16 hours or more or 12 hours or less, on average, if in fact cats truly slept 14 hours per day on average)

= 0.0836

$n = 144$

$\bar{x} = 16$

H_0: $\mu = 14$

H_A: $\mu \neq 14$