Sta 111 - Summer II 2017
Probability and Statistical Inference

8. Hypothesis testing

Lu Wang

Duke University, Department of Statistical Science

July 12, 2017

Outline

1. Use hypothesis tests to make decisions about population parameters

2. Formal testing using p-values

3. Summary

Gender discrimination experiment

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

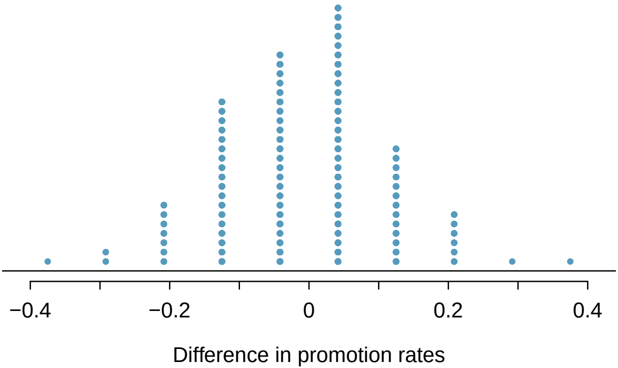
$$\hat{p}_{males} = 21/24 \approx 0.88$$

$$\hat{p}_{females} = 14/24 \approx 0.58$$

Possible explanations:

- Promotion and gender are *independent*, no gender discrimination, observed difference in proportions is simply due to chance. $\rightarrow$ *null* - (nothing is going on)
- Promotion and gender are *dependent*, there is gender discrimination, observed difference in proportions is not due to chance. $\rightarrow$ *alternative* - (something is going on)

Result



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.

- 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 methods based on the central limit theorem (coming up next...).
- If the test results suggest that the data do not provide convincing evidence for the null hypothesis, we reject the null hypothesis in favor of the alternative. If they do, then we stick with the null hypothesis.

We'll introduce the hypothesis testing framework using an example on testing a claim about a population mean.

Earlier we calculated a 95% confidence interval for the average number of exclusive relationships college students have been in to be (2.7, 3.7). Based on this confidence interval, do these data support the hypothesis that college students on average have been in more than 3 exclusive relationships.

- The associated hypotheses are:

  H_0: $\mu = 3$: College students have been in 3 exclusive relationships, on average

  H_A: $\mu > 3$: College students have been in more than 3 exclusive relationships, on average

- Since the null value is included in the interval, we do not reject the null hypothesis.
- 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.

Number of college applications

A survey asked how many colleges Duke students applied to, and 206 students responded to this question. This sample yielded an average of 9.7 college applications with a standard deviation of 7. College Board website states that counselors recommend students apply to roughly 8 colleges. Do these data provide convincing evidence that the average number of colleges all Duke students apply to is <u>higher</u> than recommended?

http://www.collegeboard.com/student/apply/the-application/151680.html

Setting the hypotheses

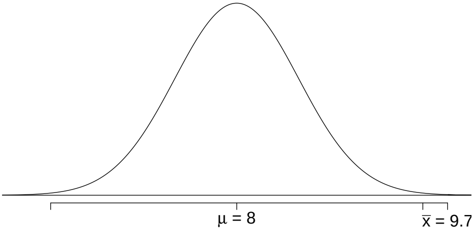
- The *parameter of interest* is the average number of schools applied to by all Duke students.
- There may be two explanations why our sample mean is higher than the recommended 8 schools.
  - The true population mean is different.
  - The true population mean is 8, and the difference between the true population mean and the sample mean is simply due to natural sampling variability.
- We start with the assumption the average number of colleges Duke students apply to is 8 (as recommended)

$$H_0 : \mu = 8$$

- We test the claim that the average number of colleges Duke students apply to is greater than 8

$$H_A : \mu > 8$$

In order to evaluate if the observed sample mean is unusual for the hypothesized sampling distribution, we determine how many standard errors away from the null it is, which is also called the *test statistic*.



$$\bar{x} \sim N\left(\mu = 8, SE = \frac{7}{\sqrt{206}} = 0.5\right)$$

$$Z = \frac{9.7 - 8}{0.5} = 3.4$$

The sample mean is 3.4 standard errors away from the hypothesized value. Is this considered unusually high? That is, is the result *statistically significant*?

- The *p-value* is the probability of getting a sample that results in a test statistic at least as favorable to the alternative hypothesis as our current data (as or more extreme than what you actually observed in the direction of alternative hypothesis) *if the null hypothesis were correct*.

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

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

Which of the following is not a condition that needs to be met to validate the above hypothesis test?

(a) Students in the sample should be independent of each other with respect to how many colleges they applied to.

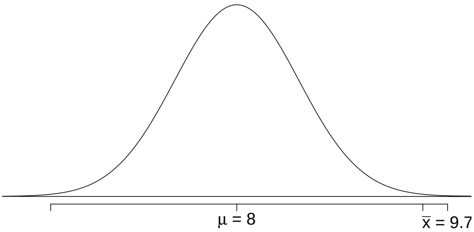
(b) Sampling should have been done randomly.

(c) The sample size should be less than 10% of the population of all Duke students.

(d) There should be at least 10 successes and 10 failures in the sample.

(e) The distribution of the number of colleges students apply to should not be extremely skewed.

p-value: probability of observing data at least as extreme as our current data set (a sample mean greater than 9.7), if in fact H_0 were true (the true population mean was 8).



$$P(\bar{x} > 9.7 \mid \mu = 8) = P(Z > 3.4) = 0.0003$$

- p-value = 0.0003
  - If the true average of the number of colleges Duke students applied to is 8, there is only 0.03% chance of observing a random sample of 206 Duke students who on average apply to 9.7 or more schools.
  - This is a pretty low probability for us to think that a sample mean of 9.7 or more schools is likely to happen simply by chance.
- Since p-value is *low* (lower than 5%) we *reject* H_0.
- The data provide convincing evidence that Duke students apply to more than 8 schools on average.
- The difference between the null value of 8 schools and observed sample mean of 9.7 schools is *not due to chance* or sampling variability.

Common misconceptions about p-value

1. P-value is the probability that the null hypothesis is true
   A p-value is the probability of getting a sample that results in a test statistic as or more extreme than what you actually observed if in fact the null hypothesis is correct. It is a conditional probability, conditioned on the null hypothesis being correct.

2. A high p-value confirms the null hypothesis.
   A high p-value means the data do not provide convincing evidence for the alternative hypothesis and hence that the null hypothesis can't be rejected.

3. A low p-value confirms the alternative hypothesis.
   A low p-value means the data provide convincing evidence for the alternative hypothesis, but not necessarily that it is confirmed.

Practice

A poll by the National Sleep Foundation found that college students on average sleep about 7 hours per night. A sample of 169 college students yielded an average of 6.88 hours, with a standard deviation of 0.94 hours. Assuming that this is a random sample representative of all college students, a hypothesis test was conducted to evaluate if college students on average sleep less than 7 hours per night. The p-value for this hypothesis test is 0.0485. Which of the following is the correct interpretation of this p-value

(a) The probability that students on average sleep less than 7 hours per night.

(b) The probability that students on average sleep less than 7 hours per night, if in fact a random sample of 169 students yields an average of 6.88 hours or less.

(c) The probability that a random sample of 169 students on average sleep 6.88 hours or less per night, if in fact students on average sleep 7 hours per night.

(d) The probability that a random sample of 169 students on average sleep 6.88 hours or more per night, if in fact students on average sleep 7 hours per night.

The p-value for this hypothesis test is 0.0485. Which of the following is correct?

(a) Fail to reject H_0, the data provide convincing evidence that college students sleep less than 7 hours on average.

(b) Reject H_0, the data provide convincing evidence that college students sleep less than 7 hours on average.

(c) Reject H_0, the data prove that college students sleep more than 7 hours on average.

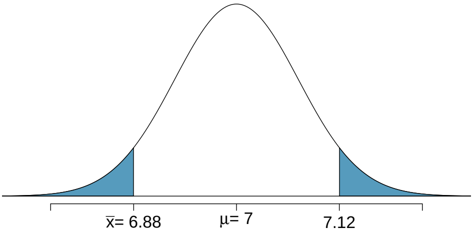
(d) Fail to reject H_0, the data do not provide convincing evidence that college students sleep less than 7 hours on average.

(e) Reject H_0, the data provide convincing evidence that college students in this sample sleep less than 7 hours on average.

- If the research question was "Do the data provide convincing evidence that the average amount of sleep college students get per night is *different* than the national average?", the alternative hypothesis would be different.

$$H_0 : \mu = 7$$
$$H_A : \mu \neq 7$$

- Hence the p-value would change as well:



p-value
$= 0.0485 \times 2$
$= 0.097$

Recap: Hypothesis testing framework

1. Set the hypotheses.
2. Check assumptions and conditions.
3. Calculate a *test statistic* and a p-value.
4. Make a decision, and interpret it in context of the research question.

Recap: Hypothesis testing for a population mean

1. Set the hypotheses
   - $H_0 : \mu =$ *null value*
   - $H_A : \mu <$ or $>$ or $\neq$ *null value*
2. Calculate the point estimate
3. Check assumptions and conditions
   - Independence: random sample/assignment, 10% condition when sampling without replacement
   - Sample size: nearly normal population or $n \geq 30$, no extreme skew
4. Calculate a *test statistic* and a p-value (draw a picture!)

$$Z = \frac{\bar{x} - \mu}{SE}, \text{ where } SE = \frac{s}{\sqrt{n}}$$

5. Make a decision, and interpret it in context
   - If p-value $< \alpha$, reject H_0, data provide evidence for H_A
   - If p-value $> \alpha$, do not reject H_0, data do not provide evidence for H_A