Hypothesis Testing

Stat 3202 @ OSU, Autumn 2018

Dalpiaz

Main Ideas and Large Sample Tests

Some Setups

Let $X_1, X_2, \ldots, X_n$ be a "large" sample from a distribution with $\text{E}[X] = \mu$ and $\text{Var}[X] = \sigma^2$. Then,

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} \overset{approx}{\sim} N(0, 1)$$

Let $X_1, X_2, \ldots, X_{n_1}$ be a "large" sample from a distribution with $\text{E}[X] = \mu_1$ and $\text{Var}[X] = \sigma_1^2$ and $Y_1, Y_2, \ldots, Y_{n_2}$ be a "large" sample from a distribution with $\text{E}[Y] = \mu_2$ and $\text{Var}[X] = \sigma_2^2$. Then,

$$\frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \overset{approx}{\sim} N(0, 1)$$

More Setups Setups

Let $X_1, X_2, \ldots, X_n$ be a "large" sample from a Bernoulli distribution with parameter p. Then,

$$\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \overset{approx}{\sim} N(0,1)$$

Let $X_1, X_2, \ldots, X_{n_1}$ be a "large" sample from a Bernoulli distribution with parameter p_1 and $Y_1, Y_2, \ldots, Y_{n_2}$ be a "large" sample from a Bernoulli distribution with parameter p_2. Then,

$$\frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \overset{approx}{\sim} N(0,1)$$

Example: One Sample Test for μ

An administrator claims that undergraduate students at Ohio State are extremely healthy. In particular, she claims that they sleep 8 or more hours a night on average. (Let μ be the true average sleep.) To test this claim, a random sample of 50 students is selected to report on the amount of sleep they obtained the previous night. They slept on average 7.72 hours, with a standard deviation of 1.63 hours. Do you believe the administrator's claim? Use a significance level of $\alpha = 0.05$ and an appropriate test.

Hypothesis Test Steps

- Develop scientific hypothesis
- Translate to statistical **hypothesis** about parameters
  - **Null** hypothesis, H_0
  - **Alternative** hypothesis, H_A or H_1.
- Set **significance level**, α
- Collect data
- Calculate **test statistic**
  - Note *distribution* of this statistic **under null hypothesis**
- Calculate **p-value** or find **rejection region**
- State the statistical **conclusion**
- Translate to scientific conclusion

Example: One Sample Test for p

- Is a coin fair?

Alex is suspicious of a particular coin so he flips it 900 times and observes an outcome of heads 477 times. Let p be the probability of obtaining heads. Perform the appropriate test using a significance level of $\alpha = 0.10$.

Example: Two Sample Test for $p_1 - p_2$

In a comparative study of two new drugs, A and B, 120 patients were treated with drug A and 150 patients with drug B, and the following results were obtained.

	Drug A	Drug B
Cured	78	111
Not	42	39

We wish to test whether drug B has a higher cure rate than drug A. Perform the appropriate test using a significance level of $\alpha = 0.05$.

Hypothesis Testing Main Ideas

Statistical Hypothesis

Statistical hypothesis: an assertion or conjecture about the distribution of one or more random variables, often specifically about a **parameter** of a distribution

- **Null hypothesis**, H_0: Hypothesis of no difference or no effect; we generally look for evidence against the null hypothesis
- **Alternative hypothesis**, H_A or H_1: A hypothesis that often complements the null; this is often what we are trying to show

Statistical Hypothesis (Left-tailed)

- $H_0 : \mu = \mu_0$ vs $H_A : \mu < \mu_0$
- $H_0 : p = p_0$ vs $H_A : p < p_0$
- $H_0 : \mu 1 = \mu 2$ vs $H_A : \mu_1 < \mu_2$
  - $H_0 : \mu 1 - \mu 2 = 0$ vs $H_A : \mu_1 - \mu_2 < 0$
- $H_0 : p_1 = p_2$ vs $H_A : p_1 < p_2$
  - $H_0 : p_1 - p_2 = 0$ vs $H_A : p_1 - p_2 < 0$

Statistical Hypothesis (Right-tailed)

- $H_0 : \mu = \mu_0$ vs $H_A : \mu > \mu_0$
- $H_0 : p = p_0$ vs $H_A : p > p_0$
- $H_0 : \mu 1 = \mu 2$ vs $H_A : \mu_1 > \mu_2$
    - $H_0 : \mu 1 - \mu 2 = 0$ vs $H_A : \mu_1 - \mu_2 > 0$
- $H_0 : p_1 = p_2$ vs $H_A : p_1 > p_2$
    - $H_0 : p_1 - p_2 = 0$ vs $H_A : p_1 - p_2 > 0$

Statistical Hypothesis (Two-tailed)

- $H_0 : \mu = \mu_0$ vs $H_A : \mu \neq \mu_0$
- $H_0 : p = p_0$ vs $H_A : p \neq p_0$
- $H_0 : \mu 1 = \mu 2$ vs $H_A : \mu_1 \neq \mu_2$
  - $H_0 : \mu 1 - \mu 2 = 0$ vs $H_A : \mu_1 - \mu_2 \neq 0$
- $H_0 : p_1 = p_2$ vs $H_A : p_1 \neq p_2$
  - $H_0 : p_1 - p_2 = 0$ vs $H_A : p_1 - p_2 \neq 0$

Statistical Conclusions

- If $p - value < \alpha$ or the test statistic is in the rejection region
  - **Reject** H_0
  - Claim "statistical significance!"
- If $p - value > \alpha$ or the test statistic is *not* in the rejection region
  - **Fail to reject** H_0
  - "Accept" H_0?

Hypotheses and Conclusions

	H_0 true	**H_0 false**
Accept H_0 **(Do NOT Reject H_0)**	☺	Type II Error
Reject H_0	Type I Error	☺

- Type I Error: “False Positive”
- Type II Error: “False Negative”

α and β

- $\alpha = P(\text{Reject } H_0 \mid H_0 \text{ True})$
  - The probability of making a Type I error
  - The probability of a false positive
  - The **significance level** of a test
- $\beta = P(\text{Accept } H_0 \mid H_0 \text{ False})$
  - The probability of making a Type II error
  - The probability of a false negative
- $1 - \beta = P(\text{Reject } H_0 \mid H_0 \text{ False})$
  - The **power** of a test

Test Statistics

$$z = \frac{\text{EST} - \text{HYP}}{\text{SE(EST)}} \overset{approx}{\sim} N(0, 1)$$

Test Statistics

$$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \overset{approx}{\sim} N(0,1)$$

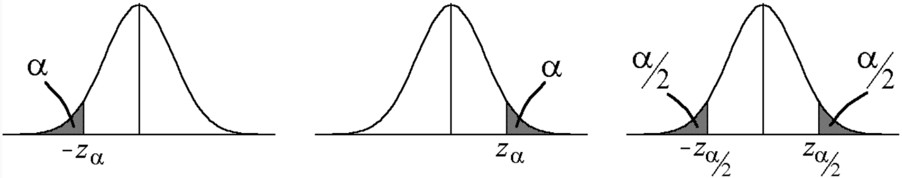
$$z = \frac{(\bar{x} - \bar{y}) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \overset{approx}{\sim} N(0,1)$$

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \overset{approx}{\sim} N(0,1)$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}}} \overset{approx}{\sim} N(0,1), \qquad \hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

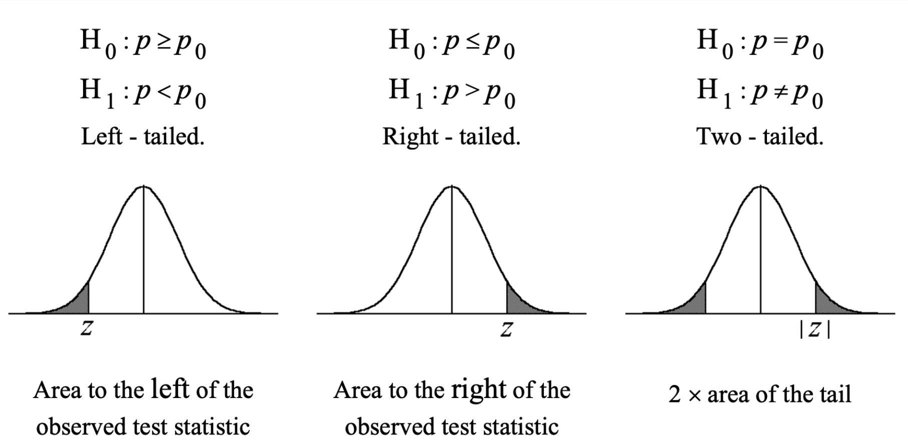
Rejection Regions

Left-tailed	Right-tailed	Two-tailed
$H_0 : p \geq p_0$	$H_0 : p \leq p_0$	$H_0 : p = p_0$
$H_1 : p < p_0$	$H_1 : p > p_0$	$H_1 : p \neq p_0$
Left - tailed.	Right - tailed.	Two - tailed.



Left-tailed	Right-tailed	Two-tailed
Reject H_0 if $z < -z_{\alpha}$	Reject H_0 if $z > z_{\alpha}$	Reject H_0 if $z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$

P-Values


$H_0: p \geq p_0$
$H_1: p < p_0$
Left - tailed.
z
Area to the left of the observed test statistic
$H_0: p \leq p_0$
$H_1: p > p_0$
Right - tailed.
z
Area to the right of the observed test statistic
$H_0: p = p_0$
$H_1: p \neq p_0$
Two - tailed.
$|z|$
$2 \times$ area of the tail

Rejection Regions and P-Values

- **Rejection Region**: potential values of the test statistic that occur with probability α if the null hypothesis is true
- **p-value:** probability of observing something (such as the test statistic) as extreme or more extreme than what we observed, assuming that the null hypothesis is true. [Note: "extreme" is defined in the direction of the alternative.]
  - **THIS IS NOT THE PROBABILITY THAT THE NULL HYPOTHESIS (OR ANY HYPOTHESIS) IS TRUE!**

Example: Two Sample Test for $\mu_1 - \mu_2$

Professor Professorson, a researcher at Greendale Community College, believes that caffeine has a negative effect on the sleep of students.

Professorson obtains a random sample of 50 students who are given 400 mg of caffeine at noon on some day. (Don't try this at home.) Professor Professorson invites these students for a sleep study and finds that they sleep an average of 6.5 hours with a standard deviation of 1.2 hours that night.

Professorson also recruits 75 students who are given a placebo, also at noon. He again monitors them during a sleep study and finds that they sleep an average of 7.3 hours with a standard deviation of 1.4 hours that night.

Perform the appropriate test using a significance level of $\alpha = 0.05$.

Small Sample Tests

(One-Sample) Small Sample Setups

Let $X_1, X_2, \ldots, X_n$ be a sample from a normal distribution with mean μ and variance σ^2. Then,

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

(Two-Sample) Small Sample Setups

- Let $X_1, X_2, \ldots, X_{n_1}$ be a sample from a normal distribution with mean μ_1 and variance σ_1^2.
- Let $Y_1, Y_2, \ldots, Y_{n_2}$ be a sample from a normal distribution with mean μ_2 and variance σ_2^2.

Then,

$$\frac{(\bar{x} - \bar{y}) - 0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1 + n_2 - 2}$$

where

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Example: One Sample Test for μ

Battery packs for an artificial heart are tested to determine their average lifetime which the manufacturer claims is over 4 years. In a random sample of 20 battery packs, the sample average was 4.05 years with a standard deviation of 0.2 years. Assume that the lifetime of the battery packs follows a normal distribution. Is there evidence to support the claim that the mean battery life exceeds 4 years a significance level of $\alpha = 0.05$?

Example: One Sample Test for σ

Consider a filler machine in a dog food production plant. From studying the process over time, we assume that the population standard deviation, σ, is 0.17, but we observe an unusual level of variability in the fill weights on a particular day. We would like to test whether the standard deviation has increased.

In a sample of 30 boxes, we find a standard deviation of 0.21 lbs. Is this evidence that the standard deviation has increased? Carry out a hypothesis test using a significance level of $\alpha = 0.05$.

Example: Two Sample Test for $\mu_1 - \mu_2$

Consider an experiment conducted on mice to examine the effect of a magnetic field on the amount of weight gain. The experimental set-up included two groups, a treatment group that was exposed to a magnetic field and a control group that was not exposed. Each group contained 10 mice. The data consist of the weight gain per mouse, and we can assume that the data in each group are normally distributed, with equal variances across groups. Carry out a hypothesis test to determine whether exposure to a magnetic field inhibits growth in mice. Use a significance level of $\alpha = 0.01$.

Paired Sample Test

A new revolutionary diet-and-exercise plan is introduced. Eight participants were weighed in the beginning of the program, and then again a week later. The results were as follows:

Participant	1	2	3	4	5	6	7	8
Weight Before	213	222	232	201	230	188	218	182
Weight After	208	220	224	200	220	185	220	184

Is there enough evidence to conclude that the diet-and-exercise plan is effective? (Use $\alpha = 0.05$.) What is the p-value of this test?