Lecture 13: Hypothesis testing

Announcements:

- Reading for this week's subject: pp. 61 - 79 in Vasilj.

Lecture 13: Hypothesis testing

Announcements:

- Reading for this week's subject: pp. 61 - 79 in Vasilj.
- Problem sets 4 through 6 are due today in class.

Announcements:

- Reading for this week's subject: pp. 61 - 79 in Vasilj.
- Problem sets 4 through 6 are due today in class.
- If you have any questions about the exams, please see me or Tomislav.

Lecture 13: Hypothesis testing

Announcements:

- Reading for this week's subject: pp. 61 - 79 in Vasilj.
- Problem sets 4 through 6 are due today in class.
- If you have any questions about the exams, please see me or Tomislav.
- Next exam will include all questions in this exam that less than 50% of people answered correctly. These are questions 9, 24, 25, 21, 16, and 5. So make sure you understand those questions!

Lecture 13: Hypothesis testing

- We have a question about a population.

Lecture 13: Hypothesis testing

- We have a question about a population.
- We take a sample, calculate a statistic.

- We have a question about a population.
- We take a sample, calculate a statistic.
- We plot that statistic on the expected probability distribution.

Lecture 13: Hypothesis testing

- We have a question about a population.
- We take a sample, calculate a statistic.
- We plot that statistic on the expected probability distribution.
- If the statistic is within the tail, we conclude that the sample is "unusual" relative to the reference population.

Hypothesis testing

Hypothesis testing: the α level, or **significance level**, of a test

α = false positive

tail region is α

$P(\text{tail}) = 0.05$

Hypothesis testing: procedure

1. We ask a yes/no question about a **population**.

Hypothesis testing: procedure

1. We ask a yes/no question about a **population**.
2. We answer the question yes, and answer the question no, using symbols for the population means.

Hypothesis testing: procedure

1. We ask a yes/no question about a **population**.
2. We answer the question yes, and answer the question no, using symbols for the population means.
3. We label one answer the **null hypothesis** and the other answer the **alternative hypothesis**.

Hypothesis testing: procedure

1. We ask a yes/no question about a **population**.
2. We answer the question yes, and answer the question no, using symbols for the population means.
3. We label one answer the **null hypothesis** and the other answer the **alternative hypothesis**.
4. We decide the criterion for rejecting the null hypothesis.

Hypothesis testing: procedure

1. We ask a yes/no question about a **population**.
2. We answer the question yes, and answer the question no, using symbols for the population means.
3. We label one answer the **null hypothesis** and the other answer the **alternative hypothesis**.
4. We decide the criterion for rejecting the null hypothesis.
5. The test is one of: **two-tailed**, **right-tailed**, or **left-tailed**.

Hypothesis testing: procedure

1. We ask a yes/no question about a **population**.
2. We answer the question yes, and answer the question no, using symbols for the population means.
3. We label one answer the **null hypothesis** and the other answer the **alternative hypothesis**.
4. We decide the criterion for rejecting the null hypothesis.
5. The test is one of: **two-tailed**, **right-tailed**, or **left-tailed**.
6. We take a sample, and calculate our **test statistic** (Z or t for now)

Hypothesis testing: procedure

1. We ask a yes/no question about a **population**.
2. We answer the question yes, and answer the question no, using symbols for the population means.
3. We label one answer the **null hypothesis** and the other answer the **alternative hypothesis**.
4. We decide the criterion for rejecting the null hypothesis.
5. The test is one of: **two-tailed**, **right-tailed**, or **left-tailed**.
6. We take a sample, and calculate our **test statistic** (Z or t for now)
7. We find if the observed test statistic is in the rejection region (**critical region** or **tail**) of the distribution.

Hypothesis testing: procedure

1. We ask a yes/no question about a **population**.
2. We answer the question yes, and answer the question no, using symbols for the population means.
3. We label one answer the **null hypothesis** and the other answer the **alternative hypothesis**.
4. We decide the criterion for rejecting the null hypothesis.
5. The test is one of: **two-tailed**, **right-tailed**, or **left-tailed**.
6. We take a sample, and calculate our **test statistic** (Z or t for now)
7. We find if the observed test statistic is in the rejection region (**critical region** or **tail**) of the distribution.
8. If the statistic is in the rejection region, we **reject** the null hypothesis and **accept** the alternative hyopthesis.

Hypothesis testing: procedure

1. We ask a yes/no question about a **population**.
2. We answer the question yes, and answer the question no, using symbols for the population means.
3. We label one answer the **null hypothesis** and the other answer the **alternative hypothesis**.
4. We decide the criterion for rejecting the null hypothesis.
5. The test is one of: **two-tailed**, **right-tailed**, or **left-tailed**.
6. We take a sample, and calculate our **test statistic** (Z or t for now)
7. We find if the observed test statistic is in the rejection region (**critical region** or **tail**) of the distribution.
8. If the statistic is in the rejection region, we **reject** the null hypothesis and **accept** the alternative hyopthesis.
9. If the statistic is not in the rejection region, we **retain** the null hypothesis, and do not accept the alternative hypothesis.

Examples

$p =$ prob. of heads

$H_0: p = 0.5$ null hypothesis

$H_A: p \neq 0.5$ alternative hypothesis

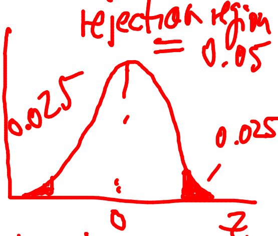
I flip a coin n times and get x heads. Is this a fair coin?

$H_A \; \neq$: 2-tailed test

$H_A \; <$: left-tailed test

$H_A \; >$: right-tailed test

$$z = \frac{\frac{x}{n} - 0.5}{\sqrt{\frac{pq}{n}}} = \frac{\frac{x}{n} - 0.5}{\sqrt{\frac{(0.5)(.5)}{n}}}$$



$|z| > 1.96$

Examples

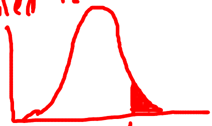
μ_n = mean of new

μ_o = mean of old

$H_A: \mu_n - \mu_o > 0$

$H_0: \mu_n - \mu_o \le 0$

Right-tailed test



t_{crit}

I create a new genotype of olive tree. Do its olives produce more oil than the old genotype?

$$t = \frac{\bar{X}_n - \bar{X}_o - 0}{\sqrt{\frac{S_n^2}{n_n} + \frac{S_o^2}{n_o}}}$$

$t > t_{crit}$

only pos. t rejects

Rules

- If the alternative hypothesis contains the $\neq$ sign, the test is two-tailed.

Rules

- If the alternative hypothesis contains the $\neq$ sign, the test is two-tailed.
- If the alternative hypothesis contains the $<$ sign, the test is left-tailed.

Rules

- If the alternative hypothesis contains the $\neq$ sign, the test is two-tailed.
- If the alternative hypothesis contains the $<$ sign, the test is left-tailed.
- If the alternative hypothesis contains the $>$ sign, the test is right-tailed.

Examples

μ_n = mean res of new
μ_o = mean res of old

$H_A: \mu_n - \mu_o < 0$
$H_0: \mu_n - \mu_o \geq 0$

left tail test

I have a new chemical pesticide. Does it leave less residue on apples than the old kind?

$-$ 0 $+$

$$t = \frac{\bar{x}_n - \bar{x}_o - 0}{\sqrt{\frac{s_n^2}{n_n} + \frac{s_o^2}{n_o}}}$$

$t < t_{crit}$

$\therefore$ reject

Examples

Hypothesis testing

p = prop. of Croatians in favor

$H_A: p > 0.5$

$H_0: p \leq 0.5$

right-tailed test

I take a poll. Are Croatians in favor of entering the EU?

$$t = \frac{\hat{p} - 0.5}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

reject if $t > t_{crit}$

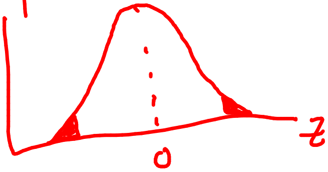
Examples

reject if $|z| > 1.96$

p = prop of ♀

$H_0: p = 0.5$

$H_A: p \neq 0.5$



Is the sex ratio of males to females 1:1 in some population?

$$Z = \frac{\hat{p} - 0.5}{\sqrt{\frac{p(1-p)}{n}}}$$

Terminology

t_{crit} or z_{crit}

- The **critical value** of a test statistic is the value that defines the rejection region of the distribution.

Terminology

$t_{crit} = t_{0.05/2}$

- The **critical value** of a test statistic is the value that defines the rejection region of the distribution.
- $t_{0.05/2}$ is the critical t-value for $\alpha = 0.05$ and a 2-tailed test. Each tail here contains 0.025 of the probability.

reject if $|t| > t_{0.05/2}$

Terminology

- The **critical value** of a test statistic is the value that defines the rejection region of the distribution.
- $t_{0.05/2}$ is the critical t-value for $\alpha = 0.05$ and a 2-tailed test. Each tail here contains 0.025 of the probability.
- $t_{0.05/1}$ is the critical t-value for $\alpha = 0.05$ and a 1-tailed test. The single tail contains 0.05 of the probability.

Terminology

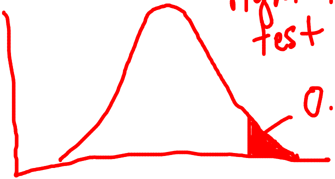
- The **critical value** of a test statistic is the value that defines the rejection region of the distribution.
- $t_{0.05/2}$ is the critical t-value for $\alpha = 0.05$ and a 2-tailed test. Each tail here contains 0.025 of the probability.
- $t_{0.05/1}$ is the critical t-value for $\alpha = 0.05$ and a 1-tailed test. The single tail contains 0.05 of the probability.
- $t_{0.01/1}$ is the critical t-value for $\alpha = 0.01$ and a 1-tailed test. The single tail contains 0.01 of the probability.

Examples

μ = mean arsenic in well

$H_A: \mu > x$

$H_0: \mu \leq x$



Is the level of arsenic in a well greater than some minimum value x?

reject if $t > t_{0.05/}$

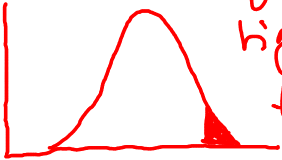
$$t = \frac{\bar{x} - x}{\sqrt{\frac{s^2}{n}}} = \frac{s}{\sqrt{n}}$$

Example using a single-sample t-test

μ_d = difference after adding fertilizer

$H_A: \mu_d > 0$

$H_0: \mu_d \leq 0$



right tailed

$t > t_{0.05/1}$

Does a particular fertilizer increase the potassium content of leaves? 16 plants were grown, and the potassium content before and after fertilizer was measured, and the difference reported. The mean difference was 5, and the standard deviation of the difference was 10.

K

$$t = \frac{5 - 0}{\sqrt{\frac{10^2}{16}}} = \frac{5}{\frac{10}{4}} = 2$$

$t_{0.05/1} = 1.75$

$2 > 1.75 \therefore$ reject

Example using a single-sample t-test

Does a particular fertilizer increase the potassium content of leaves? 16 plants were grown, and the potassium content before and after fertilizer was measured, and the difference reported. The mean difference was 5, and the standard deviation of the difference was 10.

Example using a two-sample t-test

2-tailed test

μ_f = mean f

μ_m = mean m

$H_0: \mu_m - \mu_f = 0$

$H_A: \mu_m - \mu_f \neq 0$

Our first exam was taken by 14 women and 13 men. The mean female score was 72.7, and the mean male score was 70.3. The standard deviation in score was 7.7 for females and 8.3 for males. Is the mean for the female population the same as the mean for the male population?

$70.3 - 72.7$

$|t| > t_{0.05/2}$

$$t = \frac{\bar{x}_m - \bar{x}_f - 0}{\sqrt{\frac{8.3^2}{13} + \frac{7.7^2}{14}}} = -0.78$$

$t_{0.05/2} = 2.2$

Equal variance t-test: use the pooled variance to calculate the sample standard deviation

Equal variance t-test: use the pooled variance to calculate the sample standard deviation

$$\text{pooled variance} \ =$$

Equal variance t-test: use the pooled variance to calculate the sample standard deviation

$$\text{pooled variance } = (n_1 - 1)s_i^2 +$$

Equal variance t-test: use the pooled variance to calculate the sample standard deviation

$$\text{pooled variance} = \frac{(n_1 - 1)s_i^2 + (n_2 - 1)s_2^2}{}$$

Equal variance t-test: use the pooled variance to calculate the sample standard deviation

$$\text{pooled variance} = \frac{(n_1 - 1)s_i^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Equal variance t-test: use the pooled variance to calculate the sample standard deviation

$$\text{pooled variance} = \frac{(n_1 - 1)s_i^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$s =$

Equal variance t-test: use the pooled variance to calculate the sample standard deviation

$$\text{pooled variance} = \frac{(n_1 - 1)s_i^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$s = \sqrt{\qquad\qquad\qquad}$$

Equal variance t-test: use the pooled variance to calculate the sample standard deviation

$$\text{pooled variance} = \frac{(n_1 - 1)s_i^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$s = \sqrt{\text{pooled variance}} = \sqrt{\phantom{\frac{(n_1 - 1)s_i^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}}$$

Equal variance t-test: use the pooled variance to calculate the sample standard deviation

$$\text{pooled variance} = \frac{(n_1 - 1)s_i^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$s = \sqrt{\text{pooled variance}} = \sqrt{(n_1 - 1)s_i^2 +}$$

Equal variance *t*-test: use the pooled variance to calculate the sample standard deviation

$$\text{pooled variance} = \frac{(n_1 - 1)s_i^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$s = \sqrt{\text{pooled variance}} = \sqrt{\frac{(n_1 - 1)s_i^2 + (n_2 - 1)s_2^2}{}}$$

Equal variance t-test: use the pooled variance to calculate the sample standard deviation

$$\text{pooled variance} = \frac{(n_1 - 1)s_i^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$s = \sqrt{\text{pooled variance}} = \sqrt{\frac{(n_1 - 1)s_i^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Equal variance t-test: use the pooled variance to calculate the sample standard deviation

$$\text{pooled variance} = \frac{(n_1 - 1)s_i^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$s = \sqrt{\text{pooled variance}} = \sqrt{\frac{(n_1 - 1)s_i^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$t =$$

Equal variance t-test: use the pooled variance to calculate the sample standard deviation

$$\text{pooled variance} = \frac{(n_1 - 1)s_i^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$s = \sqrt{\text{pooled variance}} = \sqrt{\frac{(n_1 - 1)s_i^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$t = \frac{\overline{x_1} - \overline{x_2}}{}$$

Equal variance t-test: use the pooled variance to calculate the sample standard deviation

$$\text{pooled variance} = \frac{(n_1 - 1)s_i^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$s = \sqrt{\text{pooled variance}} = \sqrt{\frac{(n_1 - 1)s_i^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{s^2/n_1 + s^2/n_2}} =$$

Equal variance t-test: use the pooled variance to calculate the sample standard deviation

$$\text{pooled variance} = \frac{(n_1 - 1)s_i^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$s = \sqrt{\text{pooled variance}} = \sqrt{\frac{(n_1 - 1)s_i^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{s^2/n_1 + s^2/n_2}} = \frac{\overline{x_1} - \overline{x_2}}{}$$

Equal variance t-test: use the pooled variance to calculate the sample standard deviation

$$\text{pooled variance} = \frac{(n_1 - 1)s_i^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$s = \sqrt{\text{pooled variance}} = \sqrt{\frac{(n_1 - 1)s_i^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{s^2/n_1 + s^2/n_2}} = \frac{\overline{x_1} - \overline{x_2}}{s\sqrt{1/n_1 + 1/n_2}}$$

Unequal variance t-test

Calculate separate variances for the two samples:

Unequal variance t-test

Calculate separate variances for the two samples:

$$t =$$

Unequal variance t-test

Calculate separate variances for the two samples:

$$t = \frac{\overline{x_1} - \overline{x_2}}{}$$

Unequal variance t-test

Calculate separate variances for the two samples:

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

Unequal variance t-test

Calculate separate variances for the two samples:

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

What is the degrees of freedom?

Unequal variance t-test

Calculate separate variances for the two samples:

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

What is the degrees of freedom? Either use the smallest of $n_1 - 1$ and $n_2 - 1$,

Unequal variance t-test

Calculate separate variances for the two samples:

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

What is the degrees of freedom? Either use the smallest of $n_1 - 1$ and $n_2 - 1$, or use R to calculate the following degrees of freedom:

Unequal variance t-test

Calculate separate variances for the two samples:

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

What is the degrees of freedom? Either use the smallest of $n_1 - 1$ and $n_2 - 1$, or use R to calculate the following degrees of freedom:

$$\text{d.f.} =$$

Unequal variance t-test

Calculate separate variances for the two samples:

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

What is the degrees of freedom? Either use the smallest of $n_1 - 1$ and $n_2 - 1$, or use R to calculate the following degrees of freedom:

$$\text{d.f.} = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{}$$

Unequal variance t-test

Calculate separate variances for the two samples:

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

What is the degrees of freedom? Either use the smallest of $n_1 - 1$ and $n_2 - 1$, or use R to calculate the following degrees of freedom:

$$\text{d.f.} = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)}.$$

Errors

Situation	Decision	
	Accept H_0	Reject H_0
H_0 true	$1 - \alpha$	α
H_0 false	β	$1 - \beta$

Power of a two-sample *t*-test

Our first exam was taken by 14 women and 13 men. The standard deviation in score was 7.7 for females and 8.3 for males. What is the power of a t-test on a difference in the means, if the true difference is 5 points, and $\alpha = 0.05$ for a 2-tailed test?

Power of a two-sample t-test

Our first exam was taken by 14 women and 13 men. The standard deviation in score was 7.7 for females and 8.3 for males. What is the power of a t-test on a difference in the means, if the true difference is 5 points, and $\alpha = 0.05$ for a 2-tailed test?

Power of a two-sample t-test

Our first exam was taken by 14 women and 13 men. The standard deviation in score was 7.7 for females and 8.3 for males. What is the power of a t-test on a difference in the means, if the true difference is 5 points, and $\alpha = 0.05$ for a 2-tailed test?

Power of a two-sample t-test

The standard deviation in score of our first exam was 7.7 for females and 8.3 for males. What sample size would be necessary to detect a true difference of 1 point between the sexes in either direction? Assume $\alpha = 0.05$.

Power of a two-sample t-test

The standard deviation in score of our first exam was 7.7 for females and 8.3 for males. What sample size would be necessary to detect a true difference of 1 point between the sexes in either direction? Assume $\alpha = 0.05$.

Power of a two-sample t-test

The standard deviation in score of our first exam was 7.7 for females and 8.3 for males. What sample size would be necessary to detect a true difference of 1 point between the sexes in either direction? Assume $\alpha = 0.05$.