# Topic 5: Hypothesis Testing

Optional Reading: Chapter 9

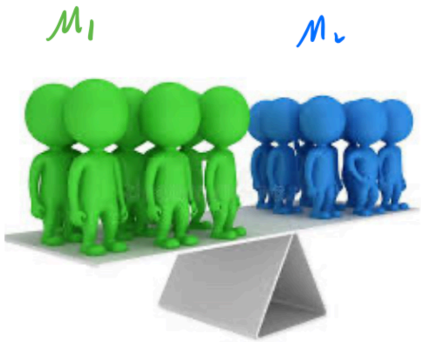
Xiner Zhou

Department of Statistics

University of California, Davis

- Test about a population mean
- Test about a population proportion
- Test about the difference between two population means
- Test about the difference between two population proportions

## Hypothesis Testing about the Difference between two population means




$\mu_1 \neq \mu_2$ ?

$\mu_1 > \mu_2$ ?

$\mu_1 < \mu_2$ ?

# Summary: Hypothesis Testing about Difference between two population means

Step 1: null hypothesis
$H_0 : \mu_1 - \mu_2 = 0$

Step 2: Alternative hypothesis

One-sided
$H_a : \mu_1 - \mu_2 > 0$ or $H_a : \mu_1 - \mu_2 < 0$

Two-sided
$H_a : \mu_1 - \mu_2 \neq 0$

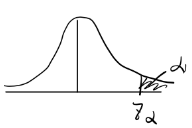
Step 3: calculate test statistic

$$Z_{obs} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

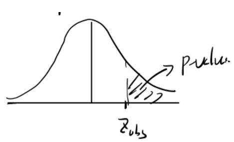
Step 4: critical value or p-value approach

One-sided:

Critical value

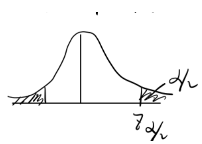



P-value

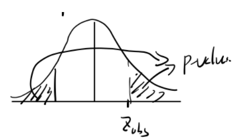



Two-sided:

Critical value




P-value




Step 5: Decision and Conclusion

Reject $H_0$ if:

$Z_{obs} \geq Z_\alpha$

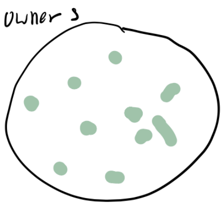
p-value $\leq \alpha$

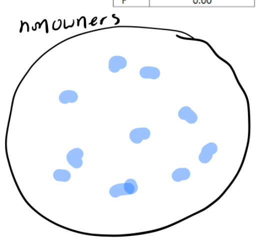
$|Z_{obs}| \geq Z_\alpha$

p-value $\leq \alpha$

Car ownership affects students' academic performance?

| Grades | Grade Point Equivalent |
| --- | --- |
| A | 4.00 |
| A- | 3.67 |
| B+ | 3.33 |
| B | 3.00 |
| B- | 2.67 |
| C+ | 2.33 |
| C | 2.00 |
| D | 1.00 |
| F | 0.00 |




$n_1 = 100$

$\bar{x}_1 = 2.7$

$s_1^2 = 0.36$

$n_2 = 100$

$\bar{x}_2 = 2.54$

$s_2^2 = 0.4$

Does the data present sufficient evidence to indicate a difference in mean GPA between car owners and non owners? Test using $\alpha = 0.05$

Step 1 : null , alternative hypotheses

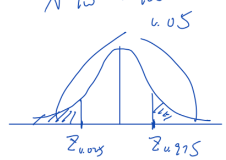
$H_0 : \mu_1 - \mu_2 = 0$ vs $H_a : \mu_1 - \mu_2 \neq 0$

Step 2 : test statistic

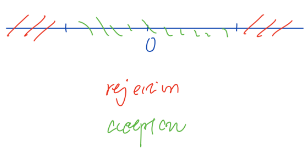
$$Z_{obs} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{2.7 - 2.54}{\sqrt{\frac{0.36}{100} + \frac{0.4}{100}}} = 1.84$$

Step 3 :

critical value approach :




$$\underbrace{Z_{0.975}}_{1.96} = \underbrace{-Z_{0.025}}_{= -1.96}$$




p-value approach : p-value $= P(|Z| \geq 1.84)$

$= 2 \times P(Z \leq -1.84)$

$= 0.0658$

Step 4: conclusion and decision:

<u>Critical value approach:</u>

Since the observed test statistic falls in the acceptance region,
We do not reject $H_0$ and conclude that, the average GPA for the two groups are not significantly different.

<u>P-value approach:</u>

Since p-value =0.0658>0.05,
We do not reject $H_0$ and conclude that, the average GPA for the two groups are not significantly different.

*Note: Observe some difference does not mean there is real difference, due to randomness and chance we could still see difference!*

Healthy Eating As Americans become more conscious about the importance of good nutrition, researchers theorize that the consumption of red meat may have decreased over the last 10 years. A researcher selects hospital nutrition records for 400 subjects surveyed 10 years ago and compares the average amount of beef consumed per year to amounts consumed by an equal number of subjects interviewed this year. The data are given in the table.

| | Ten Years Ago | This Year |
|---|---|---|
| Sample Mean | 73 | 63 |
| Sample Standard Deviation | 25 | 28 |

1. Do the data present sufficient evidence to indicate that per-capita beef consumption has decreased in the last 10 years? Test at the 1% level of significance.

Step 1 : null , alternative hypotheses

$H_0: \mu_1 - \mu_2 = 0$ vs $H_a: \mu_1 - \mu_2 > 0$

Step 2 : test statistic

$$Z_{obs} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{73-63}{\sqrt{\frac{25^2}{400} + \frac{28^2}{400}}} = 5.33$$

Step 3 :

p-value approach : $\text{p-value} = P(Z \geq 5.33)$

$= 4.91 \times 10^{-8} = 0^+$

$< 0.01$

Step 4: conclusion and decision:

P-value approach:

Since p-value =0+< 0.01,
We reject $H_0$ and conclude that the per-capita beef consumption has decrease significantly for the past 10 years.

Cheaper Airfares Looking for a great airfare? Consumer Reports [8] has several hints about how to minimize your costs, which include being flexible about travel dates and times, checking multiple websites, and knowing when to book your flight. One suggestion involved checking fares at "secondary" airports - airports that might be slightly farther from your home, but where fares are lower. For example, the average of all domestic ticket prices at Los Angeles International Airport (LAX) was quoted as \$349 compared to an average price of \$287 at nearby Ontario International Airport (ONT). Suppose that these estimates were based on random samples of 1000 domestic tickets at each airport and that the standard deviation of the prices at both airports was \$200

1. Is there sufficient evidence to indicate that the mean ticket prices differ for these two airports at the $\alpha = .05$ level of significance? Use the large-sample ztest. What is the $p$-value of this test?

2. Construct a 95% confidence interval for $(\mu_1 - \mu_2)$. Does this interval confirm your conclusions in part a?

Step 1 : null , alternative hypotheses

$H_0 : \mu_1 - \mu_2 = 0$ vs $H_a : \mu_1 - \mu_2 \neq 0$

Step 2 : test statistic

$$Z_{obs} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{349 - 287}{\sqrt{\frac{200^2}{1000} + \frac{200^2}{1000}}} = 6.93$$

Step 3 :

p-value approach : $p\text{-value} = P(|Z| \geq 6.93)$

$$= 2 \times \underbrace{P(Z \leq -6.93)}_{4.2 \times 10^{-12}}$$

$= 0^+$

Step 4: conclusion and decision:

P-value approach:

Since p-value =0+< 0.05,
We reject $H_0$ and conclude that the mean tickets price differ significantly for two airports.

95% C.I.: $(\bar{X}_1 - \bar{X}_2) \pm Z_{\frac{0.05}{2}} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

$$(349-287) \pm 1.96 \sqrt{\frac{200^2}{1000} + \frac{200^2}{1000}}$$

$$= (44.47, 79.53)$$

We’re 95% confidence that the mean tickets price differ for the two airports. (Same conclusion from 95% confidence interval)

Losing Weight In a comparison of the mean 1-month weight losses for women aged 20 - 30 years, these sample data were obtained for each of two diets:

| | Diet I | Diet II |
|---|---|---|
| Sample Size $n$ | 40 | 40 |
| Sample Mean $\bar{x}$( lb ) | 10 | 8 |
| Sample Variance $s^2$ | 4.3 | 5.7 |

Do the data provide sufficient evidence to indicate that diet I produces a greater mean weight loss than diet II? Use $\alpha = .05$.

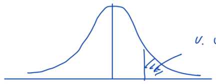
Step 1 : null , alternative hypotheses

$H_0: \mu_1 - \mu_2 = 0$ vs $H_a: \mu_1 - \mu_2 > 0$

Step 2 : test statistic

$$Z_{obs} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{10 - 8}{\sqrt{\frac{4.3^2}{40} + \frac{4.7^2}{40}}} = 4$$

Step 3 :

critical value approach :  0.05

$z_{\alpha=0.05} = 1.64$

Acceptance region rejection region

rejection region

p-value approach : p-value $= P(z \geq 4)$

$= \underbrace{0.000..}_{0+}$

Step 4: conclusion and decision:

Critical value approach:

Since the observed test statistic falls in the rejection region,
We reject $H_0$ and conclude that, Diet I produce significantly greater weight reduction than Diet II.

P-value approach:

Since p-value =0+ < 0.05,
We reject $H_0$ and conclude that, Diet I produce significantly greater weight reduction than Diet II.