GPCO 453: Quantitative Methods I

Sec 09: More on Hypothesis Testing

Shane Xinyang Xuan[1]
ShaneXuan.com

November 27, 2017

[1]Department of Political Science, UC San Diego, 9500 Gilman Drive #0521.

Contact Information

Shane Xinyang Xuan

xxuan@ucsd.edu

The teaching staff is a team!

Professor Garg	Tu	1300-1500 (RBC 1303)
Shane Xuan	M	1100-1200 (SSB 332)
	M	1530-1630 (SSB 332)
Joanna Valle-luna	Tu	1700-1800 (RBC 3131)
	Th	1300-1400 (RBC 3131)
Daniel Rust	F	1100-1230 (RBC 3213)

Comparing Two Populations

- Standard error

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \tag{1}$$

$$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \tag{2}$$

Comparing Two Populations

- Standard error

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad (1)$$

$$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \quad (2)$$

- Confidence interval

$$(\bar{x}_1 - \bar{x}_2) \pm z \times \sigma_{\bar{x}_1-\bar{x}_2} \quad (3)$$

$$(\hat{p}_1 - \hat{p}_2) \pm z \times \sigma_{\hat{p}_1-\hat{p}_2} \quad (4)$$

Some “Tweaks”

- What if σ is unknown?

Some “Tweaks”

- What if σ is unknown?
  Use s to replace σ, and calculate t-score instead of z-score

Some “Tweaks”

- What if σ is unknown?
  Use s to replace σ, and calculate t-score instead of z-score
- How to calculate t-score again?

$$t = \frac{(\overline{x}_1 - \overline{x}_2) - D_0}{\sigma_{\overline{x}_1 - \overline{x}_2}} \tag{5}$$

Some "Tweaks"

- What if σ is unknown?
  Use s to replace σ, and calculate t-score instead of z-score
- How to calculate t-score again?

$$t = \frac{(\overline{x}_1 - \overline{x}_2) - D_0}{\sigma_{\overline{x}_1 - \overline{x}_2}} \tag{5}$$

- Degree of freedom is calculated by

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2} \tag{6}$$

Some "Tweaks"

- What if σ is unknown?
  Use s to replace σ, and calculate t-score instead of z-score
- How to calculate t-score again?

$$t = \frac{(\overline{x}_1 - \overline{x}_2) - D_0}{\sigma_{\overline{x}_1 - \overline{x}_2}} \tag{5}$$

- Degree of freedom is calculated by

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2} \tag{6}$$

- If we do not know p, we use $\hat{p}$ instead:

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} \tag{7}$$

Some "Tweaks"

- What if σ is unknown?
  Use s to replace σ, and calculate t-score instead of z-score
- How to calculate t-score again?

$$t = \frac{(\overline{x}_1 - \overline{x}_2) - D_0}{\sigma_{\overline{x}_1 - \overline{x}_2}} \tag{5}$$

- Degree of freedom is calculated by

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2} \tag{6}$$

- If we do not know p, we use $\hat{p}$ instead:

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} \tag{7}$$

- Two-tailed tests

Null and Alternative Hypotheses

- The null hypothesis always contains the equal sign

Null and Alternative Hypotheses

- The null hypothesis always contains the equal sign
  - If we want you to test a claim that X is smaller than Y, then this claim is an alternative hypothesis

Null and Alternative Hypotheses

- The null hypothesis always contains the equal sign
  - If we want you to test a claim that X is smaller than Y, then this claim is an alternative hypothesis
  - If we want you to test a claim that X is smaller or equal to Y, then this claim is a null hypothesis

Null and Alternative Hypotheses

- The null hypothesis always contains the equal sign
  - If we want you to test a claim that X is smaller than Y, then this claim is an alternative hypothesis
  - If we want you to test a claim that X is smaller or equal to Y, then this claim is a null hypothesis
  - If we want you to test a claim that P is not equal to Q, then this claim is an alternative hypothesis

Null and Alternative Hypotheses

- The null hypothesis always contains the equal sign
  - If we want you to test a claim that X is smaller than Y, then this claim is an alternative hypothesis
  - If we want you to test a claim that X is smaller or equal to Y, then this claim is a null hypothesis
  - If we want you to test a claim that P is not equal to Q, then this claim is an alternative hypothesis
  - If we want you to test a claim that P is equal to Q, then this claim is a null hypothesis

Comparing Two Means

- Test the alternative hypothesis at $\alpha = 0.05$ that the average temperature in San Diego is higher than the average temperature in San Francisco. A random sample of 33 days and 37 days is obtained from San Diego and San Francisco. Note that $\overline{x}_{SD} = 72, s_{SD} = 10, \overline{x}_{SF} = 65,\ s_{SF} = 12.$

Comparing Two Means

- Test the alternative hypothesis at $\alpha = 0.05$ that the average temperature in San Diego is higher than the average temperature in San Francisco. A random sample of 33 days and 37 days is obtained from San Diego and San Francisco. Note that $\overline{x}_{SD} = 72, s_{SD} = 10, \overline{x}_{SF} = 65,\ s_{SF} = 12.$
- Right-tailed test $H_0 : \mu_{SD} - \mu_{SF} \leq 0$; $H_A : \mu_{SD} - \mu_{SF} > 0$

- Test the alternative hypothesis at $\alpha = 0.05$ that the average temperature in San Diego is higher than the average temperature in San Francisco. A random sample of 33 days and 37 days is obtained from San Diego and San Francisco. Note that $\overline{x}_{SD} = 72, s_{SD} = 10, \overline{x}_{SF} = 65,\ s_{SF} = 12.$
- Right-tailed test $H_0 : \mu_{SD} - \mu_{SF} \leq 0;\ H_A : \mu_{SD} - \mu_{SF} > 0$
- Calculate the t-statistic

$$t = \frac{\overline{x}_1 - \overline{x}_2 - D_0}{\sigma_{\overline{x}_1 - \overline{x}_2}} \tag{8}$$

$$= \frac{72 - 65 - 0}{\sqrt{\frac{10^2}{33} + \frac{12^2}{37}}} = 2.66 \tag{9}$$

Comparing Two Means

- Test the alternative hypothesis at $\alpha = 0.05$ that the average temperature in San Diego is higher than the average temperature in San Francisco. A random sample of 33 days and 37 days is obtained from San Diego and San Francisco. Note that $\overline{x}_{SD} = 72, s_{SD} = 10, \overline{x}_{SF} = 65$, $s_{SF} = 12$.
- Right-tailed test $H_0 : \mu_{SD} - \mu_{SF} \leq 0$; $H_A : \mu_{SD} - \mu_{SF} > 0$
- Calculate the t-statistic

$$t = \frac{\overline{x}_1 - \overline{x}_2 - D_0}{\sigma_{\overline{x}_1 - \overline{x}_2}} \tag{8}$$

$$= \frac{72 - 65 - 0}{\sqrt{\frac{10^2}{33} + \frac{12^2}{37}}} = 2.66 \tag{9}$$

- Find degree of freedom (d.f.=68)

Comparing Two Means: Cont.

- $\overline{x}_{SD} = 72, s_{SD} = 10, \overline{x}_{SF} = 65,\ s_{SF} = 12,\ t = 2.66$, d.f.=68

Comparing Two Means: Cont.

- $\overline{x}_{SD} = 72, s_{SD} = 10, \overline{x}_{SF} = 65,\ s_{SF} = 12,\ t = 2.66$, d.f.=68
- Find the critical vaule

Degree of Freedom	...	**Area in Upper Tail = 0.05**
⋮	⋱	⋮
68	...	1.668

Comparing Two Means: Cont.

- $\overline{x}_{SD} = 72, s_{SD} = 10, \overline{x}_{SF} = 65,\ s_{SF} = 12,\ t = 2.66$, d.f.=68
- Find the critical vaule

Degree of Freedom	...	Area in Upper Tail = 0.05
⋮	⋱	⋮
68	...	1.668

- Compare t-statistic to the critical value:

$$2.66 > 1.668 \tag{10}$$

Comparing Two Means: Cont.

- $\overline{x}_{SD} = 72, s_{SD} = 10, \overline{x}_{SF} = 65,\ s_{SF} = 12,\ t = 2.66$, d.f.=68
- Find the critical vaule

Degree of Filedom	...	**Area in Upper Tail = 0.05**
⋮	⋱	⋮
68	...	1.668

- Compare t-statistic to the critical value:

$$2.66 > 1.668 \tag{10}$$

- We conclude that we can reject the null hypothesis and the average SD temperature is higher than the average SF temperature

Comparing Two Proportions

- Test the alternative hypothesis at $\alpha = 0.01$ that the unemployment rate in San Diego is higher than the unemployment rate in New York. A random sample of 35 people is obtained in San Diego and 6 of them are unemployed. A random sample of 46 people is obtained in New York and 5 of them are unemployed.

Comparing Two Proportions

- Test the alternative hypothesis at $\alpha = 0.01$ that the unemployment rate in San Diego is higher than the unemployment rate in New York. A random sample of 35 people is obtained in San Diego and 6 of them are unemployed. A random sample of 46 people is obtained in New York and 5 of them are unemployed.
- Right-tailed test $H_0 : p_{SD} - p_{NY} \leq 0$; $H_A : p_{SD} - p_{NY} > 0$

Comparing Two Proportions

- Test the alternative hypothesis at $\alpha = 0.01$ that the unemployment rate in San Diego is higher than the unemployment rate in New York. A random sample of 35 people is obtained in San Diego and 6 of them are unemployed. A random sample of 46 people is obtained in New York and 5 of them are unemployed.
- Right-tailed test $H_0 : p_{SD} - p_{NY} \leq 0$; $H_A : p_{SD} - p_{NY} > 0$
- Note that $\hat{p}_1 = \frac{6}{35}$ and $\hat{p}_2 = \frac{5}{46}$

Comparing Two Proportions

- Test the alternative hypothesis at $\alpha = 0.01$ that the unemployment rate in San Diego is higher than the unemployment rate in New York. A random sample of 35 people is obtained in San Diego and 6 of them are unemployed. A random sample of 46 people is obtained in New York and 5 of them are unemployed.
- Right-tailed test $H_0 : p_{SD} - p_{NY} \leq 0$; $H_A : p_{SD} - p_{NY} > 0$
- Note that $\hat{p}_1 = \frac{6}{35}$ and $\hat{p}_2 = \frac{5}{46}$
- Calculate the z-statistic

$$z = \frac{\hat{p}_1 - \hat{p}_2 - D_0}{\sigma_{\hat{p}_1 - \hat{p}_2}} \tag{11}$$

$$= \frac{\frac{6}{35} - \frac{5}{46}}{\sqrt{\frac{\left(\frac{6}{35}\right)\left(1-\frac{6}{35}\right)}{35} + \frac{\left(\frac{5}{46}\right)\left(1-\frac{5}{46}\right)}{46}}} = 0.799 \tag{12}$$

Comparing Two Proportions

- Test the alternative hypothesis at $\alpha = 0.01$ that the unemployment rate in San Diego is higher than the unemployment rate in New York. A random sample of 35 people is obtained in San Diego and 6 of them are unemployed. A random sample of 46 people is obtained in New York and 5 of them are unemployed.
- Right-tailed test $H_0 : p_{SD} - p_{NY} \leq 0$; $H_A : p_{SD} - p_{NY} > 0$
- Note that $\hat{p}_1 = \frac{6}{35}$ and $\hat{p}_2 = \frac{5}{46}$
- Calculate the z-statistic

$$z = \frac{\hat{p}_1 - \hat{p}_2 - D_0}{\sigma_{\hat{p}_1 - \hat{p}_2}} \tag{11}$$

$$= \frac{\frac{6}{35} - \frac{5}{46}}{\sqrt{\frac{\left(\frac{6}{35}\right)\left(1-\frac{6}{35}\right)}{35} + \frac{\left(\frac{5}{46}\right)\left(1-\frac{5}{46}\right)}{46}}} = 0.799 \tag{12}$$

- Find p-value: $\Pr(z > 0.799) = 1 - 0.787 = 0.213 > 0.01$

Comparing Two Proportions

- Test the alternative hypothesis at $\alpha = 0.01$ that the unemployment rate in San Diego is higher than the unemployment rate in New York. A random sample of 35 people is obtained in San Diego and 6 of them are unemployed. A random sample of 46 people is obtained in New York and 5 of them are unemployed.
- Right-tailed test $H_0 : p_{SD} - p_{NY} \leq 0$; $H_A : p_{SD} - p_{NY} > 0$
- Note that $\hat{p}_1 = \frac{6}{35}$ and $\hat{p}_2 = \frac{5}{46}$
- Calculate the z-statistic

$$z = \frac{\hat{p}_1 - \hat{p}_2 - D_0}{\sigma_{\hat{p}_1 - \hat{p}_2}} \tag{11}$$

$$= \frac{\frac{6}{35} - \frac{5}{46}}{\sqrt{\frac{\left(\frac{6}{35}\right)\left(1-\frac{6}{35}\right)}{35} + \frac{\left(\frac{5}{46}\right)\left(1-\frac{5}{46}\right)}{46}}} = 0.799 \tag{12}$$

- Find p-value: $\Pr(z > 0.799) = 1 - 0.787 = 0.213 > 0.01$
- We fail to reject the null hypothesis and we do not have evidence that the SD unemployment rate is higher

Conclusion

- One population or two populations

Conclusion

- One population or two populations
- Mean or proportion

Conclusion

- One population or two populations
- Mean or proportion
- One-tailed or two-tailed tests

Conclusion

- One population or two populations
- Mean or proportion
- One-tailed or two-tailed tests
- Whether we know σ or not

Conclusion

- One population or two populations
- Mean or proportion
- One-tailed or two-tailed tests
- Whether we know σ or not
- Whether we should compute a z-statistic or t-statistic