Inferences When Comparing Two Means

Dr. Tom Ilvento
FREC 408

Thus far…

- We have made an inference from a single sample mean and proportion to a population, using
  - The sample mean (or proportion)
  - The sample standard deviation
  - Knowledge of the sampling distribution for the mean (proportion)
- And it matters if the sample size is large or small

Testing differences between two means or proportions

- The same strategy will apply for testing differences between two means or proportions
- With a few twists
  - Mean
    - Large sample
    - Small sample – pool the variance
  - Proportions
    - When testing H_0: we need to check if $p_1 = p_2$

Testing differences between two means or proportions

- We will also need to come up with:
  - An estimator of the difference of two means/proportions
  - The standard error of the sampling distribution for our estimator
- With two sample problems we have two sources of variability and sampling error
- We also must assume the samples are independent random samples

What are independent, random samples?

- Independent samples means that each sample and the resulting variables do not influence the other sample
  - If we sampled the same subjects at two different times we would not have independent samples
  - If we sampled husband and wife, they would not be independent
- However, we have a strategy to assess change over time of the same subject – paired difference test

Decision Tree for Two Means

Target	Assumptions	Test Statistic
H_0: $\propto_1-\propto_2=D$	Independent random samples Large sample size (n_1, n_2>30)	z, using sample variance
	Independent random samples Small sample size Populations appr. normal **Equal variances**	t, using pooled variance S_p^2

Decision Tree for two Proportions

Testing	Assumptions	Test Statistic
$H_0: p_1 - p_2 = 0$	Independent random samples Large sample size ($n_1, n_2 > 30$) Known that $p_1 = p_2$ under H_0	z, using **pooled sample** proportion P_p
$H_0: p_1 - p_2 = D_0$	Independent random samples Large sample size **When $D_0 \neq 0$**	z

Example Problem

- Two groups of students were surveyed about lecture notes
  - 86 students in a promotional strategy class that required the purchase of the lecture notes
  - 35 students enrolled in a sales/retailing class which didn't offer lecture notes
  - At the end of the semester, students in both classes were asked if *"Having a copy of the lecture was [would be] helpful in understanding the material."*
  - The question was measured on a nine point scale where 1= strongly disagree and 9 = strongly agree

Lecture Notes problem

Class with Lecture Notes	Class without Lecture Notes
$n_1 = 86$	$n_2 = 35$
$\bar{x}_1 = 8.48$	$\bar{x}_2 = 7.80$
$s_1^2 = .94$	$s_2^2 = 2.99$

Do the samples provide sufficient evidence to conclude that there is a difference in mean responses of the two groups?

Use $\alpha = .01$

Lecture Notes problem

Null hypothesis	$H_0: (\mu_1 - \mu_2) = ?$
Alternative	$H_a: (\mu_1 - \mu_2) \neq ?$ two-tailed test
Assumptions	Two independent samples, n is Large
Test Statistic	$z^* =$
Rejection Region	$z_{\alpha=.01/2} = 2.575$
Calculation	$z^* =$
Conclusion	z^* ? $z_{\alpha=.005}$? ?? $H_0: (\mu_1 - \mu_2) = 0$

We need to figure out the sampling distribution $(\bar{x}_1 - \bar{x}_2)$

- The mean of the sampling distribution for $(\bar{x}_1 - \bar{x}_2)$
- Will equal = $(\mu_1 - \mu_2)$
- = D_0
  - We usually designate the expected difference as D_0 under the the null hypothesis
  - Most often we think of $D_0 = 0$; no difference

Standard Error of the difference of two means

- The **Standard Error** of the difference of two means is given as:

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

Page 454

The sampling distribution of $(\bar{x}_1 - \bar{x}_2)$ is approximately normal for large samples under the **Central Limit Theorem**

The Standard Error for the difference of two means

- Is based on two **independent random samples**
- We typically use the sample estimates of σ_1 and σ_2
  - Which are s_1 and s_2

$$\sigma_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

The Test Statistic for our problem

$$z^* = \frac{(8.48 - 7.80) - 0}{\sqrt{\frac{.94}{86} + \frac{2.99}{35}}}$$

$Z^* = (.68 - 0)/.3104 = 2.1906$

Comparison of two means

Null hypothesis	$H_0: (\mu_1 - \mu_2) = 0$
Alternative	$H_a: (\mu_1 - \mu_2) \neq 0$ two-tailed test
Assumptions	Two independent samples, n is Large
Test Statistic	$z^* = (8.48-7.80-0)/[(.94/86)+(2.99/35)]^{.5}$
Rejection Region	$z_{\alpha=.01/2} = 2.575$
Calculation	$z^* = 2.19$
Conclusion	$z^* < z_{\alpha=.005}$ $2.19 < 2.575$ Cannot reject $H_0: (\mu_1 - \mu_2) = 0$

99% Confidence Interval for the Difference of Two Means

- $(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sigma_{(\bar{x}_1 - \bar{x}_2)}$
- $(8.48-7.80) \pm z_{.01/2}[(.94/86) + (2.99/35)]^{.5}$
- $.68 \pm 2.575(.3104)$
- **$.68 \pm .80$**
- **$-.12$ to 1.48**
- Notice the 99% C.I. contains the null hypothesis value, zero

Decision Tree for Two Means

Target	Assumptions	Test Statistic
H_0: $\alpha_1 - \alpha_2 = D$	Independent random samples Large sample size (n_1, $n_2 > 30$)	z, using sample variance
	Independent random samples Small sample size Populations appr. normal **Equal variances**	t, using pooled variance S_p^2

What about when n is small?

- We will use a **t-test** and the t distribution
- Assumptions
  - Both samples are **approximately normal**
  - The **<u>population variances are equal</u>**
  - **Random** samples selected **independentl**y of each other