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_o: 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: $\mu_1-\mu_2=D$	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	$z^* = (8.48\text{-}7.80\text{-}0)/[(.94/86)+(2.99/35)]^{.5}$
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

- $(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sigma_{(\bar{x}_1-\bar{x}_2)}$
- $(8.48\text{-}7.80) \pm z_{.01/2}\,[(.94/86) + (2.99/35)]^{.5}$
- $.68 \pm 2.575(.3104)$
- **$.68 \pm .80$**
- **-.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: $\mu_1-\mu_2=D$	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 **independent**ly of each other

The standard error for a small sample difference of means

- **Since we assume $\sigma_1 = \sigma_2$**
- **thus ($s_1 = s_2$)**
- We should pool our estimate of the standard error of the sampling distribution
- Using information from both sample estimates

POOLED ESTIMATE OF THE VARIANCE

- Then our formula will be a weighted average of s_1 and s_2

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

Page 458

Note: the denominator reduces to ($n_1 + n_2 - 2$) which is the d.f. for the t distribution

Next, we use the Pooled Estimate of the Variance to calculate the estimate of the standard error

$$\hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} = \sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$\hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)} = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

What does pooling do for us?

- Pooling generates a weighted average as the estimate of the variance
- The weights are the sample sizes for each sample
- A pooled estimate is thought to be a better estimate if we can assume the variances are equal
- And our degrees of freedom are larger - d.f.= $n_1 + n_2 - 2$
- Which means the t-value will be smaller

Problem, Tapeworms in sheep

- An experiment was done to compare the mean number of tapeworms in the stomachs of sheep that had been treated for worms versus those not treated.
- There were 7 sheep in the Treatment group and 7 in the Control Group
- **Is the number of tapeworms lower in the treatment group at α=.05?**

Tapeworms in sheep

- The means and standard deviations are:

Treatment	**Control**
$\bar{x}_1$ = 28.57	$\bar{x}_2$ = 40.0
s_1^2= 198.62	s_2^2= 215.33
n_1= 7	n_2= 7

Tapeworms in sheep

- We assume the variances are equal so we make a pooled estimate

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

Tapeworms in sheep

- We assume the variances are equal so we make a pooled estimate

$$s_p = \sqrt{\frac{(7-1)198.62 + (7-1)215.33}{7+7-2}} =$$

$$s_p = \sqrt{\frac{2,483.7}{12}} = 14.387$$

Tapeworms in sheep

- Then we use our pooled estimate to estimate the standard error of the sampling distribution for the difference of means

$$\hat{\sigma}_{(\bar{x}_1 - \bar{x}_2)} = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Tapeworms in sheep

- Then we use our pooled estimate to estimate the standard error of the sampling distribution for the difference of means

$$s_{(\bar{x}_1 - \bar{x}_2)} = 14.387\sqrt{\frac{1}{7} + \frac{1}{7}} =$$

$$s_{(\bar{x}_1 - \bar{x}_2)} = 14.387(.5345) = 7.69$$

Tapeworms in sheep

Null hypothesis	H_0: $(\mu_1 - \mu_2) = 0$
Alternative	H_a: $(\mu_1 - \mu_2) < 0$ **one-tailed test, lower**
Assumptions	Small independent samples, approx normal, variances are equal
Test Statistic	t* = (28.57-40.00 - 0)/7.69
Rejection Region	$-t_{.05,\ 12\ d.f.} = -1.782$
Calculation	t* =-1.48
Conclusion	$t^* > -t_{.05,\ 12\ d.f.}$ -1.48 > -1.782 We cannot reject H_0: $(\mu_1 - \mu_2) = 0$

The 90% C.I. For the Tapeworm example

- $(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2, n_1+n_2-2 d.f.} \sigma_{(\bar{x}_1 - \bar{x}_2)}$
- $(28.57-40.00) \pm t_{.01/2,\ n1+n2-2\ d.f.} [14.387\ (1/7+1/7)^{.5}]$
- -11.43 ± 1.782(7.690)
- -11.43 ± 13.704
- -25.134 to 2.273
- This C.I. contains zero – analogous to null hypothesis, one-tailed test, at α=.05

What Do I need for a Difference of Means Tests?

- Two independent random samples
- Determine if I am dealing with a mean or proportion
- If a Mean, large sample or small sample
  - Large sample I need not assume the distribution of the populations, or anything about the variances
  - I calculate the standard error and make my test using a z-value

---

- Small sample I must assume
  - Independent random samples
  - From populations that are normally distributed
  - The variances are equal
- Small Sample Steps
  - Calculate pooled estimate of Variance
  - Use pooled estimate to calculate the standard error
  - Conduct the test use a t-value with n1 + n2 – 2 df

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**

What about the difference in proportions?

- Based on large sample only
- Same strategy as for the mean
  - We calculate the difference in the two sample proportions
  - Establish the sampling distribution for our estimator
  - Calculate a standard error of this sampling distribution
  - Conduct a test

Differences of Proportions

- For the null hypothesis
  - E($\hat{p}_1 - \hat{p}_2$) = (p_1-p_2) = D_0
- The Standard Error for ($\hat{p}_1 - \hat{p}_2$)

$$\sigma_{(\hat{p}_1 - \hat{p}_2)} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

For Proportions

- Since this is a large sample problem, we could use the sample estimates of p_1 and p_2 to estimate the standard error

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

Page 476

For Proportions

- **Note: Under the Null Hypothesis where $p_1 - p_2 = 0$**

- The book suggests that we use a weighted average for p_p based on adding the total number of successes and divide by the sum of the two sample sizes (P477-78)
  - $p_p = (x_1 + x_2)/(n_1 + n_2)$
  - where x = # of successes

Note : the book uses $\hat{p}$ instead of p_p

So for Proportions

- **For a confidence interval**, use the sample estimates in this manner to generate the standard error – **since there is no assumption that the variances are equal**

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

So for Proportions

- **For a Hypothesis Test** where the proportion are equal, pool the information and use this approach
  - $p_p = (x_1 + x_2)/(n_1 + n_2)$ where x = # of successes
  - $q_p = 1 - p_p$

$$\sigma_{(p_1-p_2)} = \sqrt{\hat{p}_p\hat{q}_p\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Gender gap in politics

- Over half the votes will be cast by women
- The question is if the Democratic platform is viewed more favorably by women
- Suppose 150 men and 150 women stated their party preferences for Republicans
  - Data survey of 300 – 150 men and 150 women
  - 81 men favor Republicans
  - 70 women favor Republicans
- Conduct a test using α=.05 that a **lower proportion of women favor Republicans**

Calculate the standard error

- The sample proportions are
  - Men 81/150 = .540
  - Women 70/150 = .467

- Pooled estimate
  - p_p= (81+70)/(150+150) = .5033
  - q_p= 1 - .5033 = .4967

Calculate the standard error

- Standard Error for the problem
  - $\sigma_{(p1-p2)}$ =
    = $[(.5033)(.4967)(1/150 + 1/150)]^{.5}$

  - $\sigma_{(p1-p2)}$ = $[(.249989)(.01333)]^{.5}$
  - $\sigma_{(p1-p2)}$ = .057734

What will be the test statistic?

- **z* = [(.540-.467)- 0]/.057734**
- Where the numerator shows the difference in voting Republican for men and women
- Note that I subtract out the hypothesized value $D_o = 0$
- I usually try to keep the z* for a difference of proportion (or means) positive as a matter of convenience

Gender Gap in Politics

Null hypothesis	H_0: $(p_m-p_w) = 0$
Alternative	H_a: $(p_m-p_w) > 0$ **one-tailed test, upper**
Assumptions	Large sample proportion, use normal, assume variances equal
Test Statistic	z* = [(.540-.467)-0]/ .057734
Rejection Region	$z_{.05,} = 1.645$
Calculation	z* = 1.264
Conclusion	$z^* < z_{.05}$ 1.264 < 1.645 Cannot reject H_0: $(p_m-p_w) = 0$

Confidence Interval for the Gender in Politics Problem

- We will use α=.10 to relate to the previous one-tailed hypothesis test of α=.05
- Standard error (not assuming equal variances) is:
  - $\sigma_{(p1-p2)}$=[(.540)(.460)/150 + (.467)(.533)/150]$^{.5}$
  - **= .05758**
- 90% C.I.
  - (.540-.467)±1.645(.05758)
  - .073 ± .095
  - -.022 to .168

Section 9.3: Paired Difference Test

- When you have a situation where we record a pre and post test for the same individual
- We cannot treat the samples as independent
- In these cases we can do a **Paired Difference Test.**
  - It's called **"Matched Pairs Test"** in the book. (page467-70)

Paired Difference Test

- The strategy is relatively simple
- We simply create a new variable which is the difference of the pre-test from the post test
- This new variable can be thought as a **single** random sample
- For this new variable we calculate sample estimates of the mean and standard deviation

Paired Difference Test

- And then calculate C.I. or conduct a hypothesis test on this new variable
- Often times the mean difference is referred to as
  - $\bar{x}_D$
- And the hypothesis is often:
  - H_0: $\mu_D = 0$
- This is no different than any single mean test, large sample or small sample

Example of paired difference data

Patient	Time 1	Time 2	Difference 2 - 1
1	5	5	0
2	1	3	-2
3	0	0	0
4	1	1	0
5	0	1	-1
6	2	1	1

Alzheimer's study

- Twenty Alzheimer patients were asked to spell 24 homophone pairs given in random order
  - Homophones are words that have the same pronunciation as another word with a different meaning and different spelling
  - Nun and none; doe and dough
- The number of confusions were recorded
- The test was repeated one year later

Alzheimer's study

- The researchers posed the following question:

  Do Alzheimer's patients show a significant increase in mean homophone confusion over time?
- Use an alpha value of .05

Alzheimer's study

Statistics	Time 1	Time 2	Difference
Mean	4.15	5.80	1.65
Standard Error	0.78	0.94	0.72
Median	5.00	5.50	1.00
Mode	5.00	3.00	0.00
Standard Deviation	3.50	4.21	3.20
Sample Variance	12.24	17.75	10.24
Kurtosis	-0.85	0.13	0.08
Skewness	0.41	0.64	0.48
Range	11	16	12
Minimum	0	0	-3
Maximum	11	16	9
Sum	83	116	33
Count	20	20	20

Alzheimer's study

Null hypothesis	H_0: $\mu_D = 0$
Alternative	H_a: $\mu_D > 0$ one-tailed test, upper
Assumptions	small sample, normal
Test Statistic	$t^* = (1.65 – 0)/(3.2/\sqrt{20})$
Rejection Region	$t_{\alpha=.05,\ 19\ d.f.} = 1.729$
Calculation	$t^* = 2.31$
Conclusion	$t^* > t_{\alpha=.05,\ 19\ d.f.}$ $2.31 > 1.729$ Reject H_0: $\mu_D = 0$