FREC 408

Group Exercise 8b

On Wednesday I handed out a sheet on how to use Excel to do a difference of means test.

- I want your group to conduct a hypothesis test for a difference of means for cholesterol levels between males and females.
- Then we will explore the outputs from Excel and SAS.

The Descriptive Statistics are given on the handout.

- Briefly review the descriptive statistics and describe differences between the two groups.

Describe Differences

Cholesterol Level		
	Females	Males
Mean	200.32	196.09
Standard Error	0.88	0.97
Median	201	196
Mode	194	196
Standard Deviation	10.72	12.37
Sample Variance	114.94	153.07
Kurtosis	-0.49	0.02
Skewness	-0.11	0.09
Range	47	61
Minimum	176	166
Maximum	223	227
Sum	29647	32156
Count	148	164
Confidence Level(95.0%)	1.74	1.91

Description

- The mean level of cholesterol for females is slightly higher (200.32 versus 196.09)
- The median for each group is very close to the mean, indicating no extreme outliers and possibly a symmetrical distribution
- The spread of the data for males is larger than that of females.
  - Males have the larger range, from 166 to 227
  - Std Dev for males is 12.37 versus 10.72 for females

95% Confidence Interval

- 95% C.I. for females
  - 200.32 ± 1.96(.8813) =
  - 200.32 ± 1.73

  From Excel output 200.32 ± 1.74

- 95% C. I. for males
  - 196.09 ± 1.96(.9661) =
  - 196.09 ± 1.89

  From Excel output 196.09 ± 1.91

Hypothesis Test

- Conduct the test that females cholesterol levels are different from males.
- The null hypothesis should be that the two groups are equal.
- **Do this by hand.**
- Use $\alpha = .05$.

Hypothesis Test

Null Hypothesis	Hc: $\mu_1 - \mu_2 = 0$
Alternative	Hc: $\mu_1 - \mu_2 \neq 0$
Assumptions of	Large sample, use a z, standard normal table for
Test Statistic (z^* or t^*)	$z^* =$ (200.32 - 196.09 - 0)/[(114.94/148) + (153.07/164)]*
Rejection Region	$z_{.05/2} = 1.96$
Calculation of Test Statistics	$z^* = 8.23$
Comparison of z^*/t^* with Rejection Region	$z^* > z_{.05/2}$ 8.23 > 1.96 We can reject Hc: $\mu_1 - \mu_2 = 0$ p-value < .02

Now Look at Excel Output

- **Look at the output from Excel for the same problem.** Find the same elements for the Hypothesis Test you just calculated.
- What is the meaning of:
  - **P(T<=t)one-tail**
  - **t Critical one-tail**

Excel Output

t-Test: Two-Sample Assuming Unequal Variances

	Females	*Males*
Mean	200.32	196.09
Variance	114.94	153.07
Observations	148	164
Hypothesized Mean Difference	0	
df	310	
t Stat	3.24	
P(T<=t) one-tail	0.00	
t Critical one-tail	1.65	
P(T<=t) two-tail	0.00	
t Critical two-tail	1.97	

Excel Output

- What is the meaning of:
  - **P(T<=t)one-tail**
    - This is the p-value for the test statistic, under a one-tailed test p = .0006706
  - **t Critical one-tail 1.65**
    - This the t-value at the rejection region for a one-tailed test

SAS Output

SAS Output

- **Look at the results from SAS and find the same elements.** What do you think is the meaning of the Equality of Variances test?
  - When n is large we don't have to assume equal variances, but we would gain information if we could.
  - The equality of variance test is the ratio of the variances - we test for a ratio = 1.
  - Since p = .077, this test did not meet the criteria for significance at the alpha = .05 level. **Hence we could assume the variances are equal**

What if we assumed equal variances in Excel?

t-Test: Two-Sample Assuming Equal Variances

	Females	*Males*
Mean	200.32	196.09
Variance	114.94	153.07
Observations	148	164
Pooled Variance	134.99	
Hypothesized Mean Difference	0	
df	310	
t Stat	3.213	
P(T<=t) one-tail	0.001	
t Critical one-tail	1.650	
P(T<=t) two-tail	0.001	
t Critical two-tail	1.968	

Geneticists have identified E2F1 transcription factor as an important component of cell proliferation control.

- The researchers induced DNA synthesis in two batches of serum-starved cells.
- In one group of 92 cells (treatment), cells were micro-injected with the E2F1 gene.
- A control group of (158 cells) was not exposed to E2F1.
- After 30 hours, researchers determined the number of altered growth cells in each batch. Test to see if the proportion for the treatment group is larger than that of the control.
- Conduct the hypothesis test using $\alpha = .01$.

This is a difference of proportions test, and here is the data

	Control	E2F1 Treated
Total Cells	**158**	**92**
Number of growth altered cells	**15**	**41**
	P =	**P =**

This is a difference of proportions test, and here is the data

	Control	E2F1 Treated
Total Cells	**158**	**92**
Number of growth altered cells	**15**	**41**
	P = .0949	**P = .4457**

Estimate a "pooled" proportion based on the Null Hypothesis that the groups are equal

	Control	E2F1 Treated
Total Cells	**158**	**92**
Number of growth altered cells	**15**	**41**
	P = .0949	**P = .4457**
Pooled estimate of $p_p = (15 + 41)/(158 + 92)$ $p_p = 56/250 = .224$ $q_p = .776$		