# FREC 408

## Group Exercise 7

## Problem 1: Hypothesis test of a proportion

- **A recent survey by Opinion Research Corporation** focused on the perception of voters as to their obligation to vote. The poll, taken Oct. 16-22, covered a sample of 400 people between 18 and 30 were interviewed, with a margin of error of 5.5 percentage points. The respondents were asked if they thought voting was a choice or a constitutional duty. **For the respondents aged 18 to 30, 59% indicated that they felt voting was a choice.**
- Conduct a hypothesis test to see if the true proportion of young citizens (aged 18 to 30) who feel voting is a choice is greater than 50%.

## Voting as a Choice

| | |
|---|---|
| **Null hypothesis** | **$H_0$: p = .5** |
| **Alternative** | **$H_a$: p > .5** one-tailed test, upper |
| **Assumptions** | **Large sample, binomial = normal** |
| **Test Statistic** | **$z^* = (.59 - .5)/(.25/400)^{.5}$** |
| **Rejection Region** | **$z_{\alpha=.05} = 1.645$** |
| **Calculation** | **$z^* = 3.6$** |
| **Conclusion** | **$z^* > z_{\alpha=.05}$**<br>**3.6 > 1.645**<br>**Reject $H_0$: p = .5** |

**p-value for z* = 3.6 off the z-table, so p < .001**

## Problem

- **The current pain reliever in a hospital brings relief in 3.5 minutes on average**
- **A new pain reliever is tried on a sample of 20 people**
- **Mean time is 2.8 minutes with s=1.14 minutes**
- **Do the data provide sufficient evident that the new pain reliever was effective in reducing the mean time to relief?**
- **Use $\alpha$= .01**

## Pain Relief

| | |
|---|---|
| **Null hypothesis** | **$H_0$: $\mu$ = 3.5** |
| **Alternative** | **$H_a$: $\mu$ < 3.5** one-tailed test, lower |
| **Assumptions** | **Small sample, t distribution** |
| **Test Statistic** | **$t^* = (2.8 - 3.5)/(1.14/\sqrt{20})$** |
| **Rejection Region** | **$t_{\alpha=.01,\ 19\ d.f.} = -2.539$** |
| **Calculation** | **$t^* = -2.746$** |
| **Conclusion** | **$t^* < t_{\alpha=.01,\ 19\ d.f.}$**<br>**-2.746 < -2.539**<br>**Reject $H_0$: $\mu$ = 3.5** |

## Problem – Support for Preserving Agricultural Lands

- **States have started programs to help preserve agricultural lands and keep them from being developed**
- **One strategy is to think of the value of agricultural land as having a**
  - **Value as its use for agriculture**
  - **Value as its use for development**
- **State programs seek to purchase the development rights from the farmer and pay him/her for these rights**

## Survey of Delaware Households

- ***I support the state's efforts to preserve farmland by purchasing development rights from farmers***
- 550 **responded to this question**
  - 325 **said Strongly Agree or Agree**
- **Conduct a hypothesis Test to see if the the majority of Delawarean households agree with this statement (i.e., more than half support this program)**
- **Use** $\alpha$ = .01 **level**

## Support for Farmland Preservation

**Null hypothesis** $H_0$: p = .5
**Alternative** $H_a$: p >.5 one-tailed test,upper
**Assumptions** Large sample, binomial = normal
**Test Statistic** $z^* = (.591-.5)/(.25/550)^{.5}$
**Rejection Region** $z_{\alpha=.01} = 2.33$
**Calculation** $z^* = 4.27$
**Conclusion** $z^* > z_{\alpha=.01}$
4.27 > 2.33
Reject $H_0$: p = .5

## Golf Course Problem

- **Golf course designers are worried that the new equipment is making old courses obsolete.**
- **One designer says that courses need to be built with the expectation that players will be able to drive the ball an average of 250 yards or more.**

## Golf Course Problem

- **A sample of 135 golfers is taken and they measured their driving distance**
  - $\bar{x}$ **= 256.3 yards**
  - **s = 43.4 yards**
- **Does the sample provide enough evidence to suggest that golfers are already hitting it farther than the 250 mark? Use** $\alpha$**=.05**
- **Also, calculate the p-value for this problem**

## Golf Course Problem

**Null hypothesis** $H_0$: $\mu = 250$
**Alternative** $H_a$: $\mu > 250$ one-tailed test,upper
**Assumptions** Large sample, normal
**Test Statistic** $z^* = (256.3-250)/(43.4/\sqrt{135})$
**Rejection Region** $z_{\alpha=.05} = 1.645$
**Calculation** $z^* = 1.69$
**Conclusion** $z^* > z_{\alpha=.05}$
1.69 > 1.645
Reject $H_0$: $\mu = 250$

**p= .0455**

## Problem – 800 number service response time

- **A proposal has been made that would allow people to put a portion of their Social Security payroll taxes into personal retirement accounts that would be invested in private stocks and bonds.**
- **A random group of adults in the U.S. were asked if they favored or opposed this proposal.**
- **Of the 884 adults aged 30 and older who responded, 513 said they favored the proposal.**
- **Test to see if more than half of adults 30 years and older favor the proposal.**
- **Use** $\alpha$ **= .05**

## Customer Response time

| | |
|---|---|
| **Null hypothesis** | $H_0$: p = .5 |
| **Alternative** | $H_a$: p >.5 one-tailed test, upper |
| **Assumptions** | Large sample, binomial = normal |
| **Test Statistic** | $z^* = (.58 - .5)/(.25/884)^{.5}$ |
| **Rejection Region** | $z_{\alpha=.05} = 1.645$ |
| **Calculation** | $z^* = 4.757$ |
| **Conclusion** | $z^* > z_{\alpha=.05}$<br>4.757 < 1.645<br>Reject $H_0$: p = .5 |

## 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 | 32158 |
| **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** | **$H_0: \mu_1 - \mu_2 = 0$** |
| **Alternative** | **$H_0: \mu_1 - \mu_2 \neq 0$** |
| **Assumptions of** | **Large sample, use a z, standard normal table for** |
| **Test Statistic (z* or t*)** | **$z^* = (200.32 - 196.09 - 0)/[(114.94/148) + (153.07/164)]^{.5}$** |
| **Rejection Region** | **$z_{.05/2} = 1.96$** |
| **Calculation of Test Statistics** | **$z^* = 3.23$** |
| **Comparison of z*/t* with Rejection Region** | **$z^* > z_{.05/2}$<br>$3.23 > 1.96$<br>We can reject $H_0: \mu_1 - \mu_2 = 0$ 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

The TTEST Procedure

Statistics

| Variable | GENDER | N | Lower CL Mean | **Mean** | Upper CL Mean | Lower CL Std Dev | Std Dev | Upper CL Std Dev | Std Err |
|---|---|---|---|---|---|---|---|---|---|
| CHOLEST | 0 | 164 | 194.18 | 196.09 | 197.99 | 11.163 | 12.372 | 13.878 | 0.9661 |
| CHOLEST | 1 | 148 | 198.58 | 200.32 | 202.06 | 9.623 | 10.721 | 12.104 | 0.8813 |
| CHOLEST | Diff (1-2) | | -6.824 | **-4.232** | -1.64 | 10.772 | 11.619 | 12.611 | 1.3173 |

T-Tests

| Variable | Method | Variances | DF | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| CHOLEST | Pooled | Equal | 310 | -3.21 | 0.0015 |
| **CHOLEST** | **Satterthwaite** | **Unequal** | **310** | **-3.24** | **0.0013** |

Equality of Variances

| Variable | Method | Num DF | Den DF | F Value | Pr > F |
|---|---|---|---|---|---|
| CHOLEST | Folded F | 163 | 147 | 1.33 | 0.0770 |

## 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)<br>$p_p$ = 56/250 = .224 $q_p$ = .776 | | |

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

## Conduct a Hypothesis Test

| | |
|---|---|
| **Null Hypothesis** | **$H_0$: $p_T - p_C = 0$** |
| **Alternative** | **$H_0$: $p_T - p_C > 0$** |
| **Assumptions of** | **Proportions, large sample, equality of p under** |
| **Pooled Estimate of p** | **$p_p = (15 + 41)/(158 + 92) = 56/250 = .225$**<br>**$q_p = .776$** |
| **Test Statistic (z* or t*)** | **$z^* =$**<br>**$(.4457 - .0949 - 0)/[(.224)(.776)(1/158 + 1/92)]^{.5}$** |
| **Rejection Region** | **$z_{.01} = 2.33$** |
| **Calculation of Test** | **$z^* = 6.4159$** |
| **Comparison of z* or t* with Rejection Region** | **$z^* > z_{.01}$**<br>**$6.4159 > 2.33$**<br>**We can reject $H_0$: $p_T - p_C = 0$** |