

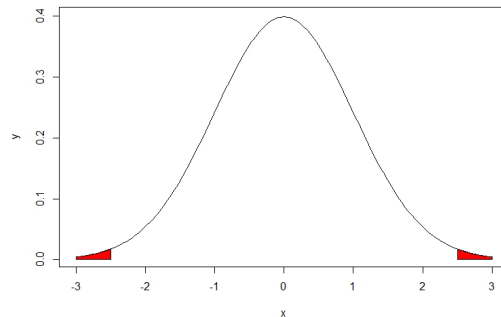
## 10.4 Hypothesis Testing: Two Independent Samples Proportion

### Example 3:

Smoking cigarettes has been known to cause cancer and other ailments. One politician believes that a higher tax should be imposed on those who smoke. An observational study was conducted in a suburb of Baltimore where 900 random people were asked the following question “Should the federal tax on cigarettes be raised to pay for health care reform?”. Of those that were asked this question, 650 respondents were non-smokers ( $NS$ ) and 250 respondents were smokers ( $S$ ). 480 non-smokers said yes whereas 173 smokers said yes. Is there a statistical difference between the opinions of those who smoke and those who do not, with a significance level of 0.01?

Summary of Information:

- $n_{NS} = 650$  and  $n_S = 250$
  - $x_{NS} = 480$  and  $x_S = 173$
  - $\hat{p}_{NS} = 480/650 = 0.7385$  and  $\hat{p}_S = 173/250 = 0.692$
1. Identify and State the Statistical Question: **Is there a statistical difference between the opinions of those who smoke and those who do not**
    - Determine the variable(s) of interest: **Opinion, Yes/No**
    - Determine the type variable(s) (i.e., quantitative or qualitative): **Qualitative, Nominal**
    - Identify and state the hypotheses (Null and Alternative Hypotheses) based on the question at hand:  $H_0 : p_{NS} = p_S$  and  $H_1 : p_{NS} \neq p_S$
  2. Identify and state level of significance  $\alpha$  (the probability of rejecting the  $H_0$  when  $H_0$  is true):  **$\alpha/2 = 0.01/2 = 0.005$**



Really IMPORTANT:

- $\alpha: 0.005$
- Critical Value:  $|-2.58| = 2.58$

3. Perform Statistical Test and Interpret Results

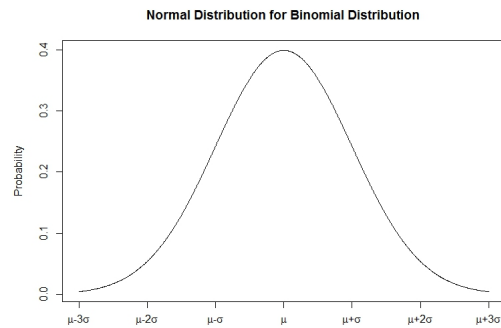
$$TS = z = \frac{0.7385 - 0.692}{\sqrt{\frac{(0.7256)(0.2744)}{650} + \frac{(0.7256)(0.2744)}{250}}}$$

where

$$\bar{p} = \frac{x_{NS} + x_S}{n_{NS} + n_S}$$

and  $\bar{q} = 1 - \bar{p}$

$$TS = z = \frac{\hat{p}_{NS} - \hat{p}_S}{\sqrt{\frac{\bar{p}\bar{q}}{n_{NS}} + \frac{\bar{p}\bar{q}}{n_S}}} = 1.4003$$



- Test Statistic: **[1.4003] = 1.4003**
- p-value: **From z table  $2(1 - 0.9192) = 2 * 0.0808 = 0.1616$**

4. State the sample, null hypothesis, test that was used, and conclusion with non-statistical terms

- $n_{NS} = 650$  and  $n_S = 250$
- No difference between opinion in raising health care tax between smokers and non-smokers
- A test for two independent sample proportions
- Fail to Reject the claim that there no difference between opinion in raising health care tax between smokers and non-smokers

## 10.5 Hypothesis Testing: Two Independent Samples Standard Deviation

We did not do a hypothesis test for one sample standard deviation due to time, however, understanding the process for hypothesis testing of two groups standard deviations (variation) is critical to a concept discussed later this quarter.

Review Assumptions:

1. Independent Samples
2. Two populations are normally distributed

**Note:** The standard deviation follows a Chi-Squared Distribution. Some small facts:

- The standard deviation is always positive  $\implies$  Chi-square Distribution is always positive
- This distribution requires degrees of freedom which is  $n - 1$

What other distribution requires degrees of freedom?

There is a relationship

**Example 4:**

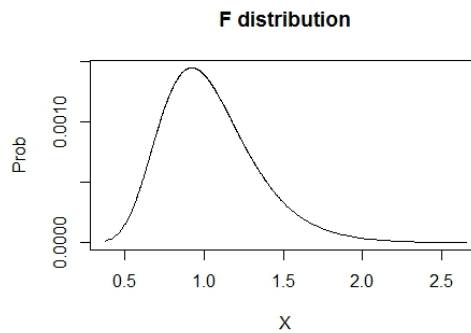
Researchers conducted a study to determine whether Treatment 1 works better than Treatment 2 in pain reduction. Use a significance level 0.05 to test the claim that those that had Treatment 1 vary more than those that had Treatment 2. **Process for Hypothesis**

Table 3: Summary of Data

|             | $n$ | $\bar{x}$ | $s$  |
|-------------|-----|-----------|------|
| Treatment 1 | 20  | 0.44      | 1.4  |
| Treatment 2 | 20  | 0.49      | 0.96 |

**Testing for this class:**

1. Identify and State the Statistical Question: **Does Treatment 1 and 2 vary at the same rate?**
  - Determine the variable(s) of interest: **Measure of Pain**
  - Determine the type variable(s) (i.e., quantitative or qualitative): **Quantitative**
  - Identify and state the hypotheses (Null and Alternative Hypotheses) based on the question at hand:  $H_0 : \sigma_1 = \sigma_2$  and  $H_1 : \sigma_1 > \sigma_2$
2. Identify and state level of significance  $\alpha$  (the probability of rejecting the  $H_0$  when  $H_0$  is true)  $\alpha = 0.05$

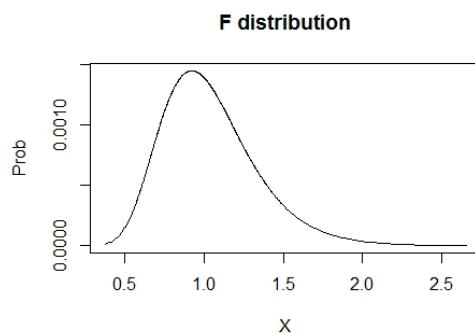


Really IMPORTANT:

- Degrees of Freedom:  $df_1 = 19, df_2 = 19$ :
- $\alpha$ :  $\alpha = 0.05$
- Critical Value = 2.155 (F table, Appendix A)

3. Perform Statistical Test and Interpret Results

$$TS = F = \frac{s_1^2}{s_2^2} = \frac{1.4}{0.96} = 1.4583$$



- Test Statistic: **1.4583**
  - p-value: **0.2092** (You would not have to do this)
4. State the sample, null hypothesis, test that was used, and conclusion with non-statistical terms
- $n_1 = 20$  and  $n_2 = 20$
  - Treatment 1 varies at the same rate as Treatment 2
  - A test for two independent sample standard deviation
  - Fail to Reject the claim that Treatment 1 varies at the same rate as Treatment 2

## 10.6 Confidence Intervals Revisited

Review of a Confidence Interval (CI)

- $PE$ : Point Estimate
- $ME$ : Margin of Error
- $CV$ : Critical Value, based  $1 - \alpha$  confidence level
- $SE$ : Standard Error
- $PE \pm ME = PE \pm CV(SE)$
- $LB = PE - ME = PE - CV(SE)$
- $UB = PE + ME = PE + CV(SE)$
- We are  $1 - \alpha$  % confident that the true parameter is between  $LB$  and  $UB$
- **NOTE:** Remember a CI is an estimate and gives the same result as an two-tailed hypothesis test.
- If 0 is within the  $LB$  and  $UB$  and you are testing if a parameter equal 0, you will always fail to reject the null hypothesis. *Remember this for next week.*

### 10.6.1 CI: Two Independent Samples Mean

**Example 5:**

Monterey and Santa Cruz Trees.

|            | $\bar{x}$ | $s$   | n  |
|------------|-----------|-------|----|
| Santa Cruz | 310.36    | 32.63 | 27 |
| Monterey   | 269.23    | 30.3  | 24 |

Create a 99 % CI.

Process:

1.  $PE = (\bar{x}_1 - \bar{x}_2) =$

2.  $SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} =$

3.  $1 - \alpha\% = \alpha =$

4.  $CV = t_{\alpha/2} =$  from t-table using  $\alpha/2, df =$

5.  $LB = PE - (CV)SE =$

6.  $UB = PE + (CV)SE =$

7. **Interpretation:** We are  $1 - \alpha\%$  confident that the difference between population means is within the  $LB =$  and  $UB =$  .

## 10.6.2 CI: Two Dependent (Paired) Samples Mean

### Example 6:

GPA Improvement Program

$\bar{x}_d = 0.3014$ ,  $s_d = 0.5065$  Create a 95 % CI.

Process:

1.  $PE = (\bar{x}_d) =$
2.  $SE = s_d/\sqrt{n} =$
3.  $1 - \alpha\% =$   $\alpha =$
4.  $CV = t_{\alpha/2} =$  from t-table using  $\alpha/2, df =$
5.  $LB = PE - (CV)SE =$
6.  $UB = PE + (CV)SE =$
7. **Interpretation:** We are  $1 - \alpha\%$  confident that the mean difference between population is within the  $LB =$  and  $UB =$  .