9.3 Hypothesis Testing: Sample Proportion

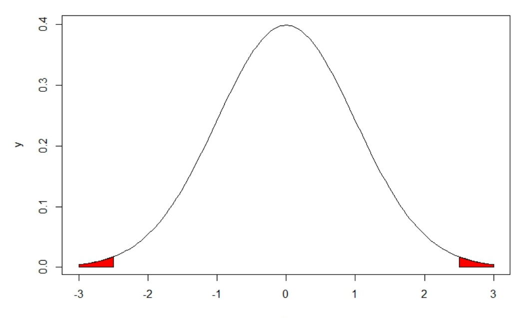
Example 2: Two Tailed

Do people believe that smoking should be allowed in Santa Cruz? In a county wide survey administered last year, 25 % of people in Santa Cruz do not believe that smoking should be permitted. A questionnaire was handed out to 60 random people in downtown Santa Cruz, in which 18 people said no. Did the sample vote differently than the Santa Cruz survey last? Use a significance level of 0.06.

Before we go through the statistical process, lets review the assumptions:

- Random Sample
- $np \geq 10$ and $nq \geq 10$

1. Identify and State the Statistical Question: **Does 25 % of people in this year's questionnaire believe that smoking should not be allowed in Santa Cruz?**
   - Determine the variable(s) of interest: **Response, Yes/No**
   - Determine the type variable(s) (i.e., quantitative or qualitative): **Qualitative, Nominal (Always use Z distribution)**
   - Identify and state the hypotheses (Null and Alternative Hypotheses): based on the question at hand: $\boldsymbol{H_0 : p = 0.25}$ and $\boldsymbol{H_1 : p \neq 0.25}$
   - Assumptions:
     - **Random Sample were asked to participate**
     - $\mathbf{60(0.25) = 15 \geq 10}$ and $\mathbf{60(0.75) = 45 \geq 10}$
2. Identify and state level of significance α (the probability of rejecting the H_0 when H_0 is true) $\alpha = 0.06$

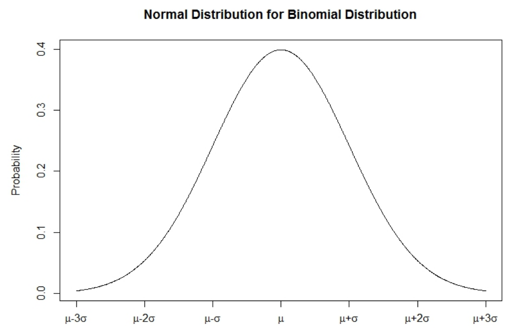


   Really IMPORTANT:
   - $\alpha =: 0.06$ (two-tails) $\implies$ $\alpha/2 = 0.03$
   - Critical Value: 1.88

   Note:
   - Point Estimate: $\hat{p} = 0.3$
   - Population Parameter: $p = 0.25$
3. Perform Statistical Test and Interpret Results

$$TS = z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{0.\hat{3} - 0.25}{\sqrt{\frac{(0.25)(0.75)}{60}}} = 0.8944$$

   Identify the CV, TS, p-value, α



- Test Statistic: 0.8944
- p-value: $1 - 0.8133 = 0.1867$
- Fail to reject H_0

4. State the sample size, null hypothesis, test that was used, and conclusion with non-statistical terms.
   - $n = 60$
   - 25 % of people in Santa Cruz do not believe in smoking in Santa Cruz
   - A test for single proportion
   - Fail to reject the claim that 25 % of people in Santa Cruz do not believe in smoking in Santa Cruz

10 Lecture 11 Notes: Hypothesis Testing for Two Samples

Never worry about what you're going lose but what you're going gain. Karim Kharbouch.

10.1 Some Theory Behind Two Sample Hypothesis Testing

Now that we understand the process of hypothesis testing, lets move to more complicated questions.

What if we want to determine if there is a difference between two populations?

We can implement the same process of hypothesis testing however various assumptions need to be made based on the data given and the test statistic will be modified. Some examples of two groups:

1. **Fires effect on redwood tree growth in Monterey**
2. Warriors points and Wizard points
3. **Smokers vs. Non-Smokers opinion on insurance**

Notice that within all of these examples, the two groups did not depend on each other. Another way of saying this is that the two groups are **independent**.

That is not always the case. The two groups may be dependent on each other meaning that there is a relationship between the elements in the first and second population. For example,

1. Patients in a study for new medication on blood pressure (before and after)
2. **Program to help students' GPA (before and after)**
3. Couples take a survey on relationship about love

Two Sample Hypothesis Testing: There are three different Scenarios that we will discussed in today's lecture.

1. Two Independent Sample Mean Test
2. Two Dependent Sample Mean Test
3. Two Independent Sample Proportion Test

The Two Independent Sample standard deviation will be discussed in a couple of weeks. However, the null and alternative hypotheses for these tests can be seen as:

A further understanding of these concepts test will be discussed in the following examples. Please pay attention.

	Two Independent Sample Mean	Two Dependent Sample Mean	Two Independent Sample Proportion
Null	$H_0 : \mu_1 = \mu_2$	$H_0 : \mu_d = 0$	$H_0 : p_1 = p_2$
Alternative (Left-Side)	$H_0 : \mu_1 < \mu_2$ **OR** $H_0 : \mu_1 - \mu_2 < 0$	$H_0 : \mu_d < 0$	$H_0 : p_1 < p_2$ **OR** $H_0 : p_1 - p_2 < 0$
Alternative (Right-Side)	$H_0 : \mu_1 > \mu_2$ **OR** $H_0 : \mu_1 - \mu_2 > 0$	$H_0 : \mu_d > 0$	$H_0 : p_1 > p_2$ **OR** $H_0 : p_1 - p_2 > 0$
Alternative (Two-Side)	$H_0 : \mu_1 \neq \mu_2$ **OR** $H_0 : \mu_1 - \mu_2 \neq 0$	$H_0 : \mu_d \neq 0$	$H_0 : p_1 \neq p_2$ **OR** $H_0 : p_1 - p_2 \neq 0$
Test Statistic	$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$	$\frac{\bar{x}_d}{s_d/\sqrt{n}}$	TBD

10.2 Hypothesis Testing: Two Independent Samples Mean

Example 1:

The fires last year had a dramatic effect on the growth redwood trees found in Santa Cruz as well as the redwood trees found in Monterey. A researcher wanted to see if there was a statistical difference between Monterey and Santa Cruz in terms of the growth of redwood trees after the fires. 27 trees were randomly sampled from Santa Cruz in which the mean was $\bar{x}_S = 310.36$ feet and standard deviation was $s_S = 30.3$ and 24 burned trees were randomly sampled from Monterey in which the mean was $\bar{x}_M = 269.23$ feet and standard deviation was $s_M = 32.63$. Did the burns cause the Monterey redwood trees to not grow as much as the Santa Cruz redwoods? Use a significance level of 0.01?

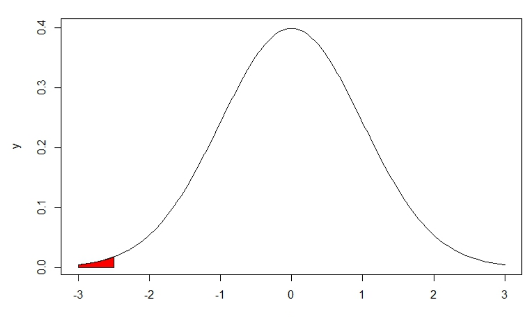
Data Summary from the problem organized:

Table 2: My caption

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

1. Identify and State the Statistical Question: **Did the burns cause the Monterey redwood trees to not grow as much as the Santa Cruz redwoods?**
   - Determine the variable(s) of interest: **Tree growth in feet and inches**
   - 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 $\boldsymbol{H_0 : \mu_M = \mu_S}$ and $\boldsymbol{H_0 : \mu_M < \mu_S}$

   I use subscripts M and S to make sense of the problem at hand.
2. Identify and state level of significance α (the probability of rejecting the H_0 when H_0 is true) $\boldsymbol{\alpha = 0.01}$



Really IMPORTANT:

- Degree of freedom: **Choose the smaller degrees of freedom of both samples**, the degrees of freedom for Santa Cruz is $n_S - 1 = 27 - 1 = 26$ and the degrees of freedom for Monterey is $n_M - 1 = 24 - 1 = 23$, therefore the degree of freedom for this problem is 23

- α: 0.01
- Critical Value: 2.5 (t-table)

3. Perform Statistical Test and Interpret Results

$$TS = \frac{\bar{x}_M - \bar{x}_S}{\sqrt{\frac{s_S^2}{n_S} + \frac{s_M^2}{n_M}}} = \frac{269.23 - 310.36}{\sqrt{\frac{30.3^2}{24} + \frac{32.63^2}{27}}} = -4.6664$$

Note: If it is believe that the two populations have the same standard deviation the standard error (the denominator) of this test statistic,

$$\sqrt{\frac{s_S^2}{n_S} + \frac{s_M^2}{n_M}}$$

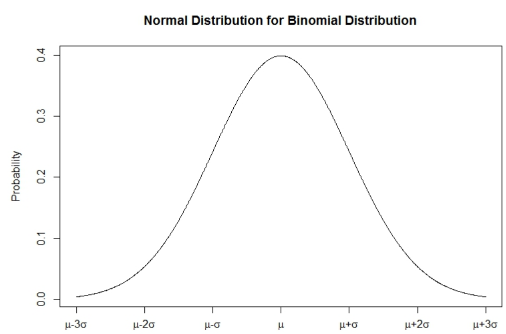
can be re-written as,

$$\sqrt{S_p^2(\frac{1}{n_S} + \frac{1}{n_M})}$$

where $df = n_M + n_S - 1$ and s_p can be calculated using the following equation

$$S_p^2 = \frac{(n_S - 1)s_S^2 + (n_M - 1)s_M^2}{n_M + n_S - 1}.$$

Identify the CV, TS, p-value, α



- Test Statistic: $|-4.6664| = 4.6664$
- p-value: < 0.0005

4. State the sample, null hypothesis, test that was used, and conclusion with non-statistical terms
   - $n_S = 27$ and $n_M = 24$
   - Monterey trees grew as high as Santa Cruz trees
   - A test for two independent sample means
   - Reject the claim that Monterey trees grew as high as Santa Cruz trees

What are some assumptions that this problem ignored?

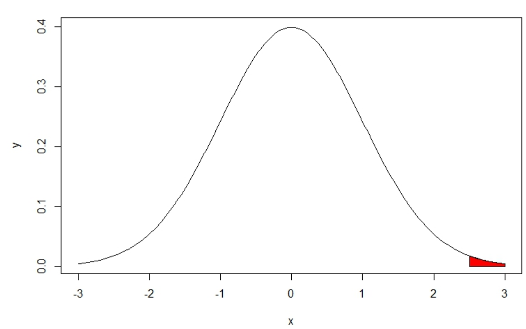
10.3 Hypothesis Testing: Two Dependent (Paired) Samples Mean

Example 2:

There is a new summer program to improve students' GPA and overall school performance at UCSC after probation students' first year. A study was conducted to determine the effectiveness of the new program. The evaluator had the GPA for 7 random students after their $3rd$ and $4th$ quarters (ignoring summer quarter). The evaluator does not use the cumulative GPA. Does the program have an impact (improvement) on these students? Use a significance level of 0.13. The data can be seen as:

								$\bar{x}$	s
Quarter 3	2.85	2.65	2.90	2.10	2.50	1.85	2.65	2.5000	0.3894
Quarter 4	3.75	2.17	3.12	2.34	2.65	2.85	2.73	2.8014	0.5233
Difference	0.9	-0.48	0.22	0.24	0.15	1.00	0.08	$\bar{x}_d$=0.3014	s_d=0.5065

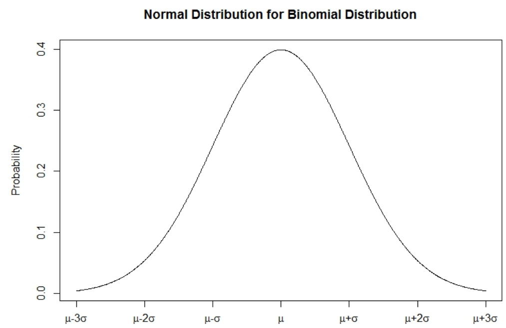
1. Identify and State the Statistical Question: **Does the program have an impact (improvement) on these students?**
   - Determine the variable(s) of interest: **GPA**
   - Determine the type variable(s) (i.e., quantitative or qualitative) : **quantitative**
   - Identify and state the hypotheses (Null and Alternative Hypotheses) based on the question at hand $\boldsymbol{H_0 : \mu_d = 0}$ and $\boldsymbol{H_0 : \mu_d > 0}$
2. Identify and state level of significance α (the probability of rejecting the H_0 when H_0 is true) $\boldsymbol{\alpha = 0.13}$



Really IMPORTANT:
   - Degree of Freedom: $n - 1 = 6$
   - α: 0.13
   - Critical Value: 1.440 (t-table)
3. Perform Statistical Test and Interpret Results

$$TS = \frac{\bar{x}_d}{s_d/\sqrt{n}} = \frac{0.3014}{0.5065/\sqrt{7}} = 1.5744$$

Identify the CV, TS, p-value, α



- Test Statistic: 1.5744
- p-value: $0.05 \leq$ p-value ≤ 0.1

4. State the sample, null hypothesis, test that was used, and conclusion with non-statistical terms
    - $n = 7$
    - The program has no impact on GPA
    - A test for two dependent (paired) sample means
    - Reject the claim that The program has no impact on GPA

What is wrong with a significance level of 0.13 in this example?