1. **You've sampled 15 units from the latest production lot to measure the weight of the parts. You calculate the sample mean to be 8.75 lbs, and the sample standard deviation to be 1.44 lbs. Calculate the 95% confidence interval for the population mean.**
   - **7.95 – 9.55**
   - 8.09 – 9.41
   - 7.31 – 10.19
   - 8.25 – 9.25

   Because we've only sampled 15 units and we only know the sample standard deviation (not the population standard deviation), we must use the t-distribution to create this confidence interval.

   With n = 15, we can calculate our degrees of freedom (n - 1) to be 14. Since this confidence interval is two-sided, we will split our alpha risk (5%) in half (2.5% or 0.025) to lookup the critical t-value of 0.975 (1 - α/2) at d.f. = 14 in the t-distribution table at 2.145.

$$Interval\ Estimate\ of\ Population\ Mean\ (unknown\ variance):\ \bar{x} \pm t_{\frac{\alpha}{2}} * \frac{s}{\sqrt{n}}$$

$$95\%\ Confidence\ Interval:\ 8.75 \pm 2.145 * \frac{1.44}{\sqrt{15}}$$

$$95\%\ Confidence\ Interval:\ 8.75 \pm 0.80$$

$$95\%\ Confidence\ Interval:\ \mathbf{7.95 - 9.55}$$

2. **You're creating a linear regression model for your data and you've calculated the values below. What is the Y-intercept for your regression model?** ($\bar{X} = 9.5,\ \bar{Y} = 180,\ \beta_1 = 4$)
   - 19
   - **142**
   - 45
   - 200

$$Y - Intercept:\ \beta_0 = \bar{Y} - \hat{\beta}_1\bar{X} = 180 - (4 * 9.5) = \mathbf{142}$$

3. **The one-way ANOVA Analysis below has 19 treatment groups with the total degrees of freedom of 25. Complete this ANOVA table and calculate the F-value.**
   - 3.66
   - 1.95
   - **2.40**
   - 0.42

Variation Source	Sum of Squares (SS)	Degrees of freedom (DF)	Mean Squares (MS)	F-Value
Treatment (Between)	370			
Error (Within)				
Total	430			

First, we can solve for the error sum of squares by simply subtracting 370 from 430, to get an error sum of Square of 60.

Then we must solve for the degrees of freedom, which is equal to the number of treatment levels (19) minus 1; so, 18 degrees of freedom for treatment. Then we can solve for the error degrees of freedom by subtracting 25 - 18; so, 7 degrees of freedom for error.

Then we can calculate the mean squares as the sum of squares, divided by the degrees of freedom.

Variation Source	Sum of Squares (SS)	Degrees of freedom (DF)	Mean Squares (MS)	F-Value
Treatment (Between)	370	= 19 - 1 = 18	= 370 / 18 = 20.56	= 20.56 / 8.57 **= 2.40**
Error (Within)	= 430 - 370 = 60	= 25 - 18 = 7	= 60 / 7 = 8.57	
Total	430	25		

4. Your manufacturing process has a historical yield loss average of 8%. You made a change to the process and you'd like to determine if this change has impacted the yield loss. You sample 70 units and find 9 units are non-conforming.

Based on this sample, can we conclude that the proportion of defects in this lot is different than the historical average using a significance level of 5%? Identify the statement below that is true.

- We must reject the null hypothesis because the z-statistics is greater than the critical z-score
- We must accept the null hypothesis because the z-statistics is greater than the critical z-score
- We must reject the alternative hypothesis because the z-statistics is greater than the critical z-score
- **We must fail to reject the null hypothesis because the z-statistics does not fall into the rejection region**

Based on the wording of the problem statement, this is a two-tail hypothesis test as we're attempting to determine if the proportion has changed from the historical average of 8%.

H_0: p = 8% (0.08) & H_a: p ≠ 8% (0.08)

Based on our significance level of 5%, and the two-tail test, we can look up our critical Z-score that creates our rejection criteria of Z_{crit} = -1.96 and Z_{crit} = 1.96.

Based on the problem statement we know that the hypothesized population proportion (P_0) is 0.08, and the sample size is 70:

$$p_O = 8\% \ (0.08) \ and \ n = 70$$

We can also calculate the sample proportion, p-hat:

$$\hat{p} = \frac{9}{70} = 13\% \ (0.13)$$

Next, we plug these into our z-transformation to calculate our test statistic:

$$Z_O = \frac{\hat{p} - p_O}{\sqrt{\frac{p_O(1 - p_O)}{n}}} = \frac{0.13 - 0.08}{\sqrt{\frac{0.08 * (1 - 0.08)}{70}}} = 1.54$$

Because our test statistic (Z_0 = 1.54) does NOT fall into the rejection region (*the rejection region is greater than Z_{crit} = 1.96 or less than Z_{crit} = -1.96*), we must fail to reject the null hypothesis.

Note: We never accept the null hypothesis. We should either reject the null hypothesis, or fail to reject the null hypothesis.

5. **You’re attempting to characterize the variation of your product over time, so you’ve setup a data collection plan to measure product across multiple months. How is this type of variation described from the multi-variate analysis perspective?**
   - Cyclical
   - **Temporal**
   - Positional
   - Common

6. **You've sampled 15 units from the latest production lot to measure the weight of the parts. You calculate the sample mean to be 4.65 lbs, and the sample standard deviation to be 0.55 lbs. Calculate the 95% confidence interval for the population standard deviation.**
   - 0.417 - 0.898
   - 0.162 - 0.752
   - 0.382 - 0.718
   - **0.403 - 0.867**

Ok, let's see what we know after reading the problem statement:

n = 15, s = 0.55 lbs, α = 0.05, x-bar = 4.65 lbs

First, we must find our critical chi-squared values with the Chi-Squared Table associated with our alpha risk (5%), sample size (15), and degrees of freedom (14):

$$Confidence\ Interval\ for\ Standard\ Deviation:\ \sqrt{\frac{(n-1)s^2}{X^2_{1-\alpha/2}}} < \sigma < \sqrt{\frac{(n-1)s^2}{X^2_{\alpha/2}}}$$

$$X^2_{\alpha/2} = X^2_{.05/2} = X^2_{.025}$$

$$X^2_{1-\alpha/2} = X^2_{1-.05/2} = X^2_{1-.025} = X^2_{.975}$$

$$X^2_{.025,14} = 5.629 \quad \& \quad X^2_{.975,14} = 26.119$$

Now we can complete the equation using these chi-squared values along with the sample size, and sample standard deviation to calculate our interval.

$$Confidence\ Interval\ for\ Standard\ Deviation:\ \sqrt{\frac{(n-1)s^2}{X^2_{1-\alpha/2}}} < \sigma < \sqrt{\frac{(n-1)s^2}{X^2_{\alpha/2}}}$$

$$\sqrt{\frac{(15-1)*0.55^2}{26.119}} < \sigma < \sqrt{\frac{(15-1)*0.55^2}{5.629}}$$

$$\mathbf{0.403 < \sigma < 0.867}$$

7. **Identify the statement below regarding hypothesis testing that is true:**
   - **Power can be thought of as the probability of avoiding a Type II error. (True)**
   - The probability of occurrence of a Type I error is defined as the ~~Beta (β)~~ (alpha (α)) risk. (False)
   - The alpha risk in hypothesis testing is analogous to the ~~consumers~~ (producers) risk in the world of acceptance sampling. (False)
   - Failing to reject the null hypothesis is analogous to **proving that the null hypothesis is true** – (False) *this is an incorrect interpretation of a rejection of the null hypothesis.*

8. **You're creating a linear regression model for your data and you've calculated the following least squares estimates (S_{yy} = 41, S_{xy} = 19, S_{xx} = 20). Based on these results, what percentage of variation in Y, can be explained by the variation in X?**
   - 15%
   - 66%
   - 2%
   - **44%**

   Recall that the *Coefficient of Determination (R^2) reflects the percentage of variation in Y that can be explained by the variation in X*.

   Below is the equation to solve for the Pearson Correlation Coefficient (r_{xy}), which then can be squared to find the coefficient of determination (R^2).

$$r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx}} * \sqrt{S_{yy}}} = \frac{19}{\sqrt{20} * \sqrt{41}} = 0.66$$

$$R^2 = r_{xy}{}^2 = 0.66^2 = 0.44 = \mathbf{44}\%$$

9. **You're performing a hypothesis test for the population mean, which you believe to be 7.50". You sample 60 parts and find the sample mean to be 7.35". The population standard deviation is known to be 0.80". What is the test statistic for this hypothesis test?**
   - 1.45
   - **-1.45**
   - -1.82
   - 1.82

   Because we've taken more than 30 samples, and we know the population standard deviation, we can use the Z-score to calculate our test statistic.

$$Z - Score = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{7.35 - 7.50}{\frac{0.80}{\sqrt{60}}} = \mathbf{-1.45}$$

10. Which distribution is used to make the accept/reject decision for the ANOVA Analysis:

- Chi-squared Distribution
- T Distribution
- Normal Distribution
- **F Distribution**

11. You're creating a linear regression model for your data and you've calculated the following values. What is the predicted value of Y when X = 12? (S_{yy} = 192, S_{xy} = 32, S_{xx} = 96, β_0 = 9)

- 111.0
- 45.0
- **13.0**
- 108.3

$$\beta_1 = \frac{S_{xy}}{S_{xx}} = \frac{32}{96} = 0.33$$

$$Y(x) = \beta_0 + \beta_1 * x$$

$$Y(12) = 9 + 0.33 * 12 = \mathbf{13.0}$$

12. You've sampled 25 units from the last production lot and found that 4 of them are non-conforming. Find the 95% confidence interval for the true population proportion of defective products.

- 0.09 < p < 0.23
- **0.02 < p < 0.30**
- 0.04 < p < 0.28
- 0 < p < 0.35

First, we can calculate the sample proportion p using n = 25, and the number of non-conformances (4):

$$Sample\ Proportion:\ p = \frac{4}{25} = 0.16$$

Then we can look up our Z-score at the 5% alpha risk: $Z_{\frac{\alpha}{2}} = Z_{\frac{0.05}{2}} = Z_{.025} = 1.96$

$$Confidence\ Interval\ (Proportion):\ p \pm Z_{\frac{\alpha}{2}}\sqrt{\frac{p*(1-p)}{n}}$$

$$Confidence\ Interval:\ 0.16 \pm 1.96 * \sqrt{\frac{0.16*(1-0.16)}{25}}$$

$$Confidence\ Interval:\ 0.16 \pm 1.96 * \sqrt{0.0054} = 0.16 \pm 0.14$$

$$Confidence\ Interval\ for\ Population\ Proportion:\ \mathbf{0.02 < p < 0.30}$$

13. Match the following terms with their appropriate location on this table of Null & Alternative Hypothesis:

- D - Correct Decision to Reject the Null Hypothesis
- B - Type II Error (Beta Risk)
- C - Type I Error (Alpha Risk)
- A - Correct Decision to Fail to Reject the Null Hypothesis

		The Truth	
		H_0 is True	H_0 is False
The Outcome of the Hypothesis Test	Fail to Reject H_0	Correct Decision	INCORRECT DECISION (Type II Error) Beta (β) Risk
	Reject H_0	INCORRECT DECISION (Type I Error) Alpha (α) risk	Correct Decision *Power (1 - β)*

14. Your manufacturing process has a particular step that is performed by two different machines that you believe to be identical. Prior to testing the mean values associated with your two processes, you want to test your assumption of the homogeneity of variances. You take a sample of 10 units from machine A and 9 samples from Machine B and measure the sample variance of each machine to be 0.17 and 0.25 respectively.

You want to use the significance level of 10% to test the hypothesis that the variances from each machine are equal. Identify the statement below that is true.

- We must reject the null hypothesis because the test statistics is greater than the critical score.
- We must reject the alternative hypothesis because the test statistics is greater than the critical score.
- We must fail to reject the null hypothesis because the test statistics is greater than the critical score.
- **We must fail to reject the null hypothesis because the test statistics does not fall into the rejection region.**

(The detailed answer is on the next page)

Since we're attempting to determine if these variances are different, here's what the null and alternative hypothesis look like. $H_0: \sigma_a^2 = \sigma_b^2$ & $H_a: \sigma_a^2 \neq \sigma_b^2$

With machine B having a greater variance than machine A, we will consider v_B to be v_1, and v_A to be v_2.

$v_1 = v_B = 9 - 1 = 8$ and $v_2 = v_A = 10 - 1 = 9$

We can look up the critical F-Values using the F-Table at the 10% alpha risk that's split between the upper and lower tail.

$$Upper\ Critical\ Value\ (Right\ Tail) = F_{\frac{\alpha}{2},(v_1,v_2)} = F_{0.05,(8,9)} = 3.230$$

$$Lower\ Critical\ Value\ (Left\ Tail) = F_{1-\frac{\alpha}{2},(v_1,v_2)} = \frac{1}{F_{\frac{\alpha}{2},(v_2,v_1)}} = \frac{1}{F_{0.05,(9,8)}} = \frac{1}{3.388} = 0.295$$

Now we can calculate the Test Statistic from our sample data:

$$F - Test\ Statistic: F = \frac{s_1^2}{s_2^2} = \frac{0.25}{0.17} = 1.47$$

Since our F-Statistic (1.47) is not greater than the upper critical value (3.230), nor is it less than the lower critical value (0.295), **we must fail to reject the null hypothesis.**

15. You're creating a linear regression model for your data and you've calculated the following least squares estimates (S_{yy} = 5,300, S_{xy} = -2,500, S_{xx} = 3,100). What is the correlation coefficient for this data set?

- 0.38
- **-0.62**
- -0.47
- 0.62

$$r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx}} * \sqrt{S_{yy}}} = \frac{-2{,}500}{\sqrt{3{,}100} * \sqrt{5{,}300}} = \mathbf{-0.62}$$

Start Time: __________ Number Correct: ___________

Stop Time: __________

Total Time: __________

Available Time: 35 Minutes Target Time: 25 Minutes

Question #	Chapter	Topic
1	17	Hypothesis Testing
2	16	Exploratory Data Analysis
3	17	Hypothesis Testing
4	17	Hypothesis Testing
5	16	Exploratory Data Analysis
6	17	Hypothesis Testing
7	17	Hypothesis Testing
8	16	Exploratory Data Analysis
9	17	Hypothesis Testing
10	17	Hypothesis Testing
11	16	Exploratory Data Analysis
12	17	Hypothesis Testing
13	17	Hypothesis Testing
14	17	Hypothesis Testing
15	16	Exploratory Data Analysis

<table>
<tr><th>Chapter</th><th>Title</th><th colspan="5">Questions</th></tr>
<tr><td>16</td><td>Exploratory Data Analysis</td><td>2</td><td>5</td><td>8</td><td>11</td><td>15</td></tr>
<tr><td rowspan="2">17</td><td rowspan="2">Hypothesis Testing</td><td>1</td><td>3</td><td>4</td><td>6</td><td>7</td></tr>
<tr><td>9</td><td>10</td><td>12</td><td>13</td><td>14</td></tr>
</table>