Module 15 Challenge

Start Assignment

Due Sunday by 11:59pm   **Points** 100   **Submitting** a text entry box or a website url

Background

A few weeks after starting his new role, Jeremy is approached by upper management about a special project. AutosRUs' newest prototype, the MechaCar, is suffering from production troubles that are blocking the manufacturing team's progress. AutosRUs' upper management has called on Jeremy and the data analytics team to review the production data for insights that may help the manufacturing team.

In this challenge, you'll help Jeremy and the data analytics team do the following:

- Perform multiple linear regression analysis to identify which variables in the dataset predict the mpg of MechaCar prototypes

- Collect summary statistics on the pounds per square inch (PSI) of the suspension coils from the manufacturing lots

- Run t-tests to determine if the manufacturing lots are statistically different from the mean population

- Design a statistical study to compare vehicle performance of the MechaCar vehicles against vehicles from other manufacturers. For each statistical analysis, you'll write a summary interpretation of the findings.

What You're Creating

This new assignment consists of three technical analysis deliverables and a proposal for further statistical study. You'll submit the following:

- Deliverable 1: Linear Regression to Predict MPG

- Deliverable 2: Summary Statistics on Suspension Coils

- Deliverable 3: T-Test on Suspension Coils

- Deliverable 4: Design a Study Comparing the MechaCar to the Competition

Files

Use the following links to download the Challenge data sets.

Download the **MechaCar MPG dataset**   **(https://2u-data-curriculum-team.s3.amazonaws.com/dataviz-online/module_15/MechaCar_mpg.csv)** .

Download the **Suspension Coil dataset**   **(https://2u-data-curriculum-team.s3.amazonaws.com/dataviz-online/module_15/Suspension_Coil.csv)** .

Before You Start

Create a new GitHub repository, "MechaCar_Statistical_Analysis," and initialize the repository with a README.

After you've completed the technical analysis for each deliverable, provide a short summary of the results in the README.md of the analysis. For the final deliverable, you'll write up a short description of the study design for additional statistical analysis. In the written summaries, we would like you to think critically about your analysis, not demonstrate proficiency of automotive manufacturing.

Deliverable 1: Linear Regression to Predict MPG (30 points)

Deliverable 1 Instructions

The `MechaCar_mpg.csv` dataset contains mpg test results for 50 prototype MechaCars. The MechaCar prototypes were produced using multiple design specifications to identify ideal vehicle performance. Multiple metrics, such as vehicle length, vehicle weight, spoiler angle, drivetrain, and ground clearance, were collected for each vehicle. Using your knowledge of R, you'll design a linear model that predicts the mpg of

MechaCar prototypes using several variables from the `MechaCar_mpg.csv` file. Then, you'll write a short interpretation of the multiple linear regression results in the README.md.

REWIND

For this deliverable, you've already done the following in this module:

- **Lesson 15.2.3:** Import and Read a CSV File

- **Lesson 15.7.2:** Use the Linear Regression Function

- **Lesson 15.7.3:** Run a Multiple Linear Regression RScript

- **Lesson 15.7.3:** Use the Summary Function for Multiple Linear Regression

---

Follow the instructions below to complete Deliverable 1.

Technical Analysis

1. Download the `MechaCar_mpg.csv` file, and place it in the active directory for your R session.

2. Create a new RScript in your R source pane, name it `MechaCarChallenge.RScript`, and save it to your active directory.

NOTE

> Create a new RScript by going to the File menu. Select "New File," followed by "RScript," or you can click the icon in the top-left corner of the RStudio window. Note that the icon looks like a white square with a plus sign in the top left corner.

---

3. Use the `library()` function to load the `dplyr` package.

4. Import and read in the `MechaCar_mpg.csv` file as a dataframe.

5. Perform linear regression using the `lm()` function. In the `lm()` function, pass in all six variables (i.e., columns), and add the dataframe you created in Step 4 as the data parameter.

6. Using the `summary()` function, determine the p-value and the r-squared value for the linear regression model.

7. Save your `MechaCarChallenge.RScript` file to your GitHub repository.

Written Summary

In your README, create a subheading, `## Linear Regression to Predict MPG`, and write a short summary using a screenshot of the output from the linear regression, and address the following questions:

- Which variables/coefficients provided a non-random amount of variance to the mpg values in the dataset?

- Is the slope of the linear model considered to be zero? Why or why not?

- Does this linear model predict mpg of MechaCar prototypes effectively? Why or why not?

Deliverable 1 Requirements

You will earn a perfect score for Deliverable 1 by completing all requirements below:

- The `MechaCar_mpg.csv` file is imported and read into a dataframe **(5 pt)**

- An RScript is written for a linear regression model to be performed on all six variables **(10 pt)**

- An RScript is written to create the statistical summary of the linear regression model with the intended p-values **(10 pt)**

- There is a summary that addresses all three questions **(5 pt)**

Deliverable 2: Create Visualizations for the Trip Analysis (30 points)

Deliverable 2 Instructions

The MechaCar `Suspension_Coil.csv` dataset contains the results from multiple production lots. In this dataset, the weight capacities of multiple suspension coils were tested to determine if the manufacturing process is consistent across production lots. Using your knowledge of R, you'll create a summary statistics table to show:

- The suspension coil's PSI continuous variable across all manufacturing lots
- The following PSI metrics for each lot: mean, median, variance, and standard deviation.

Then, in the README.md, you'll briefly detail and interpret the suspension coil summary statistics.

REWIND

For this deliverable, you've already done the following in this module:

- **Lesson 15.2.3:** Import and Read a CSV File
- **Lesson 15.2.5:** Use the `group_by()` Function
- **Lesson 15.2.5:** Use the `summarize()` Function

Follow the instructions below to complete Deliverable 2.

Technical Analysis

1. Download the `Suspension_Coil.csv` file, and place it in the active directory for your R session.
2. In your `MechaCarChallenge.RScript`, import and read in the `Suspension_Coil.csv` file as a table.
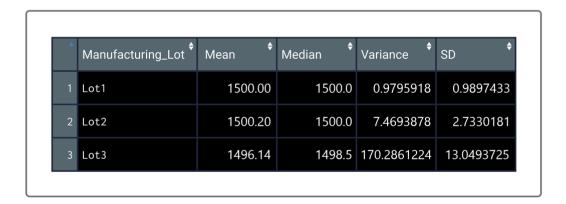
3. Write an RScript that creates a `total_summary` dataframe using the `summarize()` function to get the mean, median, variance, and standard deviation of the suspension coil's PSI column.

Your `total_summary` dataframe should look like this:

	Mean	Median	Variance	SD
1	1498.78	1500	62.29356	7.892627

4. Write an RScript that creates a `lot_summary` dataframe using the `group_by()` and the `summarize()` functions to group each manufacturing lot by the mean, median, variance, and standard deviation of the suspension coil's PSI column.

Your `lot_summary` dataframe should look like this:

	Manufacturing_Lot	Mean	Median	Variance	SD
1	Lot1	1500.00	1500.0	0.9795918	0.9897433
2	Lot2	1500.20	1500.0	7.4693878	2.7330181
3	Lot3	1496.14	1498.5	170.2861224	13.0493725

5. Save your `MechaCarChallenge.RScript` file to your GitHub repository.

Written Summary

In your README, create a subheading `## Summary Statistics on Suspension Coils`, and write a short summary using screenshots from your `total_summary` and `lot_summary` dataframes, and address the following question:

- The design specifications for the MechaCar suspension coils dictate that the variance of the suspension coils must not exceed 100 pounds per square inch. Does the current manufacturing data meet this design specification for all manufacturing lots in total and each lot individually? Why or why not?

Deliverable 2 Requirements

You will earn a perfect score for Deliverable 2 by completing all requirements below:

- The `Suspension_Coil.csv` file is imported and read into a dataframe **(5 pt)**

- An RScript is written to create a total summary dataframe that has the mean, median, variance, and standard deviation of the PSI for all manufacturing lots **(10 pt)**

- An RScript is written to create a lot summary dataframe that has the mean, median, variance, and standard deviation for each manufacturing lot **(10 pt)**

- There is a summary that addresses the design specification requirement for all the manufacturing lots and each lot individually **(5 pt)**

Deliverable 3: T-Tests on Suspension Coils (20 points)

Deliverable 3 Instructions

Using your knowledge of R, perform t-tests to determine if all manufacturing lots and each lot individually are statistically different from the population mean of 1,500 pounds per square inch.

REWIND

For this deliverable, you've already done the following in this module:

- **Lesson 15.4.1:** The Different Statistical Test Types

- **Lesson 15.6.2:** Perform a Sample T-Test

Follow the instructions below to complete Deliverable 3.

Technical Analysis

1. In your `MechaCarChallenge.RScript`, write an RScript using the `t.test()` function to determine if the PSI across all manufacturing lots is statistically different from the population mean of 1,500 pounds per square inch.

2. Next, write three more RScripts in your `MechaCarChallenge.RScript` using the `t.test()` function and its `subset()` argument to determine if the PSI for each manufacturing lot is statistically different from the population mean of 1,500 pounds per square inch.

SHOW HINT

3. Save your `MechaCarChallenge.RScript` file to your GitHub repository.

Written Summary

In your README, create a subheading `## T-Tests on Suspension Coils`, then briefly summarize your interpretation and findings for the t-test results. Include screenshots of the t-test to support your summary.

Deliverable 3 Requirements

You will earn a perfect score for Deliverable 3 by completing all requirements below:

- An RScript is written for t-test that compares all manufacturing lots against mean PSI of the population **(5 pt)**

- An RScript is written for three t-tests that compare each manufacturing lot against mean PSI of the population **(10 pt)**

- There is a summary of the t-test results across all manufacturing lots and for each lot **(5 pt)**

Deliverable 4: Design a Study Comparing the MechaCar to the Competition (20 points)

Deliverable 4 Instructions

Using your knowledge of R, design a statistical study to compare performance of the MechaCar vehicles against performance of vehicles from other manufacturers.

Follow the instructions below to complete Deliverable 4.

1. In your README, create a subheading `## Study Design: MechaCar vs Competition`.

2. Write a short description of a statistical study that can quantify how the MechaCar performs against the competition. In your study design, think critically about what metrics would be of interest to a consumer: for a few examples, cost, city or highway fuel efficiency, horse power, maintenance cost, or safety rating.

3. In your description, address the following questions:

- What metric or metrics are you going to test?

- What is the null hypothesis or alternative hypothesis?

- What statistical test would you use to test the hypothesis? And why?

- What data is needed to run the statistical test?

SHOW HINT

Follow the instructions below to complete Deliverable 4.

Deliverable 4 Requirements

You will earn a perfect score for Deliverable 4 by completing all requirements below:

- The statistical study design has the following:

    - A metric to be tested is mentioned **(5 pt)**

    - A null hypothesis or an alternative hypothesis is described **(5 pt)**

    - A statistical test is described to test the hypothesis **(5 pt)**

    - The data for the statistical test is described **(5 pt)**

Submission

Once you're ready to submit, make sure to check your work against the rubric to ensure you are meeting the requirements for this Challenge one final time. It's easy to overlook items when you're in the zone!

As a reminder, the deliverables for this Challenge are as follows:

- Deliverable 1: Linear Regression to Predict MPG

- Deliverable 2: Summary Statistics on Suspension Coils

- Deliverable 3: T-Test on Suspension Coils

- Deliverable 4: Design a Study Comparing the MechaCar to the Competition

Upload the following to your MechaCar_Statistical_Analysis GitHub repository:

- The `MechaCarChallenge.RScript` file.

- An updated README.md that has the written summaries for Deliverables 1, 2, and 3 and your design for a statistical study comparing vehicle performance of the MechaCars against other manufacturers' vehicles.

To submit your challenge assignment for grading in Bootcamp Spot, click Start Assignment, click the Website URL tab, then provide the URL of your MechaCar_Statistical_Analysis GitHub repository, and then click Submit. Comments are disabled for graded submissions in BootCampSpot. If you have questions about your feedback, please notify your instructional staff or the Student Success Manager. If you would like to resubmit your work for an improved grade, you can use the **Re-Submit Assignment** button to upload new links. You may resubmit up to 3 times for a total of 4 submissions.

IMPORTANT

Once you receive feedback on your Challenge, make any suggested updates or adjustments to your work. Then, add this week's Challenge to your professional portfolio.

NOTE

You are allowed to miss up to two Challenge assignments and still earn your certificate. If you complete all Challenge assignments, your lowest two grades will be dropped. If you wish to skip this assignment, click Next, and move on to the next Module.

Module-15 Rubric

Criteria	Ratings					Pts
Deliverable 1: Linear Regression to Predict MPG	**30 to >27.0 pts** **Demonstrating Proficiency** ✓The csv file is imported and read into a dataframe. ✓An RScript is written for a linear regression model to be performed on ALL SIX variables. ✓An RScript is written to create the statistical summary of the linear regression model with the intended p-values.	**27 to >24.0 pts** **Approaching Proficiency** ✓The csv file is imported and read into a dataframe. ✓An RScript is written for a linear regression model to be performed on ALL SIX variables. ✓An RScript is written to create the statistical summary, but the p-values are higher for some variables. ✓The	**24 to >22.0 pts** **Developing Proficiency** ✓The csv file is imported and read into a dataframe. ✓An RScript is written for a linear regression model to be performed on ALL SIX variables. ✓An RScript is written to create the statistical summary, but there is no overall statistical significance. ✓The	**22 to >0.0 pts** **Emerging** ✓The csv file is imported and read into a dataframe. ✓An RScript is written for a linear regression model to be performed on ALL SIX variables. ✓An RScript is written for the statistical summary, but there is an error and no output. ✓The	**0 pts** **Incomplete**	30 pts
Deliverable 2: Summary Statistics on Suspension Coils	**30 to >27.0 pts** **Demonstrating Proficiency** ✓The csv file is imported and read into a dataframe. ✓The total summary dataframe has ALL FOUR metrics for all the manufacturing lots. ✓The lot summary dataframe has ALL FOUR metrics for each manufacturing lot. ✓The summary addresses the design	**27 to >24.0 pts** **Approaching Proficiency** ✓The csv file is imported and read into a dataframe. ✓The total summary dataframe has ALL FOUR metrics for all the manufacturing lots. ✓The lot summary dataframe has THREE of the FOUR metrics of each manufacturing lot. ✓The summary addresses the design specification	**24 to >22.0 pts** **Developing Proficiency** ✓The csv file is imported and read into a dataframe. ✓The total summary dataframe has ALL FOUR metrics for all the manufacturing lots. ✓The lot summary dataframe has TWO of the FOUR metrics for each manufacturing lot. ✓The summary addresses the design specification	**22 to >0.0 pts** **Emerging** ✓The csv file is imported and read into a dataframe. ✓The total summary dataframe has ALL FOUR metrics for all the manufacturing lots. ✓The lot summary dataframe has ONE of the FOUR metrics for each manufacturing lot. ✓The summary addresses the design specification	**0 pts** **Incomplete**	30 pts
Deliverable 3: T-Test on Suspension Coils	**20 to >17.0 pts** **Demonstrating Proficiency** ✓An RScript is written for a t-test that compares all manufacturing lots against the mean PSI of the population. ✓An RScript is written for ALL THREE t-tests that compare each manufacturing lot against the mean PSI of the population. ✓The summary addresses the results across all manufacturing lots and ALL THREE lots.	**17 to >14.0 pts** **Approaching Proficiency** ✓An RScript is written for a t-test that compares all manufacturing lots against the mean PSI of the population. ✓An RScript is written for TWO of THREE t-tests that compare each manufacturing lot against the mean PSI of the population. ✓The summary addresses the results across all manufacturing lots and TWO of THREE lots.	**14 to >12.0 pts** **Developing Proficiency** ✓An RScript is written for a t-test that compares all manufacturing lots against the mean PSI of the population. ✓An RScript is written for ONE of THREE t-tests that compare each manufacturing lot against the mean PSI of the population. ✓The summary addresses the results across all manufacturing lots and ONE of THREE lots.	**12 to >0.0 pts** **Emerging** ✓An RScript is written for a t-test that compares all manufacturing lots against the mean PSI of the population. ✓An RScript is written for ONE of THREE t-tests that compare each manufacturing lot against the mean PSI of the population, but there is an error. ✓The summary addresses the results across all manufacturing lots OR ONE of THREE lots.	**0 pts** **Incomplete**	20 pts

Criteria	Ratings					Pts
Deliverable 4: Design a Study Comparing the MechaCar to the Competition	**20 to >18.0 pts Demonstrating Proficiency** The statistical study design has the following: ✓A metric to be tested is mentioned. ✓A null or alternative hypothesis is described. ✓A statistical test is well described to test the hypothesis. ✓The data for the statistical test is well described.	**18 to >15.0 pts Approaching Proficiency** The statistical study design has the following: ✓A metric to be tested is mentioned. ✓A null or alternative hypothesis is described. ✓The statistical test to test the hypothesis is not fully described. ✓The data for the statistical test is not fully described.	**15 to >13.0 pts Developing Proficiency** The statistical study design has the following: ✓A metric to be tested is mentioned. ✓A null or alternative hypothesis is not well described. ✓The statistical test to test the hypothesis is not well described. ✓The data for the statistical test is not well described.	**13 to >0.0 pts Emerging** The statistical study design has the following: ✓A metric to be tested is mentioned. ✓A null or alternative hypothesis is not well described. ✓The statistical test to test the hypothesis is barely mentioned. ✓The data for the statistical test is not well described.	**0 pts Incomplete**	20 pts

Total Points: 100