

# Homework 1

---

Due: Saturday, Sep 12, 2020 at 11:59 pm

Data: **Cars.csv** for Exercise 1-2 and **airquality** (R built-in) for Exercise 3.

Use the significance level of .05

**Cars.csv** will be used for Exercise 1 and 2. The variables in the data are included below in the table. The variables in the data set are the following attributes of cars in the year 2004:

- **Make** – the auto manufacturer
- **Model** – name of the vehicle
- **Type** – SUV, sedan, sports, truck, or wagon
- **Origin** – continent of the manufacturer; Europe, Asia, or USA
- **Invoice** – price (dollars) that the manufacturer sends to the dealer upon delivery of the car
- **Horsepower** – amount of the car's power
- **MPG\_City** – miles per gallon (fuel efficiency) during city driving
- **MPG\_Highway** – miles per gallon during highway driving
- **Wheelbase** – distance (inches) between the centers of the front and rear wheels
- **Length** – distance (inches) from the nose to the tail of the car

## Exercise 1: Descriptive Statistics (20 points)

- Create a combined mpg variable called **MPG\_Combo** which combines 55% of the **MPG\_City** and 45% of the **MPG\_Highway**. Obtain a box plot for **MPG\_Combo** and comment on what the plot tells us about fuel efficiencies.
- Obtain box plots for **MPG\_Combo** by **Type** and comment on any differences you notice between the different vehicle types combined fuel efficiency.
- Obtain basic descriptive statistics for **Invoice** for all vehicles. Comment on any general features and statistics of the data. Use visual and quantitative methods to comment on whether an assumption of Normality would be reasonable for **Invoice** variable.
- Use visual and quantitative methods to comment on whether an assumption of normality would be reasonable for **Invoice** variable by **Origin**. (i.e., check normality of **Invoice** from i) Europe, ii) Asian, and iii) USA cars.

## Exercise 2: Hypothesis Testing (15 points)

Perform a hypothesis test of whether cars originated in Europe have different invoice price than Asian cars, and state your conclusions

- Which test should we perform, and why? Justify your answer based on findings on Exercise 1 (d).
- Specify null and alternative hypotheses.
- State the conclusion based on the test result.

The **airquality** data will be used for Exercise 3. The information about variables in the dataset can be found in R.

### Exercise 3: Hypothesis Testing (15 points)

Perform a hypothesis test -whether Wind in July has a different speed (mph) than Wind in August.

- a) Which test should we perform, and why? See QQ-plot and perform Shapiro-Wilk test for normality check.
- b) Specify null and alternative hypotheses
- c) State the conclusion based on the test result.