# Homework 1

**Due: Thursday February 1, 2018 at 5:00pm**

See general homework tips and submit your files via the course website.

To complete this assignment, you will need to access the cars data set in **Program_HW1_Data_Spring2018.sas** on Compass. The subset of the sashelp.cars data set, which is a part of the SAShelp library, will be used for all problems. The variables in the data are included below in the table. Follow the coding below to access the data set.

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
- **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
- **Invoice** – distance (inches) from the nose to the tail of the car

## Exercise 1

a) 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.

b) Obtain box plots for **MPG_Combo** by **Type** and comment on any differences you notice between the different vehicle types combined fuel efficiency.

c) Obtain basic descriptive statistics for the **MPG_Combo** and **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 this **MPG_Combo** variable.

d) Repeat the analysis in part (c) by **Type**. Comment on these basic statistics and comment on the assumption of normality of the population for each of the vehicle types. Be sure to use quantitative and visual methods.

## Exercise 2

a) Test the null hypothesis that the true mean or median **Invoice** price is $22000 against the alternative that it is not. Based on the normality tests from the above Problem, which location test should we use and what do we conclude from that test?

b) Perform a hypothesis test of whether cars originated in Europe have significantly greater invoice price than Asian cars, and state your conclusions. Recall the test of normality from the above Problem when choosing your test.

# Exercise 3

a) Obtain the Pearson correlation matrix for the numeric variables in the data set except **MPG_City** and **MPG_Highway**. State what this tells us about the relationships between the vehicle attributes, and what we might infer about the population of the data.

b) Perform the same correlation analysis by **Type**. Comment on how the relationships between vehicle attributes differ between the various vehicle types and note any differences with what you found in part (a).