Lan Lin lql5083

Problem (a)

1. How many observations are there in this data set?

**A: There are 428 observations.**

(ii) How many numeric variables and character variables are there in the data set respectively?

**A: 10 numeric variables and 5 character variable are there.**

Problem (d)

(i) What are the mean values of RetailPrice and Weight of Asian vehicles whose DriveTrain

is All.

**A: The mean value of retail price for this type is $28,982.47 and the weight for this type of car is 3896.44 (LBS)**

1. What is the median of Weight for all European vehicles;

**A: The median weight for all European vehicles is 3585.0 (LBS)**

1. What is the median of RetailPrice for all vehicles whose DriveTrain is FRONT.

**A: The median retail price for all front drive vehicles is $22,582.50.**

Problem (e)

1. What is the percentage of USA vehicles?

**A: The percentage of USA vehicles is 34.35%**

1. What is the percentage of Asian Sedan vehicles?

**A: The percentage of Asian Sedan vehicles is 21.96%**

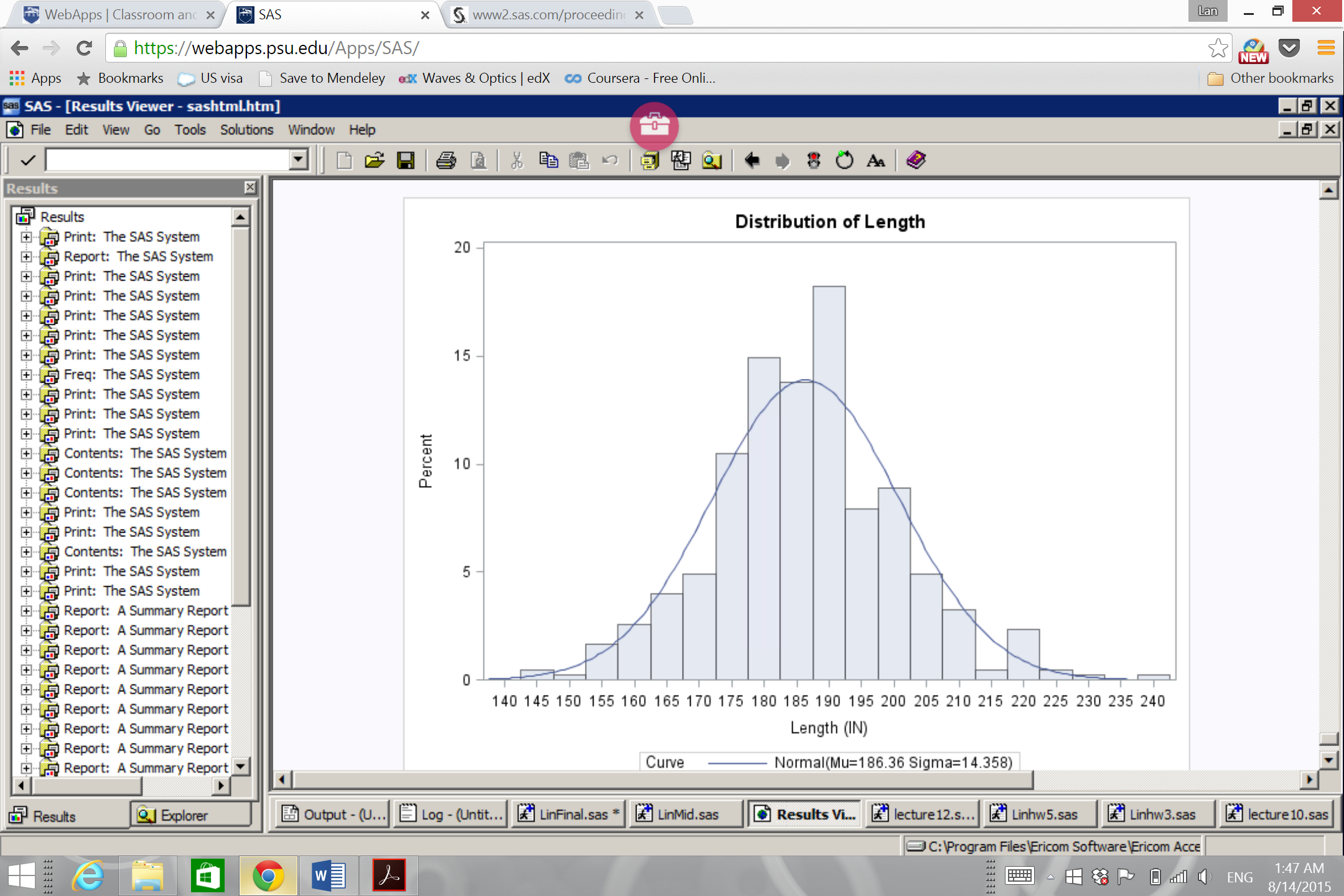
1. Among all European vehicles, what is the percentage of Sports vehicles?

**A: The percentage of Sports vehicles among all European vehicles is 18.70%.**

Problem (f)

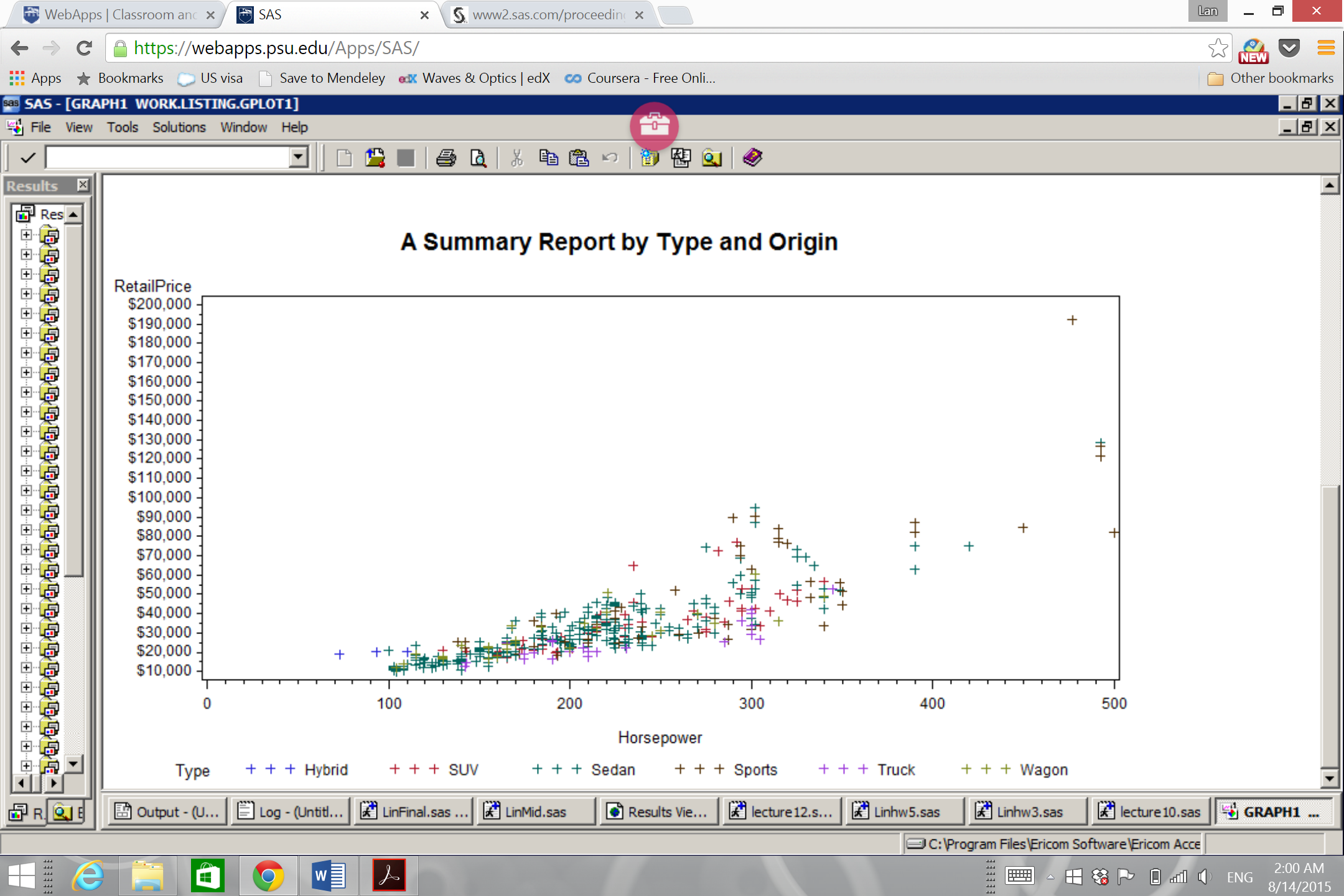
(i) Give the histogram of variable Length and the interval width is set to be 5. Then

add a normal curve to the histogram.

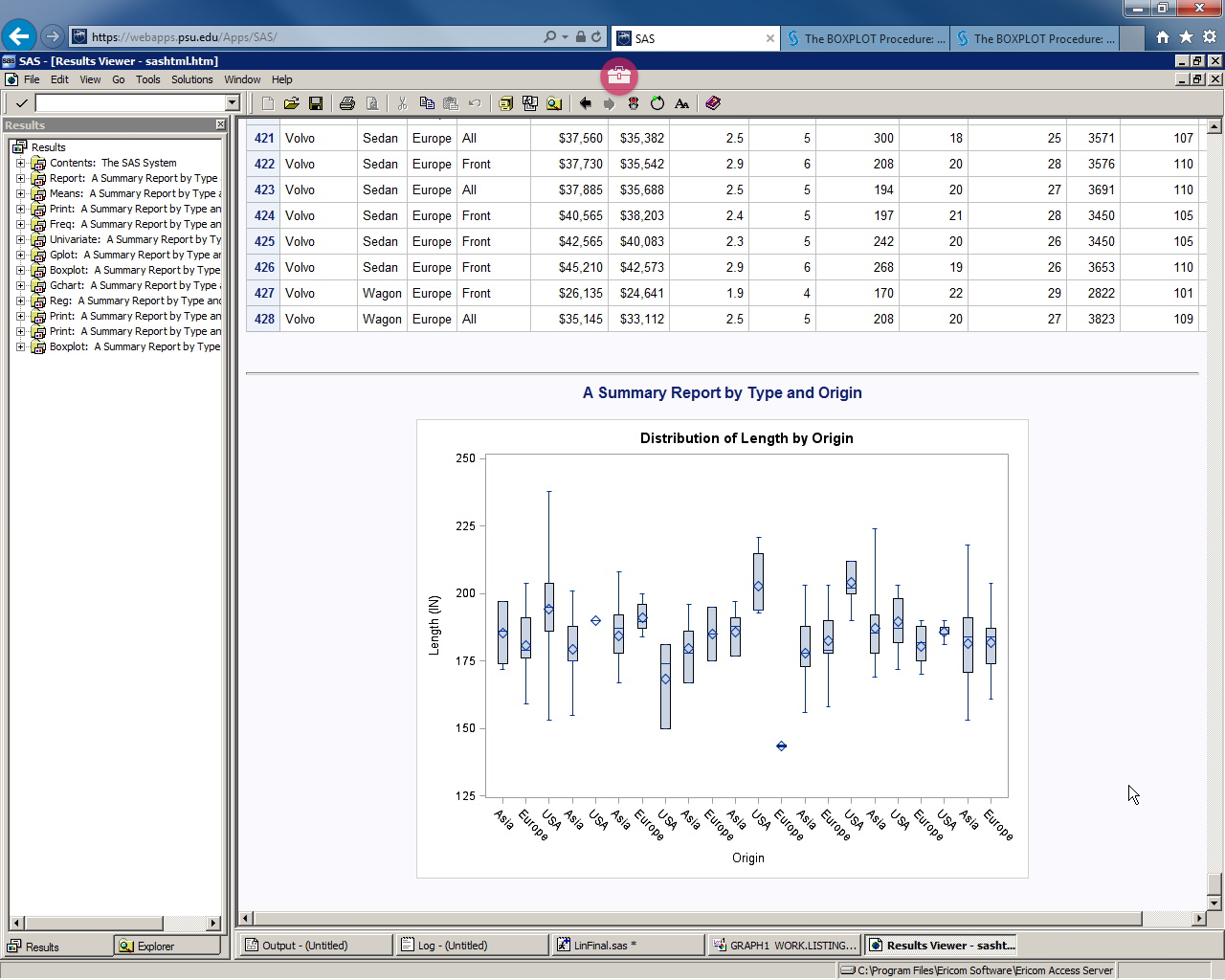


(ii) Give the scatter plot of variables RetailPrice and Horsepower (Horsepower is used

as X-axis). Different types should be labeled as different colors.

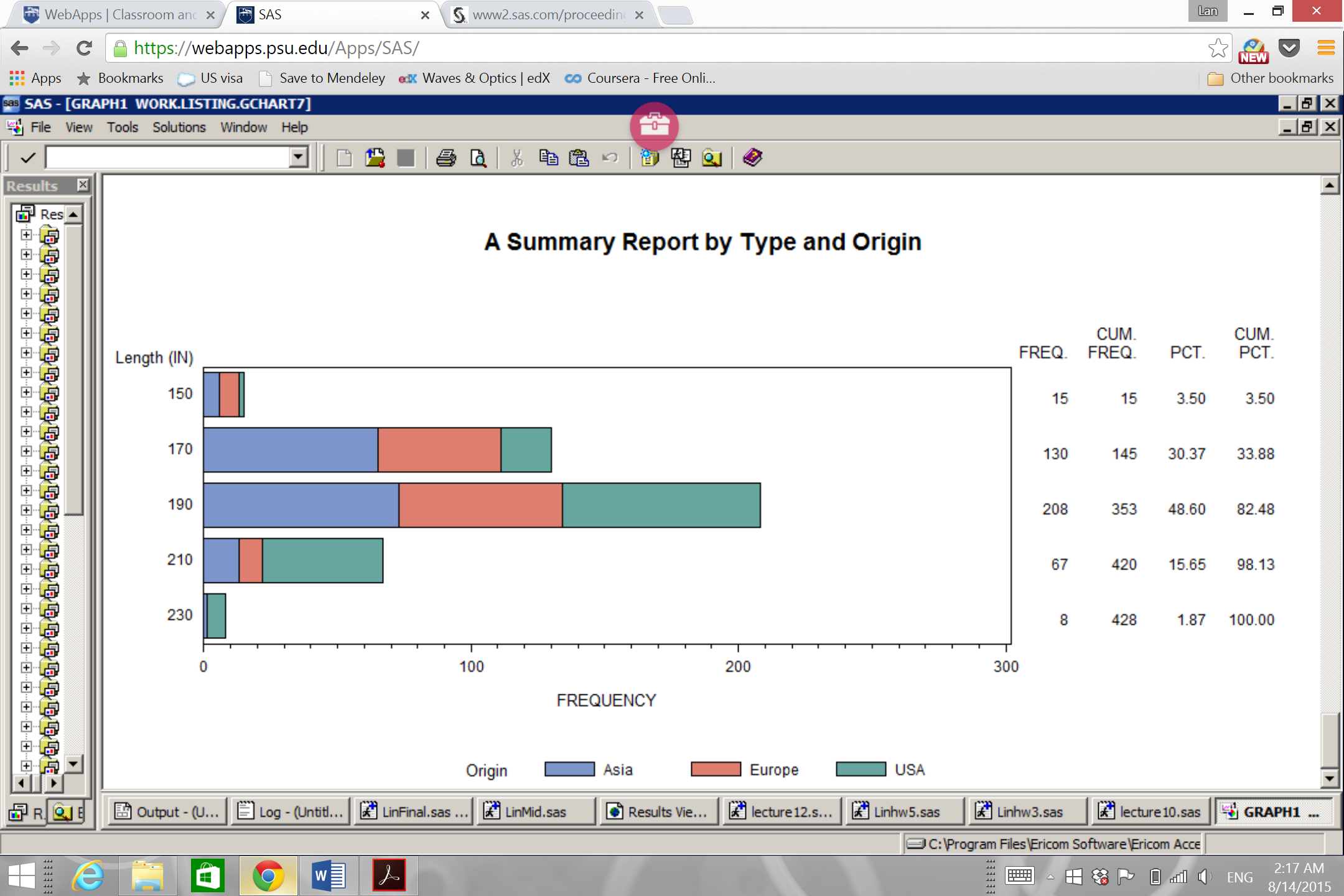


1. Give the boxplot of the variable Length.



(iv) For variable Length, use GCHART procedure to give a plot looking exactly the

same as Figure 2.



Problem (g)

(i) Give the estimated coefficients of the selected variables and write down regression

equation of your model.

**A: RetailPrice=2833.93005\*Cylinders-4820.47905\*EngineSize+245.58281\*Horsepower+589.29830\*MPG\_Highway +6.65820\*Weight-598.16006\*Wheelbase;**

(ii) Is there any multicollinearity in your model? Why?

**A: There is no multicollinearity, because the variance inflation (VIF) for these independent variable are smaller than 10.**