**Team Based Assignment (100 pts)**

The Advertising data set collects sales data of a product in a sample of 200 different markets, along with advertising budgets for the product in each of those markets for three different media: TV, radio, and newspaper. The Advertising data set lists sales, in thousands of units, and TV, radio, and newspaper budgets, in thousands of dollars.

**Question 1. (30 pts)**

Let’s imagine that you want to estimate the average number of units sold for the full population (all markets).

1. What would the 95% confidence interval be?
   1. The 95% confidence interval is 13.2994 to 14.7456 (see excel)
2. What would happen to the confidence interval at the same confidence level if you had a larger sample? Explain it qualitatively (you don’t have to provide a numerical value as answer)
   1. Increasing the sample size decreases the width of confidence intervals because it decreases the standard error.

**Question 2. (70 pts)**

Perform a regression analysis using SPSS Modeler to estimate *Sales* based on a) TV alone; b) Newspaper alone; c) Radio alone

1. What is the estimated regression equation for each regression analysis (a, b and c)?
   1. TV: Sales = TV\* 0.04754+7.033
   2. Radio: Sales = radio\*0.2025+9.312
   3. Newspaper: Sales = newspaper\*0.05469+12.35
2. Explain clearly the value of the slope coefficient you obtained in each of the three regression equations. What does the value of the y-intercept mean for the three regression equations you obtained? Does it make sense in each of the three cases?
   1. TV: slope .048, y-intercept 7.033
   2. Radio: slope .202, y-intercept 9.312
   3. Newspaper: slope 0.055, y-intercept 12.351

The y-intercept represents the number of sales if there were zero advertisements on TV, Radio or in the Newspaper.

The slope represents how much sales would increase by per advertisement.

This makes sense because the more a company advertises, the more sales they will make.

1. What would be a typical prediction error obtained from using the regression model to predict *sales* when using TV alone as a predictor? Which statistic are you using to measure this?
   * + The standard error of the estimate for the TV predictor is 3.258656
     + The statistic used is the standard error of the estimate
2. How closely do the three models fit the data? Which statistic are you using to measure this?

We use the R-squared statistics. TV fits best around 80% correct fit, while radio is about 33% and newspapers is less than 15

* 1. Newspapers is 0.052
  2. Radio is .332
  3. TV .612

1. Find a point estimate for the Sales with a TV advertising expenditure of 100 ($100,000).
   1. 7.033+0.048\*100= 11.833
2. Plot a scatterplot matrix chart to graphically depict the correlation between pairs of variables
   * + See Stream
3. Display the correlation coefficients (Pearson correlations) between each predictor and the target var (Sales) (hint: they should match the R value in the regression output)

TV: .782

Radio .576

News .228