**IT-223 - Assignment #4**

Create a Word document in which you paste all graphs and provide your discussion. For credit, your discussions of the questions asked below **MUST demonstrate that you have thought about and can therefore describe the underlying concepts**.

**Learning Objectives:**

         Demonstrate that you have been practicing and experimenting with SPSS.

         Demonstrate in your answers that you understand the underlying concepts!

**Questions:**

1. (15) Use the dataset [beer.sav](http://condor.depaul.edu/ymendels/223/beer.sav) for this problem. This dataset provides the alcohol percentage, calorie content and carbohydrate content for a series of beer brands.  You suspect that the amount of carbohydrate in a beer should be a pretty strong predictor of the calorie content. Using SPSS, create a scatterplot of carbohydrate v.s. calories.

Using SPSS:

* 1. Make a **scatterplot**. On your **scatterplot**, include the **regression line**.
  2. Calculate and report the **correlation coefficient**.
  3. **Describe this scatterplot** using the terminology we emphasized in class.
  4. Also **describe R2** in terms of how it applies to this scatterplot and dataset.

Using the output provided below:

* 1. **Report the regression model** (equation) that predicts calories from carbohydrates in the form  **y=B0 + B1x**
  2. Think of how in our last Beer-BAC model, figuring out a way to include the size of the person improved our model significantly. Try to come up with at least 1 data point or change you could make to improve this beer dataset so that you end up with a more accurate regression model.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model Summary** | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
| 1 | .803a | .645 | .641 | 16.654 |
| a. Predictors: (Constant), carbohyd | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANOVAa** | | | | | | |
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 42352.269 | 1 | 42352.269 | 152.697 | .000b |
| Residual | 23298.440 | 84 | 277.362 |  |  |
| Total | 65650.709 | 85 |  |  |  |
| a. Dependent Variable: calories | | | | | | |
| b. Predictors: (Constant), carbohyd | | | | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
| B | Std. Error | Beta |
| 1 | (Constant) | 82.660 | 5.056 |  | 16.350 | .000 |
| carbohyd | 5.277 | .427 | .803 | 12.357 | .000 |
| a. Dependent Variable: calories | | | | | | |

1. (25) Use the dataset [brain\_vs\_body\_weight.sav](http://condor.depaul.edu/ymendels/223/brain_vs_body_weight.sav) . This dataset is a list of brain weights (in kilograms) and body weights for a series of animals. You suspect (correctly!) that brain weight should be a pretty good predictor of body weight. To see if this is so,
   1. **do a regression analysis in SPSS**. When you look at the numbers alone, you should see a fairly high correlation coefficient. **Copy/paste the coefficients table or define the y=B0 + B1x equation in your submission**
   2. **Graph a scatterplot of the data**. You will see that there are some datapoints that may be either outliers, influential, or both.
      1. **Discuss which it is in your document**.
   3. Next, **explore with SPSS** and figure out how to **sort by body weight** (it’s not difficult to do….). Then delete the four highest values from your dataset (select the four rows, and clear them) and do another regression analysis. **Graph a scatterplot of this dataset as well**. **Copy/paste the coefficients table or define the y=B0 + B1x equation; copy/paste the scatterplot in your submission**
   4. Looking at your graph and at the regression analysis table provided by SPSS, **discuss** why taking out only 4 values significantly changed R/R2.