# Mathematics 220: SPSS Assignment 1

Fall 2007: **Due October 18, Thursday**

**1.** A group of 13 children and adolescents (considered healthy) participated in a psychological study designed to analyze the relationship between age (AGE) and average total sleep time (ATST). To obtain a measure for ATST (in minutes), recordings were taken on each subject on three consecutive nights and then averaged. The results obtained are displayed in the following table .

| ATST (in min/day) | AGE (in years) |
|---|---|
| 586.00 | 4.40 |
| 461.75 | 14.00 |
| 491.10 | 10.10 |
| 565.00 | 6.70 |
| 462.00 | 11.50 |
| 532.10 | 9.60 |
| 477.60 | 12.40 |
| 515.20 | 8.90 |
| 493.00 | 11.10 |
| 528.30 | 7.75 |
| 575.90 | 5.50 |
| 532.50 | 8.60 |
| 530.50 | 7.20 |

**(a).** Construct a histogram of $x$ variable (AGE) using SPSS and comment on some characteristics such as center and shape of the histogram.

**(b).** Obtain numerical summaries (such as mean, median, variance, standard deviation, IQR, etc.) for $y$ variable (ATST) using SPSS. Also obtain a boxplot of $y$ variable and comment on some characteristics (shape, center, skewness, outliers) of those plots.

**(c).** Give an interpretation of the value of the standard deviation of $y$ variable obtained in (b).

**(d).** Construct a scatterplot of $(x, y)$ data pairs using SPSS and explain why (or why not) the linear regression analysis is appropriate for the data, i.e, is there an approximate linear association between $x$ and $y$?

**(e).** Determine the sample correlation coefficient $r$ using SPSS and circle and label the number in the output. How do you interpret the value of $r$? (Hint: $r$ tells us the direction and the strength of linear association).

**(f).** Obtain a regression line of $y$ on $x$ using SPSS and write out the regression equation, i.e., $\hat{y} = a + bx$. Also interpret the value of slope $b$ of the regression line in the context of the problem. (Circle relevant output used to answer this question).

**(g).** What proportion of observed variation in ATST ($y$) can be explained by AGE ($x$)? (Circle relevant output used to answer this question).

**(h).** Compute the predicted value of the first observation (x=4.40) and obtain the residual (prediction error) for this individual.

**(i).** Predict the ATST for a person who are 25 years old. Explain why (or why not) this prediction is a valid prediction.