Problem 1. Egyptian Age-Height Example Again

(a) Obtain the Kalama data from the webpage (http://www.csub.edu/~sbehseta/math140winter06labs.htm) by clicking on Kalama.sav.

(b) Using SPSS, draw a scatterplot for age and height variables. Make sure that ”height” is in the Y axis and ”age” is in the X axis.

(c) Calculate the correlation coefficient $r$, between the two variables.

(d) Comment on linearity, strength, and direction of the relationship using both graph and the value of $r$.

(e) Let’s try fitting a linear regression line ($y = a + bx$) to these data. Notice that our response variable is ”height” and our explanatory variable is ”age”. Essentially, we need the estimates for the intercept ($a$) and the slope ($b$). First, you should obtain The intercept and the slope using the following two formulas:

\[ b = r \times \frac{S_y}{S_x} \]

\[ a = \bar{y} - (b \times \bar{x}) \]

Overlay your line on the scatterplot.

(f) Let’s try the above, using SPSS. Go to Analyze → Regression → Linear .... Put the ”height” variable in the dependent box, and the ”age” variable in the independent(s) box. Click ”OK”. Find the values for $r$ and $R^2$ in the Model Summary box. Compare those with your results. Also, in the Coefficient box, report the $B$ values for (Constant) and age. Those values should match with your estimations of $a$ and $b$ respectively.

(g) Next thing we want to do is to have SPSS plot the regression line for us. Go back to your scatterplot in the SPSS output viewer. Double click on the graph. This should open a new
environment called **SPSS chart editor**. Click on one of the data points in your chart editor. This should highlight the whole dataset. Now go to: **Chart → Options ...** and click **Total** in the **Fit Line** box. Now click on "OK". This should overlay the regression line (in red) onto your scatterplot!

**(h)** Knowing the fitted line, predict the height of a Kalama child whose age is 25.5 months. Predict the height of another child whose age is 32 months.

### Problem 2. CEO Problem Again!

**(a)** Obtain the **CEO** data from the webpage (http://www.csub.edu/~sbehseta/math140winter06labs.htm) by clicking on **CEO.sav**.

**(b)** Draw a scatterplot for **age** and **salary** variables.

**(c)** Calculate the correlation coefficient **$r$**, between the two variables.

**(d)** Comment on **linearity**, **strength**, and **direction** of the relationship using both graph and the value of **$r$**.

**(e)** Find the linear regression line for $Y :$ salary against $X :$ age (compare SPSS results with your calculations).

**(f)** Predict the salary of a CEO who is 38 years old. Predict the salary of a CEO who is 69 years old.

### Problem 3. An Interesting Concept in Linear Regressions!  

**(a)** Obtain the **Anscombe** data from the webpage (http://www.csub.edu/~sbehseta/lab140fall04.htm) by clicking on **FA.sav**. This data set has six variables, $X$, $Y_1$, $Y_2$, $Y_3$, $X_4$, and $Y_4$.

**(b)** Compute the mean and standard deviation for $X$ and $X_4$.

**(c)** Compute the mean and standard deviation for $Y_1$, $Y_2$, $Y_3$, and $Y_4$.

---

1*(Moore and McCabe, 1998)*
(d) Determine Pearson’s correlation coefficient for each of the following pairs of variables.

(a) \((X, Y_1)\)

(b) \((X, Y_2)\)

(c) \((X, Y_3)\)

(d) \((X_4, Y_4)\)

(e) Find the regression lines for each of the pairs in the above.

(f) In which of the four cases would you be willing to use the regression line to describe the dependence of \(y\) on \(x\)? Explain your answer in each case.

This is a very famous dataset that was originally designed by Frank Anscombe to demonstrate the dangers of calculating without first plotting the data.

Problem 4. Hurricane strengths

Download the file Hurricane.sav. You can obtain the file from http://www.csusb.edu/~sbehseta/math140winter06labs.htm

This is a file of tropical storm strengths (in knots) and their minimum central pressure (in millibars) for all the storms that appeared in the Atlantic Ocean in 1996.

(a) It is assumed that the strength of a hurricane is related to its central minimum pressure. In order to see this phenomenon, draw a scatter plot for between the two variables associated with the year 1996.

(b) Judging from the graph, do you see any relationship between the two variables? If there is a correlation, what kind of correlation is it? That is, comment on linearity, direction and strength of the relationship.

(c) Next, use ”wind” as the response variable and ”pressure” as the explanatory variable. Find the regression line.

(d) Based on your regression line, what is the expected ”wind” strength when the ”pressure” is 966?
(e) Now, we use "wind" as the explanatory variable and "pressure" as the response variable. Find the regression line.

(f) Based on your regression line, what is the expected pressure when the wind strength is 100?

(g) Notice that there is a difference between the values in parts b) and d). Why would you think we are getting two different lines? Explain.

**Problem 5. Normal Distributions**

According to the College Board, the scores on the SAT Math exam have a normal distribution with a mean of 500 and a standard deviation of 100.

(a) What is $P(X \leq 450)$?

(b) What is the proportion of students scored below 600 on this exam?

(c) What score would the student have to make to be in the top 1%?

**Problem 6. Sampling Distribution of $\bar{X}$**

Suppose that the data in a population follow a normal distribution with mean 200 and standard deviation 5.

(a) What is the probability of an individual in the population being less than 198?

(b) What is the probability that $\bar{X}$ (the average of a random sample of size 49) will be less than 198?

**Problem 7. Probability Distributions**

The incubation temperature to hatch ostrich eggs for a Type $SR - 50$ incubator is set at $99^\circ F$ and is allowed to vary with a standard deviation of $2^\circ F$. Each hour the temperature is recorded at 50 randomly selected times. If the average of the measurements is less than $98.5^\circ F$ or greater than $99.5^\circ F$, an alarm goes off. What is the probability that the alarm goes off, that is, what is the probability that the average is not between $98.5^\circ F$ and $99.5^\circ F$?