STA 6127: Solutions of Exercises for Reviewing Basic Regression

1. The given prediction equation is: $\hat{y} = 61.4 + 2.4x$.
   a) The y-intercept is 61.4 and the slope is 2.4.
   The slope can be interpreted as follows: For a 1 cm increase in the length of th4e femur, the predicted height of an individual increases by 2.4cm.
   b) The predicted height of an individual whose femur had length 50cm is $61.4 + 2.4(50) = 181.4$ cm.

2. The prediction equation is: $\hat{y} = 22 - 1.3x$.
   a) The y-intercept is 22. This means that for a nation which does not have any social expenditure (i.e., 0% of GDP), the predicted child poverty rate is 22%. The slope of the equation is -1.3. This means that as social expenditure (as percent of GDP) increases by 1%, the predicted child poverty rate drops by 1.3%.
   b) The correlation of -0.79 means there is a strong negative association between social expenditure (as percent of GDP) and child poverty rate. Thus, as the former increases (decreases), the latter decreases (increases) with a sharp regularity.

3. The prediction equation is: $\hat{y} = 1.26 + 0.346x$.
   a) For U.S, the predicted per capita Carbon dioxide emission when $x = 34.3$ is $\hat{y} = 1.26 + 0.346(34.3) = 13.13$. The observed per capita Carbon dioxide emission is $y = 19.7$. So the residual is y - $\hat{y}$ = 19.7 - 13.13 = 6.57. Since the residual is positive, we conclude that in the case of U.S, the given regression line "underestimates" the value of the response variable (since the observed value is higher than the predicted value).
   b) For Switzerland, the predicted per capita Carbon dioxide emission when $x = 28.1$ is $\hat{y} = 1.26 + 0.346(28.1) = 10.98$. The observed per capita Carbon dioxide emission is $y = 5.7$. So the residual corresponding to $x = 28.1$ is y - $\hat{y}$ = 5.7 - 10.98 = -5.28. Since the residual is negative, we conclude that in the case of Switzerland, the given regression line "overestimates" the value of the response variable (since the observed value is lower than the predicted value).

4. The prediction equation is given by $\hat{y} = 30 + 0.60x$.
   a) (i) The predicted Final exam score for a student whose midterm score is 100 is $\hat{y} = 30 + 0.60(100) = 90$.
   (ii) The predicted Final exam score for a student whose midterm score is 50 is $\hat{y} = 30 + 0.60(50) = 60$.
   b) We know that the correlation coefficient r is given by: $r = b(s_y/s_x)$ where b is the slope and s_y (s_x) are the standard deviations of y and x. Since, in our case both s_y and s_x are the same, we have, from the above formula, $r = b = 0.60$.
   c) If $\hat{y} = x$, the slope is 1.0 (and the y-intercept is 0), and since the standard deviations are equal, from the equation $r = b(s_y/s_x)$, $r = 1.0$ also.
   d) If $\hat{y} = 75$, then the slope $b = 0$ (and the y-intercept is 75) and so $r = 0$.

5. **a)** Testing the hypothesis of independence is equivalent to testing $H_o : \beta = 0$.
   The alternative hypothesis is $H_a : \beta \neq 0$.
   The test statistic is $t = b/\text{se}(b) = 0.43$, where se($b$) denotes the standard error of b. Since the P-value is 0.665, the evidence does not contradict H_0 and it is plausible that the variables are independent.
   b) The 95% CI for the population slope β is given by : $b \pm 1.96se(b) = 0.00739 \pm 1.96(0.01706) = 0.00739 \pm 0.0334$ = (-0.026, 0.041). Since the CI contains 0, it is plausible that there is no effect.
   c) "Beta" is the correlation between x and y (bad labelling by SPSS!). Since the value is very small (0.01554), we estimate that there is a very weak association between Political ideology and income.

6. **a)**From the given output the P-value = 0.0015. Since it is very small, there is extremely strong evidence of a positive association between these variables.
   b) The very small P-value is highly statistically significant, giving strong evidence of an association. Yet, the association appears to be quite weak, as evidence by the correlation of only 0.087.

7. **a)** For the interpretation in quotes, the explanatory variable is "heights in inches" and the response variable is "salaries in Dollars" of individuals. The slope is $789.
   b) If a inch increase in height leads to a predicted 789 dollar increase in salary, then 7 inch (6ft - 5ft 5 inch = 7 inch) increase in height should result in 789(7) Dollars = 5,523 Dollars increase in salary.

8. **a)** A pound change corresponds to a 1.9 dollar change. If a slope is b in dollars, then a one-unit change in the explanatory variable corresponds to a predicted change in Y of b dollars. Since b dollars equals $b/1.9 = 0.53b$ pounds, a one-unit change in the explanatory variable corresponds to a predicted change in Y of $0.53b$ pounds. Then new slope would be the old one divided by 1.9. (Note: If income were the *explanatory* variable, then the new slope would multiply by 1.9, since a change in income of 1 pound corresponds to a change of 1.9 dollars, so the effect gets multiplied by 1.9.)
   b) A basic property of the correlation is that is does not depend on the units of measure. It would not change.

9. No, there is not enough evidence to conclude this. We would expect some improvement, merely because of regression toward the mean, even if there were no tutoring program. For example, if the correlation is 0.5, a student who scored 50 on the first exam would be predicted to score 60 on the second exam (regression half way toward the mean score).

10. **a)** Substituting $x = \bar{x}$ into the prediction equation $\hat{y} = a + bx$, we have: $\hat{y} = a + b\bar{x}$
    Now substituting the expression $\bar{y} - b\bar{x}$ in place of a in the above expression, we have: $\hat{y} = \bar{y} - b\bar{x} + b\bar{x}$ $= \bar{y}$. Since putting $x = \bar{x}$ into the prediction equation, yields the value $\bar{y}$ for $\hat{y}$, the prediction equation passes through the point with co-ordinates $(\bar{x}, \bar{y})$.
    b) Substituting the given value of a in the prediction equation $\hat{y} = a + bx$, we have $\hat{y} = \bar{y} - b\bar{x} + bx$, or, $(\hat{y} - \bar{y}) = b(x - \bar{x})$.
    c) We have already shown that we can rewrite any prediction equation as $(\hat{y} - \bar{y}) = b(x - \bar{x})$, where b is the slope $b = r(s_y/s_x)$, where s_y (s_x) are the standard deviations of y (x) values. Since $s_y = s_x$, $b = r = 0.70$. So, we can write the prediction equation as $(\hat{y} - \bar{y}) = 0.70(x - \bar{x})$.