Problem 1. [Points = 4] Suppose that a researcher, using data on class size $CS$ average test scores from 100 third-grade classes, estimates the OLS regression,

$$\text{TestScore} = 520.4 - 5.82 \times CS,$$

$R^2 = .08$, $MSR = 11.5$.

(a) A classroom has 22 students. What is the regression’s prediction for that classroom’s average test score?

(b) Last year a classroom had 19 students, and this year it has 23 students. What is the regression’s prediction for the change in the classroom average test score?

(c) The sample average class size across the 100 classrooms is 21.4. What is the sample average of the test scores across the 100 classrooms?

(d) What is the sample standard deviation of the test scores across the 100 classrooms?

[Hint: Use the relationship between $R^2$ and $MSR$.]

Answer 1:

(a) The predicted score is $\text{TestScore}(CS = 22) = 520.4 - 5.82 \times 22 = 392.36$.

(b) The change in predicted score is $\text{TestScore}(CS = 23) - \text{TestScore}(CS = 19) = -5.82 \times (23 - 19) = -23.28$. The score is predicted to decrease.

(c) Recall the formula for the intercept term derived in the lecture: $\hat{\beta}_0 = \bar{y}_n - \hat{\beta}_1 \bar{x}_n$. This implies that $\bar{y}_n = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_n$. In this problem $\hat{\beta}_0 = 520.4, \hat{\beta}_1 = -5.82$ and $\bar{x}_n = \text{average class size} = 21.4$. Therefore, average test score is $\bar{y}_n = 395.85$. 

1
(d) Recall that $R^2 = ESS/TSS = 1 - RSS/TSS$. Therefore, $TSS = RSS/(1 - R^2) = (n - 2)MSR/(1 - R^2)$. So, you can get the answer as follows

$s.d. (TestScore) = s.d. (y) = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y}_n)^2} = \sqrt{\frac{TSS}{n-1}} = \sqrt{\frac{(n-2)}{(n-1)} \times \frac{MSR}{1-R^2}} = 3.52$.

Problem 2. [Points = 6] A regression of average weekly earnings ($AWE$, measured in USD) on age (measured in years) using a random sample of college-educated full-time workers aged 25-65 yields the following:

$\hat{AWE} = 696.7 + 9.6 \times Age,$

$R^2 = .023, MSR = 624.1.$

(a) Explain what the coefficient values 696.7 and 9.6 mean?

(b) What is unit of measurement for the $MSR$ - USD, years, or is it unit-free?

(c) What is unit of measurement for the $R^2$ - USD, years, or is it unit-free?

(d) What is the regression’s predicted earnings for a 25-year-old worker? A 45-year-old worker?

(e) Will the regression give reliable predictions for a 99-year-old worker? Why or why not?

(f) The average age in the sample is 41.6 years. What is the average value of $AWE$ in the sample?

Answer 2:

(a) Let us start with the interpretation of the intercept. Literally it means that at $Age = 0$ the predicted $AWE$ is 696.7. Of course, this is foolish. Actually, you can look at it from
another angle. Suppose that there is another factor – your country (or a combination of multiple factors) – that also affect the AWE, but is uncorrelated with your AGE. Then the intercept term is 696.7 tries to capture the contribution of this other factor (at its average). The slope 9.6, on the other hand, means that with each year of increase in Age, the average weekly wage increases by 9.6 USD. This is typically the effect of work-experience or increased qualifications. Once again, there is a caveat – such an increase cannot go on indefinitely, the person has to retire! To make sense of the result, it is extremely important to consider the range of the explanatory variable - Age.

(b) The unit is USD\(^2\).

(c) \(R^2\) is unit-free. Why? It is a ratio of quantities measure by the same unit.

(d) For a 25 year old the prediction is \(\hat{AWE}(25) = 696.7 + 9.6 \times 25 = 936.7\) in USD. For a 45 year old the prediction is \(\hat{AWE}(45) = 696.7 + 9.6 \times 45 = 1128.7\) in USD.

(e) This is related to (a). The regression, as a tool, will mechanically predict \(\hat{AWE}(99) = 696.7 + 9.6 \times 99 = 1647.1\) in USD. However, this does not make sense because the sample of observations does not have any information on people aged over 65. So predicting average weekly earnings for a 99 year old is a wild extrapolation that you should always try to avoid.

(f) Recall the formula for the intercept term derived in the lecture: \(\hat{\beta}_0 = \bar{y}_n - \hat{\beta}_1 \bar{x}_n\). This implies that \(\bar{y}_n = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_n\). In this problem \(\hat{\beta}_0 = 696.7, \hat{\beta}_1 = 9.6\) and \(\bar{x}_n = \text{average age} = 41.6\). Therefore, average weekly earnings score is \(\bar{y}_n = 1096.06\) (USD).

Problem 3. [Points = 5] Consider a regression model: \(y = \beta_0 + \beta_1 X + \epsilon\). Suppose that you have observations \((y_1, X_1), \ldots, (y_n, X_n)\) and you obtain the following estimates of \(\beta_0\) and
$\hat{\beta}_1$: 

$$
\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X}_n)y_i}{\sum_{i=1}^{n}(X_i - \bar{X}_n)X_i}, \\
\hat{\beta}_0 = \bar{y}_n - \hat{\beta}_1\bar{X}_n.
$$

(a) What is the residual of the regression, i.e. what is $\hat{\epsilon}_i = y_i - \hat{y}_i$?

(b) Show that $\sum_{i=1}^{n}\hat{\epsilon}_i = 0$?

(c) Show that $\sum_{i=1}^{n}X_i\hat{\epsilon}_i = 0$?

**Answer 3:**

(a) The residual is $\hat{\epsilon}_i = y_i - \hat{y}_i$. To gain a better understanding of the structure, let us examine the form of $\hat{y}_i$. Note that 

$$
\hat{y}_i = \hat{\beta} + \hat{\beta}_1X_i = (\bar{y}_n - \hat{\beta}_1\bar{X}_n) + \hat{\beta}_1X_i = \bar{y}_n + \hat{\beta}_1(X_i - \bar{X}_n)
$$

and, therefore, the residual is 

$$
\hat{\epsilon}_i = (y_i - \bar{y}_n) - \hat{\beta}_1(X_i - \bar{X}_n). \tag{1}
$$

(b) From (1), it is easy to see that 

$$
\sum_{i=1}^{n}\hat{\epsilon}_i = \sum_{i=1}^{n}[(y_i - \bar{y}_n) - \hat{\beta}_1(X_i - \bar{X}_n)] = \sum_{i=1}^{n}(y_i - \bar{y}_n) - \hat{\beta}_1\sum_{i=1}^{n}(X_i - \bar{X}_n) = 0.
$$

Note that $\sum_{i=1}^{n}(y_i - \bar{y}_n) = \sum_{i=1}^{n}y_i - \sum_{i=1}^{n}\bar{y}_n = n\bar{y}_n - n\bar{y}_n = 0$, and similarly $\sum_{i=1}^{n}(X_i - \bar{X}_n) = 0$.

(c) From (1), and further noting that $\sum_{i=1}^{n}X_i(y_i - \bar{y}_n) \equiv \sum_{i=1}^{n}(X_i - \bar{X}_n)y_i$ (Why?), you
can get
\[
\sum_{i=1}^{n} X_i \hat{e}_i = \sum_{i=1}^{n} X_i \left[ (y_i - \bar{y}_n) - \hat{\beta}_1 (X_i - \bar{X}_n) \right]
\]
\[
= \sum_{i=1}^{n} X_i (y_i - \bar{y}_n) - \hat{\beta}_1 \sum_{i=1}^{n} X_i (X_i - \bar{X}_n)
\]
\[
= \sum_{i=1}^{n} (X_i - \bar{X}_n) y_i - \hat{\beta}_1 \sum_{i=1}^{n} (X_i - \bar{X}_n) X_i
\]
\[
= \sum_{i=1}^{n} (X_i - \bar{X}_n) y_i - \frac{\sum_{i=1}^{n} (X_i - \bar{X}_n) y_i}{\sum_{i=1}^{n} (X_i - \bar{X}_n) X_i} \sum_{i=1}^{n} (X_i - \bar{X}_n) X_i
\]
\[
= 0.
\]

**Problem 4. [Points = 5]** The file hw2.dta contains data on course evaluations, course characteristics, and professor characteristics for 463 courses at the University of Texas at Austin. One of the characteristics is an index of the professors’ "beauty" as rated by a panel of 6 judges. In this exercise you will investigate how course evaluations are related to the professor’s beauty.

(a) Construct a scatter-plot of average course evaluations on the professor’s beauty. Does there appear to be a relationship between the variables?

(b) Run a regression of average course evaluations on the professor’s beauty. What is the estimated intercept? What is the estimated slope? Explain why the estimated intercept is equal to the sample mean of course evaluations? (Hint: Read the label/description of beauty carefully.)

(c) Suppose Professor A has an average value of beauty, while Professor B’s value of beauty is one standard deviation above the average. Predict Professor A’s and Professor B’s course evaluations.
(d) Comment on the size of the regression’s slope. Is the estimated effect of beauty on course evaluation large or small? Explain what you mean by "large" or "small".

(e) Does beauty explain a large fraction of the variation in evaluations across courses? Explain.

Answer 4:

(a) The scatter-plot along with the fitted regression line (with 95 % confidence band) is given in Figure-1. First focus on the scatter-plot and ignore the rest (i.e. the predicted red line and the confidence bands). Some interesting observations from the scatter-plot are: (i) the worst evaluations are for teachers who are also rated the least attractive physically, (ii) the teachers who are rated highest in terms of physical appearance
always avoid bad course evaluations, (iii) at the highest level of course evaluations, physical appearance seem to have less effect.

(b) The estimated intercept from the regression of course evaluation on beauty is 3.9 (equal to sample mean because sample mean of beauty is 0). The estimated slope is .13 with standard error .03. Both these parameters are quite precisely estimated. Note that the slope coefficient implies that an infinitesimal increase in physical beauty (as judged by the students here) increase the course evaluations by .13 points. To take a concrete example, consider a professor whose current course evaluation is 1 (i.e. very unsatisfactory). Also suppose that this professor’s physical appearance is considered very unattractive (say, -5) by the students. Now suppose (by magic) the professor turns physically attractive (say, +5) one fine morning. According to our regression, this event alone will increase the course evaluation to 2.3 (i.e. a 130 % increase!).

(c) Note that $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X = \bar{y}_n + \hat{\beta}_1(X - \bar{X}_n)$. Therefore, for Professor A, whose beauty is average (i.e., $X = \bar{X}_n$) is predicted to have average course evaluation (i.e., $\hat{y} = \bar{y}_n$) 3.9. On the other hand, Professor B has beauty one standard deviation (i.e., .79, obtained by the "sum" command) above average is predicted to have course evaluation $3.9 + .13 \times .79 = 4.0$.

(d) What do you think? Later we will further analyze the data and try to find out the other factors (that should not be related to the Professor’s teaching skills) that also affect the course evaluation.

(e) The $R^2$ is .036 meaning beauty explains about 3.6 % of the variation in course evaluation. Does this seem large to you?