Problem Set 2  
ECON 30331  
(Due by the start of class, Wednesday, January 27, 2016)  
(Problems marked with a * are former test questions)

Bill Evans  
Spring 2016

1. Suppose a researcher is interested in estimating the linear regression model $Y_i = \beta_0 + X_i \beta_1 + \varepsilon_i$ and in a sample of 101 observations, the following descriptive statistics are generated:

$$
\bar{x} = 30, \ \bar{y} = 20, \ \sum_{i=1}^{n} (x_i - \bar{x})^2 = 250, \ \sum_{i=1}^{n} (y_i - \bar{y})^2 = 160
$$

$$
\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = 125, R^2=0.2
$$

a) What is $\hat{\sigma}_x^2$?

b) What is $\hat{\rho}(x, y)$?

c) What is the OLS estimate of $\hat{\beta}_1$?

d) What is the OLS estimates of $\hat{\beta}_0$?

e) What is the SSE for this model?

2. *Using data from the 2004 baseball season, a researcher collects data on the number of wins a team had during the year and payroll in millions of dollars. The researcher wants to estimate a model to examine whether the size of the payroll alters wins, so they want to consider an OLS model of the form $\text{wins}_i = \beta_0 + \text{payroll}_i \beta_1 + \varepsilon_i$. The author gets as far as getting descriptive statistics and the correlation coefficient between wins and payroll (presented below), then their computer crashes. Using the data below:

a) Calculate the estimate for $\hat{\beta}_0$.

b) For $\hat{\beta}_1$.

c) Interpret the results for $\hat{\beta}_1$. According to the model estimates, by how much will wins increase if a team spends $15 million more on salary?

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>wins</td>
<td>30</td>
<td>80.9667</td>
<td>13.36615</td>
<td>43</td>
<td>101</td>
</tr>
<tr>
<td>payroll</td>
<td>30</td>
<td>70.13708</td>
<td>27.26755</td>
<td>19.63</td>
<td>149.711</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>corr wins payroll (obs=30)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>wins payroll</td>
<td>1.0000</td>
<td>0.4176</td>
</tr>
<tr>
<td>payroll</td>
<td>0.4176</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
3. *Below are STATA results from a CORrelation, SUMmary, and REGression statement, but some of the results have been whited-out. Please provide estimates for A, B, C, D, and E

```
sum x y

Variable | Obs  Mean   Std. Dev.   Min   Max
---------|-------|----------|----------|-----|-----
        x | 48    .5     .5052912  0   1
        y | 48    D   .1177725  7.674153 8.037867

corr x y
(obs=48)

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
|       y  | 0.9285  1.0000

reg y x

Source | SS         df | MS         Number of obs = 48
--------|--------------|------------|-----------------|-----------------|
        Model | A   1.562059213 1  | Prob > F = 0.0000  |
        Residual | .08984759 46  .001953208  | R-squared = B  |
        Total | .651906803 47  .013870358  | Adj R-squared = 0.8592  |

---------|----------|----------|----------|---------|----------|
|       x  |   Coef.  Std. Err.  t  P>|t|  [95% Conf. Interval]
---------|----------|----------|----------|---------|----------|
            |   C      0.012758    16.96  0.000   0.1907409    0.2421021|
            | _cons   7.771764    0.0090213  861.49  0.000   7.753605    7.789923
```

4. Given an OLS model of the form $Y_i = \beta_0 + X_i \beta_1 + \epsilon_i$, suppose that $\hat{\beta}_1 = 0$.

a) What is the estimate for $\hat{\beta}_0$?
b) What does $R^2$ equal in this case?

5. *Given an OLS model of the form $Y_i = \beta_0 + X_i \beta_1 + \epsilon_i$, show that $\bar{Y} = \bar{Y}$. 

6. Suppose a researcher is interested in estimating the impact of gasoline taxes ($X_i$) on per capital gallons of gasoline consumed per year ($Y_i$). Assume tax is measured in cents per gallon. The researcher has data from 51 states for a 10 year period for a total of 510 observations. The researcher estimates the linear OLS model $Y_i = \beta_0 + X_i \beta_1 + \epsilon_i$ and calculates $\hat{\beta}_1 = -0.90$.

a) Suppose instead of measuring taxes in cents per gallon, the researcher measures taxes in dollars per gallon where the new model is $Y_i = \gamma_0 + X_i \gamma_1 + \epsilon_i$ and $X_i = X/100$. What will be the estimate on the coefficient on $\gamma_1$?

b) Suppose taxes are measured in cents as in the first case, but consumption is measured as gallons consumed per month, where $Y_i = Y_i/12$. The model now is of the form $Y_i = \alpha_0 + X_i \alpha_1 + \epsilon_i$. What will be the estimate on $\alpha_1$?

[You have a data set with Q and taxes but for a different product. Check your answer with the data set we have used in class.]

7. On the assignments page for the class is a STATA data set called meps_senior.dta that has information on total medical care expenditures for a sample of elderly respondents (aged 65 and over) from the Medical Expenditure Panel Survey (MEPS). There are many variables in the data set but for now, run a regression using totalexp (total annual medical expenditures) as the dependent variable, and as the independent covariate, use age (age in years), so the regression is $\text{totalexp}_i = \beta_0 + \text{age}_i \beta_1 + \epsilon_i$.

A) What are the estimates for $\hat{\beta}_0$ and $\hat{\beta}_1$?

B) Interpret the coefficient for $\hat{\beta}_1$, what is $\frac{\partial \text{total exp}}{\partial \text{age}}$?

C) Using the estimates for $\hat{\beta}_0$ and $\hat{\beta}_1$ what is predicted spending for someone 70 years of age?

D) Using the estimates for $\hat{\beta}_0$ and $\hat{\beta}_1$ what is predicted spending for someone 71 years of age?

E) Generate the difference between the answers in part D) and C) – does this make sense?


a. What is the $R^2$ in the previous problem?

b. Run a regression of the form $\text{age}_i = \gamma_0 + \text{totalexp}_i \gamma_1 + \epsilon_i$. What is the $R^2$ from this regression?

c. Show that in general, the $R^2$ from a regression of $Y$ on $X$ is the same as a regression of the $R^2$ from a regression of $X$ on $Y$.

[Hint: $R^2 = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2}$, substitute the definition of $\hat{y}_i - \bar{y}$. In the second model $\text{age}_i = \gamma_0 + \text{totalexp}_i \gamma_1 + \epsilon_i$ and which we can write as $x = \gamma_0 + y \gamma_1 + \epsilon_i$ we know that $\hat{\gamma}_1 = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (y_i - \bar{y})^2}$ and the $R^2$ in this case equals $R^2 = \frac{\sum_i (\hat{x}_i - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}$]