Second Midterm Exam
You must answer all the questions. The exam is closed book and closed notes you may use calculators.

**You must show your work to receive full credit**

1. We are interested in the relationship between how many times a month a person plays golf and their score in a golf tournament.

<table>
<thead>
<tr>
<th>X=Games month</th>
<th>Y=Tournament score</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>75</td>
</tr>
<tr>
<td>12</td>
<td>64</td>
</tr>
<tr>
<td>3</td>
<td>82</td>
</tr>
<tr>
<td>9</td>
<td>73</td>
</tr>
</tbody>
</table>

\[ \text{Tourn\_Score} = B_0 + B_1 \text{games\_month} + u \]

a. Compute \( \hat{B}_0 \) and \( \hat{B}_1 \)

b. What would we expect a person who plays 8 times a month to get in the tournament?
c. Will playing one additional game of golf per month necessarily change the golfers’ score by $\hat{B}_1$ on average? Why or why not?

2. We are interested in the determinants of CEOs salary. We think the CEOs tenure (years) as CEO, their age and the company sales may be important determinants of the CEOs salary. We run the following regression.

$$\log(\text{CEO\_salary}) = B_0 + B_1 \text{sales} + B_2 \text{CEO\_tenure} + B_3 \text{Age} + u$$

a. Before looking at the regression results below what sign do you expect for $B_1$ and why?

We run the following regression in Stata

```
. reg lsalary sales ceoten age
```

```
Source |       SS       df       MS              Number of obs =     177
--------+------------------------------           F(  3,   173) =   13.40
Model | 12.1860321     3  4.06201069           Prob > F      =  0.0000
Residual | 52.4601895   173  .303238089           R-squared     =  0.1885
--------+------------------------------           Adj R-squared =  0.1744
Total | 64.6462215   176  .367308077           Root MSE      =  .55067

------------------------------------------------------------------------------
 lsalary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 sales | .000042   6.92e-06     XXXX   XXXXX    XXXXXXXXX    XXXXXXXX
 ceoten | .0125175   .0062134     XXXX   XXXXX    XXXXXXXXX    XXXXXXXX
 age | -.0009396   .0053066     XXXX   XXXXX    XXXXXXXXX    XXXXXXXX
 _cons |   6.388217   .2863196     XXXX   XXXXX    XXXXXXXXX    XXXXXXXX
------------------------------------------------------------------------------
```
b. State the null and alternative hypothesis then conduct a 1-sided t-test at the 3.01% level for the coefficient for $B_2$? Which table should you look your results up on?

c. Interpret the coefficient for $B_2$ and determine if it is practically significant.

d. Interpret the $R^2$ in this regression.
3. We are interested in checking if larger cities have higher rates of influenza than we would expect just from them being larger (Does a 1% increase in population result in a more than 1% increase in Influenza). We have a dataset with information on the number of flu cases, the population of the town and the average temperature in December. We get the following regression results.

\[
\text{Log(Influenza)} = 170.4 + 1.2 \text{ log(city\_population)} - 0.03 \text{ Dec\_temp}
\]

\[
\begin{align*}
(17.3) & & (.08) & & (.001) \\
\text{n} = 129, \ R^2 = 0.34
\end{align*}
\]

Test if the elasticity (Coefficient on city population) of influenza with respect to population is one. Conduct a one sided test at the 1% level.

a. Please write out the null and alternative hypothesis for a one sided test

b. Compute the t-statistic and determine if it is statistically significant. Please include a picture.
c. Compute the p-value for the coefficient on population. Where the p-value is the probability (under the null hypothesis) of getting a t-statistic larger than the one observed. It may be useful to draw a picture.

d. Why did I ask you to do a one sided test.
4. We are interested in the determinants of crime in a community and think that local economic conditions may play a significant role. We run the following regression.

```
. reg crimes pop officers pcinc unem

Source |       SS       df       MS              Number of obs =      92
-------------+------------------------------           F(  4,    87) =  104.68
Model |  6.6425e+10     4  1.6606e+10           Prob > F      =  0.0000
    Residual |  1.3802e+10    87   158645388           R-squared     =  0.8280
    Total |  8.0227e+10    91   881620871           Adj R-squared =  0.8201
-------------+------------------------------           Root MSE      =   12595

------------------------------------------------------------------------------
crimes |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+---------------------------------------------------------------
population |    .064322   .0107904     XXXX   XXXXX    XXXXXXXXX    XXXXXXXX
policemen  |   15.16236   3.596209     XXXX   XXXXX    XXXXXXXXX    XXXXXXXX
        pcinc |   .4778875   .8211205     XXXX   XXXXX    XXXXXXXXX    XXXXXXXX
    unem |  -63.81206   551.9349     XXXX   XXXXX    XXXXXXXXX    XXXXXXXX
     _cons |  -3523.062   10580.21     XXXX   XXXXX    XXXXXXXXX    XXXXXXXX
------------------------------------------------------------------------------
```

a. Is per capita income significant at the 5% level for a two sided test. Write out the null hypothesis the alternative hypothesis and calculate the t-statistic. Which distribution do you check the statistic against?

b. Is unemployment significant at the 5% level for a one sided test. Write out the null hypothesis the alternative hypothesis and calculate the t-statistic. Which distribution do you check the statistic against?
c. Are the two measures of the economy, per capita income and unemployment, jointly significant at the 5% level?

```
. reg crimes pop officers

Source |       SS       df       MS              Number of obs =      92
--------+------------------------------           F(  2,    89) =  211.80
Model |  6.6298e+10     2  3.3149e+10           Prob > F      =  0.0000
Residual |  1.3930e+10    89   156513952           R-squared     =  0.8264
        |                     Adj R-squared =  0.8225
Total |  8.0227e+10    91   881620871           Root MSE      =   12511

------------------------------------------------------------------------------
crimes |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------+----------------------------------------------------------------
   pop |   0.0663665   .0104739     6.34   0.000      .045555    .0871779
  officers |  14.41498    3.44767     4.18   0.000     7.564535    21.26543
       _cons |  111.4153   2464.613     0.05   0.964    -4785.718    5008.549
------------------------------------------------------------------------------
```

Extra Credit: What is the impact on $\hat{B}$ and the $\text{se}(\hat{B})$ of adding an additional independent variable to the regression that is uncorrelated with the dependent variable $y$ but is correlated with all of the other independent variables ($x$s)?

Useful Formulas

$$
\hat{B}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}
$$

$$
\hat{B}_0 = \bar{y} - \hat{B}_1 \bar{x}
$$

$$
F = \frac{(SSR_{\text{res}} - SSR_{\text{adj}})/ q}{SSR_{\text{adj}}/(n-k-1)}
$$

$$
F = \frac{(R^2_{\text{adj}} - R^2_{\text{res}})/ q}{(1 - R^2_{\text{adj}})/(n-k-1)}
$$