1.(a) Consider the time series plot for the natural logarithm of $y$ given at the top of the last page. What is the approximate annual growth rate in the level of $y$? Explain your answer.

(b) An investment takes four years to double. What is the approximate annual rate of return for the investment? Explain your answer.

(c) Give the formula for a four-period moving average in the series $y_t$.

(d) What is a convenient source for time series data on government bond interest rates?

(e) What does the following Stata code produce: `graph twoway (line y year) (line x year)`?

(f) Suppose we have four observations with $(x, y) = (2, 6), (3, 1), (5, 1)$ and $(6, 4)$. Compute the least squares intercept and slope coefficients from the formulae. Show all workings.
QUESTIONS 2-3 USE STATA OUTPUT GIVEN AT THE END OF THIS EXAM. For some questions the answer is given directly in the output. For other questions you will need to use the output plus additional computation.

2. This question uses data for hospitals in New York state in 2011.
   - meancost is the average cost of these knee replacements
   - meancharge is the average charge for knee replacements.
Note: the hospital charge is the initial price that the hospital charges. The hospital usually receives less than this after negotiating with the patient or the patient’s insurance company.

(a) How does the mean charge change when mean cost increases by one thousand dollars?

(b) Give a 95 percent confidence interval for the population slope coefficient.

(c) Give a 99 percent confidence interval for the population slope coefficient.

(d) The claim is made that the mean charge is not associated with the mean cost. Test this claim at significance level 0.05. State clearly the null and alternative hypotheses and your conclusion.

(e) The claim is made that the mean charge increases with the mean cost. Test this claim at significance level 0.05. State clearly the null and alternative hypotheses and your conclusion.

(f) The claim is made that the mean charge increases by $1 with each extra $1 of mean cost. Test this claim at significance level 0.01. State clearly the null and alternative hypotheses and your conclusion.
3. This question continues from the previous question. If the Stata output given at the end of the exam is insufficient to answer the question then say so.

(a) Give the predicted mean hospital charge when the mean cost is $20,000. Show any workings.

(b) Suppose we regressed `meancost` on an intercept and `meancharge`. What value do you expect for the slope coefficient? Explain your answer.

(c) Suppose we regressed `meancharge` on only an intercept. What value do you expect for the intercept coefficient? Explain your answer.

4. You are given the following information following regression of $y$ on an intercept and $x$:

Regression sum of squares = 40
Total sum of squares = 160
Number of observations = 10

(a) Give the $R^2$ for this regression. Show all workings.

(b) Give the correlation coefficient between $x$ and $y$. Show all workings.

(c) Give the standard error of the residual for this regression.

(d) This part unrelated to parts a-c
Following `regress y x`, what does the following Stata command produce?
`predict z, resid`
Multiple Choice Questions (1 point each)

1. If you regress monthly data for the Dow Jones Industrial average on the Standard and Poors 500 then the $R^2$ is
   
   a. between 0.00 and 0.25  
   b. between 0.25 and 0.50  
   c. between 0.50 and 0.75  
   d. between 0.75 and 1.00.

2. The sample correlation coefficient between $y$ and $x$ equals
   
   a. the slope coefficient from regression of $y$ on $x$  
   b. the slope coefficient from regression of $x$ on $y$  
   c. the slope coefficient from regression of $(y - \bar{y})/s_y$ on $(x - \bar{x})/s_x$  
   d. all of the above  
   e. none of the above.

3. Assumptions necessary for the least squares estimator to be unbiased for $\beta_2$ include
   
   a. $y = \beta_1 + \beta_2 x + u$  
   b. $\text{Var}[u|x] = \sigma_u^2$  
   c. neither a. nor b.  
   d. both a. and b.

4. If we rescale $y$ by multiplying by 10 then
   
   a. the correlation of $y$ and $x$ is 10 times larger  
   b. the covariance of $y$ and $x$ is 10 times larger  
   c. neither a. nor b.  
   d. both a. and b.

5. The width of a confidence interval for the slope coefficient
   
   a. increases with increasing confidence level  
   b. increases with increasing sample size  
   c. both a. and b.  
   d. neither a. nor b.
**SOME USEFUL FORMULAS**

### Univariate Data

\[ \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad \text{and} \quad s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \]

\[ \bar{x} \pm t_{\alpha/2,n-1} \times (s_x/\sqrt{n}) \quad \text{and} \quad t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \]

\[ \text{ttail}(df, t) = \Pr[T > t] \text{ where } T \sim t(df) \]

t_{\alpha/2} \text{ such that } \Pr[|T| > t_{\alpha/2}] = \alpha \text{ is calculated using invttail}(df, \alpha/2).\]

### Bivariate Data

\[ r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}} = \frac{s_{xy}}{\sqrt{s_{xx} \times s_{yy}}} \quad \text{[Here } s_{xx} = s_x^2 \text{ and } s_{yy} = s_y^2\text{]} \]

\[ \hat{y} = b_1 + b_2x \quad \text{with} \quad b_2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad \text{and} \quad b_1 = \bar{y} - b_2 \bar{x} \]

\[ \text{TSS} = \sum_{i=1}^{n} (y_i - \bar{y})^2 \quad \text{ErrorSS}=\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad \text{RegSS} = \text{TSS} - \text{ErrorSS} \]

\[ R^2 = 1 - \text{ErrorSS}/\text{TSS} \]

\[ b_2 \pm t_{\alpha/2,n-2} \times s_{b_2} \]

\[ t = \frac{b_2 - \beta_{20}}{s_{b_2}} \quad s_{b_2}^2 = \frac{s_e^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad s_e^2 = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]

\[ y|x = x^* \in \mu_1 + b_2x^* \pm t_{\alpha/2,n-2} \times s_e \times \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \times 1} \]

\[ \text{E}[y|x = x^*] \in \mu_1 + b_2x^* \pm t_{\alpha/2,n-2} \times s_e \times \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \]

### Multivariate Data

\[ \hat{y} = b_1 + b_2x_{2i} + \cdots + b_kx_{ki} \]

\[ R^2 = 1 - \text{ErrorSS}/\text{TSS} \quad \bar{R}^2 = R^2 - \frac{k-1}{n-k} (1 - R^2) \]

\[ b_j \pm t_{\alpha/2,n-k} \times s_{b_j} \quad \text{and} \quad t = \frac{b_j - \beta_{j0}}{s_{b_j}} \]

\[ F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \quad \text{and} \quad F = \frac{\text{SSE}_e - \text{SSE}_u}{\text{SSE}_u/(n-k)} \]

\[ \text{Ftail}(df_1, df_2, f) = \Pr[F > f] \text{ where } F \text{ is } F(df_1, df_2) \text{ distributed.} \]

\[ F_\alpha \text{ such that } \Pr[F > f_\alpha] = \alpha \text{ is calculated using invFtail}(df_1, df_2, \alpha). \]
. summarize meancharge meancost

<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>meancharge</td>
<td>169</td>
<td>47957.27</td>
<td>22248.97</td>
<td>14953</td>
<td>123131</td>
</tr>
<tr>
<td>meancost</td>
<td>169</td>
<td>21007.51</td>
<td>10376.75</td>
<td>7021</td>
<td>86730</td>
</tr>
</tbody>
</table>

. regress meancharge meancost

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 169</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>3.1277e+10</td>
<td>1</td>
<td>3.1277e+10</td>
<td>F( 1, 167) = 100.67</td>
</tr>
<tr>
<td>Residual</td>
<td>5.1886e+10</td>
<td>167</td>
<td>310694328</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>8.3163e+10</td>
<td>168</td>
<td>495016535</td>
<td>R-squared = 0.3761</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adj R-squared = 0.3724</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Root MSE = 17627</td>
</tr>
</tbody>
</table>

| meancharge | Coef.  | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|------------|--------|-----------|------|-----|----------------------|
| meancost   | 1.314908 | .1310541  | 10.03| 0.000| 1.056171 1.573644 |
| _cons      | 20334.33 | 3068.893  | 6.63 | 0.000| 14275.5 26393.15 |

. display "t_.005;167 = " invttail(167,.005) " t_.01;167 = " invttail(167,.01) \\n>    _n "t_.025;167 = " invttail(167,.025) " t_.05;167 = " invttail(167,.05) \\n>    " t_.10;167 = " invttail(167,.10) 
> t_.005;167 = 2.6055891 t_.01;167 = 2.3488842 
> t_.025;167 = 1.974271 t_.05;167 = 1.6540291 t_.10;167 = 1.2866415