Economics 102: Analysis of Economic Data
Cameron Spring 2015
Department of Economics, U.C.-Davis

Final Exam (A) Saturday June 6

Compulsory. Closed book. Total of 56 points and worth 45% of course grade.
Read question carefully so you answer the question.

**Question scores**

| Question | 1a | 1b | 1c | 2a | 2b | 2c | 2d | 2e | 2f | 2g | 2h | 2i | 2j | 2k | 3a | 3b | 3c | 3d | 3e | 3f | 3g |
|----------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Points   | 1  | 2  | 3  | 1  | 1  | 1  | 2  | 3  | 1  | 1  | 3  | 1  | 1  | 1  | 2  | 1  | 1  | 1  | 2  | 3  | 1  | 1  |

<table>
<thead>
<tr>
<th>Question</th>
<th>4a</th>
<th>4b</th>
<th>4c</th>
<th>5a</th>
<th>5b</th>
<th>5c</th>
<th>5d</th>
<th>5e</th>
<th>Mult Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Points</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>10</td>
</tr>
</tbody>
</table>

1. a b c d e
2. a b c d e
3. a b c d e
4. a b c d e
5. a b c d e
6. a b c d e
7. a b c d e
8. a b c d e
9. a b c d e
10. a b c d e
Questions 1-4
Consider data on price and size of diamond rings of commercial-grade for the mass market.

**Dependent Variable**
- \(\text{price} = \) price in dollars
- \(\ln\text{price} = \) Natural logarithm of price.

**Regressors**
- \(\text{size} = \) Size in tenths of a carat (one carat = 200 milligrams = 0.007055 ounces)
- \(\text{sizesq} = \) Square of size
- \(\ln\text{size} = \) Natural logarithm of size
- \(d_1 = 1\) if above-average quality and \(= 0\) otherwise
- \(d_2 = 1\) if average quality and \(= 0\) otherwise
- \(d_3 = 1\) if below-average quality and \(= 0\) otherwise.

Use the two pages of output provided at the end of this exam on:
- \(t\) critical values, summary statistics, correlations and regressions.

Part of the following questions involves deciding which output to use.
You can use the output that gets the correct answer in the quickest possible way.

1. (a) How would you use Stata to determine whether or not \(\text{price}\) is normally distributed?

(b) The prices are from one supplier. Now consider the population of diamond ring suppliers providing prices for mass market diamond rings with diamonds in the range 0.12 to 0.35 carats. Give a 95% confidence interval for the population mean price of diamond rings.

(c) Perform a test at significance level .05 of the claim that the population mean ring price exceeds $450.
State clearly the null and alternative hypotheses of your test, and your conclusion.
2. In this question the regression studied is a linear regression of price on size.

(a) According to the regression results, by how much does diamond ring price change if size increases by 0.1 carats (equals one unit of variable size).

(b) Are you surprised by the value of the intercept coefficient? **Explain.**

(c) Give a 95 percent confidence interval for the population slope parameter.

(d) Give a 99 percent confidence interval for the population slope parameter.

(e) Test the hypothesis at significance level 1% that the population slope coefficient is equal to 350. **State clearly** the null and alternative hypothesis in terms of population parameters and state your conclusion.
(f) Is it meaningful to use the results of this regression to estimate the effect on price of an increase of ten carats in the size of the diamond? Explain.

(g) Predict the conditional mean price for a 0.25 carat ring (so variable size = 2.5).

(h) Give a 95 percent confidence interval for the conditional mean price for a 0.25 carat ring. Give your answer as an expression involving numbers only. You need not complete all the calculations to get a final answer.

(i) The output for model LINHET was obtained by command `reg price size, vce(robust)` What does this command do?

(j) The original data had variable carats that equalled size in carats. Give the Stata command that created variable size from variable carats.

(k) Suppose we regress size on price rather than price on size? How well will the model fit? Explain.
3. In this question consider models **LINEAR**, **QUAD** and **DUMMIES**. In these models the dependent variable is **price** and the pairs of numbers given are the OLS coefficients and their standard errors.

(a) In model **QUAD** what is the marginal effect at the mean on ring price of increasing diamond size by 0.1 carats (equals one unit of variable **size**)?

(b) After controlling for the size of the diamond, what is the difference in price between a ring of average quality and a ring of below-average quality?

(c) After controlling for the size of the diamond, what is the difference in price between a ring of above-average quality and a ring of average quality?

(d) Are all the regressors in model **DUMMIES** jointly statistically significant at significance level 0.05? Perform an appropriate test. **State clearly** the null and alternative hypotheses of your test, and your conclusion.

(e) Are the indicator variables d1 and d2 in model **DUMMIES** jointly statistically significant at significance level 0.05? Perform an appropriate test. **State clearly** the null and alternative hypotheses of your test, and your conclusion.

(f) Do you see any problems in adding the variable d3 as a regressor in model **DUMMIES**? **Explain**.

(g) Using a measure of model fit that controls for the size of the model, which of the three models best explains the data? **Explain** your answer.
4. In this question consider models LOGLIN and LOGLOG. In these models the dependent variable is lnprice and the pairs of numbers given are the OLS coefficient and their standard errors.

(a) Provide a meaningful interpretation of the effect of variable size on price (not lnprice) in model LOGLIN.

(b) Provide a meaningful interpretation of the effect of variable size on price (not lnprice) in model LOGLOG.

(c) Suppose we use model LOGLIN. Do you see any problems in using \( \exp(4.7849 + 0.6787 \times \text{size}) \) to predict price. **Explain.**
5. This question has various unrelated parts (though (d) and (e) use the same information).

(a) Suppose $X = 1$ with probability 0.6, $X = 2$ with probability 0.3 and $X = 3$ with probability 0.1. What is the variance of $X$? **Show all workings.**

(b) Suppose for $X \sim (400, 30^2)$ we form 500 samples of size 100 and obtain 500 sample means $\bar{x}$. What approximately do you expect the average of the $\bar{x}$ to equal? **Explain.**
What approximately do you expect the standard deviation of the $\bar{x}$ to equal? **Explain.**

(c) State the four population model assumptions (not data assumptions) for bivariate OLS regression. (1 point for each correct assumption).

(d)-(e) You run the following code, where `runiform(-1,1)` draws random variables that have the uniform distribution on the interval -1 to 1 and have mean zero.

```-stata
clear
set seed 10101
program myprogram, rclass
    drop _all
    quietly set obs 1000
    generate x = runiform(-1,1)
    generate u = runiform(-1,1)
    generate y = 1 + 2*x + u
    regress y x
    return scalar mystery = _b[x]
end
simulate mystery=r(mystery), seed(10101) reps(500): myprogram
summarize mystery
histogram mystery
```

(d) What do you expect the sample mean of variable **mystery** to be? **Explain.**

(e) What distribution do you expect variable **mystery** to have? **Explain.**
Multiple choice questions (1 point each)

1. FRED is most useful
   a. for obtaining panel data
   b. for obtaining time-series data
   c. for obtaining cross-section data
   d. FRED is not a source of data.

2. The Stata command `generate y = x[n-1]` is used to create variable `y` equal to
   a. `x[n-1]`
   b. the previous observation of `x`
   c. a random variable between 1 and `n`
   d. none of the above.

3. Variable `x` increased by 10 percent. It follows that `ln x` increased by approximately
   a. `exp(10)`
   b. `exp(0.1)`
   c. 10
   d. 0.1
   e. none of the above.

4. Prices doubled over seven years. It follows that the annual inflation rate is
   a. less than 10 percent
   b. between 10 and 12 percent
   c. between 12 and 14 percent
   d. between 14 and 16 percent
   e. more than 16 percent.

5. What approximate value do you expect for the Stata function `ttail(200,1.0)`?
   a. more than 0.4
   b. between 0.4 and 0.3
   c. between 0.3 and 0.2
   d. between 0.2 and 0.1
   e. between 0.1 and 0.0.
6. For hypothesis testing
   a. test size is the probability of a type I error
   b. test power is one minus the probability of a type II error
   c. both a. and b.
   d. neither a. nor b.

7. Let \( \hat{y}_i = b_1x_{1i} + b_2x_{2i} + \cdots + b_kx_{ki} \). Then the explained sum of squares is
   a. \( \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \)
   b. \( \sum_{i=1}^{n} (y_i - \bar{y})^2 \)
   c. \( \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \)
   d. none of the above

8. For forecasting an actual value from a multiple regression from model that satisfies assumptions 1 to 4, the half-width of the 95 percent confidence interval for the forecast is approximately
   a. zero as the sample size goes to infinity
   b. two times the standard error of the regression as the sample size goes to infinity
   c. neither of the above

9. When an irrelevant variable is included in a regression
   a. OLS is unbiased and efficient
   b. OLS is unbiased and inefficient
   c. OLS is biased and efficient
   d. OLS is biased and inefficient

10. When a variable correlated with the error term is omitted from an equation
    a. OLS is unbiased and efficient
    b. OLS is unbiased and inefficient
    c. OLS is biased and efficient
    d. OLS is biased and inefficient
Univariate Data
\[ x = \frac{1}{n} \sum_{i=1}^{n} x_i \quad \text{and} \quad s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \]
\[ x \pm t_{\alpha/2, n-1} \times (s_x / \sqrt{n}) \quad \text{and} \quad t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} \]
\[ \text{ttail}(df, t) = \Pr[T > t] \text{ where } T \sim t(df) \]
\[ t_{\alpha/2} \text{ such that } \Pr[|T| > t_{\alpha/2}] = \alpha \text{ is calculated using invttail}(df, \alpha/2). \]

Bivariate Data
\[ r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \times \sum_{i=1}^{n} (y_i - \bar{y})^2}} = \frac{s_{xy}}{s_x \times s_y} \quad [\text{Here } s_{xx} = s_x^2 \text{ and } s_{yy} = s_y^2]. \]
\[ \hat{y} = b_1 + b_2 x_i \quad b_2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad b_1 = \bar{y} - b_2 \bar{x} \]
\[ \text{TSS} = \sum_{i=1}^{n} (y_i - \bar{y})^2 \quad \text{Residual SS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad \text{Explained SS} = \text{TSS} - \text{Residual SS} \]
\[ R^2 = 1 - \frac{\text{Residual SS}}{\text{TSS}} \]
\[ b_2 \pm t_{\alpha/2, n-2} \times s_{b_2} \quad t = \frac{b_2 - \beta_{20}}{s_{b_2}} \quad s_{b_2}^2 = \frac{s_e^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad s_e^2 = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]
\[ y|x = x^* \in b_1 + b_2 x^* \pm t_{\alpha/2, n-2} \times s_e \times \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} + 1} \]
\[ E[y|x = x^*] \in b_1 + b_2 x^* \pm t_{\alpha/2, n-2} \times s_e \times \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \]

Multiple Regression
\[ \hat{y} = b_1 + b_2 x_{2i} + \cdots + b_k x_{ki} \]
\[ R^2 = 1 - \frac{\text{Residual SS}}{\text{TSS}} \quad \bar{R}^2 = R^2 - \frac{k-1}{n-k}(1 - R^2) \]
\[ b_j \pm t_{\alpha/2, n-k} \times s_{b_j} \quad \text{and} \quad t = \frac{b_j - \beta_{j0}}{s_{b_j}} \]
\[ F = \frac{R^2/(k-1)}{(1-\bar{R}^2)/(n-k)} \sim F(k-1, n-k) \]
\[ F = \frac{(\text{ResSS}_r - \text{ResSS}_u)/(k-g)}{\text{ResSS}_u/(n-k)} \sim F(k-g, n-k) \]
\[ \text{Ftail}(df1, df2, f) = \Pr[F > f] \text{ where } F \text{ is } F(df1, df2) \text{ distributed.} \]
\[ F_{\alpha} \text{ such that } \Pr[F > f_{\alpha}] = \alpha \text{ is calculated using invFtail}(df1, df2, \alpha). \]
\[d_3\] 0.0341 0.0665 0.0421 0.0201 0.0633 -0.7868 -0.4402 1.0000

\[d_2\] -0.0485 -0.0530 -0.0496 -0.0524 -0.0445 -0.2078 1.0000

\[d_1\] -0.0038 -0.0360 -0.0118 -0.0141 -0.0383 1.0000

\[\ln(size)\] 0.9780 0.9852 0.9924 0.9710 1.0000

\[\text{sizesq}\] 0.9850 0.9519 0.9928 1.0000

\[\text{size}\] 0.9891 0.9519 0.9928 1.0000

\[\ln(price)\] 0.9845 1.0000

\[\text{price}\] 1.0000

\[\text{sum price lnprice size sizesq lnsize d1 d2 d3}\]

<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>price</td>
<td>48</td>
<td>500.0833</td>
<td>213.6428</td>
<td>223</td>
<td>1086</td>
</tr>
<tr>
<td>\ln(price)</td>
<td>48</td>
<td>6.134642</td>
<td>.3950927</td>
<td>5.407172</td>
<td>6.990256</td>
</tr>
<tr>
<td>size</td>
<td>48</td>
<td>2.041667</td>
<td>.5678752</td>
<td>1.2</td>
<td>3.5</td>
</tr>
<tr>
<td>sizesq</td>
<td>48</td>
<td>4.484167</td>
<td>2.632861</td>
<td>1.44</td>
<td>12.25</td>
</tr>
<tr>
<td>\ln(size)</td>
<td>48</td>
<td>.6792832</td>
<td>.2597902</td>
<td>.1823215</td>
<td>1.252763</td>
</tr>
</tbody>
</table>

| d1       | 48  | .2708333 | .4490929 | 0   | 1   |
| d2       | 48  | .1041667 | .3087093 | 0   | 1   |
| d3       | 48  | .625     | .4892461 | 0   | 1   |

\[\text{correlate price lnprice size sizesq lnsize d1 d2 d3}\]

(\obs=48)

<table>
<thead>
<tr>
<th></th>
<th>price</th>
<th>\ln(price)</th>
<th>size</th>
<th>sizesq</th>
<th>\ln(size)</th>
<th>d1</th>
<th>d2</th>
<th>d3</th>
</tr>
</thead>
<tbody>
<tr>
<td>price</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>\ln(price)</td>
<td>0.9845</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>size</td>
<td>0.9891</td>
<td>0.9755</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sizesq</td>
<td>0.9850</td>
<td>0.9519</td>
<td>0.9928</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>\ln(size)</td>
<td>0.9780</td>
<td>0.9852</td>
<td>0.9924</td>
<td>0.9710</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d1</td>
<td>-0.0038</td>
<td>-0.0360</td>
<td>-0.0118</td>
<td>0.0141</td>
<td>-0.0383</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d2</td>
<td>-0.0485</td>
<td>-0.0530</td>
<td>-0.0496</td>
<td>-0.0524</td>
<td>-0.0445</td>
<td>-0.2078</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>d3</td>
<td>0.0341</td>
<td>0.0665</td>
<td>0.0421</td>
<td>0.0201</td>
<td>0.0633</td>
<td>-0.7868</td>
<td>-0.4402</td>
<td>1.0000</td>
</tr>
</tbody>
</table>
. reg price size

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs</th>
<th>48</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>2098596</td>
<td>1</td>
<td>2098596</td>
<td>F(1, 46)</td>
<td>2069.99</td>
</tr>
<tr>
<td>Residual</td>
<td>46635.671</td>
<td>46</td>
<td>1013.81894</td>
<td>Prob &gt; F</td>
<td>0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>2145231.67</td>
<td>47</td>
<td>45643.227</td>
<td>R-squared</td>
<td>0.9783</td>
</tr>
</tbody>
</table>

Adj R-squared  = 0.9778
Root MSE      = 31.841

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs</th>
<th>48</th>
</tr>
</thead>
<tbody>
<tr>
<td>price</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>_cons</td>
<td>-259.6259</td>
<td>17.31866</td>
<td>-14.99</td>
<td>0.000</td>
<td>-294.487</td>
</tr>
</tbody>
</table>

. est table LINEAR LINHET QUAD DUMMIES LOGLIN LOGLOG, b(%10.3f) se stats(N F r2 r2_a rmse rss)

<table>
<thead>
<tr>
<th>Variable</th>
<th>LINEAR</th>
<th>LINHET</th>
<th>QUAD</th>
<th>DUMMIES</th>
<th>LOGLIN</th>
<th>LOGLOG</th>
</tr>
</thead>
<tbody>
<tr>
<td>size</td>
<td>372.102</td>
<td>372.102</td>
<td>292.013</td>
<td>372.182</td>
<td>0.679</td>
<td></td>
</tr>
<tr>
<td>sizesq</td>
<td>8.179</td>
<td>7.775</td>
<td>68.130</td>
<td>8.362</td>
<td>0.023</td>
<td></td>
</tr>
<tr>
<td>d1</td>
<td></td>
<td></td>
<td>3.982</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d2</td>
<td></td>
<td></td>
<td>1.552</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lnsize</td>
<td></td>
<td></td>
<td>1.498</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>_cons</td>
<td>-259.626</td>
<td>-259.626</td>
<td>-174.130</td>
<td>-261.027</td>
<td>4.749</td>
<td></td>
</tr>
<tr>
<td></td>
<td>17.319</td>
<td>15.856</td>
<td>74.238</td>
<td>18.219</td>
<td>0.048</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>48</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>2069.991</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>r2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.978</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>r2_a</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.978</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>rmse</th>
</tr>
</thead>
<tbody>
<tr>
<td>31.841</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>rss</th>
</tr>
</thead>
<tbody>
<tr>
<td>46635.671</td>
</tr>
</tbody>
</table>

legend: b/se