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	2	3	1	1

Question	4a	4b	4c	5a	5b	5c	5d	5e	Mult Choice
Points	1	1	1	2	2	4	1	1	10

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

`price` = price in dollars
`lnprice` = Natural logarithm of `price`.

Regressors

`size` = Size **in tenths of a carat** (one carat = 200 milligrams = 0.007055 ounces)
`sizesq` = Square of size
`lnsize` = Natural logarithm of `size`
`d1` = 1 if above-average quality and = 0 otherwise
`d2` = 1 if average quality and = 0 otherwise
`d3` = 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 `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 $\exp(4.7849 + 0.6787 \times \texttt{size})$ 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.

```
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 $\widehat{y}_i = b_1 x_{1i} + b_2 x_{2i} + \cdots + b_k x_{ki}$. Then the explained sum of squares is

a. $\sum_{i=1}^{n} (\widehat{y}_i - \bar{y})^2$

b. $\sum_{i=1}^{n} (y_i - \bar{y})^2$

c. $\sum_{i=1}^{n} (y_i - \widehat{y}_i)^2$

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

Cameron: Department of Economics, U.C.-Davis

SOME USEFUL FORMULAS

Univariate Data

$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ *and* $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

$\bar{x} \pm t_{\alpha/2;n-1} \times (s_x/\sqrt{n})$ *and* $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$

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

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

Bivariate Data

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

$\widehat{y} = b_1 + b_2 x_i$ $\quad b_2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ $\quad b_1 = \bar{y} - b_2\bar{x}$

$\text{TSS} = \sum_{i=1}^{n}(y_i - \bar{y}_i)^2$ $\quad \text{ResidualSS} = \sum_{i=1}^{n}(y_i - \widehat{y}_i)^2$ $\quad$ Explained SS = TSS - Residual SS

$R^2 = 1 -$ ResidualSS/TSS

$b_2 \pm t_{\alpha/2;n-2} \times s_{b_2}$

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

$y|x = x^* \in b_1 + b_2 x^* \pm t_{\alpha/2;n-2} \times s_e \times \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_i (x_i - \bar{x})^2} + 1}$

$\text{E}[y|x = x^*] \in b_1 + b_2 x^* \pm t_{\alpha/2;n-2} \times s_e \times \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}}$

Multiple Regression

$\widehat{y} = b_1 + b_2 x_{2i} + \cdots + b_k x_{ki}$

$R^2 = 1 -$ ResidualSS/TSS $\quad \bar{R}^2 = R^2 - \frac{k-1}{n-k}(1 - R^2)$

$b_j \pm t_{\alpha/2;n-k} \times s_{bj}$ *and* $t = \dfrac{b_j - \beta_{j0}}{s_{b_j}}$

$F = \dfrac{R^2/(k-1)}{(1 - R^2)/(n-k)} \sim F(k-1, n-k)$

$F = \dfrac{(\text{ResSS}_r - \text{ResSS}_u)/(k-g)}{\text{ResSS}_u/(n-k)} \sim F(k-g, n-k)$

$\text{Ftail}(df1, df2, f) = \Pr[F > f]$ where F is $F(df1, df2)$ distributed.

F_α such that $\Pr[F > f_\alpha] = \alpha$ is calculated using $\text{invFtail}(df1, df2, \alpha)$.

```
t_.05,v for v = 48     v = 47     v = 46     v = 45     v = 44     v = 43
            1.6772242  1.6779267  1.6786604  1.6794274  1.68023     1.6810707
t_.025,v for v = 48    v = 47     v = 46     v = 45     v = 45     v = 43
            2.0106348  2.0117405  2.0128956  2.0141034  2.0141034   2.0166922
t_.01,v for v = 48     v = 47     v = 46     v = 45     v = 45     v = 43
            2.4065813  2.4083451  2.4101881  2.4121159  2.4121159   2.4162501
t_.005,v for v = 48    v = 47     v = 46     v = 45     v = 45     v = 43
            2.682204   2.6845556  2.6870135  2.689585   2.689585    2.6951021

F_.05,v1,v2 for  v1,v2=2,48  v1,v2=2,47  v1,v2=2,46  v1,v2=2,45  v1,v2=2,44  v1,v2=2,43
                 3.1907273   3.1950563   3.1995817   3.2043173   3.209278    3.2144803
F_.05,v1,v2 for  v1,v2=3,48  v1,v2=3,47  v1,v2=3,46  v1,v2=3,45  v1,v2=3,44  v1,v2=3,43
                 2.7980606   2.8023552   2.8068449   2.8115435   2.8164658   2.8216282
```

. sum price lnprice size sizesq lnsize d1 d2 d3

Variable	Obs	Mean	Std. Dev.	Min	Max
price	48	500.0833	213.6428	223	1086
lnprice	48	6.134642	.3950927	5.407172	6.990256
size	48	2.041667	.5678752	1.2	3.5
sizesq	48	4.484167	2.632861	1.44	12.25
lnsize	48	.6792832	.2597902	.1823215	1.252763
d1	48	.2708333	.4490929	0	1
d2	48	.1041667	.3087093	0	1
d3	48	.625	.4892461	0	1

. correlate price lnprice size sizesq lnsize d1 d2 d3
(obs=48)

	price	lnprice	size	sizesq	lnsize	d1	d2	d3
price	1.0000							
lnprice	0.9845	1.0000						
size	0.9891	0.9755	1.0000					
sizesq	0.9850	0.9519	0.9928	1.0000				
lnsize	0.9780	0.9852	0.9924	0.9710	1.0000			
d1	-0.0038	-0.0360	-0.0118	0.0141	-0.0383	1.0000		
d2	-0.0485	-0.0530	-0.0496	-0.0524	-0.0445	-0.2078	1.0000	
d3	0.0341	0.0665	0.0421	0.0201	0.0633	-0.7868	-0.4402	1.0000

```
. reg price size

      Source |       SS       df       MS              Number of obs =      48
-------------+------------------------------           F(1, 46)      = 2069.99
       Model |     2098596     1     2098596           Prob > F      =  0.0000
    Residual |  46635.6711    46  1013.81894           R-squared     =  0.9783
-------------+------------------------------           Adj R-squared =  0.9778
       Total |  2145231.67    47   45643.227           Root MSE      =  31.841

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        size |   372.1025   8.178589    45.50   0.000     355.6398    388.5651
       _cons |  -259.6259   17.31886   -14.99   0.000     -294.487   -224.7649
------------------------------------------------------------------------------

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

------------------------------------------------------------------------------------------
    Variable |   LINEAR       LINHET        QUAD        DUMMIES       LOGLIN       LOGLOG
-------------+----------------------------------------------------------------------------
        size |    372.102      372.102      292.013      372.182        0.679
             |      8.179        7.775       68.130        8.362        0.023
      sizesq |                               17.399
             |                               14.695
          d1 |                                             3.982
             |                                            10.796
          d2 |                                             1.552
             |                                            15.724
      lnsize |                                                                       1.498
             |                                                                       0.038
       _cons |   -259.626     -259.626     -174.130     -261.027        4.749        5.117
             |     17.319       15.856       74.238       18.219        0.048        0.028
-------------+----------------------------------------------------------------------------
           N |         48           48           48           48           48           48
           F |   2069.991     2290.555     1044.740      662.090      906.175     1515.544
          r2 |      0.978        0.978        0.979        0.978        0.952        0.971
        r2_a |      0.978        0.978        0.978        0.977        0.951        0.970
        rmse |     31.841       31.841       31.702       32.506        0.088        0.069
         rss |  46635.671    46635.671    45226.677    46491.431        0.354        0.216
------------------------------------------------------------------------------------------
                                                                            legend: b/se

.
. di "F is overall F-test      r2 is R-suared     r2_a is adjusted R-squared"  _n ///
>   "rmse is Root MSE     rss is Residual sum of squares"
F is overall F-test      r2 is R-suared     r2_a is adjusted R-squared
rmse is Root MSE     rss is Residual sum of squares
```