Economics 102_A: Analysis of Economic Data
Cameron Fall 22 Department of Economics, U.C.-Davis

Second Midterm Exam (Nov 3) Version A

Compulsory. Closed book. Total of 35 points and worth 20% of course grade.
Read question carefully so you answer the question.
For computations final answers should be to at least four significant digits.

Question scores

Question	1a	1b	1c	1d	1e	2a	2b	2c	2d	2e	2e	3a	3b	3c	4a	4b	4c	4d	Mult.choice
Points	2	1	1	4	1	1	1	3	2	1	3	1	1	1	2	2	1	2	5

1.(a) Suppose we have a sample with three observations with (x, y) equal to (1,5), (2,2) and (3,2). Calculate $\sum_{i=1}^{3}(x_i - \bar{x})(y_i - \bar{y})$. Show all your calculations.

(b) Variables x and y have variances of, respectively, 100 and 25, and their covariance is 5. What is the sample correlation between the two variables?

(c) Continuing with the same data as in part (b). What is the R-squared from regression of y on an intercept and x.

(d) Provide the four population assumptions used for the linear regression model.
(1 point per correct assumption).

(e) Which of the assumptions given in part (d) are necessary for the OLS estimates to be unbiased?

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.

The sample is one of 51 states (including District of Columbia as a separate "state").
The data from the Statistical Abstract of the United States are:

`ENERGY` = State Residential energy consumption in millions of BTU in 2000
(a British Thermal Unit is a measure of energy)

`INCOME` = State Personal income in billions of dollars in 2000.

2.(a) According to the regression output, when state personal income increases by $1,000,000,000, by how much does state residential energy consumption change? [Hint: pay attention to the units of measurement].

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

(c) Give a **90 percent** confidence interval for the population slope coefficient.

(d) Perform a two-sided test **at significance level .05** of whether or not there is a relationship between energy use and income. **State clearly the null and alternative hypotheses of your test, and your conclusion.**

(e) Give the exact Stata command, including any relevant numbers, that will compute the p-value of your test in part (d).

(f) According to a recent paper on energy efficiency use, across European Union countries a billion dollar increase in personal income in a country is associated with more than a one million BTU increase in residential energy consumption. Test this claim using an appropriate statistical test **at significance level .01. State clearly the null and alternative hypotheses of your test, and your conclusion.**

3. Question 3 uses data from the same page of output.

(a) How well does state income explain variability in state residential energy consumption? Explain your answer.

(b) What is the predicted residential energy consumption for the state of Alaska which had income of 18.8 billion dollars.

(c) What quantity does the entry `root MSE = 122.77` measure?

4.(a) We wish to test $H_0 : \theta = 10$ against $H_a : \theta \neq 10$. Based on a very large sample we obtain estimate $\widehat{\theta} = 19$ with standard error of $\widehat{\theta} = 5$. Do we reject H_0 at significance level 5%? **Explain.**

(b) Provide a 95% confidence interval for θ.

(c) Regression of y on an intercept and x yields residual sum of squares equal to 100 and explained sum of squares equal to 50. Compute the value of R^2.

(d) You are given the following Stata code.

```
clear
quietly set obs 120
generate a = rnormal(2,8)
generate b = rnormal(0,16)
generate c = 3 + 4*a + b
regress c a
```

Suppose you run this code 500 times leading to 500 estimates of the slope coefficients. What do you expect the average of these 500 slopes to approximately equal?

Multiple Choice Questions (1 point each)

1. The OLS estimator

a. minimizes $\sum_{i=1}^{n}(y_i - b_1 - b_2 x_i)^2$

b. minimizes $\sum_{i=1}^{n}(y_i - \beta_1 - \beta_2 x_i)^2$

c. both a. and b.

d. neither a. nor b.

2. Which of the following statements is true

a. The sample correlation between x and y equals that between y and x

b. The sample correlation between x and y lies between 0 and 1

c. neither a. nor b.

d. both a. and b.

3. We regress y on an intercept and x and find that $b_2 = 10$ with $R^2 = 0.5$. Then we expect that

a. regression of x on an intercept and y has a slope coefficient 0.10

b. regression of x on an intercept and y has $R^2 = 0.5$

c. both a. and b.

d. neither a. nor b.

4. To fit a nonparametric regression curve to the relationship between y and x we can use Stata command

a. `lowess`

b. `lpoly`

c. neither a. nor b.

d. both a. and b.

5. Suppose the population model is $y = 2 + 4x + u$ and $E[u|x] = 0$ and the sample leads to fitted regression line $\widehat{y} = 3 + 4x$. Then for observation $(x, y) = (1, 8)$

a. the error term is 2

b. the fitted value is 7

c. both a. and b.

d. neither a. nor b.

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

SOME USEFUL FORMULAS FOR EXAMS

Univariate Data

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

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

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

$t_{\alpha/2}$ such that $\Pr[|T_{df}| > t_{\alpha/2}] = \alpha$ is calculated using 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 $s_{yy} = s_y^2$].

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

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

$R^2 = 1-$ ResidualSS/TSS

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

$t = \dfrac{b_2 - \beta_2}{s_{b_2}}$ $s_{b_2}^2 = \dfrac{s_e^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$ $s_e^2 = \frac{1}{n-2}\sum_{i=1}^n (y_i - \widehat{y}_i)^2$

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

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

Multivariate Data

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

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

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

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

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

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 invFtail$(df1, df2, \alpha)$.

```
. summarize ENERGY INCOME

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      ENERGY |         51     356.451    333.8133         24       1333
      INCOME |         51    164.6765    197.9608       13.8     1100.7

.
. regress ENERGY INCOME

      Source |       SS           df       MS      Number of obs   =        51
-------------+----------------------------------   F(1, 49)        =    320.62
       Model |  4832957.99         1  4832957.99   Prob > F        =    0.0000
    Residual |  738608.642        49  15073.6457   R-squared       =    0.8674
-------------+----------------------------------   Adj R-squared   =    0.8647
       Total |  5571566.63        50  111431.333   Root MSE        =    122.77

------------------------------------------------------------------------------
      ENERGY |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      INCOME |   1.570516   .0877092    17.91   0.000     1.394258    1.746774
       _cons |   97.82392   22.45396     4.36   0.000       52.701    142.9468
------------------------------------------------------------------------------

.
. di _n "KEY CRITICAL VALUES FOR THIS EXAM" _n _n       ///
>       "t_" dof ",.005 = " %5.3f invttail(dof,.005) _n ///
>       "t_" dof ",.01  = " %5.3f invttail(dof,.01) _n  ///
>       "t_" dof ",.025 = " %5.3f invttail(dof,.025) _n ///
>       "t_" dof ",.05  = " %5.3f invttail(dof,.05) _n  ///
>       "t_" dof ",.10  = " %5.3f invttail(dof,.10) _n

KEY CRITICAL VALUES FOR THIS EXAM

t_49,.005 = 2.680
t_49,.01  = 2.405
t_49,.025 = 2.010
t_49,.05  = 1.677
t_49,.10  = 1.299
```