Economics 102_A: Analysis of Economic Data
Cameron Winter 2019 Department of Economics, U.C.-Davis

Second Midterm Exam (Feb 21) Version A

Compulsory. Closed book. Total of 35 points and worth 22.5% 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	1*a*	1*b*	1*c*	1*d*	1*e*	2*a*	2*b*	2*c*	2*d*	2*e*	2*f*	3*a*	3*b*	3*c*	3*d*	4*a*	4*b*	4*c*	*Mult.choice*
Points	1	1	2	4	1	1	1	3	2	1	3	1	1	1	2	2	2	1	5

1.(a) Variables x and y have variances of, respectively, 25 and 16, and their covariance is 4. What is the sample correlation between the two variables?

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

(c) Suppose we have a sample with three observations with (x, y) equal to (1,4), (2,1) and (3,1). Calculate the least squares slope estimate. Show all your calculations.

(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 data are annual data (fiscal year) from 1984-85 to 2007-2008 from the California Legislative Analysts Office on California state government spending, specifically:

`HigherEd` = govt. spending on Higher Education as a percentage of total govt. spending.
`CrimJustice` = govt. spending on Criminal Justice as a percentage of total govt. spending.

2.(a) When state spending on criminal justice as a share of total government spending increases by one percentage point, what is the change in state spending on higher education as a share of total government spending?

(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 changes in state spending on criminal justice are associated with changes in state spending on higher education (as a fraction of total government spending). Test this claim using an appropriate statistical test at **significance level .05. 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) The claim is made that increased state spending on criminal justice comes entirely at the expense of spending on higher education. Specifically a one percentage point increase in spending on criminal justice is associated with a one percentage point decrease in state spending on higher education. Test this claim using an appropriate statistical test at **significance level .05. State clearly the null and alternative hypotheses of your test, and your conclusion.**

3. Continue with the Stata output.

(a) Does change in spending on criminal justice necessarily cause spending on higher education? **Explain your answer.**

(b) Give the standard error of the regression.

(c) Provide a prediction of `HigherEd` when `CrimJustice` equals 12.68.

(d) Give the Stata command that will produce an (X,Y) scatterplot for these data with a regression line superimposed.

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) Now suppose that for the part (a) example the claim is made that θ exceeds 10. Perform a test of this claim at 5%, stating clearly the null and alternative hypotheses and your conclusion.

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

Multiple Choice Questions (1 point each)

1. Suppose based on a sample of 100 observations we obtain an estimate of 500 with standard error of the estimate equal to 100. Then a 95% confidence interval for the parameter of interest is approximately

a. (400, 600)

b. (300, 700)

c. (490, 510)

d. (480, 520)

e. none of the above.

2. The OLS estimator

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

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

c. both a. and b.

d. neither a. nor b.

3. We regress y on x and find that $b_2 = 10$ with standard error 2. Given only this information

a. the regressor x is highly statistically significant and highly economically significant

b. the regressor x is highly statistically significant

c. the regressor x is highly economically significant

d. none of the above.

4. The standard error of the regression

a. is normally distributed

b. is the estimated standard deviation of the slope coefficient

c. both a. and b.

d. neither a. nor b.

5. The standard error of the slope coefficient

a. increases with increase in the sample size

b. decreases with increase in the variability of the regressors

c. both a. and b.

d. neither a. nor b.

SOME USEFUL FORMULAS FOR EXAMS

Univariate Data

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

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

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

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \times \sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{s_{xy}}{s_x \times s_y}$$ [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

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

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

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

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

Multivariate Data

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

$R^2 = 1-$ 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* $t = \dfrac{b_j - \beta_{j0}}{s_{b_j}}$

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

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

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

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

```
. summarize HigherEd CrimJustice

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
    HigherEd |         24    12.94958    1.555748      11.03      15.86
 CrimJustice |         24    9.447083    1.915694       5.01      12.68

.
. regress HigherEd CrimJustice

      Source |       SS           df       MS      Number of obs   =        24
-------------+----------------------------------   F(1, 22)        =    135.03
       Model |  47.8687887         1  47.8687887   Prob > F        =    0.0000
    Residual |  7.79930545        22  .354513884   R-squared       =    0.8599
-------------+----------------------------------   Adj R-squared   =    0.8535
       Total |  55.6680941        23  2.42035192   Root MSE        =    .59541

------------------------------------------------------------------------------
    HigherEd |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 CrimJustice |  -.7530715   .0648077   -11.62   0.000    -.8874745   -.6186686
       _cons |   20.06391   .6241906    32.14   0.000     18.76942     21.3584
------------------------------------------------------------------------------

.
. 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_22,.005 = 2.819
t_22,.01  = 2.508
t_22,.025 = 2.074
t_22,.05  = 1.717
t_22,.10  = 1.321
```