Economics 102: Analysis of Economic Data
Cameron Spring 2016 May 12
Department of Economics, U.C.-Davis

Second Midterm Exam (Version A)

Compulsory. Closed book. Total of 30 points and worth 22.5% of course grade.
Read question carefully so you answer the question.
You are to use only **simple calculations (+, -, /, *, square root) and show all workings**.
Do not use calculators with graphical or statistical analysis capabilities.
For computations final answers should be to at least four significant digits.
You may remove the formula sheet and the Stata output sheet(s) at end of exam.

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*	3*e*	3*f*	*Mult. choice*
Points	4	1	1	1	1	1	1	2	2	3	1	1	2	1	1	1	1	5

1.(a) List the four population assumptions for the linear regression model.
(1 point per correct assumption).

(b) Failure of which of assumptions 1-4 will warrant using Stata command `regress y x, vce(robust)` rather than the simpler command `regress y x`?

(c) Under assumptions 1-4 the OLS estimator for β_2 is best linear unbiased. In what sense is the term "best" being used here. A simple answer will do (there is no need for any algebra).

(d) You regress the z-scores for variable y on the z-scores for variable x and obtain a slope estimate of 0.6. Provide a simple interpretation of this slope coefficient estimate.

(e) Provide the general definition of a type 1 error.

QUESTION 2 USES 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 for 200 regional markets

`sales` = sales in units

`newspaper` = newspaper advertising in thousands of dollars

Note: pay attention to the units of measurement.

2.(a) How do sales change when newspaper advertising expenditure increases by one thousand dollars? (Note: 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) The claim is made that sales are not associated with newspaper advertising. Test this claim at significance **level 0.05**. **State clearly the null and alternative hypotheses and your conclusion.**

(e) The claim is made that sales change by at least 30 units when advertising expenditure increases by one thousand dollars? (Note: pay attention to the units of measurement). Test this claim at significance **level 0.05**. **State clearly the null and alternative hypotheses and your conclusion.**

(f) Suppose $100,000 is spent on newspaper advertising. What level of sales do we expect?

3. You are given the following, where $\widehat{y}_i$ denote fitted values after OLS regression.

$$\begin{aligned}
\sum_{i=1}^{10}(x_i-\bar{x})^2 &= 10\\
\sum_{i=1}^{10}(x_i-\bar{x})(y_i-\bar{y}) &= 20\\
\sum_{i=1}^{10}(y_i-\bar{y})^2 &= 90\\
\sum_{i=1}^{10}(y_i-\widehat{y}_i)^2 &= 50\\
\bar{x} &= 1\\
\bar{y} &= 20
\end{aligned}$$

(a) Calculate the sample variance of y.

(b) Calculate the OLS intercept and slope coefficients.

(c) Calculate the standard error of the regression.

(d) Calculate the R-squared of the regression.

(e) Calculate the correlation coefficient between x and y.

(f) This part is unrelated to the preceding.
Suppose 1,000 times we do the following.

```
quietly set obs 100
generate x = rnormal(5,1)
generate u = rnormal(0,4)
generate y = 1 + 3*x + u
regress y x
```

What value do you expect for the average of the 1,000 slope estimates obtained? **Explain your answer.**

Multiple Choice Questions (1 point each)

1. If the sample covariance is positive then

a. the sample correlation coefficient is necessarily positive

b. the sample correlation coefficient is most likely positive but could be negative

c. the sample correlation coefficient could easily be positive or negative.

2. The standard error of the regression is a measure of

a. the standard deviation of the slope coefficient

b. the standard deviation of the intercept coefficient

c. the standard deviation of the dependent variable

d. the standard deviation of the error

e. none of the above.

3. Regression of y on x yields slope coefficient 0.50 and correlation coefficient 0.40. It follows that regression of x on y using the same data yields

a. slope coefficient 2.0

b. correlation coefficient 0.40

c. both a. and b.

d. neither a. nor b.

4. In order for $(b_2 - \beta_2)/se(b_2)$ to be exactly $T(n-2)$ distributed it is necessary that

a. assumptions 1-4 hold

b. assumptions 1-4 hold and the error term is normally distributed

c. assumptions 1-4 hold and the error term is $T(n-2)$ distributed.

5. Compare predicting the conditional mean of y given $x = x^*$ to predicting the actual value of y given x^*. In both cases we use prediction $\widehat{y} = b_1 + b_2 x^*$. Then a 95% confidence interval for the conditional mean of y given $x = x^*$ is

a. narrower than a 95% confidence interval for the actual value of y given $x = x^*$

b. wider than a 95% confidence interval for the actual value of y given $x = x^*$

c. possibly wider or narrower, depending on the value of x^*.

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* $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$

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 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

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

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

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

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

Multivariate Data

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

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

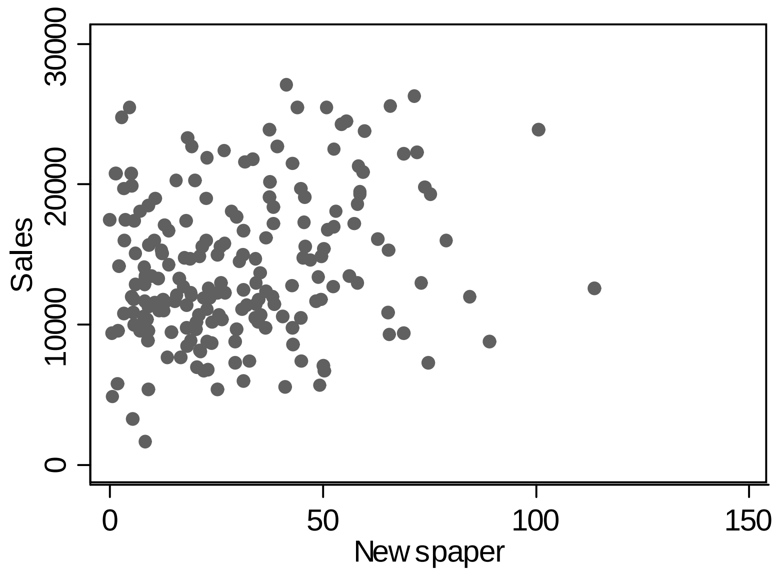
$b_j \pm t_{\alpha/2;n-k} \times s_{bj}$ *and* $t = \dfrac{b_j - \beta_{j0}}{s_{b_j}}$

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

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

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$(df1, df2, \alpha)$.




```
. sum sales newspaper

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       sales |        200     14022.5    5217.457       1600      27000
   newspaper |        200      30.554    21.77862         .3        114

. regress sales newspaper

      Source |       SS           df       MS      Number of obs   =       200
-------------+----------------------------------   F(1, 198)       =     10.89
       Model |   282344204         1   282344204   Prob > F        =    0.0011
    Residual |  5.1348e+09       198  25933356.3   R-squared       =    0.0521
-------------+----------------------------------   Adj R-squared   =    0.0473
       Total |  5.4171e+09       199    27221853   Root MSE        =    5092.5

------------------------------------------------------------------------------
       sales |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   newspaper |    54.6931   16.57572     3.30   0.001     22.00548    87.38071
       _cons |   12351.41   621.4202    19.88   0.000     11125.96    13576.86
------------------------------------------------------------------------------

. display _n "t_198,.005 = " invttail(198,.005) _n " t_198,.01 = " invttail(198,.01) ///
>  _n "t_198,.025 = " invttail(198,.025) _n  " t_198,.05 = " invttail(198,.05)  ///
>  _n " t_198,.10 = " invttail(198,.10) _n

t_198,.005 = 2.6008873
 t_198,.01 = 2.3453283
t_198,.025 = 1.9720175
 t_198,.05 = 1.6525858
 t_198,.10 = 1.2858418
```