Economics 102_A: Analysis of Economic Data
Cameron Fall 2021
Department of Economics, U.C.-Davis

Final Exam (A) Monday December 6

Compulsory. Closed book. Total of 60 points and worth 40% of course grade.
Read question carefully so you answer the question.

Question scores

Question	1a	1b	1c	1d	1e	1f	1g	1h	2a	2b	2c	2d	2e	3a	3b	3c	3d	3e	3f	3g
Points	1	1	2	2	1	1	1	1	2	2	3	1	2	1	1	1	2	2	1	2

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

Consider data on various characteristics of different models and makes of gasoline-fuelled automobiles sold in the U.S. in 1995.

Dependent Variable

`mpg` = Miles per gallon (a measure of fuel consumption).
`lmpg` = Natural logarithm of `mpg`.

Regressors

`curbwt` = Curb weight (weight of car in pounds with full fuel tank but no passengers or cargo)
`hp` = Horsepower (a measure of engine power)
`torque` = Engine torque (in foot pounds)
`accel` = Acceleration of car (seconds to reach 60 miles per hour)
`lcurbwt` = Natural logarithm of `curbwt`
`lhp` = Natural logarithm of `hp`
`ltorque` = Natural logarithm of `torque`
`dUS` = 1 if U.S. company and = 0 otherwise
`dASIA` = 1 if Asian company and = 0 otherwise
`dEUR` = 1 if European company and = 0 otherwise
Note: `dUS`+`dASIA`+`dEUR` = 1 for all observations.

Use the two pages of output provided at the end of this exam on:

1. Various t critical values.
2. Various descriptive statistics output and correlations for all variables.
3. Three 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) What Stata command will yield a smoothed histogram of `hp`?

(b) What option of the Stata `summarize` command will give the symmetry statistic.

(c) Suppose variable `hp` is approximately normally distributed. What range of values of `hp` do you expect 95% of the observations to fall in? Explain.

(d) Perform a test at significance level .05 of the claim that the population mean horsepower equals 150.
State clearly the null and alternative hypotheses of your test, and your conclusion.

(e) Give the complete Stata command that will give the p-value for your test statistic computed in part (d).

(f) For the test in part (c) give the probability of making a type 1 error.

(g) If we regressed `hp` on just one of `curbwt`, `torque` and `accel`, which of these variables would best explain `hp`? Explain.

(h) What would R^2 be if `hp` was regressed on `curbwt`? Explain.

2. In this question the regression studied is a linear regression of `mpg` on `hp`.

(a) According to the regression results, by how much does fuel consumption change if horsepower increases by one standard deviation of `hp`?

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

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

(d) Test the hypothesis at significance level 5% that the population slope coefficient is equal to $-.05$.
State clearly the null and alternative hypothesis in terms of population parameters and state your conclusion.

(e) Predict the conditional mean miles per gallon of a car with sample average horsepower.

3. In this question consider the second regression with `mpg` the dependent variable.

(a) Provide an interpretation of the coefficient of variable `hp`.

(b) Provide an interpretation of the coefficient of variable `dEUR`.

(c) Should the regression also include variable `dUS`? Explain.

(d) Which regressors are individually statistically significant at 5 percent?

(e) Are the regressors jointly statistically significant at 5 percent in the multivariate regression? State clearly the null and alternative hypotheses of your test, and your conclusion.

(f) Suppose you wanted to test whether the regressors other than `hp` are jointly statistically significant at 5%. Give the complete Stata command.

(g) Which model explains the data better on the basis of goodness of fit, using a measure that includes a penalty for model size. This model, or the model with just hp as a regressor? Explain.

4. In this question consider the regression where `lmpg` is the dependent variable.

(a) What is the impact of acceleration on level of miles per gallon (not log of `mpg`)?

(b) Provide a meaningful interpretation of the estimated coefficient for `lcurbwt`.

(c) The cars in this sample are made by 24 distinct manufacturers (Audi, Ford, ..., Volvo). So the sample includes multiple cars made by the same manufacturer. Do you see any problems in using this regression and, if so, how would you adjust the analysis.

(d) Do you see any problems in using this regression to predict population mean miles per gallon? Explain.

(e) Do you see any problems in using this regression if the model error is correlated with variable `lhp`? Explain.

(f) Do you see any problems in using this regression if `accel` is correlated with variable `lmpg`? Explain.

(g) Give the complete Stata command that creates variable `ltorque` from variable `torque`.

5. This question has various unrelated parts.

(a) Suppose $X = 10$ with probability 0.7 and $X = 20$ with probability 0.3. What is the variance of X? **Show all workings.**

(b) The natural logarithm of GDP in Argentina (measured in US dollars at official exchange rates) was 25.4 in 2002 and 26.9 in 2012. What was the approximate annual growth rate in Argentinian GDP over this period?
You are to answer this question in the simplest way possible.

(c) An investment receives a return of 4% per year. Approximately how many years will it take for the investment to double.

(d) Suppose in a sample the average hourly wage for men equals 24 and the average hourly wage for women is 21. We regress `wage` (hourly wage) on an intercept and an indicator variable `gender` (equal to one if male and equal to zero if female) yielding estimate $\widehat{\texttt{wage}} = b_1 + b_2 \times \texttt{gender}$.
Give the values for b_1 and b_2.

(e) Suppose we know that $y = 10 + 3x + u$ where $E[u|x] = 0$. A sample gives estimates $\widehat{y} = 8 + 5x$. Consider the observation $(x, y) = (3, 30)$. **Show any workings.**

(i) What is the conditional mean for this observation?

(ii) What is the value of the error for this observation?

(iii) What is the predicted value for this observation?

(iv) What is the value of the residual for this observation?

Multiple choice questions (1 point each)

1. A useful visual tool for finding outlying observations on a variable is

a. kernel density estimate

b. pie chart

c. box chart

d. line chart

2. A key feature of statistical inference on the mean is that

a. $\overline{X}$ is treated as fixed and μ is treated as a random variable

b. $\overline{X}$ is treated as fixed and μ is treated as fixed

c. $\overline{X}$ is treated as a random variable and μ is treated as a random variable

d. $\overline{X}$ is treated as a random variable and μ is treated as fixed

3. A type 1 error of a statistical test of H_0 against H_a occurs if we

a. reject H_0 given H_0 true

b. reject H_0 given H_a true

c. do not reject H_0 given H_0 true

d. do not reject H_0 given H_a true.

4. The Stata command `use mydata` is used to

a. read in data from a text spreadsheet file `mydata.csv`

b. read in data from an Excel formatted spreadsheet file `mydata.xlsx`

c. read in data from a Stata dataset `mydata.dta`

d. none of the above.

5. The Stata command `mean y` and the Stata command `regress y`

a. give the same estimate

b. give the same standard error

c. both a. and b.

d. neither a. nor b.

6. Suppose $\sum_{i=1}^{n}(x_i - \bar{x})^2 = 4$, $\sum_{i=1}^{n}(y_i - \bar{y})^2 = 10$ and $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = 20$. Then the slope coefficient from OLS regression equals

a. 0.5

b. 2

c. 2.5

d. 5

e. none of the above.

7. Suppose we obtain model estimates $\widehat{y}_i = 2 + 3x_i + x_i^2 + 5z_i$ where $\bar{x} = 2$ and $\bar{z} = 3$. Then the marginal effect of a change in x evaluated at the mean is

a. less than 4

b. 4 or more but less than 6

c. 6 or more but less than 8

d. 8 or more.

8. A linear regression of the academic performance index (y) for California high schools on various regressors finds that the most important explanators are

a. socioeconomic background measures such as fraction of students eligible for free meals

b. average educational attainment of parents

c. teacher quality measures such as whether they have teaching credentials

d. none of the above.

9. Suppose we obtain estimates $\widehat{y}_i = 2 + 3x_i$ and want to both (1) predict $E[y|x = 20]$ and (2) forecast $y|x = 20$. Then

a. we use the same estimate for both the prediction and the forecast

b. we use the same standard error for both the prediction and the forecast

c. both a. and b.

d. neither a. nor b.

10. Multicollinearity is a problem that arises when

a. the dependent variable is highly correlated with the regressors

b. the error is highly correlated with the regressors

c. the error is highly correlated with the dependent variable

d. the regressors are highly correlated with each other

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

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

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

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

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

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

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* $t = \dfrac{b_j - \beta_j^*}{s_{b_j}}$

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

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

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

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

```
t_v,.05 for  v = 94  v = 93   v = 92   v = 91   v = 90   v = 89   v = 88
             1.661   1.661    1.662    1.662    1.662    1.662    1.662
t_v,.025 for v = 94  v = 93   v = 92   v = 91   v = 90   v = 89   v = 88
             1.986   1.986    1.986    1.986    1.987    1.987    1.987
t_v,.005 for v = 94  v = 93   v = 92   v = 91   v = 90   v = 89   v = 88
             2.629   2.630    2.630    2.631    2.632    2.632    2.633

. summarize mpg hp curbwt torque accel lhp lmpg lcurbwt dUS dEUR dASIA

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
         mpg |         95    28.80105    5.055616       17.3       38.6
          hp |         95    161.6842    57.68573         70        389
      curbwt |         95    3035.537    489.6085       2075       4463
      torque |         95    229.4442    86.67477      100.3      569.5
       accel |         95    9.877033    1.512222    6.27578   13.65402
-------------+---------------------------------------------------------
         lhp |         95    5.030046    .3280396   4.248495   5.963579
        lmpg |         95    3.345303    .1747336   2.850706   3.653252
     lcurbwt |         95    8.005289    .1615053   7.637716   8.403577
         dUS |         95    .4736842    .5019559          0          1
        dEUR |         95    .1052632    .3085203          0          1
-------------+---------------------------------------------------------
       dASIA |         95    .4210526    .4963472          0          1

. summarize mpg, detail

                      Miles per gallon
-------------------------------------------------------------
      Percentiles      Smallest
 1%         17.3           17.3
 5%         22.8           20.3
10%           23           21.2       Obs                  95
25%         24.4           21.9       Sum of wgt.          95

50%         28.1                      Mean           28.80105
                        Largest       Std. dev.      5.055616
75%         32.2           38.2
90%           37           38.5       Variance       25.55925
95%         37.9           38.6       Skewness       .3684942
99%         38.6           38.6       Kurtosis       2.280529

. correlate mpg hp curbwt torque accel
(obs=95)

             |      mpg       hp   curbwt   torque    accel
-------------+---------------------------------------------
         mpg |   1.0000
          hp |  -0.8085   1.0000
      curbwt |  -0.8960   0.8354   1.0000
      torque |  -0.8131   0.9479   0.8795   1.0000
       accel |   0.6491  -0.9017  -0.6065  -0.7843   1.0000
```

```
. * Regressions for exam
. regress mpg hp, vce(robust)

Linear regression                               Number of obs     =         95
                                                F(1, 93)          =     137.02
                                                Prob > F          =     0.0000
                                                R-squared         =     0.6538
                                                Root MSE          =     2.9908

------------------------------------------------------------------------------
             |               Robust
         mpg | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
          hp |  -.0708618   .0060537   -11.71   0.000    -.0828833   -.0588404
       _cons |   40.25829   1.075757    37.42   0.000     38.12205    42.39453
------------------------------------------------------------------------------

. regress mpg hp curbwt torque accel dEUR dASIA, vce(robust)

Linear regression                               Number of obs     =         95
                                                F(6, 88)          =      67.99
                                                Prob > F          =     0.0000
                                                R-squared         =     0.8334
                                                Root MSE          =     2.1329

------------------------------------------------------------------------------
             |               Robust
         mpg | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
          hp |   .0071719   .0231857     0.31   0.758    -.0389049    .0532487
      curbwt |  -.0090926   .0010608    -8.57   0.000    -.0112007   -.0069845
      torque |   .0083239   .0094324     0.88   0.380    -.0104209    .0270688
       accel |    .986283   .5829988     1.69   0.094    -.1723046    2.144871
        dEUR |  -1.653195     .67888    -2.44   0.017    -3.002326   -.3040634
       dASIA |    .285032   .5395841     0.53   0.599    -.7872781    1.357342
       _cons |   43.64497   7.132522     6.12   0.000     29.47058    57.81936
------------------------------------------------------------------------------

. regress lmpg lhp lcurbwt ltorque accel, vce(robust)

Linear regression                               Number of obs     =         95
                                                F(4, 90)          =     149.31
                                                Prob > F          =     0.0000
                                                R-squared         =     0.8680
                                                Root MSE          =     .06487

------------------------------------------------------------------------------
             |               Robust
        lmpg | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         lhp |  -.7777867   .2789421    -2.79   0.006    -1.331954   -.2236196
     lcurbwt |  -.0402201   .2593707    -0.16   0.877    -.5555053     .475065
     ltorque |  -.0264077   .0644809    -0.41   0.683    -.1545102    .1016948
       accel |  -.0842394   .0369613    -2.28   0.025    -.1576694   -.0108094
       _cons |   8.553553   .6638139    12.89   0.000     7.234771    9.872336
------------------------------------------------------------------------------
```