Economics 102: Analysis of Economic Data
Cameron Winter 2014
Department of Economics, U.C.-Davis

Final Exam (A) Wednesday March 12

Compulsory. Closed book. Total of 55 points and worth 45% of course grade.
Read question carefully so you answer the question.

Question scores

Question	1a	1b	1c	1d	1e	1f	2a	2b	2c	2d	2e	3a	3b	3c	3d	3e	3f
Points	1	2	2	3	1	1	2	2	3	1	2	1	1	2	1	4	1

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

Questions 1-4

Consider data on various characteristics of 330 different models and makes of gasoline-fuelled automobiles sold in the U.S. in 2006.

Dependent Variable

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

Regressors

`hp` = Horsepower (a measure of engine power)
`curbwt` = Curb weight (weight of car in pounds with full fuel tank but no passengers or cargo)
`torque` = Torque (a measure of turning force of the engine)
`disp` = Engine displacement (volume of engine cylinders in liters)
`lncurbwt` = Natural logarithm of `curbwt`
`lndisp` = Natural logarithm of `disp`

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

Critical T 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) From the output, is variable `mpg` approximately normally distributed? Explain.

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

(c) Give a 95% confidence interval for the population mean miles per gallon.

(d) Perform a test at significance level .05 of the claim that the population mean horsepower exceeds 260.
State clearly the null and alternative hypotheses of your test, and your conclusion.

(e) If we regressed `mpg` on just one of `hp`, `curbwt`, `torque` and `disp`, which of these variables would worst explain mpg? Explain.

(f) 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) Test the hypothesis at significance level 1% 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.

(d) Predict the conditional mean miles per gallon of cars in 2006 if they had the same average horsepower as they did in 1990, which was 140.

(e) Give a 95 percent confidence interval for the conditional mean miles per gallon when horsepower is 140.
Give your answer as an expression involving numbers only. You need not complete all the calculations.

3. In this question consider both the regressions where `mpg` is the dependent variable.

(a) Do any of the coefficients in the larger model have unexpected sign?

(b) What is the impact of horsepower on gasoline consumption, controlling for weight, torque and displacement?

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

(d) Using a measure of model fit that controls for model size, which model explains the data better - the multivariate regression or the bivariate regression? Explain your answer.

(e) Are `curbwt`, `torque` and `disp` jointly statistically significant at 5 percent in the multivariate regression? Perform an appropriate test. State clearly the null and alternative hypotheses of your test, and your conclusion. You can use as critical value 2.63.

(f) Do you think multicollinearity is a problem for the larger model? Explain.

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

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

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

(c) The cars in this sample are made by 25 distinct manufacturers (Audi, Ford, ..., Volvo). Suppose we created 25 indicator variables for the 25 manufacturers

```
d1 = 1 if mfr == "Audi" and d1 = 0 otherwise
  ⋮    ⋮    ⋮    ⋮    ⋮    ⋮
d25 = 1 if mfr == "Volvo" and d25 = 0 otherwise
```

Do you see any problems in giving the following Stata command: `regress lnmpg d1-d25?`
Explain your answer.
If there is a problem, how would you solve it?

(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 `disp` is correlated with variable `lnmpg`? Explain.

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

(g) Give the Stata command that creates variable `lnmpg` from variable `mpg`.

5. This question has various unrelated parts.

(a) Consider a simple random sample of size n drawn from a random variable with mean μ and variance σ^2.
Give the mean and variance of $\bar{X}$.

(b) Suppose $X = 10$ with probability 0.4 and $X = 20$ with probability 0.6. What is the variance of X? **Show all workings.**

(c) A regression of `wage` (hourly wage) on an intercept and an indicator variable `gender` (equal to one if female and equal to zero if male) leads to estimate $\widehat{\texttt{wage}} = 20 - 4 \times \texttt{gender}$.
What are average wages for men and for women in the sample?

(d) 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 annual growth rate in Argentinian GDP over this period?
You are to answer this question in the simplest way possible.

(e) Regression of y on an intercept and x with a sample of 50 observations leads to results that includes the explained sum of squares equal to 400 and the residual sum of squares equal to 600.
Compute R^2.
Compute the standard error of the regression.

Multiple choice questions (1 point each)

1. In statistical inference the goal is to

a. infer population behavior from sample data

b. infer sample behavior from population data

c. both of the above

d. none of the above

2. For a simple random sample, $(\bar{X} - \mu)/(\sigma/\sqrt{n})$ is

a. standardized to have mean 0 and variance 1 always

b. normally distributed as $n \to \infty$

c. neither of the above

d. both of the above

3. 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

4. Which assumptions are essential for the OLS estimator to be unbiased

a. The population model is $y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k + u$.

b. The error has mean zero and is not correlated with the regressor.

c. The errors for different observations have the same variance, denoted σ^2.

d. a . and b.

e. a., b. and c.

5. In linear OLS regression a major problem arises if

a. important regressors are omitted

b. unnecessary (or irrelevant) regressors are included

c. neither a. nor b.

d. both a. and b.

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

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

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

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

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

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

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

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

Multivariate Data

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

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

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

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

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

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$(df1, df2, \alpha)$.

```
t_.05,v for v = 330    v = 329    v = 328    v = 327    v = 326    v = 325
            1.6494842  1.6494983  1.6495125  1.6495268  1.6495412  1.6495556
t_.025,v for v = 330   v = 329    v = 328    v = 327    v = 326    v = 325
            1.9671787  1.9672007  1.9672228  1.9672451  1.9672675  1.9672901
t_.005,v for v = 330   v = 329    v = 328    v = 327    v = 326    v = 325
            2.5908094  2.5908552  2.5909012  2.5909476  2.5909942  2.5910411

. summarize mpg hp curbwt torque disp lnmpg lncurbwt lntorque lndisp

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         mpg |       330    25.94315    5.435451      13.05       45.8
          hp |       330    268.5318    105.8087        103        617
      curbwt |       330    3583.002    591.6002       2093       6200
      torque |       330    358.0046    151.3064        137       1001
        disp |       330     3.45353    1.257165        1.3        8.3
-------------+--------------------------------------------------------
       lnmpg |       330    3.233925    .2118677   2.568788   3.824284
    lncurbwt |       330    8.170979    .1604684   7.646354   8.732305
    lntorque |       330    5.802813    .3889465   4.919981   6.908755
      lndisp |       330    1.174897    .3613186   .2623642   2.116256

. summarize mpg, detail

                          (mean) mpg
-------------------------------------------------------------
      Percentiles      Smallest
 1%         14.5          13.05
 5%           17           13.8
10%       19.325           13.8       Obs                 330
25%         23.1           14.5       Sum of Wgt.         330

50%           26                      Mean           25.94315
                        Largest       Std. Dev.      5.435451
75%           28          41.85
90%         32.7           44.9       Variance       29.54413
95%         36.6           45.8       Skewness       .5433534
99%        41.85           45.8       Kurtosis       4.157237

. correlate mpg hp curbwt torque disp
(obs=330)

             |      mpg       hp   curbwt   torque     disp
-------------+---------------------------------------------
         mpg |   1.0000
          hp |  -0.8131   1.0000
      curbwt |  -0.6705   0.6383   1.0000
      torque |  -0.7563   0.9314   0.7137   1.0000
        disp |  -0.7596   0.8560   0.7297   0.8942   1.0000
```

```
. * Regressions for exam
. regress mpg hp

      Source |       SS       df       MS              Number of obs =     330
-------------+------------------------------           F(  1,   328) =  640.14
       Model |  6426.92135     1  6426.92135           Prob > F      =  0.0000
    Residual |  3293.09826   328  10.0399337           R-squared     =  0.6612
-------------+------------------------------           Adj R-squared =  0.6602
       Total |   9720.0196   329  29.5441325           Root MSE      =  3.1686

------------------------------------------------------------------------------
         mpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          hp |  -.0417717    .001651   -25.30   0.000    -.0450196   -.0385238
       _cons |   37.16018   .4764228    78.00   0.000     36.22295    38.09741
------------------------------------------------------------------------------

. regress mpg hp curbwt torque disp

      Source |       SS       df       MS              Number of obs =     330
-------------+------------------------------           F(  4,   325) =  204.45
       Model |  6955.79742     4  1738.94935           Prob > F      =  0.0000
    Residual |  2764.22219   325  8.50529904           R-squared     =  0.7156
-------------+------------------------------           Adj R-squared =  0.7121
       Total |   9720.0196   329  29.5441325           Root MSE      =  2.9164

------------------------------------------------------------------------------
         mpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          hp |  -.0432345   .0042664   -10.13   0.000    -.0516277   -.0348412
      curbwt |  -.0025332   .0004105    -6.17   0.000    -.0033408   -.0017256
      torque |   .0142477   .0035139     4.05   0.000     .0073348    .0211606
        disp |  -.8329362   .3037788    -2.74   0.006    -1.430557   -.2353152
       _cons |   44.40531   1.119392    39.67   0.000     42.20314    46.60748
------------------------------------------------------------------------------

. regress lnmpg hp lncurbwt torque disp

      Source |       SS       df       MS              Number of obs =     330
-------------+------------------------------           F(  4,   325) =  283.62
       Model |  11.4795648     4  2.86989119           Prob > F      =  0.0000
    Residual |  3.28855632   325  .010118635           R-squared     =  0.7773
-------------+------------------------------           Adj R-squared =  0.7746
       Total |  14.7681211   329  .044887906           Root MSE      =  .10059

------------------------------------------------------------------------------
       lnmpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          hp |  -.0016124   .0001463   -11.02   0.000    -.0019003   -.0013246
    lncurbwt |  -.3158294    .052377    -6.03   0.000    -.4188702   -.2127885
      torque |   .0004298     .00012     3.58   0.000     .0001937    .0006658
        disp |  -.0353481   .0105357    -3.36   0.001    -.0560749   -.0146213
       _cons |   6.215758   .4119311    15.09   0.000      5.40537    7.026146
------------------------------------------------------------------------------
```