EEP 118 / IAS 118
University of California at Berkeley

Elisabeth Sadoulet and Daley Kutzman
Fall 2012

Introductory Applied Econometrics
Midterm examination

Scores add up to 50.

Your name:____ SID:__________________________

SID:______________________________

1. (5 points) In a poll of 1000 persons, you find that 44% of the respondents state that they are in favor of a land tax. Build a 95% confidence interval for the percentage of the population that is in favor of the land tax.

2. (10 points) Using data from 49 US states on traffic fatalities in year 2000, we can investigate whether primary seatbelt laws make roads safer. Consider the regression of state-wide traffic fatalities (*fatalities*) in number of deaths, on state population (*pop*) in 1000s of people, and a dummy variable for whether the state had a primary seatbelt law in place (*primary*):

$$\widehat{fatalities} = \hat{\beta}_0 + \hat{\beta}_1 pop + \hat{\beta}_2 primary$$

a) Do you think that the estimated effect of primary seatbelt laws on traffic fatalities (controlling for population) will be biased, and if yes, in what direction? Make a precise argument to justify your answer.

b) Given your answer, do you think that any of the assumptions MLR.1-4 are violated? Why or why not?

SID:________________________________

3. (5 points) The p-value for the test H0: $\beta_1 = 0$ against H1: $\beta_1 \neq 0$ is 0.06. Would you reject the null hypothesis at the 1% significance level? And at the 10% significance level? Explain.

4. (20 points) A city has been publicizing its new compost collection service with flyers and billboards in order to increase use of the service. The results reported in Table 1 (last page of exam) show a regression of compost collected per neighborhood (*compost*) in 100lbs on expenditures for *flyers* and *billboards* in $100's.

a) Formally test whether the effect of billboards on collected compost is different from zero at the 5% significant level.

SID:________________________________

b) Interpret the results on the effect of flyer spending.

c) Calculate the 95% confidence interval for the effect of flyer spending.

d) Calculate and interpret the R-squared for this regression.

SID:________________________________

5. (10 points) From a sample of 200 households, we estimated the following two models of gasoline consumption (t-statistics in parentheses):

$$gas = 34.2 + 10.5 suv + 0.25 inc - 0.00005 inc^2 \qquad R^2 = 0.356$$
$$(2.3) \quad (3.1) \quad (1.7) \quad (1.8)$$

$$gas = 22.2 + 15.3 suv \qquad R^2 = 0.323$$
$$(2.3) \quad (3.1)$$

where *gas* gives the number of gallons per month, *suv* is a dummy variable for whether the household owns an SUV, and *inc* is the annual household income in thousands of $.

a) Using the estimated parameters in the first equation, how does gasoline consumption vary with income?

b) Are the two income variables jointly significant at the 5% level?

SID:________________________________

Formulae

Statistics

Covariance between two variables: $\text{cov}(x,y) = \frac{1}{n}\sum_i (x_i - \bar{x})(y_i - \bar{y})$

$\text{cov}(a_1 x + b_1, a_2 y + b_2) = a_1 a_2 \text{ cov}(x,y)$

$\text{var}(ax + by) = a^2 \text{ var } x + b^2 \text{ var } y + 2ab\,\text{cov}(x,y)$

Variance for the difference in means of two independent samples:

$\text{var}(\bar{x}_1 - \bar{x}_2) = \text{var}(\bar{x}_1) + \text{var}(\bar{x}_2)$

When y is a binary variable with probability prob(y = 1) = p, its variance is p (1– p)

OLS estimator

$\hat{\beta}_1 = \frac{\text{cov}(x,y)}{\text{var } x}$ with $\text{var}(\hat{\beta}_1) = \frac{\sigma^2}{SST_x}$

For multiple regression: $\text{var}(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}$

$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$, $SSE = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$, and $SSR = \sum_{i=1}^{n}\hat{u}_i^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

Test statistics:

F statistic for q restrictions in a regression done with n observations and k exogenous variables:

$$\frac{(R_{UR}^2 - R_R^2)/q}{(1 - R_{UR}^2)/(n-k-1)} \sim F(q, n-k-1)$$

Variances:

Let x be a random variable of unknown mean μ and variance σ^2.

Using a sample of n observations:

The "observed" variance: $\text{var}(x) = \frac{1}{n}\sum_i (x_i - \bar{x})^2$ is a biased estimator of σ^2

The "sample" variance: $s^2 = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2$ is an unbiased estimator of σ^2

Recall that the Stata *summarize* command reports s under the column "standard deviation"

Table 1 for question 4

```
. reg compost flyers billboards

      Source |       SS       df       MS              Number of obs =     123
-------------+------------------------------           F(  2,   120) =   37.76
       Model |  500.170986     2  250.085493           Prob > F      =  0.0000
    Residual |   794.74383   120  6.62286525           R-squared     =
-------------+------------------------------           Adj R-squared =  0.3760
       Total |  1294.91482   122  10.6140559           Root MSE      =  2.5735

------------------------------------------------------------------------------
     compost |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      flyers |   1.376345                8.58   0.000
  billboards |   .4267414   .4139366
       _cons |   10.25937   2.202563     4.66   0.000     5.898452    14.62029
------------------------------------------------------------------------------
```