EEP 118 / IAS 118 Elisabeth Sadoulet and Kyle Emerick
University of California at Berkeley Fall 2011

Introductory Applied Econometrics
Midterm examination

Scores add up to 50.

Your name:____ SID:__________________________

SID:________________________________

1. (5 points) A researcher is interested in testing whether mean incomes for men and women in India are the same. He collects incomes from a sample of 1000 men and 1000 women and tests the null hypothesis of equal mean incomes against the two-sided alternative. He decides to conduct the test at the 5% significance level and thus picks the critical value c from the corresponding statistical table. Explain *in words* what the meaning of the significance level is.

2. (5 points) We are interested in finding the relationship between hours of training (*train*) received by employees in a large firm and their productivity (*productivity*). Consider the equation:

$$productivity = \beta_0 + \beta_1 train + u$$

Knowing that employees are assigned to training based on their ability level when hired, is $\hat{\beta}_1$ biased? If you think it is biased, is it biased upward or downward? Why? (Give a very precise argument for the direction of the bias if you feel there is a bias)

SID:______________________________

3. (10 points) From a sample of 136 cities around the world, you estimated an equation relating air pollution (emission of sulfur dioxide (SO2) in microgram per cubic meter) to the local income per capita (Y) in dollars:

$$\widehat{SO2} = 0.5 + 4Y - 0.0005Y^2$$

a. What is the marginal impact of an increase in income per capita on predicted air pollution?

b. In which range of income per capita is pollution increasing with income per capita? In which range of income per capita is pollution decreasing with income per capita? Give arguments that could explain this relationship between income per capita and pollution.

4. (20 points) Data from the U.S. Department of Agriculture for consumption from 1960 to 1982 were used to estimate the following model of chicken consumption:

$$\ln C = \beta_0 + \beta_{inc} \log Inc + \beta_{pc} \log p_c + \beta_{pb} \log p_b + \beta_{pp} \log p_p + u$$

where C is per capita consumption of chicken, in lbs. per person, Inc is annual real income per capita in \$, and p_c, p_b, and p_p are the retail prices of chicken, beef, and pork, in \$/lb.

(a) What is the economic interpretation of the parameters β_{inc}, β_{pc}, and β_{pb}?

The estimated model is:

$$\widehat{\log C} = 2.19 + 0.42\log Inc - 0.37\log p_c + 0.149\log p_b + 0.09\log p_p \qquad R^2 = 0.9632$$
$$(0.16) \quad (0.08) \qquad (0.11) \qquad (0.100) \qquad (0.10) \qquad n = 23$$

where n is the number of years of observation.

(b) What is the predicted increase of chicken consumption for a household with a \$20,000 income that obtains a raise of \$1000?

(c) Perform the test for the hypothesis of $\beta_{pc} = -1$ against $\beta_{pc} > -1$ at the 1 percent significance level. (Make sure to use the 5 steps of hypothesis testing)

We now re-estimate the model without the prices of pork and beef.

$$\widehat{\log C} = 2.03 + 0.42\log Inc - 0.37\log p_c \qquad R^2 = 0.9605$$
$$(0.12)\,(0.08) \qquad (0.11) \qquad n = 23$$

(d) Comparing the two estimated models, would you say that chicken consumption is affected by the prices of beef or pork, considered together, at the 5% significance level? (Perform a joint test of the two parameters).

SID:________________________________

5. (10 points). The following is the wage regression that we have extensively studied in class.

```
. reg lwage educ exper female;

      Source |       SS       df       MS              Number of obs =    2000
-------------+------------------------------           F(  3,  1996) =  247.50
       Model |   152.35726     3  50.7857535           Prob > F      =  0.0000
    Residual |  520.219607  1996  .260631065           R-squared     =
-------------+------------------------------           Adj R-squared =  0.2700
       Total |  672.576867  1999  .336456662           Root MSE      =  .49558

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .1167441   .0095157    12.27   0.000
       exper |   .0109089    .001869            0.000     .0072457    .0145721
      female |  -.2543189   .0222067   -11.45   0.000    -.2978696   -.2107682
       _cons |   1.055792   .0757381    13.94   0.000     .9072576    1.204326
------------------------------------------------------------------------------
```

1. Fill in the blank for the missing t-stat in the regression output.
   Fill in the blank for the R-squared value.

2. Fill in the blank for the missing 95% confidence interval

SID:________________________________

Formulae

Statistics

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

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

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

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 $SSR = \sum_{i=1}^{n}\hat{u}_i^2$

Test statistics:

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

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