SID ______________________________

EEP 118 / IAS 118
University of California at Berkeley

Elisabeth Sadoulet and Kelly Jones
Fall 2008

Introductory Applied Econometrics
Midterm examination

1. (5 points) You conduct a survey among 400 residents of your city about their electricity consumption. The results show electricity consumption per household to vary from 5,500 kWh to 60,000 kWh with mean value 10,656 kWh and standard deviation 4,200 kWh. Construct a 95% confidence interval for the average electricity consumption per household in your city.

2. (5 points) In 2004, Florida voted 47.1% for the democratic candidate to the presidency. A poll of 765 likely voters conducted this week shows that 51% of them favor the democratic candidate. Is there evidence that the current support for the democratic candidate among Florida voters is higher than 4 years ago, at the 5% significance level?

3. (15 points) From a random sample of young employees, we have estimated the following equation for hourly wage:

$$\widehat{\log(\text{wage})} = \underset{(.11)}{0.29} + \underset{(.007)}{.092}\text{ education} + \underset{(.0017)}{.0041}\text{ experience} - \underset{(.064)}{.006}\text{ nonwhite} \qquad R^2 = .32,\ n = 526$$

a. Interpret the results on the effect of education on wage.

b. How would the estimated effect of education be affected by not including the nonwhite variable in the equation? Justify your answer.

c. How would the estimated effect of education be affected by not including the experience variable in the equation? Justify your answer.

4. (15 points) Consider the following model of car sales:

$$\ln(qnc) = \beta_0 + \beta_p \ln(price) + \beta_{inc} \ln(income) + u$$

where *qnc* is the number of cars sold (in thousands), *price* is the average price of new cars (in \$1,000), and *income* is the per capita income (in \$1,000).

a. Give the economic interpretation of the parameters β_p and β_{inc}. What sign do you expect these parameters to have?

The estimation of the model with quarterly car sales in the U.S. from 1975 to 1990 gives:

```
. reg lqnc lprice lincome

      Source |       SS       df       MS              Number of obs =      64
-------------+------------------------------           F(  2,    61) =   12.21
       Model |  .32720224     2   .16360112            Prob > F      =  0.0000
    Residual |  .817286587    61  .013398141           R-squared     =  0.2859
-------------+------------------------------           Adj R-squared =  0.2625
       Total |  1.14448883    63  .018166489           Root MSE      =  .11575

------------------------------------------------------------------------------
        lqnc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lprice |  -.8280926   .1838504    -4.50   0.000    -1.195724   -.4604611
     lincome |   2.399991   .4860261     4.94   0.000     1.428121     3.37186
       _cons |    5.92543   .4843662    12.23   0.000      4.95688     6.89398
------------------------------------------------------------------------------
```

b. Based on the parameter estimates, what is the predicted effect of a 10% increase in price on the *number* of cars sold? What would be the effect of that price increase on the *value* of car sales?

SID ______________________________

c. Test at the 5% significance level the null hypothesis that $\beta_p = -1$ against the alternative that $\beta_p > -1$. Now, revisiting your answer to question b, what would you say is the predicted effect of the price increase on the value of car sales?

5. (10 points) Using the same data as in question 4 above, we estimated the following model:

$$\ln(qnc) = \beta_0 + \beta_p \ln(price) + \beta_{inc} \ln(income) + \beta_W winter + \beta_{Sp} spring + \beta_S summer + u$$

where *winter* = 1 for the winter quarter and 0 otherwise, *spring* = 1 for the spring quarter and 0 otherwise, and *summer* = 1 for the summer quarter and 0 otherwise.

```
. reg lqnc price income winter spring summer

      Source |       SS       df       MS              Number of obs =      64
-------------+------------------------------           F(  5,    58) =    9.32
       Model |  .509752677     5  .101950535           Prob > F      =  0.0000
    Residual |  .634736151    58  .010943727           R-squared     =  0.4454
-------------+------------------------------           Adj R-squared =  0.3976
       Total |  1.14448883    63  .018166489           Root MSE      =  .10461

------------------------------------------------------------------------------
        lqnc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       price |  -.0103586   .0020172    -5.14   0.000    -.0143966   -.0063207
      income |   .2528685   .0451826     5.60   0.000     .1624256    .3433113
      winter |   .0232618   .0142710     1.63   0.102    -.0047094    .0512330
      spring |   .1329125   .0370159     1.59   0.140    -.030929     .2967543
      summer |   .0529875   .0370027     1.43   0.110    -.0210815    .1270564
       _cons |   6.084057   .3034456    20.05   0.000     5.476644    6.691469
------------------------------------------------------------------------------
```

a. Should we have added the fall dummy variable also? Why or why not?

b. Perform a joint test of significance on the parameters on the seasonal variables? Do you find evidence of seasonality in car sales?

Formulae

Statistics

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

$\operatorname{cov}(a_1x+b_1, a_2y+b_2)=a_1a_2\operatorname{cov}(x,y)$

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

Variance for the difference in means of two independent samples:
$\operatorname{var}(\bar{x}_1-\bar{x}_2)=\operatorname{var}(\bar{x}_1)+\operatorname{var}(\bar{x}_2)$

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

OLS estimator

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

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

Test statistics:

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

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