Midterm II - Answers

Econ 253 Spring 2002

You have 75 minutes to answer the following 4 questions. The maximum number of points on this exam is 75. You may use a calculator and a single-sided sheet of paper with notes and formulas. Any collaboration is considered a violation of the honor code. Good luck!

1. There were 23 responses to the first survey we collected in this class. The sample average of the number of hours spent watching TV per week was 4.6 with a sample standard deviation of 5.3.

   (a) Assuming that hours spent watching TV are normally distributed, construct the 95% confidence interval for the number of hours students spend watching TV per week.

   Answer: This is case 2 in the handout: confidence interval for population mean, variance unknown. $P(-A < \frac{\overline{x}-\mu}{s/\sqrt{n}} < A) = 0.95$
   $P(\overline{x} - As/\sqrt{n} < \mu < \overline{x} + As/\sqrt{n}) = 0.95$ where A is a number so that 97.5% of the t_{22} distribution lies to the left. In the tables we find $A = 2.069$, hence the confidence interval is (2.3, 6.9).

   (b) How many students would we have to survey to obtain a 95% confidence interval with ± 1 hour precision? Assume that the sample standard deviation would not change.

   Answer: We need $\frac{A \cdot 5.3}{\sqrt{n}} = 1$. Since A may change depending on n, we find the answer through a very short iterative process. If n is greater than 120, then A=1.96 and n=108. But if $n = 108$, A should be 1.98. With A=1.98, n comes to 110. We would therefore need to collect information from 110 students.

2. In the class survey 8 female students reported to have spent an average of \$28 dollars on a haircut with a sample standard deviation of \$13 dollars. 15 male students reported to have spent only \$9 dollars on average with a sample standard deviation \$8 dollars. Answer the following questions:

   (a) Is the cost of a haircut for females significantly greater than the cost of haircut for males? (Ignore the fact that we have a very small number of observations in each group.)

   Answer: This is the difference in means test.
   $n_{female} = 8, \overline{x}_{female} = 28, s_{female} = 13$

$n_{male} = 15, \overline{x}_{male} = 9, s_{male} = 8$

$H_0 : \mu_{female} = \mu_{male}$

$H_1 : \mu_{female} > \mu_{male}$

The test statistic is $Z = \frac{(\overline{x}_1 - \overline{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ which is distributed N(0,1) if n_1 and n_2 are greater than 30. You are instructed to ignore this so that we can use our own data. The test statistic $Z = \frac{28-9}{5} = 3.8$. The critical region for this one sided test is $(1.64, \infty)$. The test statistic falls into the critical region and we reject H_0 that the cost of a female and male haircut is the same and accept the alternative that the cost of a female haircut is greater.

(b) Is the variance in haircuts among females significantly greater than among males?

Answer: This is an F-test for equality of variances.

$H_0 : \sigma_f = \sigma_m$

$H_1 : \sigma_f > \sigma_m$

The test statistic $F = \frac{s_f^2}{s_m^2} = \frac{169}{64} = 2.64$ is distributed F with 7 and 14 degrees of freedom. The 5% critical region is $(2.76, \infty)$. Our test statistic fall into the acceptance region and hence can not reject H_0. We find no evidence that the variance of haircut costs among females and males in our class is statistically different.

3. In Principles of Microeconomics you learned that expenditures on a commodity should be proportional to income. This relationship is known as the Engel curve. Suppose we are interested in the Engel curve for health care expenditures:

$$HLTHEXP_t = B_1 + B_2 \cdot INCOME_t + u_t$$

I used data from all 50 U.S. states and the District of Columbia on aggregate health care expenditures in billions of dollars, and data on personal income in billions of dollars. This data is available in the *Statistical Abstract of the U.S.* I estimated the above equation with the following results:

$$\begin{aligned} \widehat{HLTHEXP}_t &= \underset{(0.47)}{0.18} + \underset{(0.01)}{0.14} \cdot INCOME_t \end{aligned}$$

ESS=317 RSS=10

where standard errors of coefficient estimates are in parentheses, ESS stands for estimated sum of squares ($\sum(\widehat{Y}_i - \overline{Y})^2$ and RSS stands for residual sum of squares ($\sum e_i^2$). Answer the following questions:

(a) What is the interpretation of the B_1 coefficient? Test the hypothesis that B_1 is equal to zero against the alternative that it is different from zero. Use a 5% level of significance. Is your result consistent with economic theory?

Answer: B_1 is the expected expenditure on health care given that a state has zero income.

$H_0 : B_1 = 0$

$H_1 : B_1 \neq 0$

The test statistic $t = \frac{b_1}{s_{b_1}}$ is distributed t with 49 degrees of freedom. The rejection region at 5% level of significance is $(-\infty, -2)$ and $(2, \infty)$. Calculating the statistic, $t = \frac{0.18}{0.47} = 0.38$, we see that it is in the acceptance region. Hence, we accept the null hypothesis that B_1 is equal to zero. This is consistent with economic theory as well as common sense: with zero income a state should have zero health care expenditures.

(b) Interpret the estimate of the B_2 coefficient.

Answer: For every one dollar increase in aggregate income of a state, its health care expenditure will increase by 14 cents.

(c) Test the hypothesis that the B_2 coefficient is 0.25 against the alternative that it is less than 0.25

Answer: $H_0 : B_2 = 0.25$

$H_1 : B_2 < 0.25$

If the null hypothesis is true, the test statistic $t = \frac{b2-0.25}{s_{b_2}}$ is distributed t with 49 degrees of freedom. The critical/rejection region is in the left tail of the distribution $(-\infty, -2)$. The test statistic, $t = \frac{b2-0.25}{s_{b_2}} = -11$, is in the critical region, hence we reject the null hypothesis that B_2 is equal to 0.25 and accept the hypothesis that is is less than 0.25.

(d) What percentage of the variance in health care expenditures across states can be explained by personal income?

Answer: $R^2 = \frac{ESS}{TSS} = \frac{ESS}{ESS+TSS} = \frac{317}{317+10} = 0.97$. Hence, 97% of the variance in health care expenditures can be explained by personal income.

4. Suppose you are interested in investigating women's labor force participation. You collect data using the 1990 census from 50 U.S. states on the following variables:

- wlfp = persons 16 years and over–percent in labor force who are female (Range 42.6 - 66.4)
- yf = median earnings (in thousands of dollars) by females 15 years and over with income in 1989 (Range 14.271 - 25.62)
- ym = median earnings (in thousands of dollars) by males 15 years and over with income in 1989 (Range 21.425 - 35.622)
- ue = civilian labor force–percent unemployed (Range 3.5 - 9.6)
- mr = female population 15 and over–percent now married (excluding separated), Range 46.88 - 60.92

Suppose further that you ran a multiple regression of female labor force participation on several explanatory variables:

$$wlfp_i = B_1 + B_2 \cdot yf_i + B_3 \cdot ym_i + B_4 \cdot ue_i + B_5 \cdot mr_i + u_i$$

```
   Source |       SS       df       MS              Number of obs =      50
----------+------------------------------           F(  4,    45) =   25.25
    Model |  611.942691     4  152.985673           Prob > F      =  0.0000
 Residual |  ...A...       45  6.05807957           R-squared = 0.6918
----------+------------------------------           Adj R-squared =  0.6644
    Total |  884.556271    49  18.0521688           Root MSE      =  2.4613

------------------------------------------------------------------------------
     wlfp |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
----------+-------------------------------------------------------------------
       yf |   1.660993    .388177    ...B...   0.000       .8791646    2.442822
       ym |  -.5376039    .291862    -1.842   0.072        ...C...     ...D...
       ue |  -1.466509    ...E...    -5.353   0.000      -2.018276   -.9147426
       mr |   .3880147   .1289237     3.010   0.004       .1283491    .6476804
    _cons |   29.81552   8.964006     3.326   0.002       11.76109    47.86995
------------------------------------------------------------------------------
```

Answer the following questions:

(a) Fill in the blanks A, B, C, D and E in the computer printout above. (Answer in your booklet what A, B, C, D and E are.)

Answers:

$A = RSS = TSS - ESS = 884 - 611 = 273$

$B = \frac{b}{s_b} = \frac{1.66}{0.388} = 4.28$

$C = b_2 - t_{49} \cdot s_{b_2} = -0.537 - 2.02 \cdot 0.291 = -1.12$

$D = b_2 + t_{49} \cdot s_{b_2} = -0.537 + 2.02 \cdot 0.291 = 0.05$

$E = \frac{b_2}{t} = \frac{-1.46}{-5.35} = .2739$

(b) Interpret estimated coefficients on yf, ym and mr. Are the estimated signs and magnitudes consistent with your intuition? Are these coefficients statistically significant? At what level of confidence?

Answer:

yf - holding male earnings, unemployment and marriage rates constant, a thousand dollar increase in median females earnings increases expected women's labor force participation by 1.66 percentage points. The higher the earnings by females, the greater the incentive to participate in the labor force. On the other hand, as income rises, workers desire more leisure. This is the "income effect." At the prevailing wages the income effect is likely to be smaller, and on balance we would expect a positive sign. The coefficient is statistically significant at the 5% level of confidence.

ym - holding female earnings, unemployment and marriage rates constant, a thousand dollar increase in median male earnings decrease expected women's labor force participation by 0.5 percentage points. As husbands earn more, their wives may not need to work as much. The coefficient is statistically significant at the 10% level of confidence but not at 5%.

mr - holding female and male earnings and the unemployment rate constant, a one percentage point increase in the marriage rate increases labor force participation by 0.3 percentage points. This is somewhat inconsistent with our intuition. Married women ,compared to unmarried women, tend to stay home to care for children. Thus, we would expect a negative coefficient. The coefficient is statistically significant at 5% level of confidence.

(c) In economic theory, there are two hypotheses regarding the effects of unemployment on labor force participation. The "discouraged worker" hypothesis suggests that a high level of unemployment is an indication that searches for employment may be futile. When unemployment is high, chances of getting hired are low, and hence it is not worth searching for a job and thus participating in the labor force. The other hypothesis is called the "added worker" hypothesis. It suggests that with high levels of unemployment, some members of a family may lose their job and hence there may be a need for other members of the family to look for a job to replace the lost income. Which of the two hypothesis does our estimation support?

Answer: For every percentage point increase in the unemployment rate, the women's labor force participation rate *decreases* by 1.46 percentage points. The coefficient is statistically significant at the 5% level of confidence. It strongly supports the discouraged worker hypothesis.

(d) Suppose Massachusetts experiences an economic boom. Median incomes of females increase from \$23,000 to \$23,500, median incomes of males increase from \$26,000 to \$27,000, and the unemployment rate drops from 6% to 4%. The marriage rate stays constant. What will be the effect of this boom on women's labor force participation?

Answer: Change in expected women's labor force participation rate is $\Delta\widehat{wlfp} = b_2 \cdot \Delta yf + b_3 \cdot \Delta ym + b_4 \cdot \Delta ue = 1.66 \cdot 0.5 - 0.53 \cdot 1 - 1.46 \cdot (-2) = 0.83 - 0.53 + 2.92 = 3.22$. The women's labor force participation rate will increase by 3.22 percentage points.

(e) Suppose that we omitted the ym variable from our regression. What kind of bias would we be introducing? How would the interpretation of the yf coefficient change?

Answer: If we do not control for male earnings, the yf variable will measure the overall level of incomes in the state rather than incomes relative to male incomes. The interpretation of the yf coefficient would be the following: holding unemployment and marriage rates constant, a rise in female incomes female labor force participation increases.