AGEC5213: *ECONOMETRIC METHODS*
Spring 2019

Rijesh Mishra

PROBLEM SET NO. 2
Date Due: February 25, 2019

Now let's leave the computer for a moment. For this week's your homework assignment, you don't need to touch your computer keyboard. Just like old-timers, we are going to use pen and pencil (and maybe eraser).

I. R-Square, Interval Estimate, and Hypothesis Testing (5 points)

Consider the model:

$$y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + e_t$$

and suppose that application of least squares to 20 observations on these variables yields the following estimates, and their variances, and covariances.

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 0.96587 \\ 0.69914 \\ 1.7769 \end{bmatrix}, \text{cov}\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 0.21812 & 0.019195 & -0.050301 \\ 0.019195 & 0.048526 & -0.031223 \\ -0.050301 & -0.031223 & 0.037120 \end{bmatrix}$$

$\hat{\sigma}^2 = \frac{SSE}{T-K}$

$$\hat{\sigma}^2 = 2.5193 \qquad R^2 = 0.9466$$

(A) Find the SST (total variation), SSE (unexplained variation), and SSR (explained variation) for the model (**Hint:** start from the relationship between SSE and variance of errors, then use the formula of R^2).

(B) Find 95% interval estimates for β_2 and β_3.

(C) Use a t-test to test the hypothesis $H_0: \beta_2 \geq 1$ against the alternative $H_1: \beta_2 < 1$.

(D) Use your answers in part (A) to test the joint hypothesis $H_0: \beta_2 = 0, \beta_3 = 0$ (**Hint**: compute F-statistic using SSR and SSE, and their degrees of freedom,
i.e., in this case, you are testing if all slope coefficients except the intercept are zero or not,

$$F = \frac{SSR/(k-1)}{SSE/(T-k)} \sim F_{k-1,T-k}.$$

(E) Test each of the following hypotheses and state the conclusion.

(i) $\beta_1 = 0$

(ii) $\beta_2 + 2\,\beta_3 = 5$

$R^2 = 1 - \frac{SSE}{SST}$ $\quad SST = SSE + SSR$

II. Dummy Variables (5 points)

Consider the following regression results on annual salary with graduate degree in agricultural economics at OSU. The data set consists of information on 2,000 respondents. The estimated equation is:

Journal Art   F=1, M=0   Experience (Yrs)   NE   MID WEST   SOUTH

$$AS_i = 7.78 + 1.44J_i - 0.74F_i + 0.33E_i + 0.57N_i + 0.32M_i - 0.75S_i + \varepsilon_i,$$

(3.16) (0.21) (0.20) (0.04) (0.30) (0.28) (0.26)   SE

W → West (Dropped).

where AS is annual salary in 10,000 dollars; J is the total number of journal articles published; F is binary variable (1 if female, 0 if male); E is the number of experience (in years); N, M, and S are binary variables representing regions (N = 1 if region=Northeast, 0 otherwise; M = 1 if region = Midwest, 0 otherwise; S = 1 if region = South, 0 otherwise), and the West region has been dropped out.

1 (A) Interpret coefficients of J and N.

1 (B) Test the hypothesis that female workers are paid less than male workers. Show your null and alternative hypotheses, test these hypotheses at the 5% level, and state the conclusion.

1 (C) How would you test if the annual salary differs by the region? Discuss the test procedure including null and alternative hypotheses, degree of freedom, and distribution you would use for this test; state your conclusion at the 5% level when your calculated statistic is 15.15.

1 (D) One can include all four regional dummy variables in the model, while suppressing the intercept. Compute coefficients of all four regional dummy variables for this model.

(1)

(A) $\hat{\sigma}^2 = \dfrac{SSE}{T-K}$

$2.5193 = \dfrac{SSE}{20-3}$; $SSE = 42.8281$

$R^2 = 1 - \dfrac{SSE}{SST}$

$0.9466 = 1 - \dfrac{42.8281}{SST}$; $SST = 42.8281/(1-0.9466)$

$= 802.0243446$

$SST = SSE + SSR$

$802.0243 = 42.8281 + SSR$

$SSR = 759.1962446$

(B) $CI = b \pm S.E. \times 2.11$ for 5% CI df (17)

SE

$SE(b_2) = \sqrt{0.048526} = 0.2202861775$

$SE(b_3) = \sqrt{0.037120} = 0.192665133$

β_2, 95% CI $= 0.69914 \pm 2.11 \times 0.2202861775$

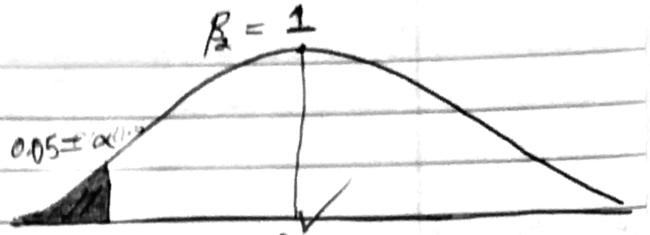
$= [-0.36185, 0.501675]$ $[\quad , 1.1639]$

$[0.234336, 1.163944]$

β_3, 95% CI $= 1.7769 \pm 2.11 \times 0.192665133$

$= [$[illegible], [illegible]$]$ $[1.370376, 2.183424]$

$t_{\alpha} = 2.11$



$t_a = -1.729$ ✓

(c) $H_0 : \beta_2 \geq 1$

$H_A : \beta < 1$

$$t_{calc} = \frac{0.69914 - 1}{0.220286} = -1.365769954$$ ✓

✓

Since $t_{calc} > t_{critic}$, we failed to reject null hypothesis.

(D) $SSR = 759.196$

$SSE = 42.82811$

$H_0 : \beta_2 = 0, \beta_3 = 0$ H_A: Atleast one of them are not equal to zero.

$$F_{(2,17)_{calc}} = \frac{(759.196)/(3-1)}{42.8281/(20-3)} = \frac{379.598}{2.5193} = 150.67598$$ ✓

✓

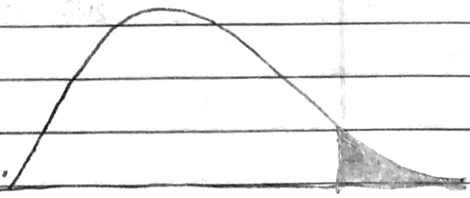
$f_{\alpha=0.05\,critic}(2,17) = 3.59$

$$F = \frac{SSR/(T-1)}{SSE/(T-K)} \sim F_{(T-1,\,T-K)}$$

$f_{calc(2,17)} > f_{critic\ \alpha=0.05(2,17)}$

We reject null hypothesis in favor of alternative hypothesis.

β_2 & β_3 are different.



$\alpha = 0.05$

$f_{critic\ \alpha=0.05}(2,17) = 3.59$

6_{17}

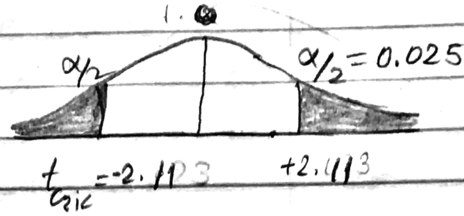
(E) (I) $H_0: \beta_1 = 0$ $\qquad \cancel{t = \frac{b-\beta}{SE(b)}}$

$H_A: \beta_1 \neq 0$

$\beta_1 = 0.96587 \qquad SE(\beta_1) = \sqrt{0.21812} = 0.4670331894$

$t_{calc} = \dfrac{0.96587 - 0}{0.4670} = 2.068244111$

$t_{crit\,\alpha/2} = 2.113$



$t_{calc_{\alpha_1}} < t_{crit_{\alpha/2}}$ we fail to reject null hypothesis.

(II) $\beta_2 + 2\beta_3 = 0.69914 + 2 \times 1.7769 = 4.25294$ $\qquad \cancel{H_0: \beta_H = 0}$

$H_0: \beta_2 + 2\beta_3 = 5 \qquad H_A: \beta_2 + 2\beta_3 \neq 5$ $\qquad \cancel{H_1: \beta_H < 0}$

~~t value~~ $\cancel{= \frac{b_H - \beta_H}{se(b_H)} = \frac{0.04}{0.2}}$

$\text{Variance}(\beta_2 + 2\beta_3) = (\beta_2 + 2\beta_3)^2$

$= \beta_2^2 + 8\beta_3\beta_2 + 4\beta_3^2$

$= \hat{Var}(\beta_2^2) + 4\,\hat{covar}(\beta_2\beta_3) + 4\hat{Var}(\beta_3^2)$

$= 0.048526 + 4 \cdot (-0.031223) + 0.037120 \times 4$

$= 0.072114$

$SE(\beta_2 + 2\beta_2) = \sqrt{0.072114} = 0.2685404997$

~~⇒ since t value is lower than critical value we reject H_0 at significance level.~~

$t_{calc} = (4.25294 - 5) / 0.268540 = -2.781932$

$t_{crit\,\alpha/2} = |2.113|$

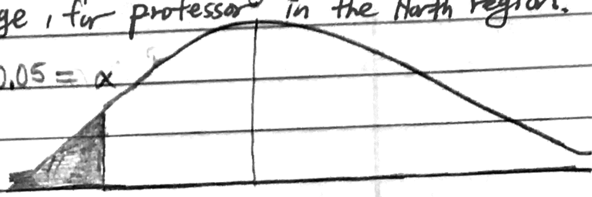
Since $t_{calc} < |t_{crit\,\alpha/2}|$ we reject null hypothesis.

(II) -0.1

(A) J ⇒ with the one unit increase in the # of journal articles published, the annual salary (in 10K) with Ag. Econ. graduate degree from OSU is expected to increase by 1.44 unit.

N ⇒ Having a region as Northeast, ~~would increase~~ the annual salary (in 10K) [of a person] with graduate degree in agriculture economics at OSU increase by 0.57 unit compared to not having greater Northeast region. For H, Annual salary (expressed in 10,000 dollars) is expected to be 0.64 unit greater, on average, for professors in the North region.

⇒ West (reference group)



(B) $H_0: \beta_F = 0$

$H_A: \beta_F < 0$

$t_{\alpha=0.05} = 1.645$

$$t_{calc} = \frac{\hat{\beta}_F - \beta_F}{SE(\hat{\beta}_F)}$$

$$= \frac{-0.74 - 0}{0.20} = -3.7$$

Since $t_{calc} < t_{cric}$, we reject the null hypothesis in favor of alternative hypothesis that female workers are paid less than male workers.

(c) $H_0 : \beta_N = \beta_m = \beta_S = 0$

H_a : Atleast one of them differs (are not equal to zero)

We can use f statistics to test the hypothesis, using restricted and unresteicted model.

Under H_0 : $F_{(S, T-K)} \approx \dfrac{(SSE_R - SSE_U)/S}{SSE_U/(T-K)}$

SSE ⇒ Sq. Sum of errors.
S ⇒ # of hypothesis to test
T = Total respondents
K = Total variables

Unrestsicted model is given by

$AS = 7.78 + 1.44J_i - 0.74F_i + 0.33E_i + 0.57N_i + 0.32M_i - 0.75S_i + \mathcal{E}_i$

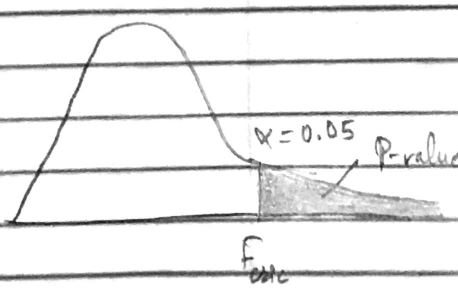
Restsicted model is given by plugging 0 for all testing βs

$AS = 7.78 + 1.44J_i - 0.74F_i + 0.33E_i + \mathcal{E}_i$

We can Test hypothesis using above formula to calculate test statistics. If H_0 is rejected at given significance level, there is evidence of difference in salary by regions.

$f_{cal}(3, 2000-7) = 15.15$

$f_{cric}(3, 1993)$ at $\alpha = 0.05$ = 2.61



Since $f_{cal} > f_{cric}$, I reject null hypothesis in favor of alternative hypothesis.
The region has impact on annual salory.

(D) $AS = 7.78 + 1.44J - 0.74F + 0.33E + 0.57N + 0.32M - 0.75S + \varepsilon$

$N + M + S + W = 1$ or $W = 1 - N - M - S$

without intercept & including all dummies,

$$
\begin{aligned}
AS &= \beta_1 J + \beta_2 F + \beta_3 E + \beta_4 N + \beta_5 M + \beta_6 S + \beta_7 W \\
&= \beta_1 J + \beta_2 F + \beta_3 E + \beta_4 N + \beta_5 M + \beta_6 S + \beta_7 (1 - N - M - S) \\
&= \beta_1 J + \beta_2 F + \beta_3 E + \beta_4 N + \beta_5 M + \beta_6 S + \beta_7 - \beta_7 N - \beta_7 M - \beta_7 S \\
&= \beta_1 J + \beta_2 F + \beta_3 E + (\beta_4 - \beta_7) N + (\beta_5 - \beta_7) M + (\beta_6 - \beta_7) S + \beta_7 \\
AS - \beta_7 &= \beta_1 J + \beta_2 F + \beta_3 E + (\beta_4 - \beta_7) N + (\beta_5 - \beta_7) M + (\beta_6 - \beta_7) S
\end{aligned}
$$

Intercept is the coefficient of W.

$\beta_w = 7.78$ ✓

$\beta_N = 0.57 + 7.78 = 8.35$ ✓ 8.35

$\beta_m = 0.32 + 7.78 = 8.10$ ✓

$\beta_s = -0.75 + 7.78 = 7.03$ ✓

9.9/10