Exercise 2 Answers- Hypothesis Testing in the General Linear Model

1. Given the information in Question 1, exercise 1, find the standard errors of the slope coefficient estimates in the 2 variable and 3 variable models. Test the hypotheses that the coefficient on X is zero in each model.

Given $\mathrm{Y}_i = \beta_0 + \beta_1 X_i + u_i$

In the 2-variable model we know that $\mathrm{var}(\hat{\beta}_i) = \frac{\hat{\sigma}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \sqrt{\frac{\hat{\sigma}^2}{\sum_i x_i^2}}$

where $\hat{\sigma}^2 = \frac{\sum_i \hat{u}^2}{N-2}$ is an unbiased estimator of the true residual variance and $\sum_i \hat{u}^2 = \sum_i y_i^2 - \hat{\beta}_i^2 \sum_i x_i^2 = 1/16 - (-0.09)^2(100) = 0.35$

(see lecture notes for explanation of mean deviation notation)

Hence $\hat{\sigma}^2 = \frac{0.35}{67-2} = 0.0054$

and so the standard error of $\hat{\beta}_1 = \sqrt{\frac{\hat{\sigma}^2}{\sum_i x_i^2}} = \sqrt{\frac{0.0054}{100}} = 0.0074$

the t test of the null hypothesis that H0: $\beta_1 = 0$
$\implies \hat{t} = \frac{\hat{\beta}_1}{s.e.(\hat{\beta}_1)} = \frac{-0.09}{0.0074} = -12.2$

using a 5% significance level test then since $\left|\hat{t}\right| > t_{N-k}^{\alpha/2} = t_{65}^{0.025} = 2.0$

reject the null hypothesis that the variable has no explanatory power
In the 3-variable model the matrix equivalent formulation for the variance of an individual parameter estimate is given by
$\mathrm{Var}(\hat{\beta}_i) = \hat{\sigma}^2 (x'x)_{ii}^{-1}$

(where $(x'x)_{ii}^{-1}$ is the ith element on the main diagonal of the mean deviation x'x inverse matrix

Also now $\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{N-k}$

we know that the OLS residuals û=y-x$\tilde{\beta}$ (in mean deviation form)

$\therefore \hat{u}'\hat{u} = y'y - \tilde{\beta}' x'x\tilde{\beta}$

(cross-product terms vanish - see lecture notes)

since $\tilde{\beta} = (x'x)^{-1}x'y$

$\implies \hat{u}'\hat{u} = y'y - \tilde{\beta}'(x'x)(x'x)^{-1}x'y$
$= \text{y'y-}\tilde{\beta}'x'y$

using the estimated β values from exercise 1

$= 1.16 - [-0.028 \quad 0.078] \begin{matrix} -9 \\ 10 \end{matrix} = 1.16 - 1.032$

$\therefore \hat{u}'\hat{u} = 0.128$

and so $\hat{\sigma}^2 = \frac{0.128}{67-3} = 0.002$
since the variances of variable x and z are the same in this question then $Var(\tilde{\beta}_1) = Var(\tilde{\beta}_2) = 0.002 * \frac{10}{360}$
and so the standard errors $s.e.(\tilde{\beta}_1) = s.e.(\tilde{\beta}_2) = \sqrt{0.00006} = 0.007$

and a t test of the hypothesis that the coefficient on the variable x is zero (ie that x has no explanatory power in determining y)

$\hat{t}^{\beta_1} = \frac{\hat{\beta}_1 - \hat{\beta}_1^0}{s.e.(\hat{\beta}_1)} = \frac{-0.028}{0.144} = -0.195$
Now since $\left|\hat{t}\right| < t_{N-k}^{\alpha/2} = t_{64}^{0.025} = 2.0$

can't reject the null of no effect

Hence the addition of other variables into a model can often change the size and significance of an estimate

2. Show that $s^2 = \frac{\hat{u}'\hat{u}}{N-k}$
ie the estimated residual sum of squares divided by the sample size minus the number of variables in the model) is an unbiased estimate of the (unobserved) true residual variance

Since $\hat{u} = y - X\hat{\beta} = y - X(X'X)^{-1}X'y$
$= \left[I - X(X'X)^{-1}X'\right]y$
$= My$

where M is a "residual maker" symmetric, idempotent matrix with the properties that

$MX = 0 \qquad My = \hat{u} \qquad M\hat{u} = \hat{u}$
$\implies \left[I - X\{X'X)^{-1}X'\right][y - X\beta] = \hat{u}$

$\therefore y - X(X'X)^{-1}X'y - X\beta + X\beta = \hat{u}$
ie $\quad y - X\hat{\beta} = \hat{u}$

$\therefore \hat{u}'\hat{u} = u'Mu$

and so E$(\hat{u}'\hat{u}) = E(u'Mu)$

Since $u'Mu$ is a scalar re 1_x1

the use the result that the trace of a scalar = sum of elements on the main diagonal which in this case is just the value of the scalar

so E$(\hat{u}'\hat{u}) = E(u'Mu) = E[tr(u'Mu)] = E[tr(M\,\mathrm{u}')]$
- using the result that tr(ABC)=tr(BAC)=tr(BCA)=tr(CBA)

Since M is assumed to be non-stochastic can take it outside the expectation

so E$(\hat{u}'\hat{u}) = tr[ME(u\,\mathrm{u}')]$
$= \sigma^2 tr(M)$
$= \sigma^2[tr(I_N) - tr[X(X'X)^{-1}X']]$
$= \sigma^2[tr(I_N) - tr[(X'X)^{-1}X'X]]$
$= \sigma^2[tr(I_N) - tr[I_k]]$

since the trace is the sum of the elements on the main diagonal

E$(\hat{u}'\hat{u}) = \sigma^2[N-k]$

Hence it follows that s$^2 = \frac{\hat{u}'\hat{u}}{N-k}$ will be an unbiased estimator of σ^2
E$\left(\frac{\hat{u}'\hat{u}}{N-k}\right) = \frac{E(\hat{u}'\hat{u})}{N-k} = \frac{\sigma^2[N-k]}{N-k} = \sigma^2$

(Note that the rank of an idempotent matrix equals its trace
$\rho(M) = tr[I - X(X'X)^{-1}X'] \quad = N-k$

3. Given the following information on a sample of 10 observations
$\Sigma Y = 20$ $\Sigma Y^2 = 88.2$ $\Sigma X_1 = 30$ $\Sigma {X_1}^2 = 92$ $\Sigma X_2 = 40$ $\Sigma {X_2}^2 = 163$ $\Sigma YX_1 = 59$ $\Sigma YX_2 = 88$ $\Sigma X_1 X_2 = 119$

Find the OLS estimates of
a) the slope coefficients on X_1 and X_2 and b) the intercept term
c) Test the hypothesis that the coefficient on X_2 is zero

1st get the 2 slope parameters by writing the model in mean deviation matrix form

Y$= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U$

becomes

y = $\beta_1 x_1 + \beta_2 x_2 + u$

and the OLS estimates $\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} x_1'x_1 & x_1'x_2 \\ x_2'x_1 & x_2'x_2 \end{bmatrix}^{-1} \begin{bmatrix} x_1'y \\ x_2'y \end{bmatrix}$ (1)

where $x_1'x_1 = \sum_i x_{1i}^2 = \sum_i (X_{1i}^2 - \bar{X}_1^2) = \sum_i X_{1i}^2 - N\bar{X}_1^2 = \sum_i X_{1i}^2 - N(\sum_i X_{1i}/N)^2 = \sum_i X_{1i}^2 - (\bar{X}_1^2/N)$

so $x_1'x_1$= 92 - (30)^2/10 = 2

Similarly $x_2'x_2 = \sum_i x_{2i}^2 = \sum_i X_{2i}^2 - (\bar{X}_2^2/N) = 163 - (40)^2/1 = 3$

x$_1'x_2 = \sum_i x_1 x_2 = \sum_i X_1 X_2 - (\sum_i X_{1i}X_{2i} - (\sum_i X_1 \sum_i X_2)/N$
= 119 - (30*40)/10
= -1

(as does x$_1'x_2$ since the matrix is symmetric)

x$_1'y = \sum_i x_1 y = \sum_i X_{1i}y_i - (\sum_i X_{1i} \sum_i y_i)/N$
= 59 - (30*20)/10
= -1

x$_2'y = \sum_i x_2 y = \sum_i X_2 y - (\sum_i X_2 \sum_i y)/N$
= 88 - (40*20)/10
= 8

Now substitute all these terms into (1)

$$\implies \hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 3 \end{bmatrix}^{-1} \begin{bmatrix} -1 \\ 8 \end{bmatrix}$$

$$= \tfrac{1}{5} \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} -1 \\ 8 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

It follows that the OLS estimate of the intercept is given by

$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X} - \hat{\beta}_2\bar{X}$

with

$\bar{Y} = \sum_i y_i/N = 20/10 = 2$

$\bar{X}_1 = \sum_i X_{1i}/N = 30/10 = 3$

$\bar{X}_2 = \sum_i X_{2i}/N = 40/10 = 4$

So $\hat{\beta}_0 = 2 - (1 * 3) - (3 * 4) = -13$

To test the hypothesis that $\beta_2 = 0$ need s.e.$(\hat{\beta}_2) = \sqrt{s^2(x'x)^{-1}_{22}}$

(where the subscript indicates the appropriate element on the main diagonal)
where s^2 = $\frac{\hat{u}'\hat{u}}{N-k} = \frac{y'y-\hat{\beta}'x'x\hat{\beta}}{N-k} = \frac{y'y-\hat{\beta}'x'y}{N-k}$

Since y'y = $\sum_i y_i^2 == \sum_i y_i^2 - (\overline{y}^2/N) = 88.2 - (20)^2/10 = 48.2$

and $\hat{\beta}'x'y = \begin{bmatrix} 1 & 3 \end{bmatrix} \begin{bmatrix} -1 \\ 8 \end{bmatrix} = 23$
then s^2 = $\frac{48.2-23}{10-3} = 3.6$

and since (x'x)$^{-1}$ = $\begin{bmatrix} 3/5 & 1/5 \\ 1/5 & 2/5 \end{bmatrix}$

then s.e.$(\hat{\beta}_2) = \sqrt{3.6 * (2/5)} = 1.2$

To test the null that $\beta_2 = 0$

use $\hat{t} = \frac{\hat{\beta}_2 - \beta_2^{null}}{s.e.(\hat{\beta}_2)} = \frac{3-0}{1.2}$

Given N=10 and k=3 the critical value at the 5% level t$_{n-k}^{\alpha/2}$ =t$_7^{0.025}$ = 2.37

(2-tailed test)

So estimated $\hat{t} > t_{critical}$
Hence reject the null that $\hat{\beta}_2 = 0$ at the 5% level

4. Given the model

y = b$_1 + b_2X_2 + b_3X_3 + u$

$$\hat{\beta} = \begin{bmatrix} 10 & 10 \\ 10 & 30 \end{bmatrix}^{-1} \begin{bmatrix} 7 \\ -7 \end{bmatrix} = (X'X)^{-1}X'y$$

$$= \frac{1}{300-100} \begin{bmatrix} 30 & -10 \\ -10 & 10 \end{bmatrix} \begin{bmatrix} 7 \\ -7 \end{bmatrix} = \begin{bmatrix} 210+70/200 \\ -70-70/200 \end{bmatrix} = \begin{bmatrix} 1.4 \\ -0.7 \end{bmatrix} = \begin{bmatrix} \hat{b}_2 \\ \hat{b}_3 \end{bmatrix}$$

To test the hypotheses need to find the standard errors

Var$(\hat{\beta}) = s^2(X'X)^{-1}$ *where* $s^2 = \frac{\hat{u}'\hat{u}}{N-k} = \frac{y'y-\hat{\beta}'(X'X)\hat{\beta}}{N-k} = \frac{y'y-\hat{\beta}'X'y}{N-k}$

$$= \frac{60 - \begin{bmatrix} 1.4 & -0.7 \end{bmatrix} \begin{bmatrix} 7 \\ -7 \end{bmatrix}}{53-3}$$

= (60-(9.8+4.9))/50

$s^2 = .906$

and so $\text{Var}(\hat{\beta}) = .906 \begin{bmatrix} 30/200 & -10/200 \\ -10/200 & 10/200 \end{bmatrix}$

and $\text{var}(\hat{b_2}) = .906 * 30/200 = 0.136$

$\text{var}(\hat{b_3}) = .906 * 10/200 = 0.045$

so to test i) $b_2 = 0$

use $t = \frac{\hat{b_2} - 0}{s.e.(\hat{b_2})} = \frac{1.4-0}{\sqrt{.136}} = 3.77$

to test ii) $b_3 = 0$

use $t = \frac{\hat{b_3} - 0}{s.e.(\hat{b_3})} = \frac{-0.7-0}{\sqrt{.045}} = -3.29$

Since the critical value at the 5% level for a 2-tailed t test with N-k = 50 degrees of freedom is $t_{50}^{.025} = 2.01$

then in both cases reject the null that coefficient is zero at the 5% level

The 95% confidence interval is given by

$$\Pr\left[\hat{b_2} - t_{n-k}^{\alpha/2} * s.e.(\hat{b_2}) <= b_2 <= \hat{b_2} + t_{n-k}^{\alpha/2} * s.e.(\hat{b_2})\right] = 0.95$$

$\Longrightarrow \Pr[1.4 - 2.01 * .367 <= b_2 <= 1.4 + 2.01 * .367] = 0.95$

$\Longrightarrow \Pr[0.665 <= b_2 <= 2.134] = 0.95$

Similarly

$\Pr[-0.7 - 2.01 * .213 <= b_3 <= -0.7 + 2.01 * .213] = 0.95$

$\Longrightarrow \Pr[-1.126 <= b_3 <= -0.273] = 0.95$

To test whether b_2 and b_3 are equal and opposite ie $b_2 - b_3 = 0$

$$F = \frac{(R\hat{\beta} - r)' \left[s^2 R(X'X)^{-1} R'\right]^{-1} (R\hat{\beta} - r)}{q} \sim F[q, N-k]$$

where now R=[1 -1] r =0 and q=1

and pre and post-multiplying $(X'X)^{-1}$ by R gives a row vector whose elements are a sum of the elements in the 2nd and 3rd rows and columns of X'X

$$s^2 R(X'X)^{-1}R' = s^2[x_{11} - 2x_{23} + x_{33}] = \text{var}(\hat{b}_2) + \text{var}(\hat{b}_3) - 2Cov(\hat{b}_2\hat{b}_3)$$
$$= \text{Var}\left(\hat{b}_2 - \hat{b}_3\right)$$

Hence F becomes $\quad F = \frac{\left(\hat{b}_2 - \hat{b}_3\right)^2}{Var\left(\hat{b}_2 - \hat{b}_3\right)} \tilde{} F[1, N-k]$

(or since the F statistic is the square of the t statistic then could use $t = \frac{\left(\hat{b}_2 - \hat{b}_3\right)}{s.e\left(\hat{b}_2 - \hat{b}_3\right)}$

So $F = \frac{(1.4 - -0.7)^2}{.136 + .045 - 2*(-.045)} = \frac{4.41}{.271} = 16.27 \tilde{} F[1, 50]$

(and the t value = 4.03)
which since from tables the critical value at the 5% level $F_{1,50}^{.05} = 4.05$

then reject the null hypothesis that the coefficients are equal.

5. On the assumption that the relationship between consumption of food, (C), its price, (P) and the income of the consumer, (I) is given by $C = AP^{b1}I^{b2}$, where A is a constant, estimate the income elasticity of demand and compute a 95% confidence interval for this coefficient given a sample size of 20 under the null hypothesis that the coefficient is zero.

The matrix of products and cross-products of the log of these variables (in mean deviation form) is

	LnC	*LnP*	*LnI*
LnC	7.59	3.12	26.99
LnP		29.16	30.80
LnI			133.0

To estimate the income elasticity need to first make the model log linear
$LnC = LnA + b_1 LnP + b_2 LnI + u$

and hence the income elasticity of demand $\frac{\%\Delta Demand}{\%\Delta Income} = \frac{\Delta C/C}{\Delta I/I} = \frac{\delta LnC}{\delta LnI} = b_2$

From the covariance matrix

$$X'X = \begin{bmatrix} 29.16 & 30.8 \\ 30.8 & 133.0 \end{bmatrix} \quad \text{and } X'y = \begin{bmatrix} 3.12 \\ 26.99 \end{bmatrix}$$

so can find OLS estimates of slope parameters by estimating model in mean deviation form $\hat{\beta} = (x'x)^{-1} x'y$

$$= \frac{1}{(29.16*133-(30.8)^2)} \begin{bmatrix} 133 & -30.8 \\ -30.8 & 29.16 \end{bmatrix} \begin{bmatrix} 3.12 \\ 26.99 \end{bmatrix}$$

$$\Longrightarrow \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \end{bmatrix} = \begin{bmatrix} -0.142 \\ 0.236 \end{bmatrix}$$

ie the income elasticity is 0.236 (a 1% increase in income raises consumption demand by 0.236%)

To test whether this effect is significantly different from zero use t $= \frac{\left(\hat{b}_2 - b_2^{null}\right)}{s.e\left(\hat{b}_2\right)}$

where var$\left(\hat{b}_2\right) = s^2(x'x)^{-1}_{22}$

ie the appropriate element on the main diagonal of (x'x)$^{-1}$

s$^2 = \frac{\hat{u}'\hat{u}}{N-k} = \frac{y'y-\hat{\beta}'x'x\hat{\beta}}{N-k} = \frac{y'y-\hat{\beta}'x'y}{N-k}$

$$= \frac{7.59 - \begin{bmatrix} -0.142 & 0.236 \end{bmatrix} \begin{bmatrix} 3.12 \\ 26.99 \end{bmatrix}}{20-3} \qquad = 0.098$$

and $(x'x)^{-1}_{22} = \frac{29.16}{(29.16*133-(30.8)^2)} = 0.0099$

Hence t $= \frac{0.236-0}{\sqrt{0.098*0.0099}} = 7.57$

Given critical value at 5% level in 2-tailed test with N-k degrees of freedom in this case is t$^{.025}_{20-3} = 2.11$

and estimated t > t$_{critical}$ so reject null that true income elasticity is zero

Follows that 95% confidence level given by

$$\Pr\left[\hat{b}_2 - t^{\alpha/2}_{n-k} * s.e.(\hat{b}_2) <= b_2 <= \hat{b}_2 + t^{\alpha/2}_{n-k} * s.e.(\hat{b}_2)\right] = 0.95$$

$$\Longrightarrow \Pr[0.236 - 2.11 * .099 <= b_2 <= 0.236 - 2.11 * .099] = 0.95$$
$$\Longrightarrow \Pr[0.170 <= b_2 <= 0.302] = 0.95$$