SC472 Econometrics Data analysis Using R **Date** – April 8, 2023

Assignment – 1

Name – Dev Jadav

Student ID - 202001016

3 The following table contains the *ACT* scores and the *GPA* (grade point average) for eight college students. Grade point average is based on a four-point scale and has been rounded to one digit after the decimal.

Student	*GPA*	*ACT*
1	2.8	21
2	3.4	24
3	3.0	26
4	3.5	27
5	3.6	29
6	3.0	25
7	2.7	25
8	3.7	30

(i) Estimate the relationship between *GPA* and *ACT* using OLS; that is, obtain the intercept and slope estimates in the equation

$$\widehat{GPA} = \hat{\beta}_0 + \hat{\beta}_1 ACT.$$

Comment on the direction of the relationship. Does the intercept have a useful interpretation here? Explain. How much higher is the *GPA* predicted to be if the *ACT* score is increased by five points?

$y_i \Rightarrow GPA_i$

$x_i \Rightarrow ACT_i$

here $n = 8$

$\Sigma x_i = 21 + 24 + 26 + 27 + 25 + 29 + 25 + 30$

$\bar{x} = \dfrac{\Sigma x_i}{n}$

$= \dfrac{207}{8} = \boxed{25.875}$

Similarly $\bar{y} = \dfrac{\Sigma y_i}{n} = \dfrac{27.5}{8} = 3.2125$

$\Rightarrow \Sigma(x - x_i)(y - y_i)$

$\Rightarrow \Sigma(x_i - 25.875)(y_i - 3.2125)$

$= \boxed{5.8125}$

$\Sigma(x_i - x)^2 = 56.875$

$\hat{\beta}_1 = \dfrac{\mathrm{cov}(x, y)}{\mathrm{var}(x)} = \dfrac{\Sigma(x_i - \bar{x})(y_i - \hat{y})}{\Sigma(x_i - \bar{x})^2}$

$= \dfrac{5.8125}{56.875} = \boxed{0.1022}$

(ii) Compute the fitted values and residuals for each observation, and verify that the residuals (approximately) sum to zero.

No	GPA	$\widehat{GPA}$	$\hat{u}_i$
1	2.8	2.7143	00857
2	3.4	3.0209	0.3791
3	3.0	3.2253	- 0.2253
4	3.5	3.3275	0.1725
5	3.6	3.5319	0.0681
6	3.0	3.1231	- 0.1231
7	2.7	3.1231	- 0.4321
8	3.7	3.6341	0.0659

$\Sigma \hat{u}_i = -0.002 \approx 0$

(iii) What is the predicted value of *GPA* when *ACT* = 20?

iii)

$ACT = 20$

$\widehat{GPA} = \hat{B}_0 + \hat{B}_i ACT$

$= 0.5681 + 0.1022(20)$

$= \boxed{2.61}$

(iv) How much of the variation in *GPA* for these eight students is explained by *ACT*? Explain.

$$R^2 = 1 - \frac{0.4347}{1.0258}$$

$$\approx \boxed{0.577}$$

$\Rightarrow$ 57.7% Variation in GPA is explained by ACT

4 The data set BWGHT.RAW contains data on births to women in the United States. Two variables of interest are the dependent variable, infant birth weight in ounces (*bwght*), and an explanatory variable, average number of cigarettes the mother smoked per day during pregnancy (*cigs*). The following simple regression was estimated using data on $n = 1{,}388$ births:

$$\widehat{bwght} = 119.77 - 0.514\, cigs$$

(i) What is the predicted birth weight when $cigs = 0$? What about when $cigs = 20$ (one pack per day)? Comment on the difference.

i)

cigs = 0

bweight = $\boxed{119.77 \text{ ounces}}$

cigs = 20

bweight = 119.77 − 0.514(20)

= $\boxed{109.49 \text{ ounces}}$

There is difference of 8.6% of baby's weight when one pack of cigarette is smoked on avg daily.

(ii) Does this simple regression necessarily capture a causal relationship between the child's birth weight and the mother's smoking habits? Explain.

ii) This simple regression capture a causal relationship between the child's birth weight and mother's smoking habits to some extent. But there are other factors involved as well. eg genes, prenatal care etc.

(iii) To predict a birth weight of 125 ounces, what would *cigs* have to be? Comment.

iii)

Birth weight => 125 ounces

$125 = 119.77 - 0.514\ cigs$

$\boxed{cigs \approx -10}$

=> We are modelling on single variable and childbirth is much more complex variable. Also our dataset is small

$Weight_{max} \approx 119.77$ ounces.

(iv) The proportion of women in the sample who do not smoke while pregnant is about .85. Does this help reconcile your finding from part (iii)?

iv)

A better model can be prepared, if we are provided data regarding => proportion of women in sample who do not smoke while pregnant is 0.85.

It will be more robust.

5 In the linear consumption function

$$\widehat{cons} = \hat{\beta}_0 + \hat{\beta}_1 inc,$$

the (estimated) *marginal propensity to consume* (MPC) out of income is simply the slope, $\hat{\beta}_1$, while the *average propensity to consume* (APC) is $\widehat{cons}/inc = \hat{\beta}_0/inc + \hat{\beta}_1$. Using observations for 100 families on annual income and consumption (both measured in dollars), the following equation is obtained:

$$\widehat{cons} = -124.84 + 0.853\ inc$$

$$n = 100,\ R^2 = 0.692.$$

(i) Interpret the intercept in this equation, and comment on its sign and magnitude.

Q-5

i)

If we say Income is $0 then the consumption is $ -124.48. But consumption is never negative.

So the Consumption function might not be a very good predictor for lower income levels.

(ii) What is the predicted consumption when family income is $30,000?

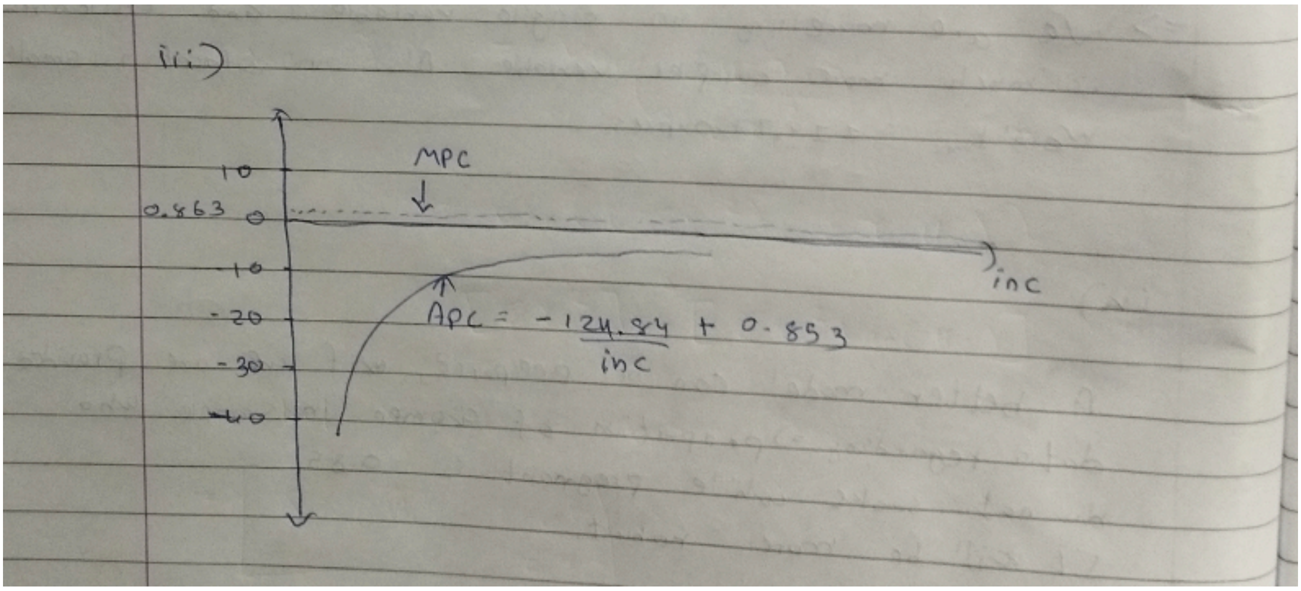
ii)

Family income = $30,000

$\widehat{Cons} = -124.84 + 0.853(30,000)$

$= \boxed{\$\ 25465.16}$

(iii) With *inc* on the *x*-axis, draw a graph of the estimated MPC and APC.



8 Consider the standard simple regression model $y = \beta_0 + \beta_1 x + u$ under the Gauss-Markov Assumptions SLR.1 through SLR.5. The usual OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased for their respective population parameters. Let $\tilde{\beta}_1$ be the estimator of β_1 obtained by assuming the intercept is zero (see Section 2.6).

(i) Find $E(\tilde{\beta}_1)$ in terms of the x_i, β_0, and β_1. Verify that $\tilde{\beta}_1$ is unbiased for β_1 when the population intercept (β_0) is zero. Are there other cases where $\tilde{\beta}_1$ is unbiased?

Q-8

$\tilde{y} = \tilde{\beta}_1 x$

We minimize sum of Residuals to obtain best slope

squared sum

$$\min\left(\Sigma\left(y_i - \tilde{\beta}_1 x_i\right)^2\right)$$

$$w = \Sigma\left(y_i - \tilde{\beta}_1 x_i\right)^2$$

$$\frac{dw}{d\beta_1} = (-2)\ \Sigma x_i\left(y_i - \tilde{\beta}_1 x_i\right) = 0$$

$$\Sigma x_i\left(y_i - \tilde{\beta}_1 x_i\right) = 0$$

$$\boxed{\tilde{\beta}_1 = \frac{\Sigma x_i y_i}{\Sigma x_i^2}}$$

$$E(\tilde{\beta}_1) = \frac{\beta_0 \Sigma x_i}{\Sigma x_i^2} + \beta_1$$

When $\beta_0 = 0$ $\tilde{\beta}_1$ is unbiased

Also for $\Sigma x_i = 0$ it is unbiased

(ii) Find the variance of $\tilde{\beta}_1$. (*Hint:* The variance does not depend on β_0.)

ii)

$$Var(\tilde{\beta}_1) = Var\left(\frac{\Sigma u_i x_i}{\Sigma x_i^2}\right)$$

$$= var((\Sigma x_i^2)^{-1}(\Sigma x_i u_i))$$

$$= [\Sigma x_i^2]^{-2} [\Sigma x_i^2 Var(u_i)]$$

Let $var(u_i) = \sigma^2$

$$\boxed{var(\tilde{\beta}_1) = \frac{\sigma^2}{\Sigma x_i^2}}$$

(iii) Show that $\operatorname{Var}(\tilde{\beta}_1) \leq \operatorname{Var}(\hat{\beta}_1)$. [*Hint:* For any sample of data, $\sum_{i=1}^{n} x_i^2 \geq \sum_{i=1}^{n}(x_i - \bar{x})^2$, with strict inequality unless $\bar{x} = 0$.]

iii)

$$Var(\tilde{\beta}_1) \leq Var(\hat{\beta}_1)$$

$$var(\hat{\beta}_1) = \frac{\sigma^2}{\Sigma (x_i - \bar{x})^2}$$

$$Var(\tilde{\beta}_1) = \frac{\sigma^2}{\Sigma x_i^2}$$

As $\Sigma x_i^2 \geq \Sigma (x_i - \bar{x})^2$

$$\frac{1}{\Sigma x_i^2} \leq \frac{1}{\Sigma (x_i - \bar{x})^2}$$

$$\frac{\sigma^2}{\Sigma x_i^2} \leq \frac{\sigma^2}{\Sigma (x_i - \bar{x})^2}$$

$$var(\tilde{\beta}_1) \leq var(\hat{\beta}_1)$$

(iv) Comment on the tradeoff between bias and variance when choosing between $\hat{\beta}_1$ and $\tilde{\beta}_1$.

iv)

Bias in $\tilde{\beta}_1$ increases as the mean increases, But the variance of $\hat{\beta}_1$ also increases when β_0 is small) bias in $\tilde{\beta}_1$ is also small (for the given sample size)

10 Let $\hat{\beta}_0$ and $\hat{\beta}_1$ be the OLS intercept and slope estimators, respectively, and let $\bar{u}$ be the sample average of the errors (not the residuals!).

(i) Show that $\hat{\beta}_1$ can be written as $\hat{\beta}_1 = \beta_1 + \sum_{i=1}^{n} w_i u_i$ where $w_i = d_i/\mathrm{SST}_x$ and $d_i = x_i - \bar{x}$.

Q-10)

i)

$$\hat{\beta}_1 = \beta_1 + \sum_{i=1}^{n} w_i u_i$$

$w_i = \dfrac{d_i}{SST_x}$, $d_i = x_i - \bar{x}$

$$\hat{\beta}_1 = \frac{Cov(x_i, y_i)}{Var(x_i)}$$

$$= \frac{\Sigma (x_i - \bar{x}) y_i}{\Sigma (x_i - \bar{x})^2}$$

$$= \frac{\Sigma (x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i)}{\Sigma (x_i - \bar{x})^2}$$

$\Rightarrow \Sigma (x_i - \bar{x})^2 = SST_x$

$$\hat{\beta}_1 = \frac{\Sigma (x_i - \bar{x})\beta_0 + \Sigma (x_i - \bar{x})\beta_1 x_i + \Sigma (x_i - \bar{x}) u_i}{SST_x}$$

$\Rightarrow \beta_0 \Sigma (x_i - \bar{x}) = 0$

$$\hat{\beta}_1 = \frac{\beta_1 \Sigma (x_i - \bar{x}) x_i + \Sigma (x_i - \bar{x}) u_i}{SST_x}$$

$\Rightarrow \Sigma (x_i - \bar{x}) x_i = \Sigma (x_i - \bar{x})^2 = SST$

$$\hat{\beta}_1 = \frac{\beta_1 SST_x + \Sigma d_i u_i}{SST_x}$$

$$\boxed{\hat{\beta}_1 = \beta_1 + \Sigma w_i u_i}$$

(ii) Use part (i), along with $\sum_{i=1}^{n} w_i = 0$, to show that $\hat{\beta}_1$ and $\bar{u}$ are uncorrelated. [*Hint*: You are being asked to show that $E[(\hat{\beta}_1 - \beta_1) \cdot \bar{u}] = 0$.]

ii)

$\mathcal{E} w_i = 0$

$\hat{\beta}_1$ and $\bar{\mu}$ are uncorrelated

$$\mathrm{corr}(\hat{\beta}, \bar{u}) = \frac{E[(\hat{\beta}_1 - \beta_1)(\bar{\mu} - \bar{\bar{u}})]}{\sigma \hat{\beta}_1 \, \sigma \bar{\mu}}$$

$\because \bar{u} = \bar{\bar{\mu}}$

$$E[(\hat{\beta}_1 - \beta_1)(\bar{u} - \bar{\bar{u}})] = 0$$

$$\boxed{\mathrm{corr}(\hat{\beta}, \bar{\mu}) = 0}$$

iii)

$$\hat{\beta}_0 = \beta_0 + \bar{u} - (\hat{\beta}_1 - \beta_1)\bar{x}$$

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{u}_i$$

$$\bar{y} = \beta_0 + \beta_1 \bar{x} + \bar{u} \quad -(1)$$

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \quad -(2)$$

(iii) Show that $\hat{\beta}_0$ can be written as $\hat{\beta}_0 = \beta_0 + \bar{u} - (\hat{\beta}_1 - \beta_1)\bar{x}$.

iii)

$$\hat{\beta}_0 = \beta_0 + \bar{u} - (\hat{\beta}_1 - \beta_1)\bar{x}$$

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{u}_i$$

$$\bar{y} = \beta_0 + \beta_1 \bar{x} + \bar{u} \quad -(1)$$

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \quad -(2)$$

From equations (1) & (2)

$$\hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \beta_0 + \beta_1 \bar{x} + \bar{u}$$

$$\hat{\beta}_0 = \beta_0 + \bar{u} + (\beta_1 - \hat{\beta}_1)\bar{x}$$

(iv) Use parts (ii) and (iii) to show that $\mathrm{Var}(\hat{\beta}_0) = \sigma^2/n + \sigma^2(\bar{x})^2/\mathrm{SST}_x$.

iv)

$$\mathrm{Var}(\hat{\beta}_0) = \frac{\sigma^2}{n} + \frac{\sigma^2(\bar{x})^2}{SST_x}$$

$$\hat{\beta}_0 = \beta_0 + \bar{u} - \bar{x}(\hat{\beta}_1 - \beta_1)$$

$$\mathrm{var}(\hat{\beta}_0) = \mathrm{Var}(\beta_0) + \mathrm{var}(\bar{u}) + \mathrm{var}(-\bar{x}(\hat{\beta}_1 - \beta_1))$$

$$= 0 + \mathrm{var}(\bar{u}) + (-\bar{x})^2 \mathrm{var}(\hat{\beta}_1 - \beta_1)$$

$$= \frac{\sigma^2}{n} + (\bar{x})^2 \frac{\sigma^2}{SST_x}$$

$$\boxed{\therefore \mathrm{Var}(\hat{\beta}_0) = \frac{\sigma^2}{n} + \frac{\sigma^2(\bar{x})^2}{SST_x}}$$

(v) Do the algebra to simplify the expression in part (iv) to equation (2.58).
[*Hint*: $\mathrm{SST}_x/n = n^{-1}\sum_{i=1}^{n} x_i^2 - (\bar{x})^2$.]

v)

$$SST_x = \Sigma(x_i - \bar{x})^2$$

$$= \Sigma(x_i)^2 - 2\,\Sigma(\bar{x})\,x_i + \Sigma(\bar{x})^2$$

$$= \Sigma(x_i)^2 - 2\bar{x}\,\Sigma x_i + n(\bar{x})^2$$

$$= \Sigma(x_i)^2 - n(\bar{x})^2$$

$$\mathrm{Var}(\hat{\beta}_0) = \sigma^2\left[\frac{1}{n} + \frac{(\bar{x})^2}{SST_x}\right]$$

$$\mathrm{Var}(\hat{\beta}_0) = \sigma^2\left[\frac{SST_x + n(\bar{x})^2}{n\,SST_x}\right]$$

$$SST_x = \Sigma(x_i)^2 - n(\bar{x})^2$$

$$\mathrm{Var}(\hat{\beta}_0) = \sigma^2 n^{-1}\left[\frac{\Sigma x_i^2 - n\bar{x}^2 + n\bar{x}^2}{SST_x}\right]$$

$$= \frac{\sigma^2\,\Sigma x_i^2}{n \cdot SST_x}$$

$$\mathrm{Var}(\hat{\beta}_0) = \frac{\sigma^2 n^{-1}\,\Sigma x_i^2}{\Sigma(x_i - \bar{x})^2}$$