Econ 102_A01-A04 (Analysis of Economic Data): Cameron Fall 2022
Solutions to Second Midterm Exam

Version A

1.(a) $\bar{x} = (1+2+3)/3 = 2$ and $\bar{y} = (5+2+2)/3 = 3$.

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
1	5	-1	2	-2
2	2	0	-1	0
3	2	1	-1	-1

$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = -2 + 0 - 1 = -3$.
(b) $r_{xy} = s_{xy}/\sqrt{s_{xx} \times s_{yy}} = 5/\sqrt{100 \times 25} = 5/50 = 0.1$.
(c) $R^2 = r_{xy}^2 = 0.1^2 = 0.01$.
(d) 1. Model $y_i = \beta_1 + \beta_2 x_i + u_i$
2. Zero conditional mean error. $\mathrm{E}[u_i|x_i] = 0$ for all i.
3. Constant conditional variance error. $\mathrm{Var}[u_i|x_i] = \sigma_u^2$ for all i.
4. Independent errors. u_i independent of u_j for all $i \neq j$.
(e) Assumptions 1 and 2 are necessary for unbiasedness.

2.(a) The regression is of energy in millions of BTU against income in billions of dollars.
So the slope gives the increase in energy in millions of BTU for a \$1 billion dollar increase in income.
Hence a \$1,000,000,000 or \$1 billion increase in income is associated with 1.571 million more BTU consumed.
(b) A 95% confidence interval for the population slope parameter is $(1.394, 1.747)$ from the output.
(c) A 90% confidence interval for the population slope parameter is $b_2 \pm t_{49;.05} \times s_{b_2} = 1.571 \pm invttail(49, .05) \times 0.0877 = 1.571 \pm 1.677 \times 0.0877 = 1.571 \pm 0.147 = (1.424, 1.718)$.
(d) $H_0 : \beta_2 = 0$ against $H_a : \beta_2 \neq 0$.
From output p-value= $0.000 < 0.05$. So reject H_0.
Conclude there is a statistically significant relationship at level 0.05.
(e) `display 2*ttail(49,17.91)`
(f) $H_0 : \beta_2 \leq 1$ against $H_a : \beta_2 > 1$. Claim is the alternative hypothesis.
$t = (b_2 - 1)/s_{b2} = (1.5705 - 1.0)/0.0877 = 0.5705/0.0877 = 6.505$.
Since $t = 6.505 > t_{49;.01} = invttail(49, .01) = 2.405$ reject H_0 at level 0.01. Support the claim.

3.(a) Quite well. $R^2 = 0.867$ is quite high.
(b) For Alaska predicted $y = 97.82 + 1.571 \times 18.8 = 127.35$ million BTU.
(c) This is s_e the standard error of the regression (estimate of the standard deviation σ_u) or the root mean square error.

4.(a) $t = (\widehat{\theta} - \theta^*)/se(\widehat{\theta}) = (19 - 10)/5 = 9/5 = 1.8$ so $|t| < 1.96$ and do not reject H_0.
(b) 95% CI is $\widehat{\theta} \pm 1.96 \times se(\widehat{\theta}) = 19 \pm 1.96 \times 5 = 19 \pm 9.8 = (9.2, 28.8)$.
or 95% CI is $\widehat{\theta} \pm 2 \times se(\widehat{\theta}) = 19 \pm 2 \times 5 = 19 \pm 10 = (9, 29)$.
(c) TSS = ExplainedSS + ResidualSS = $100 + 50 = 150$.
R^2 = ExplainedSS/TSS = $50/150 = 0.333$. (or use $R^2 = 1-$ Residual/TSS= $1 - 100/150 = 0.333$).
(d) 4.

Versions A and B: Multiple Choice

Question	**1.**	**2.**	**3.**	**4.**	**5.**
Answer Version A	*a*	*a*	*b*	*d*	*c*
Answer Version B	*b*	*b*	*a*	*a*	*d*

Econ 102_A01-A04 Version B

1.(a) 1. Model $y_i = \beta_1 + \beta_2 x_i + u_i$
2. Zero conditional mean error. $\mathrm{E}[u_i|x_i] = 0$ for all i.
3. Constant conditional variance error. $\mathrm{Var}[u_i|x_i] = \sigma_u^2$ for all i.
4. Independent errors. u_i independent of u_j for all $i \neq j$.
(b) All of assumptions 1 to 4 are necessary.
(c) $\bar{x} = (1+2+3)/3 = 2$ and $\bar{y} = (7+1+1)/3 = -3$.

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
1	7	-1	4	-4	1
2	1	0	-2	0	0
3	1	1	-2	-2	1

$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = -4 + 0 - 2 = -6$. $\sum_{i=1}^n (x_i - \bar{x})^2 = 2$. $b = -6/2 = 3$.
(d) $r_{xy} = s_{xy}/\sqrt{s_{xx} \times s_{yy}} = 10/\sqrt{100 \times 25} = 10/50 = 0.2$.
(e) $R^2 = r_{xy}^2 = 0.2^2 = 0.04$.

2.(a) The regression is of energy in millions of BTU against income in billions of dollars.
So the slope give the increase in energy in millions of BTU for a \$1 billion dollar increase in income.
A \$1,000,000,000 or \$1 billion increase in income is associated with 1.467 million more BTU consumed.
(b) A 95% confidence interval for the population slope parameter is $(1.273, 1.660)$ from the output.
(c) A 99% confidence interval for the population slope parameter is $b_2 \pm t_{38;.005} \times s_{b_2} = 1.467 \pm invttail(38, .005) \times 0.0956 = 1.467 \pm 2.712 \times 0.0956 = 1.467 \pm 0.259 = (1.208, 1.726)$.
(d) $H_0 : \beta_2 = 0$ against $H_a : \beta_2 \neq 0$. From output p-value= $0.000 < 0.05$. So reject H_0.
Conclude there is a statistically significant relationship at level 0.05.
(e) `display 2*ttail(38,15.34)`
(f) $H_0 : \beta_2 \leq 1$ against $H_a : \beta_2 > 1$. Claim is the alternative hypothesis.
$t = (b_2 - 1)/s_{b2} = (1.467 - 1.0)/0.0956 = 0.467/0.0956 = 4.88$.
Since $t = 4.88 > t_{38;.05} = invttail(38, .05) = 1.686$ reject H_0 at level 0.05. Suppport the claim.

3.(a) Quite well. $R^2 = 0.861$ is quite high.
(b) For Alaska predicted $y = 109.85 + 1.466 \times 18.8 = 137.41$ million BTU.
(c) This is s_e the standard error of the regression (estimate of the standard deviation σ_u) or the root mean square error..

4.(a) $t = (\widehat{\theta} - \theta^*)/se(\widehat{\theta}) = (39 - 20)/10 = 19/10 = 1.9$ so $|t| < 1.96$ and do not reject H_0.
(b) 95% CI is $\widehat{\theta} \pm 1.96 \times se(\widehat{\theta}) = 39 \pm 1.96 \times 10 = 39 \pm 19.6 = (19.4, 58.6)$.
or 95% CI is $\widehat{\theta} \pm 2 \times se(\widehat{\theta}) = 39 \pm 2 \times 10 = 19 \pm 20 = (19, 59)$.
(c) 3.
(d) TSS = ExplainedSS + ResidualSS = 200 + 50 = 250.
R^2 = ExplainedSS/TSS = 200/250 = 0.8. (or use R^2 = 1− Residual/TSS= $1 - 50/250 = 0.8$).

The course grade is based on a curve from the combined scores of midterm 1 (20%), midterm 2 (20%), final (40%), and assignments (20%). This suggestive curve for this exam has average GPA 2.81 for this exam.

Scores out of	35
75*th* percentile	30 (86%)
Median	27.5 (79%)
25*th* percentile	22.5 (64%)

A+	34 and above	C+	23.5 and above
A	30 and above	C	21.5and above
A-	29 and above	C-	20 and above
B+	27.5 and above	D+	18.5 and above
B	26.5 and above	D	16.5 and above
B-	25 and above	D-	15 and above