Economics 102 A01-A04: Analysis of Economic Data Cameron Fall 2022 U.C.-Davis
Answers to Final Exam Versions A and B

Version A

1.(a) The graph is one that is right-skewed as mean of 677 exceeds the median of 574 and skewness=2.67. The scale is from 0 to approximately 4000. Half the area under the curve is left of the median and half is to the right of the median.
(b) 95% confidence interval directly from computer output is (651.1, 702.4).
(c) $H_0 : \mu \geq 700$ versus $H_a : \mu < 700$ where claim is the alternative hypothesis.
From output $se(\bar{x}) = 13.069$.
So $t = (\bar{x} - \mu_0)/se(\bar{x}) = (676.799 - 700)/13.069 = -23.201/13.069 = -1.775$.
Reject H_0 that $\mu \geq 700$ at level .05 as $t = -1.775 < -t_{.05;1114} = -1.646$.
So support the claim that population mean food expenditure is less than 700 Rand.
(d) `fedu` is most highly correlated with `food` with correlation coefficient `0.1413`.
(e) $b_1 = 676.5$. Reason: $\widehat{y} = b_1 + b_2 d$ equals b_1 if $d = 0$ so $b_1 = \bar{y} = 676.5$ for those with $d = 0$.
$b_2 = 681.7 - 676.5 = 5.2$. Reason: $\widehat{y} = b_1 + b_2 d$ equals $b_1 + b_2$ if $d = 1$ so $b_1 + b_2 = \bar{y} = 681.7$ for those with $d = 1$.

2.(a) Food expenditure increases by $0.08478 \times 1000 = 84.78$ Rand.
(b) 95% conf. interval for β_2 is (0.07004, 0.0995) from output.
(c) 99% conf. interval for β_2 is $b_2 \pm t_{.005;1115-2} \times s_{b_2} = 0.08478 \pm 2.580 \times 0.007510 = 0.08478 \pm 0.01938 = (0.0654, 0.1046)$.
(d) $H_0 : \beta_2 = 0.09$ against $H_a : \beta_2 \neq 0.09$.
$t = (b_2 - 0.09)/s_b = (0.08478 - 0.09)/0.007510 = -0.695$.
$|t| = 0.695 < t_{.025;1115-2} = 1.962$.
Do not reject H_0. Conclude that at level 5% the `income` coefficient does not differ from 0.09.
(e) Most likely errors are heteroskedastic. One clue is that `food` is right skewed. And we expect more variability around the line for high values of `food` and `income` than around low values.
(f) `regress food income, vce(robust)`

3.(a) Prediction for `food` is $\widehat{y}^* = b_1 + b_2 x^* = 533.127 + .0847798 \times 1000 = 617.91$. Food expenditure is 618 Rand.
(b) The 95% confidence interval is at least $\widehat{y}^* \pm 1.96 \times s_e = \widehat{y}^* \pm 1.962 \times 413.57 = \widehat{y}^* \pm 811$. So the width of the confidence interval is at least $2 \times 811 = 1622$, which is greater than 500.
(c) Four assumptions:
1. $y_i = \beta_1 + \beta_2 x_i + u_i$ where here y is Food and x is Income
2. u_i has mean 0 and is uncorrelated with x
3. u_i has variance σ^2 that does not vary with x
4. u_i is uncorrelated with u_j (errors for different observations are uncorrelated)

(d) Yes. [Reason: To get the exact result that tests and confidence intervals are based on the $T(14)$ distribution requires the additional assumption of normally distributed errors not just i.i.d. errors.]
(e) Yes. [Reason: As long as assumptions 1-2 still holds OLS remains unbiased. We would have to use cluster-robust standard errors to get the correct standard errors of the OLS coefficients].

Version A (continued)

4.(a) A one standard deviation increase in mother's education is associated with a $8.47 \times 3.60 = 30.5$ increase in food spending.
(b) Yes, the five regressors are jointly significant at 5% since the F-statistic of overall fit equals 32.39 with $p = 0.0000 < 0.05$.
(c) Yes, the two regressors are jointly statistically significant at 5% since `test metro rural` gives $F = 5.84$ with $p = 0.0030 < 0.05$.
(d) There are several methods. One method given in the course is to confirm that each of the three variables only takes value 0 or 1 and then generate `total=metro+urban+rural` and show it takes value 1 for all observations.
(e) Yes. The only change is the choice of which of the three mutually exclusive dummies is the omitted category. So the coefficients of the other regressors, including `income`, are unchanged.
(f) Use $\overline{R}^2$ (adjusted R^2) as this penalizes larger models. Since $0.1235 > 0.1019$ the larger model fits better.
Alternatively favor model with lower s_e^2. Since $408.57 < 413.57$ the larger model fits better.

5.(a) A 1% increase in income is associated with a 0.220% increase in food expenditure.
(b) One more year of mother's education is associated with a proportionate increase of 0.0163 or a 1.63% increase in food expenditure.

6.(a) $\sum_{i=1}^{n} z_i = \sum_{i=1}^{3} 1 + 2i^2 = (1 + 2 \times 1^2) + (1 + 2 \times 2^2) + (1 + 2 \times 3^2) = 3 + 9 + 19 = 31$.
(b) $Y_i \sim (40, 100)$ so $\bar{Y} \sim (40, 100/n) \sim (40, 100/100) = (40, 1)$.
Expect mean 40, standard deviation 1 and a normal distribution by central limit theorem.
(c) $E[X] = 0.4 \times 1 + 0.2 \times 3 + 0.4 \times 5 = 3.0$.
$Var[X] = 0.4 \times (1-3)^2 + 0.2 \times (3-3)^2 + 0.4 \times (5-3)^2 = 1.6 + 0 + 1.6 = 3.2$.
(d) $\bar{x} = (15 + 20 + 40)/3 = 75/3 = 25$.
$\sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1) = \{(-10)^2 + (-5)^2 + (15)^2\}/2 = 350/2 = 175$.
Standard deviation is $\sqrt{175} = 13.23$.

7.(a) Missing entry (A): $R^2 = ModelSS/TotalSS = 810/1080 = 0.75$.
(b) Missing entry (B): RMSE$= \sqrt{ResSS/(n-3)} = \sqrt{(1080-810)/18} = \sqrt{15} = 3.873$.
(c) Missing entry (C): $t = b/se$ so $se = b/t = 9/1.5 = 6$.
(d) Missing entry (D): $t = b/se = 6/3.0 = 2$ has $p = \Pr[|T_{18}| > 2.0] \simeq 0.5$.

Multiple Choice Version A:
1. b (=72/4) **2. c** **3. b** **4. b** **5. a** **6. c** **7. d** (=20/4) **8. a** **9. d** **10. b**

The average GPA for the curve for this exam is 2.75.
Course grade is determined by curve based on the combined course score.

Scores out of	60	
75th percentile	52	(87%)
Median	44	(73%)
25th percentile	37	(62%)

		C+	37 and above
A	52 and above	C	34 and above
A-	49 and above	C-	31 and above
B+	46 and above	D+	28 and above
B	43 and above	D	26 and above
B-	40 and above	D-	24 and above