Econ 102 (Analysis of Economic Data): Cameron Spring 2015
Solutions to Second Midterm Exam

Version A

1.(a) Here $se(\bar{x}) = sd(x)/\sqrt{n} = 400/\sqrt{100} = 40$.Approximate 95% CI is $50 \pm 2 \times 40 = (-30, 130)$.
(b) 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 $i \neq j$.
(c) Assumptions 1 and 2.
(d) $r_{xy} = s_{xy}/[s_x \times s_y] = 1/[\sqrt{4} \times \sqrt{1}] = 1/2 = 0.5$.
(e) $b_2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})/\frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2 = s_{xy}/s_x^2 = 1/4 = 0.25$.
2.(a) The mean out-of-state fees and tuition increase by $1000b_2 = 1000 \times 1.373512 = 1,374$ dollars.
(b) A 95% confidence interval for β_2 is (0.9039, 1.8432) from the regression output.
(c) A 99% confidence interval for β_2 is $b_2 \pm t_{.005,.48} \times s_{b_2}$
$= 1.3735 \pm invttail(48, .005) \times .2336 = 1.3735 \pm 2.6822 \times .2336 = 1.3735 \pm .6266 = (.7469, 2.0001)$.
(d) $H_0 : \beta_2 = 0$ against $H_a : \beta_2 \neq 0$.
Using p-value approach: $p = 0.000 < 0.05$. So reject H_0.
Alternatively and equivalently $|t| = 5.88 > t_{.025,48} = invttail(48, .025) = 2.011$. So reject H_0.
Conclude there is a statistically significant relationship at level 0.05.
(e) $H_0 : \beta_2 \leq 1$ against $H_a : \beta_2 > 1$. (Make the claim the alternative)
$t = (b_2 - 1)/s_{b2} = (1.3735 - 1)/.2336 = 1.599$.
Since $t = 1.599 < t_{.01,48} = invttail(48, .01) = 2.407$ we do not reject H_0.
Conclude that out-of-state tuition does not rise by more than \$1 for each \$1 increase in in-state tuition.
(f) $\widehat{y} = b_1 + b_2 x = 12919.9 + 1.3734 \times 12972 =$ \$30,736< actual \$35,850.
Berkeley's out-of-state is \$5,000 higher than predicted by the model.

3.(a) $b_2 = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})/\sum_{i=1}^n (x_i - \bar{x})^2 = (28 + 0 + 42)/(4 + 1 + 9) = 70/14 = 5$.
$\bar{x} = (1 + 2 + 6)/3 = 3$ and $\bar{y} = (14 + 28 + 42) = 28$.
So $b_1 = \bar{y} - b_2\bar{x} = 28 - 5 \times 3 = 13$.
(b) $e_1 = y_1 - \widehat{y}_1 = 14 - (13 + 5 \times 1) = 14 - 18 = -4$.
$e_2 = y_2 - \widehat{y}_2 = 28 - (13 + 5 \times 2) = 28 - 23 = 5$.
$e_3 = y_3 - \widehat{y}_3 = 42 - (13 + 5 \times 6) = 42 - 43 = -1$.
$s_e^2 = \sum_{i=1}^n (y_i - \widehat{y}_i)^2/(n-2) = [(-4)^2 + 5^2 + (-1)^2)/([3-2] = 42$.
$s_e = \sqrt{42} = 6.48$.
(c) $R^2 = 1 - ResidSS/TotalSS = 1 - \sum_{i=1}^n (y_i - \widehat{y}_i)^2/\sum_{i=1}^n (y_i - \bar{y})^2 = 1 - 42/[196 + 0 + 196] = 1 - 42/392 = 0.893$.

4.(a) Expect approximately 2. `mystery` is the slope coefficient from estimation of the model $y = 1 + 2x + u$ where the error satisfies $\mathrm{E}[u|x] = 0$ so that b_2 is unbiased.
(b) Normal distribution (with mean 2 and unknown variance). The sample size is 1000 so by the central limit theorem the slope coefficient is normally distributed.

Multiple Choice for Versions A and B

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

For 4. it is sum OF SQUARES of vertical deviances that is minimized - so neither a. nor b.
For 5. vce(robust) option gives the same coefficient estimates but standard errors, p-values, t-statistics and CI's are different.

Version B

1.(a) Here $se(\bar{x}) = sd(x)/\sqrt{n} = 100/\sqrt{400} = 5$.Approximate 95% CI is $150 \pm 2 \times 5 = (140, 160)$.
(b) 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 $i \neq j$.
(c) Assumptions 1 and 2.
(d) $r_{xy} = s_{xy}/[s_x \times s_y] = 4/[\sqrt{4} \times \sqrt{25}] = 4/10 = 0.4$.
(e) $b_2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) / \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2 = 4/4 = 1$.

2.(a) The mean out-of-state fees and tuition increase by $1000 b_2 = 1000 \times 2.170259 = 2,170$ dollars.
(b) A 95% confidence interval for β_2 is (1.551, 2.789) from the regression output.
(c) A 99% confidence interval for β_2 is $b_2 \pm t_{.005,.40} \times s_{b_2}$
$= 2.1703 \pm invttail(40, .005) \times .3064 = 2.1703 \pm 2.7045 \times .3064 = 2.1703 \pm .8287 = (1.3416, 2.999)$.
(d) $H_0 : \beta_2 = 0$ against $H_a : \beta_2 \neq 0$.
Using p-value approach: $p = 0.000 < 0.05$. So reject H_0.
Alternatively and equivalently $|t| = 7.08 > t_{.025,40} = invttail(40, .025) = 2.021$. So reject H_0.
Conclude there is a statistically significant relationship at level 0.05.
(e) $H_0 : \beta_2 \leq 1$ against $H_a : \beta_2 > 1$. (Make the claim the alternative)
$t = (b_2 - 1)/s_{b2} = (2.1703 - 1)/.3064 = 3.819$.
Since $t = 3.819 > t_{.01,40} = invttail(40, .01) = 2.423$ we reject H_0.
Conclude that out-of-state tuition rises by more than \$1 for each \$1 increase in in-state tuition.
(f) $\widehat{y} = b_1 + b_2 x = 6066.1 + 2.1703 \times 12972 =$ \$34,219< actual \$35,850.
Berkeley's out-of-state is \$1,600 higher than predicted by the model.

3.(a) $b_2 = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) / \sum_{i=1}^n (x_i - \bar{x})^2 = (28 + 0 + 42)/(4 + 1 + 9) = 70/14 = 5$.
$\bar{x} = (2 + 3 + 7)/3 = 4$ and $\bar{y} = (11 + 25 + 39) = 25$.
So $b_1 = \bar{y} - b_2\bar{x} = 25 - 5 \times 4 = 5$.
(b) $e_1 = y_1 - \widehat{y}_1 = 11 - (5 + 5 \times 2) = 11 - 15 = -4$.
$e_2 = y_2 - \widehat{y}_2 = 25 - (5 + 5 \times 3) = 25 - 20 = 5$.
$e_3 = y_3 - \widehat{y}_3 = 39 - (5 + 5 \times 7) = 39 - 40 = -1$.
$s_e^2 = \sum_{i=1}^n (y_i - \widehat{y}_i)^2/(n-2) = [(-4)^2 + 5^2 + (-1)^2)/([3-2] = 42$.
$s_e = \sqrt{42} = 6.48$.
(c) $R^2 = 1 - ResidSS/TotalSS = 1 - \sum_{i=1}^n (y_i - \widehat{y}_i)^2 / \sum_{i=1}^n (y_i - \bar{y})^2 = 1 - 42/[196 + 0 + 196] = 1 - 42/392 = 0.893$.

4.(a) Expect approximately 3. `mystery` is the slope coefficient from estimation of the model $y = 2 + 3x + u$ where the error satisfies $\mathrm{E}[u|x] = 0$ so that b_2 is unbiased.
(b) Normal distribution (with mean 0 and unknown variance). The sample size is 1000 so by the central limit theorem the slope coefficient is normally distributed.

The course grade will be based on a curve from the combined scores of midterm 1 (22.5%), midterm 2 (22.5%), final (45%) and assignments (10%). **The curve for this exam is only a guide**. Suggested average GPA for this course is 2.4. Curve below has average GPA 2.43 for this exam.

Scores out of	30	
$75th$ percentile	22	(73%)
Median	18.5	(62%)
$25th$ percentile	16	(53%)

A+	27 and above	C+	18 and above
A	23 and above	C	17 and above
A-	22 and above	C-	16 and above
B+	21 and above	D+	15 and above
B	20 and above	D	14 and above
B-	19 and above	D-	13 and above