**Economics 102: Analysis of Economic Data**
**Cameron Spring 2016**
**Department of Economics, U.C.-Davis**
**Answers to Final Exam**

**Version A**

**1.(a)** 95% confidence interval for population mean sales
$= \bar{x} \pm t_{.025;n-1} s/\sqrt{n} = 14022.5 \pm t_{.025;199} \times 5217.457/\sqrt{200} = 14022.5 \pm 1.9720 \times 368.93$
$= 14022.5 \pm 727.53 = (13295,\ 14750)$.
**(b)** $H_0 : \mu \leq 13000$ versus $H_a : \mu > 13000$ where claim is the alternative hypothesis.
$t = (\bar{x} - \mu_0)/(s/\sqrt{n}) = (14022.5 - 13000)/(5217.457/\sqrt{200}) = 1022.5/368.93 = 2.772$.
Reject $H_0$ at level .05 if $t > t_{.05;199} = 1.6525$. So reject $H_0$ as $2.772 > 1.6525$.
Conclude that (at sig. level 0.05) population mean sales exceed 13,000 units.
**(c)** percentiles, median, skewness, kurtosis (also will allow smallest values and largest values).

**2.(a)** A 95% confidence interval for $\beta_{sales}$ is $(42.23, 52.84)$ from regression output.
**(b)** 99% confidence interval for population slope parameter
$= \bar{x} \pm t_{.005;n-2} \times s_{b_2} = 47.537 \pm t_{.005;198} \times 2.691 = 47.537 \pm 2.6009 \times 2.691$
$= 47.537 \pm 6.999 = (40.54,\ 54.54)$.
**(c)** $H_0 : \beta_2 = 50$ against $H_a : \beta_2 \neq 50$.
$t = (b_2 - 50)/s_{b_2} = (47.537 - 50)/2.691 = -.915$.
$|t| = .915 < t_{.005;198} = 2.691$.
Do not reject $H_0$ at level 0.01. Conclude that the slope coefficient does not differ from 50.
**(d)** Prediction is $\widehat{y} = b_1 + b_2 x^* = 7032.6 + 47.537 \times 100 = 11{,}786$.
(`tv`= 100 when TV advertising = \$100,000).
**(e)** Easiest is to note that the standard error of the prediction for the actual value is at least $s_e$.
So the confidence interval width is at least $\pm t_{.025;198} \times s_e = 1.9720 \times 3258.7$ or $\pm 6426$.
The statistician is correct.
**(f)** We get the same $R^2$. So $R^2 = 0.6119$.

**3.(a)** $\widehat{y} = 4246 + 38.807 \times TV + \cdots + 0.2010 \times TV \times news$
$d\widehat{y}/dx = 38.807 + 0.2010 \times news$.
Evaluate at $\overline{news} = 38.807 + 0.2010 \times 30.554 = 44.95$.
**(b)** Controlling for the other regressors, sales in region 1 are 404.5 units lower than in region 3.
**(c)** $H_0 : \beta_3 = 0, ..., \beta_7 = 0$ against $H_a$ : At least one of $\beta_3, ..., \beta_7 \neq 0$.
Use rss in tables. Residual $\text{SS}_{unrestricted} = 518350292 = 5.1834 \times 10^8$; Residual $\text{SS}_{restricted} = 21.025 \times 10^8$.
$F = \frac{(\text{Residual SS}_r - \text{Residual SS}_u)/(k-g)}{\text{Residual SS}_u/(n-k)} = \frac{(21.025 \times 10^8 - 5.1834 \times 10^8)/5}{5.1834 \times 10^8/193} = \frac{15.842/5}{0.02686} = 117.96$.
Critical value $F_{.05;5,193} = 2.261$ was given in question. Reject $H_0$ as $117.96 > 2.261$.
Conclude that the additional regressors are jointly statistically insignificant at level 0.05.
**(d)** The only change will be for the coefficient estimates, standard errors, etc. of the region variables. Nothing else changes.

**Version A (continued)**

**4.(a)** This is a semielasticity. A one unit increase in `tv` (a $1,000 change) is associated with a 0.00379 proportionate increase or 0.379 percent increase in number of units sold.
**(b)** No. Not a good prediction. Using $\widehat{y} = \exp(\widehat{\ln y})$ has retransformation bias, as discussed in class and the notes.
**(c)** No. There is not enough information. It is not meaningful to compare $R^2$ of models with different dependent variable. (Both models do seem to fit well).

**5.(a)** $\sum_{i=1}^{n} z_i = \sum_{i=1}^{3} 6/i = (6/1) + (6/2) + (6/3) = 6 + 3 + 2 = 11.$
**(b)** $Y_i \sim (10, 100)$ so $\bar{Y} \sim (10, 100/n) \sim (10, 100/100) = (10, 1)$.
Expect mean 10, standard deviation 1 and a normal distribution by central limit theorem.
**(c)** $E[X] = 0.5 \times 1 + 0.3 \times 2 + 0.2 \times 3 = 1.7.$
**(d)** $\bar{x} = (18 + 20 + 28 + 30)/4 = 96/4 = 24.$
$\sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1) = \{(-6)^2 + (-4)^2 + (4)^2 + (6^2)\}/3 = 104/3 = 34.67.$
Standard deviation is $\sqrt{34.67} = 5.89$.

**6.(a)** Missing entry (A): $t = b/se$ so $se = b/t = 3/1.5 = 2$.
**(b)** Missing entry (B): $R^2 = ModelSS/ResidualSS = 540/720 = 0.25$.
**(c)** Missing entry (C): Either from formula sheet $F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} = \frac{0.75/2}{0.25/18} = 27$.
or $F = \frac{(\text{Res SS}_r - \text{Res SS}_u)/(k-g)}{\text{Res SS}_u/(n-k)} = \frac{(720-180)/(3-1)}{180/18} = 27$.
**(d)** The $F$ statistic is the square of the $t$ statistics. $F = t^2 = (2/1)^2 = 4$.

**7.(a)** Unbiased. (Unbiased does not depend on sample size. Estimates will be less precise, however).
**(b)** Unbiased. (Irrelevant regressor included).
**(c)** Biased. (Omitted regressor. Will be unbiased only if $w$ is uncorrelated with $x$ and $z$.).
**(d)** Unbiased. (Multicollinearity effects precision but not bias).
**(e)** Unbiased. (Heteroskedasticity will effect the standard errors).

**Multiple Choice Version A:**
**1. a 2. a 3. d 4. d 5. b 6. c 7. c 8. b 9. b 10. a**

The average GPA for the curve for this exam is 2.70.
**Course grade is determined by curve based on combined course score.**

| | | |
|---|---|---|
| Scores out of | 58 | |
| 75*th* percentile | 42.5 | (73%) |
| Median | 36 | (62%) |
| 25*th* percentile | 32.5 | (56%) |

| | |
|---|---|
| A+ | 55 and above |
| A | 43 and above |
| A- | 41 and above |
| B+ | 39 and above |
| B | 37 and above |
| B- | 35 and above |
| C+ | 33 and above |
| C | 31 and above |
| C- | 29 and above |
| D+ | 27 and above |
| D | 25 and above |
| D- | 23 and above |

**Economics 102: Analysis of Economic Data**
**Cameron Spring 2016**
**Department of Economics, U.C.-Davis**

## Answers to Final Exam

### Version B

**1.(a)** 95% confidence interval for population mean sales
$= \bar{x} \pm t_{.025;n-1} s/\sqrt{n} = 14022.2 \pm t_{.025;179} \times 5139.36/\sqrt{180} = 14022.2 \pm 1.9733 \times 383.06$
$= 14022.52 \pm 755.89 = (13266,\ 14778)$.
**(b)** $H_0 : \mu \leq 13500$ versus $H_a : \mu > 13500$ where claim is the alternative hypothesis.
$t = (\bar{x} - \mu_0)/(s/\sqrt{n}) = (14022.2 - 13500)/(5139.36/\sqrt{180}) = 522.2/383.06 = 1.363$.
Reject $H_0$ at level .05 if $t > t_{.05;179} = 1.6543$. So do not reject $H_0$ as $1.363 < 1.6543$.
Conclude that (at sig. level 0.05) population mean sales do not exceed 13,500 units.
**(c)** percentiles, median, skewness, kurtosis (also will allow smallest values and largest values).

**2.(a)** A 95% confidence interval for $\beta_{sales}$ is $(41.31, 52.47)$ from regression output.
**(b)** 90% confidence interval for population slope parameter
$= \bar{x} \pm t_{.05;n-2} \times s_{b_2} = 46.893 \pm t_{.05;178} \times 2.827 = 46.893 \pm 1.6535 \times 2.827$
$= 46.893 \pm 4.674 = (42.22,\ 51.57)$.
**(c)** $H_0 : \beta_2 = 40$ against $H_a : \beta_2 \neq 40$.
$t = (b_2 - 50)/s_{b_2} = (46.893 - 50)/2.827 = 1.099$.
$|t| = 1.099 < t_{.05;178} = 1.6535$.
Do not reject $H_0$ at level 0.10. Conclude that the slope coefficient does not differ from 50.
**(d)** Prediction is $\widehat{y} = b_1 + b_2 x^* = 7164.9 + 46.893 \times 200 = 16{,}544$.
(`tv`= 200 when TV advertising = \$200,000).
**(e)** Easiest is to note that the standard error of the prediction for the actual value is at least $s_e$.
So the confidence interval width is at least $\pm t_{.025;178} \times s_e = 1.9734\times$ 3229.9 or $\pm 6374$.
The statistician is correct.
**(f)** We get the same $R^2$. So $R^2 = 0.6072$.

**3.(a)** $\widehat{y} = 3860.533 + 40.339 \times TV + \cdots + 0.1757 \times TV \times news$
$d\widehat{y}/dx = 40.339 + 0.1757 \times news$.
Evaluate at $\overline{news} = 40.339 + 0.1757 \times 31.644 = 45.90$.
**(b)** Controlling for the other regressors, sales in region 2 are 174.6 units lower than in region 3.
**(c)** $H_0 : \beta_3 = 0, ..., \beta_7 = 0$ against $H_a$ : At least one of $\beta_3, ..., \beta_7 \neq 0$.
Use rss in tables. Residual $\text{SS}_{unrestricted} = 484651098 = 4.8465 \times 10^8$; Residual $\text{SS}_{restricted} = 18.569 \times 10^8$.
$F = \frac{(\text{Residual SS}_r - \text{Residual SS}_u)/(k-g)}{\text{Residual SS}_u/(n-k)} = \frac{(18.569\times10^8 - 4.8465\times10^8)/5}{4.8465\times10^8/173} = \frac{13.7225/5}{0.02801} = 97.983$.
Critical value $F_{.05;5,173} = 2.266$ was given in question. Reject $H_0$ as $97.983 > 2.266$.
Conclude that the additional regressors are jointly statistically insignificant at level 0.05.
**(d)** The only change will be for the coefficient estimates, standard errors, etc. of the region variables. Nothing else changes.

**Version B (continued)**

**4.(a)** This is a semielasticity. A one unit increase in `tv` (a \$1,000 change) is associated with a 0.00377 proportionate increase or 0.377 percent increase in number of units sold.
**(b)** No. Not a good prediction. Using $\widehat{y} = \exp(\widehat{\ln y})$ has retransformation bias, as discussed in class and the notes.
**(c)** No. There is not enough information. It is not meaningful to compare $R^2$ of models with different dependent variable. (Both models do seem to fit well).

**5.(a)** $\sum_{i=1}^{n} z_i = \sum_{i=1}^{4} 12/i = (12/1) + (12/2) + (12/3) + (12/4) = 12 + 6 + 4 + 3 = 25.$
**(b)** $Y_i \sim (20, 400)$ so $\bar{Y} \sim (20, 400/n) \sim (20, 400/100) = (20, 2)$.
Expect mean 20, standard deviation 2 and a normal distribution by central limit theorem.
**(c)** $E[X] = 0.4 \times 3 + 0.2 \times 4 + 0.4 \times 5 = 4.0$.
**(d)** $\bar{x} = (18 + 20 + 28)/3 = 66/3 = 22$.
$\sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1) = \{(-4)^2 + (-2)^2 + (6^2)\}/2 = 56/2 = 28$.
Standard deviation is $\sqrt{28} = 5.29$.

**6.(a)** Missing entry (A): $R^2 = ModelSS/ResidualSS = 540/720 = 0.25$.
**(b)** Missing entry (B): Either from formula sheet $F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} = \frac{0.75/2}{0.25/18} = 27$.
or $F = \frac{(\text{Res SS}_r - \text{Res SS}_u)/(k-g)}{\text{Res SS}_u/(n-k)} = \frac{(720-180)/(3-1)}{180/18} = 27$.
**(c)** Missing entry (C): $t = b/se$ so $se = b/t = 3/1.5 = 2$.
**(d)** The $F$ statistic is the square of the $t$ statistics. $F = t^2 = (2/1)^2 = 4$.

**7.(a)** Biased. (Omitted regressor. Will be unbiased only if $w$ is uncorrelated with $x$ and $z$.).
**(b)** Unbiased. (Irrelevant regressor included).
**(c)** Unbiased. (Heteroskedasticity will effect the standard errors).
**(d)** Unbiased. (Multicollinearity effects precision but not bias).
**(e)** Unbiased. (Unbiased does not depend on sample size. Estimates will be less precise, however).

**Multiple Choice Version A:**
**1. b  2. d  3. d  4. b  5. a  6. c  7. a  8. d  9. a  10. c**

The average GPA for the curve for this exam is 2.70.
**Course grade is determined by curve based on combined course score.**

| | | |
|---|---|---|
| Scores out of | 58 | |
| 75*th* percentile | 42.5 | (73%) |
| Median | 36 | (62%) |
| 25*th* percentile | 32.5 | (56%) |

| | |
|---|---|
| A+ | 55 and above |
| A | 43 and above |
| A- | 41 and above |
| B+ | 39 and above |
| B | 37 and above |
| B- | 35 and above |
| C+ | 33 and above |
| C | 31 and above |
| C- | 29 and above |
| D+ | 27 and above |
| D | 25 and above |
| D- | 23 and above |