Economics 102: Analysis of Economic Data
Cameron Spring 2015
Department of Economics, U.C.-Davis
Answers to Final Exam

Version A

1.(a) Use a histogram or kernel density estimate to see if it looks like normal.
Or use summary statistics and see if symmetry$\simeq 0$ and kurtosis $\simeq 3$.
(b) 95% confidence interval for population mean salary
$= \bar{x} \pm t_{.025;n-1}s/\sqrt{n} = 500.083 \pm t_{.025;47} \times 213.6428/\sqrt{48} = 500.083 \pm 2.012 \times 30.837$
$= 500.083 \pm 62.044 = (438.0,\ 562.1)$.
(c) $H_0 : \mu \leq 450$ versus $H_a : \mu > 450$ where claim is the alternative hypothesis.
$t = (\bar{x} - \mu_0)/(s/\sqrt{n}) = (500.083 - 450)/(213.6428/\sqrt{48}) = 50.083/30.837 = 1.625$.
Reject H_0 at level .05 if $t > t_{.05;47} = 1.678$. So do not reject H_0 as $1.625 < 1.678$.
Conclude that (at sig. level 0.05) population mean price does not exceed 450.

2.(a) This is just the slope: price increases by $372.10.
(b) Yes. A negative coefficient means that if we extrapolate back to zero size a diamond ring would be sold for a negative price (-$259.10).
(c) This is directly given in regression output.
A 95% confidence interval for β_{size} is (355.64, 388.56).
(d) 99% confidence interval for population mean salary
$= \bar{x} \pm t_{.005;n-2}s/\sqrt{n} = 372.10 \pm t_{.005;46} \times 8.179 = 372.10 \pm 2.687 \times 8.179$
$= 372.10 \pm 21.977 = (350.1,\ 394.1)$.
(e) $H_0 : \beta_2 = 350$ against $H_a : \beta_2 \neq 350$.
$t = (b_2 - 350)/s_{b_2} = (372.10 - 350)/8.179 = 22.10/8.179 = 2.702$.
$|t| = 2.702 > t_{.005;46} = 2.687$.
Reject H_0 at level 0.01. Conclude that the slope coefficient differs from 350.
(f) It is not meaningful. The data range from 0.12 to 0.35 carats (from summary statistics for `size`).
A 10 unit carat is way outside this range. A great extrapolation from the observed regressors.
(g) Prediction is $\widehat{y} = b_1 + b_2 x^* = -259.62 + 372.10 \times 2.5 = 670.63$.
(h) $\mathrm{E}[y|x = 2.5] \in b_1 + b_2 \times 2.5 \pm t_{.025;n-2} \times s_e \times \sqrt{\frac{1}{n} + \frac{(2.5-\bar{x})^2}{\sum_i(x_i-\bar{x})^2}}$
$\mathrm{E}[y|x = 2.5] \in 670.63 \pm 2.013 \times 31.841 \times \sqrt{\frac{1}{48} + \frac{(2.5-2.042)^2}{47 \times 0.5679^2}}$
[Not necessary but $\mathrm{E}[y|x = 2.5] \in 670.63 \pm 64.096 \times 0.18620 \in 670.63 \pm 11.935 \in (658.7, 682.6)$]
(i) This does OLS and gives same $b1$ and $b2$ but gets heteroskedastic-robust standard errors that differ from the default standard errors.
(j) `generate size = 10*carats`
(k) Very well. $R^2 = 0.9783$ as same as R^2 from regress `price` on `size`.
(Or R^2 equals correlation coefficient squared $= 0.9891^2 = 0.9783$.

3.(a) $\widehat{y} = b_1 + b_2 x + b_3 x^2$ so $d\widehat{y}/dx = b_2 + 2b_3 x = 292.0 + 2 \times 17.399 \times x = 292.0 + 34.798 \times x = \ .$
MEM $= 292.0 + 2 \times 17.399 \times \bar{x} = 292.0 + 34.798 \times 2.042 = 292.0 + 71.1 = 363.1$.
(b) d1 and d2 coefficients are relative to the omitted category which is d3 (below-average).
So this is just the coefficient of d2: $1.55 more.
(c) The above-average is $3.98 more than below-average and the average is $1.55 more than below-average so the above-average sells for $2.53 ($= 3.98 - 1.55$) than the average ring.
(d) $H_0 : \beta_{size} = 0, \beta_{d1} = 0, \beta_{d2} = 0$ against $H_a :$ At least one of $\beta_{size}, \beta_{d1}, \beta_{d2} \neq 0$.
Use F statistic in table. $F = 662.090$.
Critical value $F_{.05;3,44} = 2.816$. Reject H_0 as $662.090 > 2.816$.
Conclude that the regressors are jointly statistically significant at level 0.05.

3.(e) $H_0 : \beta_{d1} = 0, \beta_{d2} = 0$ against H_a : At least one of $\beta_{d1}, \beta_{d2} \neq 0$.

Use rss in table. Residual $SS_{unrestricted} = 46491.431$; Residual $SS_{restricted} = 46635.671$.

$F = \frac{(\text{Residual SS}_r - \text{Residual SS}_u)/(k-g)}{\text{Residual SS}_u/(n-k)} = \frac{(46635.671 - 46491.431)/2}{46491.431/44} = \frac{72.12}{1056.62} = 0.068$.

Critical value $F_{.05;2,44} = 3.209$. Do not reject H_0 as $0.068 < 3.209$.

Conclude that d1 and d2 are jointly statistically insignificant at level 0.05.

(f) Yes. Not all parameters can be estimated. It will not be identified due to the dummy variable trap as d1+d2+d3=1 by construction.

(g) Model LINEAR or QUAD as they have the (equal) largest adjusted R-squared ($\bar{R}^2 = 0.978$).

4.(a) This is a semielasticity. A one unit change in `size` is associated with a 0.679 proportionate increase or 67.9 percent increase in `price`.

(b) This is an elasticity. A one percent change in `size` is associated with a 1.498 percent increase in `price`.

(c) This is the naive estimate that ignores retransformation bias.

(If errors were independent homoskedastic normal we scale up by $\exp(s_e^2/2)$.)

5.(a) $E[X] = 0.6 \times 1 + 0.3 \times 2 + 0.1 \times 3 = 1.5$.

$Var[X] = E[(X - \mu)^2] = 0.6 \times (1 - 1.5)^2 + 0.3 \times (2 - 1.5)^2 + 0.1 \times (3 - 1.5)^2$
$= 0.6 \times 0.25 + 0.3 \times 0.25 + 0.1 \times 2.25 = 0.45$.

(b) Mean of $\bar{x} \simeq 400$ since $E[\bar{X}] = E[X] = 400$.

Standard deviation of $\bar{x} \simeq 3.0$ since $SD[\bar{X}] = SD[X]/\sqrt{100} = 30/\sqrt{100} = 3.0$.

(c) The population model is $y_i = \beta_1 + \beta_2 x_i + u_i$ for all i.

The error for the ith observation has zero mean conditional on the regressor.

The error for the ith observation has constant variance conditional on the regressor:

The errors for different observations are statistically independent.

(d) Expect approximately 2. (Reason: `mystery` is the slope coefficient from estimation of the model $y = 1 + 2x + u$ where the error satisfies $E[u|x] = 0$ so that b_2 is unbiased.

(e) Normal distribution (with mean 2 and unknown variance). Reason: the sample size is 1000 so by the central limit theorem the slope coefficient is normally distributed.

Multiple Choice Version A:

1. b   2. b   3. d   4. b   5. d   6. c   7. a   8. b   9. b   10. d

4. Price double every $72/r$ years $72/r = 7$ so $r = 72/7 = 10.3$. Inflation rate is 10.3%.

5. $T(200)$ is essentially standard normal. This is probability $z > 1$. Standard normal has area 2/3 within one standard deviation so area 1/6 in right tail above 1. Answer is approximately 1/6=0.18.

This exam was more difficult than final exams in recent years.

The average GPA for the curve for this exam is 2.50.

Course grade is determined by curve based on combined course score.

	A+	50 and above
	A	42 and above
	A-	39 and above
	B+	37 and above
Scores out of 56	B	35 and above
75th percentile 38.5 (69%)	B-	33 and above
Median 33.5 (60%)	C+	31 and above
25th percentile 29.5 (53%)	C	29.5 and above
	C-	28 and above
	D+	26.5 and above
	D	25 and above
	D-	23.5 and above

Economics 102: Analysis of Economic Data
Cameron Spring 2015
Department of Economics, U.C.-Davis
Answers to Final Exam

Version B

1.(a) Use a histogram or kernel density estimate to see if it looks like normal.
Or use summary statistics and see if symmetry$\simeq 0$ and kurtosis $\simeq 3$.
(b) 99% confidence interval for population mean salary
$= \bar{x} \pm t_{.005;n-1}s/\sqrt{n} = 500.083 \pm t_{.005;47} \times 213.6428/\sqrt{48} = 500.083 \pm 2.685 \times 30.837$
$= 500.083 \pm 82.797 = (417.3, 582.9)$.
(c) $H_0 : \mu \leq 420$ versus $H_a : \mu > 420$ where claim is the alternative hypothesis.
$t = (\bar{x} - \mu_0)/(s/\sqrt{n}) = (500.083 - 420)/(213.6428/\sqrt{48}) = 80.083/30.837 = 2.597$.
Reject H_0 at level .05 if $t > t_{.05;47} = 1.678$. So reject H_0 as $2.597 > 1.678$.
Conclude that (at sig. level 0.05) population mean price exceeds 420.

2.(a) This is just the slope: price increases by $372.10.
(b) Yes. A negative coefficient means that if we extrapolate back to zero size a diamond ring would
be sold for a negative price (-$259.10).
(c) This is directly given in regression output.
A 95% confidence interval for β_{size} is (355.64, 388.56).
(d) 90% confidence interval for population mean salary
$= \bar{x} \pm t_{.05;n-2}s/\sqrt{n} = 372.10 \pm t_{.05;46} \times 8.179 = 372.10 \pm 1.679 \times 8.179$
$= 372.10 \pm 13.733 = (358.4, 385.8)$.
(e) $H_0 : \beta_2 = 360$ against $H_a : \beta_2 \neq 360$.
$t = (b_2 - 360)/s_{b_2} = (372.10 - 360)/8.179 = 12.10/8.179 = 1.479$.
$|t| = 1.479 < t_{.005;46} = 2.687$.
Do not reject H_0 at level 0.01. Conclude that the slope coefficient does not differ from 360.
(f) It is not meaningful. The data range from 0.12 to 0.35 carats (from summary statistics for `size`).
A 10 unit carat is way outside this range. A great extrapolation from the observed regressors.
(g) Prediction is $\widehat{y} = b_1 + b_2 x^* = -259.62 + 372.10 \times 3.5 = 1042.73$.
(h) $E[y|x = 3.5] \in b_1 + b_2 \times 3.5 \pm t_{.025;n-2} \times s_e \times \sqrt{\frac{1}{n} + \frac{(2.5-\bar{x})^2}{\sum_i (x_i - \bar{x})^2}}$
$E[y|x = 3.5] \in 1042.73 \pm 2.013 \times 31.841 \times \sqrt{\frac{1}{48} + \frac{(3.5-2.042)^2}{47 \times 0.5679^2}}$
[Not necessary but $E[y|x = 3.5] \in 1042.73 \pm 64.096 \times 0.40134 \in 1042.73 \pm 25.724 \in (1017.0, 1068.5)$]
(i) This does OLS and gives same $b1$ and $b2$ but gets heteroskedastic-robust standard errors that
differ from the default standard errors.
(j) `generate size = 10*carats`
(k) Very well. $R^2 = 0.9783$ as same as R^2 from regress `price` on `size`.
(Or R^2 equals correlation coefficient squared $= 0.9891^2 = 0.9783$.

3.(a) $\widehat{y} = b_1 + b_2 x + b_3 x^2$ so $d\widehat{y}/dx = b_2 + 2b_3 x = 292.0 + 2 \times 17.399 \times x = 292.0 + 34.798 \times x = .$
MEM $= 292.0 + 2 \times 17.399 \times \bar{x} = 292.0 + 34.798 \times 2.042 = 292.0 + 71.1 = 363.1$.
(b) d1 and d2 coefficients are relative to the omitted category which is d3 (below-average).
So this is just the coefficient of d2: $7.31 more.
(c) The above-average is $3.98 more than below-average and the average is $1.55 more than below-
average so the above-average sells for $2.04 $(= 9.35 - 7.31)$ than the average ring.
(d) $H_0 : \beta_{size} = 0, \beta_{d1} = 0, \beta_{d2} = 0$ against H_a : At least one of $\beta_{size}, \beta_{d1}, \beta_{d2} \neq 0$.
Use F statistic in table. $F = 672.611$.
Critical value $F_{.05;3,44} = 2.816$. Reject H_0 as $672.611 > 2.816$.
Conclude that the regressors are jointly statistically significant at level 0.05.

Version B (continued)

3.(e) $H_0 : \beta_{d1} = 0, \beta_{d2} = 0$ against H_a : At least one of $\beta_{d1}, \beta_{d2} \neq 0$.
Use rss in table. Residual SS$_{unrestricted}$ = 45779.723; Residual SS$_{restricted}$ = 46635.671.
$F = \frac{(\text{Residual SS}_r - \text{Residual SS}_u)/(k-g)}{\text{Residual SS}_u/(n-k)} = \frac{(46635.671 - 45779.723)/2}{45779.723/44} = \frac{427.97}{1040.45} = 0.411$.
Critical value $F_{.05;2,44} = 3.209$. Do not reject H_0 as $0.404 < 3.209$.
Conclude that d1 and d2 are jointly statistically insignificant at level 0.05.
(f) Yes. Not all parameters can be estimated. It will not be identified due to the dummy variable trap as d1+d2+d3=1 by construction.
(g) Models LINEAR or QUAD as both have the (equal) largest adjusted R-squared ($\bar{R}^2 = 0.978$).

4.(a) This is an elasticity. A one percent change in `size` is associated with a 1.498 percent increase in `price`.
(b) This is a semielasticity. A one unit change in `size` is associated with a 0.679 proportionate increase or 67.9 percent increase in `price`.
(c) This is the naive estimate that ignores retransformation bias.
(If errors were independent homoskedastic normal we scale up by $\exp(s_e^2/2)$).

5.(a) $E[X] = 0.4 \times 1 + 0.5 \times 2 + 0.1 \times 6 = 2.0$.
$Var[X] = E[(X - \mu)^2] = 0.4 \times (1 - 2)^2 + 0.5 \times (2 - 2)^2 + 0.1 \times (6 - 2)^2$
$= 0.4 \times 1 + 0.5 \times 0 + 0.1 \times 16 = 2.0$.
(b) Mean of $\bar{x} \simeq 500$ since $E[\bar{X}] = E[X] = 500$.
Standard deviation of $\bar{x} \simeq 4.0$ since $SD[\bar{X}] = SD[X]/\sqrt{400} = 80/\sqrt{400} = 4.0$.
(c) The population model is $y_i = \beta_1 + \beta_2 x_i + u_i$ for all i.
The error for the ith observation has zero mean conditional on the regressor.
The error for the ith observation has constant variance conditional on the regressor:
The errors for different observations are statistically independent.
(d) Expect approximately 3. (Reason: `mystery` is the slope coefficient from estimation of the model $y = 2 + 3x + u$ where the error satisfies $E[u|x] = 0$ so that b_2 is unbiased.
(e) Normal distribution (with mean 3 and unknown variance). Reason: the sample size is 10000 so by the central limit theorem the slope coefficient is normally distributed.

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

This exam was more difficult than final exams in recent years.
The average GPA for the curve for this exam is 2.50.
Course grade is determined by curve based on combined course score.

			A+	50 and above
			A	42 and above
			A-	39 and above
			B+	37 and above
Scores out of	56		B	35 and above
75th percentile	38.5	(69%)	B-	33 and above
Median	33.5	(60%)	C+	31 and above
25th percentile	29.5	(53%)	C	29.5 and above
			C-	28 and above
			D+	26.5 and above
			D	25 and above
			D-	23.5 and above