

Economics 102: Analysis of Economic Data
Cameron Winter 2014
Department of Economics, U.C.-Davis
Answers to Final Exam (Versions A and B)

Version A

1.(a) There is little skewness (skew= 0.5 and median \simeq mean) and kurtosis is close to 3. So approximately normal. (This involves judgement - key is to consider skewness and kurtosis).

(b) Expect 95% within two standard deviations of mean: $\bar{x} \pm 2s = 25.94 \pm 2 \times 5.44 = (15.06, 36.82)$.

(c) 95% confidence interval for population mean miles per gallon
 $= \bar{x} \pm t_{.025; n-1} s / \sqrt{n} = 25.94 \pm t_{.025; 329} \times 5.44 / \sqrt{330} = 25.94 \pm 1.967 \times .299$
 $= 25.94 \pm .59 = (25.35, 26.53)$.

(d) $H_0 : \mu \leq 260$ versus $H_a : \mu > 260$ where claim is the alternative hypothesis.
 $t = (\bar{x} - \mu_0) / (s / \sqrt{n}) = (268.53 - 260) / (105.81 / \sqrt{330}) = 8.53 / 5.82 = 1.47$.
 Reject H_0 at level .05 if $t > t_{.05; 330-1} = 1.649$. So do not reject H_0 as $1.47 < 1.649$.
 Conclude that (at sig. level 0.05) population mean horsepower does not exceed 260.

(e) Regressing mpg on curbwt will worst explain mpg since from the correlation output it is the variable that is the least highly correlated with mpg.

(f) R^2 for hp on curbwt equals squared correlation coefficient $= (0.6383)^2 = 0.407$.

2.(a) The slope coefficient is -.04177. A standard deviation of hp is 105.8 .
 So mpg changes by $-.04177 \times 105.8 = -4.42$. A decrease of 4.42 miles per gallon.

(b) This is directly given in regression output.
 A 95% confidence interval for β_{hp} is $(-.045, -.038)$.

(c) $H_0 : \beta_2 = -.05$ against $H_a : \beta_2 \neq -.05$.
 $t = (b_2 - -.05) / se(b_{2b}) = (-.04177 + .05) / .00165 = .00823 / .00165 = 4.98$.
 $|t| = 4.98 > t_{.005; 330-2} = 2.591$.

Reject H_0 at level 0.01. Conclude that the slope coefficient differs from $-.05$.

(d) Prediction is $\hat{y} = b_1 + b_2 x^* = 37.16 - .04177 \times 140 = 31.31$.

(e) $E[y|x = 140] \in b_1 + b_2 x \pm t_{.025; n-2} \times s_e \times \sqrt{\frac{1}{n} + \frac{(140 - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}}$
 $E[y|x = 140] \in 31.31 \pm 1.967 \times 3.1686 \times \sqrt{\frac{1}{330} + \frac{(140 - 268.53)^2}{329 \times 105.81^2}}$
 [Not necessary but $E[y|x = 600] \in 31.31 \pm 1.967 \times 0.275 \in 31.31 \pm .54 \in (30.77, 31.85)$]

3.(a) Yes. Expect fuel consumption to increase with torque, so mpg falls (the case for the raw correlations). But in the regression torque has a positive coefficient.

(b) A one horsepower increase is associated with a 0.0432 decrease in miles per gallon.

(c) $H_0 : \beta_2 = 0, \dots, \beta_5 = 0$ against $H_a : \text{At least one of } \beta_2, \dots, \beta_5 \neq 0$.
 From the regression output $F = 204.45$ has p value of $0.0000 < .05$.
 Reject H_0 . Conclude that the regressors are jointly significant at level 0.05.

(d) Use \bar{R}^2 . Favors larger model as $0.7121 > 0.6602$.

(e) $H_0 : \beta_3 = 0, \beta_4 = 0, \beta_5 = 0$ against $H_a : \text{At least one of } \beta_3, \beta_4, \beta_5 \neq 0$.
 $F = \frac{(\text{ResSS}_r - \text{ResSS}_u) / (k - g)}{\text{ResSS}_u / (n - k)} = \frac{(3293.1 - 2764.2) / (5 - 2)}{2764.2 / (330 - 5)} = \frac{528.9 / 3}{2764.2 / 325} = \frac{176.3}{8.52} = 20.69$.
 Since $20.69 > 2.63$ we reject H_0 at level 0.05.

Conclude that the three additional regressors are jointly significant.

(f) No. The individual coefficients were reasonably precisely estimated (enough to be statistically significant at 5%).

Version A (continued)

4.(a) Miles per gallon is a proportion .0016 lower, or 0.16% lower for each one unit increase in horsepower.

(b) This is now an elasticity as now log-log. A one percent increase in curb weight is associated with a 0.31 percent decrease in miles per gallon.

(c) This leads to the dummy variable trap as the 25 dummies sum to one and an intercept is also included in the model. One of the coefficients cannot be estimated. Drop the intercept or drop one of the indicator variables.

(d) Yes, because the dependent variable is a log. We cannot simply use the model to get $\widehat{\ln y}$ and then let $\widehat{y} = \exp(\widehat{\ln y})$ due to retransformation bias.

(e) No. The reason we included it as a regressor is because it is correlated with `lnmpg`.

(f) Yes. OLS is biased and inconsistent if the error is correlated with one of the regressors.

(g) generate `lnmpg = ln(mpg)`

5.(a) Mean of \bar{X} is μ and variance is σ^2/n .

(b) $E[X] = 0.4 \times 10 + 0.6 \times 20 = 4 + 12 = 16$.

$\text{Var}[X] = E[(X - \mu)^2] = 0.4 \times (10 - 16)^2 + 0.6 \times (20 - 16)^2 = 0.4 \times 6^2 + 0.6 \times 4^2 = 14.4 + 9.6 = 24$.

(c) $y = b_1 + b_2d$ then $y = b_1$ when $d = 0$ and $y = b_1 + b_2$ when $d = 1$.

So here average for men is $b_1 = 20$ and average for women is $b_1 + b_2 = 20 - 4 = 16$.

(d) Annual growth is $\Delta \ln y / \Delta x = (26.9 - 25.4) / (2012 - 2002) = 1.5 / 10 = 0.15$.

Annual growth is 15% per year.

(e) $R^2 = \text{explained SS} / \text{total SS} = 400 / (400 + 600) = 0.4$.

Standard error of regression = $\sqrt{\text{ResidualSS} / (n - k)} = \sqrt{600 / 48} = \sqrt{12.5} = 3.53$.

Multiple Choice Version A:

1. a 2. d 3. a 4. d 5. a

	A	48 and above
	A-	47 and above
	B+	45 and above
	B	43 and above
	B-	41 and above
	C+	39 and above
	C	37 and above
	C-	35 and above
	D+	33 and above
	D	31 and above
	D-	29 and above
Scores out of	55	
75th percentile	45	(82%)
Median	40.5	(74%)
25th percentile	35.5	(65%)

Economics 102: Analysis of Economic Data
Cameron Winter 2014
Department of Economics, U.C.-Davis
Answers to Final Exam

Version B

1.(a) There is little skewness (skew= 0.5 and median \simeq mean) and kurtosis is close to 3. So approximately normal. (This involves judgement - key is to consider skewness and kurtosis).

(b) Expect 95% within two standard deviations of mean: $\bar{x} \pm 2s = 26.11 \pm 2 \times 5.57 = (14.97, 37.25)$.

(c) 95% confidence interval for population mean miles per gallon
 $= \bar{x} \pm t_{.025; n-1} s / \sqrt{n} = 26.11 \pm t_{.025; 229} \times 5.57 / \sqrt{230} = 26.11 \pm 1.970 \times .363$
 $= 26.11 \pm .72 = (25.39, 26.83)$.

(d) $H_0 : \mu \leq 3500$ versus $H_a : \mu > 3500$ where claim is the alternative hypothesis.

$t = (\bar{x} - \mu_0) / (s / \sqrt{n}) = (3574.0 - 3500) / (605.5 / \sqrt{230}) = 74.0 / 39.9 = 1.85$.

Reject H_0 at level .05 if $t > t_{.05; 230-1} = 1.651$. So reject H_0 as $1.85 > 1.651$.

Conclude that (at sig. level 0.05) population mean mpg exceeds 260.

(e) Regressing **mpg** on **hp** will best explain **mpg** since from the correlation output it is the variable that is the most highly correlated with **mpg**.

(f) R^2 from regress **hp** on **disp** equals squared correlation coefficient $= (.8614)^2 = 0.742$.

2.(a) The slope coefficient is $-.04311$. A standard deviation increase of **hp** is 105.3.

So **mpg** changes by $-.04311 \times 105.3 = -4.54$. A decrease of 4.54 miles per gallon.

(b) This is directly given in regression output.

A 95% confidence interval for β_{hp} is $(-0.47, -0.39)$.

(c) $H_0 : \beta_2 = -.04$ against $H_a : \beta_2 \neq -.04$.

$t = (b_2 - -.04) / se(b_{2b}) = (-.04311 + .04) / .00203 = -.00311 / .00203 = 1.53$.

$|t| = 1.53 < t_{.005; 230-2} = 2.597$.

Do not reject H_0 at level 0.05. Conclude that the slope coefficient does not differ from $-.04$.

(d) Prediction is $\hat{y} = b_1 + b_2 x^* = 37.50 - .04311 \times 170 = 30.17$.

(e) $E[y|x = 160] \in b_1 + b_2 x \pm t_{.025; n-2} \times s_e \times \sqrt{\frac{1}{n} + \frac{(170-\bar{x})^2}{\sum_i (x_i - \bar{x})^2}}$

$E[y|x = 160] \in 30.17 \pm 1.970 \times 3.2332 \times \sqrt{\frac{1}{230} + \frac{(170-264.06)^2}{229 \times 105.30^2}}$

[Not necessary but $E[y|x = 600] \in 30.17 \pm 1.970 \times 0.286 \in 30.17 \pm .56 \in (29.61, 30.73)$]

3.(a) Yes. Expect fuel consumption to increase with torque, so mpg falls (the case for the raw correlations). But in the regression **torque** has a positive coefficient.

(b) A one horsepower increase is associated with a 0.0419 decrease in miles per gallon.

(c) $H_0 : \beta_2 = 0, \dots, \beta_5 = 0$ against $H_a : \text{At least one of } \beta_2, \dots, \beta_5 \neq 0$.

From the regression output $F = 140.71$ has p value of $0.0000 < .05$.

Reject H_0 . Conclude that the regressors are jointly significant at level 0.05.

(d) Use \bar{R}^2 . Favors larger model as $0.7093 > 0.6630$.

(e) $H_0 : \beta_3 = 0, \beta_4 = 0, \beta_5 = 0$ against $H_a : \text{At least one of } \beta_3, \beta_4, \beta_5 \neq 0$.

$F = \frac{(\text{ResSS}_r - \text{ResSS}_u) / (k - g)}{\text{ResSS}_u / (n - k)} = \frac{(2383.4 - 2028.6) / (5 - 2)}{2028.6 / (230 - 5)} = \frac{354.8 / 3}{2028.6 / 225} = \frac{118.3}{9.02} = 13.11$.

Since $13.11 > 2.65$ we reject H_0 at level 0.05.

Conclude that the three additional regressors are jointly significant.

(f) No. The individual coefficients were reasonably precisely estimated (enough to be statistically significant at 5%).

Version B (continued)

4.(a) Miles per gallon is a proportion .0015 lower, or 0.15% lower for each one unit increase in horsepower.

(b) This is now an elasticity as now log-log. A one percent increase in curb weight is associated with a 0.35 percent decrease in miles per gallon.

(c) This leads to the dummy variable trap as the 24 dummies sum to one and an intercept is also included in the model. One of the coefficients cannot be estimated. Drop the intercept or drop one of the indicator variables.

(d) Yes, because the dependent variable is a log. We cannot simply use the model to get $\widehat{\ln y}$ and then let $\widehat{y} = \exp(\widehat{\ln y})$ due to retransformation bias.

(e) Yes. OLS is biased and inconsistent if the error is correlated with one of the regressors.

(f) No. The reason we included it as a regressor is because it is correlated with `lnmpg`.

(g) `generate lnmpg = ln(mpg)`

5.(a) $E[X] = 0.7 \times 10 + 0.3 \times 20 = 13$.

$\text{Var}[X] = E[(X - \mu)^2] = 0.7 \times (10 - 13)^2 + 0.3 \times (20 - 13)^2 = 0.7 \times 3^2 + 0.3 \times 7^2 = 6.3 + 14.7 = 21$.

(b) Mean of \bar{X} is μ and variance is σ^2/n .

(c) $y = b_1 + b_2d$ then $y = b_1$ when $d = 0$ and $y = b_1 + b_2$ when $d = 1$.

So here average for women is $b_1 = 20$ and average for men is $b_1 + b_2 = 20 + 2 = 22$.

(d) Annual growth is $\Delta \ln y / \Delta x = (26.7 - 24.7) / (2012 - 1992) = 2.0 / 20 = 0.10$.

Annual growth is 10% per year.

(e) $R^2 = \text{explained SS} / \text{total SS} = 600 / (600 + 400) = 0.60$.

Standard error of regression = $\sqrt{\text{ResidualSS} / (n - k)} = \sqrt{400 / 48} = \sqrt{14.29} = 3.78$.

Multiple Choice Version B:

1. b 2. a 3. b 4. c 5. a

		A	48 and above
		A-	47 and above
		B+	45 and above
		B	43 and above
		B-	41 and above
Scores out of	55	C+	39 and above
75th percentile	45 (82%)	C	37 and above
Median	40.5 (74%)	C-	35 and above
25th percentile	35.5 (65%)	D+	33 and above
		D	31 and above
		D-	29 and above