Chapter 17

© A. Colin Cameron: Answers to Exercises

Chapter Outline

1. Answers: Bivariate Regression
2. Answers: Multivariate Regression

17.1 Bivariate Regression

17.1.1 Hypothesis Testing

1. Regression of y on x with $n = 35$ yields $b_2 = 7$ and $s_{b_2} = 4$.
(a) Two-sided test of stat. sign. of x at 5 percent using critical value approach.
$H_0 : \beta_2 = 0$ against $H_0 : \beta_2 \neq 0$.
$t = (b_2 - 0)/s_{b_2} = 7/4 = 1.75$.
Critical value is $t_{.025;35-2} = TINV(.05, 33) = 2.035$.
Do not reject as $|1.75| < 2.035$. So not statistically significant at 5%.
(b) Two-sided test of stat. sign. of x at 5 percent using p value approach.
Same setup as part (a) with $t = 1.75$.
p value is $Pr[|T_{35-2}| > 1.75) = TDIST(1.75, 33, 2) = 0.089$.
Do not reject as $0.089 > 0.05$. So not statistically significant at 5%.
(c) One-sided test of positive stat. sign. of x at 5 percent using p value approach.
$H_0 : \beta_2 \leq 0$ against $H_0 : \beta_2 > 0$.
Same test statistic as part (a) with $t = 1.75$.

Critical value is $t_{.05;35-2} = TINV(.10, 33) = 1.692$.
Reject as $1.75 > 1.692$. So statistically significant at 5%.
(d) One-sided test of stat. sign. of x at 5 percent using p value approach.
Same setup as part (c) with $t = 1.75$.
p value is $Pr[T_{35-2} > 1.75) = TDIST(1.75, 33, 1) = 0.045$.
Reject as $0.045 < 0.05$. So statistically significant at 5%.
(e) Two-sided test at 10 percent of whether or not slope coefficient equals 1.
$H_0 : \beta_2 = 1$ against $H_0 : \beta_2 \neq 1$. $t = (b_2 - 1)/s_{b_2} = (7 - 1)/4 = 1.50$.
Critical value is $t_{.05;35-2} = TINV(.10, 33) = 1.692$.
Do not reject as $|1.50| < 1.692$. So not statistically different from 1 at 10%.

17.1.2 Coefficient estimate, standard error and t statistic

2.(a) $b_2 = 10$ and $t = 4$. Calculate s_{b_2}, the slope coefficient.
Since $t = b_2/s_{b_2}$ it must be that $s_{b_2} = b_2/t = 10/4 = 2.5$.
(b) $b_2 = 10$ and $p = 0.0575$. Calculate the t statistic. $n = 35$.
The p value is the combined area in the left-hand and right-hand tails. This is 0.0575.
Now $t_{.0575;35-2} = TINV(.0575, 33) = 1.968$.
The t statistic is 1.968.

17.1.3 R-Squared

3. $n = 37$ and $R^2 = 0.4$.
(a) For regression of y on just one regressor x the R^2 equals the squared correlation coefficient. Thus the correlation coefficient $r = \pm\sqrt{0.4} = \pm 0.632$. Note that we do not know the sign of the correlation coefficient without information in addition to R^2.
(b) 40 percent of the variance of y is explained by regression on x.
(c) If $s_y^2 = 100$ what is TSS, the total sum of squares of y?
In general $s_y^2 = \frac{1}{n-1}\sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n-1}TSS$.
Here $100 = \frac{1}{36}TSS$ implies $TSS = 36 \times 100 = 3600$.
(d) Given $R^2 = 0.40$ and part (c), calculate ESS, the error sums of squares.
In general $R^2 = 1 - ESS/TSS$.
Here $0.4 = 1 - ESS/3600$ implies $ESS/3600 = 0.6$ implies $ESS = 0.6 \times 3600 = 2160$.

17.1.4 Inflation and Money Supply Growth

Annual data from 1960 to 1998. $n = 39$.
$y =$INFL. $x =$GM2. $\bar{y} = 4.19$, $s_y^2 = 6.19$. $\bar{x} = 7.12$, $s_x^2 = 9.32$.
$y = 2.094 + 0.295x$. $R^2 = 0.13$.
1. $GM2_t = 100 \times (M2_t - M2_{t-1})/M2_{t-1}$.
2. Money supply growth is more volatile as it has larger variance, since $9.32 > 6.19$.
3. Yes. Significant at 5 percent since $p = 0.024 < 0.05$.
Or more work is $|t| = 2.361 > t_{.025;39-2} = TINV(0.05, 37) = 2.026$.
4. $dy/dx = dINFL/dGM2 = b_2 = 0.295$. A one percent increase in the growth rate of M2 is associated with a 0.295 percent increase in the inflation rate.
5. Question is asking whether or not the slope coefficient equals one.
Test $H_0 : \beta_2 = 1$ against $H_0 : \beta_2 \neq 1$ at 5 percent.
$t = (b_2 - 1)/s_{b_2} = (0.295 - 1)/0.1249 = -0.705/0.1249 = -5.645$.
Critical value is $t_{.05;39-2} = TINV(.05, 37) = 2.026$. Reject H_0 as $5.645 > 2.026$.
Or $p = \Pr(|T_{n-2}| > 5.645) = TDIST(5.645, 37, 2) = 1.90 \times 10^{-6} < 0.05$. Reject H_0.
Either way yields reject H_0. Conclude that statistically different from one.
6. No. $R^2 = .13$ and $r = \sqrt{0.13} = 0.36$ are low. Not highly correlated.
7. The regression cannot discriminate between the two theories as it does not establish causation.

17.1.5 Okun's Law

Annual data from 1960 to 1998. $n = 39$.
$y =$GDP growth rate. $x =$CHANGE IN UNEM. $\bar{y} = 4.19$, $s_y^2 = 6.19$. $\bar{x} = 7.12$, $s_x^2 = 9.32$.
$y = 0.0318 - 1.951x$.
1. $GDPgrowth_t = 100 \times (GDP_t - GDP_{t-1})/GDP_{t-1}$.
2. Yes. Significant at 5 percent since $p = 4.43 \times 10^{-11} < 0.05$.
Or more work is $|t| = 9.184 > t_{.025;39-2} = TDIST(0.05, 37) = 2.026$.
Is the change in unemployment statistically significant at 5 percent?
3. Question is asking whether or not the slope coefficient equals -2.
Test $H_0 : \beta_2 = -2$ against $H_0 : \beta_2 \neq -2$ at 5 percent.
$t = (b_2 - -2)/s_{b_2} = (-1.951 - -2)/0.212 = -0.049/0.212 = -0.23$.

Critical value is $t_{.05;39-2} = TINV(.05, 37) = 2.026$.
Do not reject H_0 as $0.23 < 2.026$.
Or $p = \Pr(|T_{n-2}| > 0.23) = TDIST(0.23, 37, 2) = 0.82 > 0.05$.
Do not reject H_0.
Either way do not reject H_0. Conclude that not statistically different from -2.
4. Increase in unemployment rate from 5 percent to 7 percent in a year implies change in unemployment rate equals $7 - 5 = 2$.
Predicted $\widehat{y} = 0.0318 - 1.951 \times 2 = -3.87$.
Predicted reduction in the growth rate of real GDP is 3.87 percent.

17.2 Multivariate Regression

17.2.1 Hypothesis Testing

1. $y = b_1 + b_2x2 + b_3x3$ with $b_2 = 20$, $t_2 = 2.5$, $b_3 = 40$, $t_3 = 1.5$. $n = 40$.
(a) Two-sided test of stat. sign. $x2$ and $x3$ at 5% using critical value approach.
$H_0 : \beta_j = 0$ against $H_0 : \beta_j \neq 0$.
Critical value is $t_{.025;40-3} = TINV(.05, 37) = 2.026$.
For $x2$, $t_2 = 2.5$ and $|2.5| > 2.026$. So $x2$ statistically significant at 5%.
For $x3$, $t_3 = 1.5$ and $|1.5| < 2.026$. So $x3$ not statistically significant at 5%.
(b) Two-sided test of stat. sign. of $x2$and $x3$ at 5% using p value approach.
For $x2$, p value is $Pr[|T_{40-3}| > 2.5) = TDIST(2.5, 37, 2) = 0.017 < 0.05$.
For $x3$, p value is $Pr[|T_{40-3}| > 1.5) = TDIST(1.5, 37, 2) = 0.142 > 0.05$.
So $x2$ statistically significant at 5% and $x3$ not statistically significant at 5%..
(c) Test $H_0 : \beta_3 = 20$ against $H_0 : \beta_3 \neq 20$ at 10 percent.
$t = (b_3 - 20)/s_{b_3}$.
The problem is what is s_{b_3}? Recall that $t_3 = b_3/s_{b_3}$, so here $1.5 = 40/s_{b_3}$. It follows that $s_{b_3} = 40/1.5 = 26.67$.
Thus $t = (40 - 20)/26.67 = 0.75$.
Critical value is $t_{.05;39-2} = TINV(.05, 37) = 2.026$.
Do not reject H_0 as $0.75 < 2.026$.
Or $p = \Pr(|T_{n-2}| > 0.75) = TDIST(0.75, 37, 2) = 0.458 > 0.05$.
Do not reject H_0.
Either way do not reject H_0. Conclude that not statistically different from -2.

17.2.2 F-test, R-squared and Adjusted R-squared

2. $n = 35$. Regress y on $x2$ yields $R^2 = 0.40$ with $k = 2$. Regress y on $x2$ and $x3$ yields $R^2 = 0.41$ with $k = 3$.
(a) F test of whether or not $x2$ stat. sign. at 5% in regression of y on $x2$.
$H_0 : \beta_2 = 0$ against $H_a : \beta_2 \neq 0$.
$F = [R^2/(k-1)]/[(1-R^2)/(n-k)] = [0.40/(2-1)]/[(1-0.40)/(35-2)] = [0.40]/[0.60/33] = 22$.
(b) F test of whether or not $x2$ and $x3$ are jointly stat. sign. at 5% in regression of y on $x2$ and $x3$.
$H_0 : \beta_2 = 0, \beta_3 = 0$ against H_a :At least one of $\beta_2 \neq 0, \beta_3 \neq 0$.
$F = [R^2/(k-1)]/[(1-R^2)/(n-k)] = [0.41/(3-1)]/[(1-0.39)/(35-3)] = [0.41/2]/[0.59/32] = 13.83$.
(c) In genera $\bar{R}^2 = R^2 - [(k-1)/(n-k)](1-R^2)$.
Just $x2 : \bar{R}^2 = 0.40-[(2-1)/(35-2)](1-0.40) = 0.40-[1/33]\times 0.60 = 0.382$.
Both $x2$ and $x3 : \bar{R}^2 = 0.41-[(3-1)/(35-3)](1-0.41) = 0.41-[2/32]\times 0.59 = 0.373$.
Prefer the model with just $x2$ on the basis of $\bar{R}^2$, as it has higher $\bar{R}^2$.

17.2.3 Republican Voting

$n = 50$.

$$VOTEREPUB = \underset{(20.63)}{58.5814} + \underset{(3.22)}{0.27829} \times POPCH - \underset{(-2.24)}{0.0014034} \times SCHOOL \qquad R^2 = 0.258.$$

1. Yes. The two regressor variables are statistically significant at 5 percent. Critical value is $t_{.025;50-3} = TINV(.05, 47) = 2.012$.
2. Yes. The coefficients of POCCH and SCHOOL have the expected signs and are statistically significant at 5 percent. (Here best to use a one-sided test in which case the critical value is $t_{.05;50-3} = TINV(.10, 47) = 1.678$).
3. Test $H_0 : \beta_2 = 0, \beta_3 = 0$ against H_a :at least one $\beta_2, \beta_3 \neq 0$.
From formula sheet $F = [R^2/(k-1)]/[(1-R^2)/(n-k)] = 8.17$.
Yes. Reject H_0 at 5 percent as $F = 8.17 > FINV(0.05, 2, 47) = 3.195$.
or reject H_0 at 5 percent as $p = FDIST(8.171, 2, 47) = 0.0009 < 0.05$.
4. Prediction for Canada is $VOTEREPUB = 58.5814 + 0.27829 \times 10 - 0.0014034 \times 4000 = 55.75 > 50.00$.
Yes. The model predicts that more than 50 percent would vote Republican.

[A better answer would formally test whether the difference from 50 is statistically significant. There is not enough information here to do that.]
5. Positive: on average as income increases people are more likely to vote Republican. [This result may not necessarily hold once one controls for POPCH and SCHOOL, but I expect it would.]
6. R^2 always increases (or more precisely never decreases) when additional regressors are added. So the increase is not a surprise. The question is whether going from 0.250 to 0.270 is a big or small decrease.
7. Use $\bar{R}^2 = R^2 - [(k-1)/(n-k)](1-R^2)$ which adjusts for model size.
Without income $\bar{R}^2 = 0.258 - [2/47] \times 0.742 = 0.226$.
With income $\bar{R}^2 = 0.270 - [3/47] \times 0.730 = 0.223$.
The model without income is preferred as $\bar{R}^2$ is larger.