Week 5 Code

Ryan Martin

January 31, 2018

Chapter 5 Notes - Multiple Regression

Key Ideas

Coefficients to linear regressions interpreted as partial derivatives.

explanatory variables are the x's.

Assume. MR1. linear model is correct MR2. mean 0 errors MR3. homoskedastic errors MR4. uncorrelated errors MR5. no multicollinearity in x's MR6. (Sometimes) Errors are $N(0, \sigma^2)$

Consequences 1. $E(y|x)$ is everything but e 2. Var(y|x) = var(e) 3. $cov(y_i, y_j|x) = 0$ 4. $y|x \sim N(\beta_0 \sum \beta_i x_i, \sigma^2)$

Interpolation vs Extrapolation 1. Extrapolation is often unlikely to be accurate 2. Interpolation is safer, but may not be sensible if high degree polynomial model or if overfit.

$$\hat{\sigma}^2 = \frac{1}{N-K} \sum \hat{e_i}^2$$

where K is the number of β parameters

Gauss-Markov - for MR model, if MR1-MR5 hold then OLS is BLUE

Recall variance covariance matrices

$se(b_k) := \sqrt{(\widehat{var(b_k)})}$

same rules for linear combos as before

Delta Method Mentioned in 5.6!

- Delta method is an approximation formula for complicated coefficients (e.g. things that we can't easily compute the variance for)
- If $\lambda = f(\beta_1, \beta_2)$ then

$$var(\lambda) \approx (\frac{\partial f}{\partial \beta_1})^2 var(\beta_1) + (\frac{\partial f}{\partial \beta_2})^2 var(\beta_2) + 2(\frac{\partial f}{\partial \beta_1})(\frac{\partial f}{\partial \beta_1}) cov(\beta_1, \beta_2)$$

See appendix 5b.5 for more.

5.7 - Interaction Terms

- So partial of 1 variable may depend on value of another
- log linear models still have same interpretation as before

5.8 Measuring Goodness of Fit - R^2

$R^2 = SSR/SST = \frac{(\sum \hat{y_i} - \bar{y})^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{\sum \hat{e_i}^2}{\sum (y_i - \bar{y})^2}$ where the last equality comes from adding and subtracting y_i from the sum for all i and then a little algebra (called decomposing the square). Note, the last equality depends on an intercept term being in the regression! If don't have an intercept, also the below interpretation is not correct! If no intercept, don't discuss the R^2 (go back to plots)

Whenever we have an R^2, the interpretation is "R^2 percent of the variation in y is explained by the variation in the x variables"

A small but worthwile point to make is that explaining variation is not the same as explaining! I may be able to explain a good deal of variation in car max speed with variation in price, but that doesn't mean price explains max speed.

Questions

5.19

```
Use the data in \texttt{cps4\_small.dat} to estimate the following wage equation:   $$\ln(WAGE) = \beta_
```

a

Report the regression results. Interpret the estimates for non-intercept terms. Are they significantly different from 0?

```r
my_wd <- "C:/Users/ryanj/Dropbox/TA/Econ 103/Winter 2018/Data/s4poe_statadata"
my_file <- paste(my_wd, "cps4_small.dta", sep = "/")
library(haven)
dat <- read_stata(my_file)
```

b

Test the hypothesis that an extra year fo education increases the wage rate by at least 10% against the alternative that is less than 10%

c

Find a 90% interval estimate for the percentage increase in wage from working an addtional hour per week

d

Re-estimate the model with the additional variables $EDUC \times EXPER$, $EDUC^2$ and $EXPER^2$. Report the results. Are the estimated coefficients significantly different from zero?

e

For the new model, find expressions for the marginal effects $\frac{\partial \ln(WAGE)}{\partial EDUC}$ and $\frac{\partial \ln(WAGE)}{\partial EXPER}$

f

Estimate the marginal effects $\frac{\partial \ln(WAGE)}{\partial EDUC}$ for two workers Jill and Wendy. Jill has 16 years of education and 10 years of experience while Wendy ahs 12 years of education and 10 years of experience. What can you say about the marginal effect of education as education increases?

g

Test, as an alternative hypothesis, that Jill's marginal effect of education is greater than that of Wendy. Use a 5% significance level.

h

Estimate the marginal effects $\frac{\partial \ln(WAGE)}{\partial EXPER}$ for two workers Chris and Dave. Chris has 16 years of education and 20 years of experience while Wendy Dave has 16 years of education and 30 years of experience. What can you say about the marginal effect of experience as experience increases?

i

For someone with 16 years of education, find a 95% interval estimate for the number of years of experience after which the marginal effect of experience becomes negative.