Chapter 4: Additional concepts









Formal definition of ***interaction*** (from text): Two variables x1 and x2 are said to ***interact*** if the change in E(Y) for a 1-unit change in x1 (when x2 is held fixed) is dependent on the value of x2.

**Example:** Y= medical costs during 2019: x1=disease severity, x2=age

For young patients, a unit change in disease severity might result in an increase of $500 while for older patients, a unit change in disease severity might result in an increase of $5,0000 in medical costs during 2019.

**Extra Sums of Squares in R**

**>** #Peform a multiple regression using the quasar data

**>** lmod1 <- lm(RFEWIDTH ~ REDSHIFT+LINEFLUX+LUMINOSITY+AB1450+ABSMAG,data=ex496)

**>** summary(lmod1)

**Call:**

**lm(formula = RFEWIDTH ~ REDSHIFT + LINEFLUX + LUMINOSITY + AB1450 + ABSMAG, data = ex496)**

**Residuals:**

**Min 1Q Median 3Q Max**

**-19.911 -8.696 -2.370 2.092 48.732**

**Coefficients:**

**Estimate Std. Error t value Pr(>|t|)**

**(Intercept) 18989.03 26569.36 0.715 0.483**

**REDSHIFT 113.71 102.19 1.113 0.280**

**LINEFLUX 471.36 830.50 0.568 0.577**

**LUMINOSITY -254.50 825.25 -0.308 0.761**

**AB1450 20.78 573.36 0.036 0.971**

**ABSMAG 64.59 570.65 0.113 0.911**

**Residual standard error: 15.81 on 19 degrees of freedom**

**Multiple R-squared: 0.9119, Adjusted R-squared: 0.8887**

**F-statistic: 39.34 on 5 and 19 DF, p-value: 2.218e-09**

**> anova(lmod1)**

**Analysis of Variance Table**

**Response: RFEWIDTH**

**Df Sum Sq Mean Sq F value Pr(>F)**

**REDSHIFT 1 274 274 1.0942 0.308661**

**LINEFLUX 1 2218 2218 8.8720 0.007719 \*\***

**LUMINOSITY 1 2342 2342 9.3701 0.006429 \*\***

**AB1450 1 44329 44329 177.3323 4.38e-11 \*\*\***

**ABSMAG 1 3 3 0.0128 0.911065**

**Residuals 19 4750 250**

**---**

**Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1**

The ANOVA table given by R provides the extra sum of squares for each predictor variable, given that the previous predictors are already in the model. **These are known as Type I Sums of Squares.**

For example: SSR1(ABSMAG) = SSR(ABSMAG| REDSHIFT LINEFLUX LUMINOSITY AB1450) = 3

A **Type III Sum of Squares** for a variable X is equal to SSR(X|all other X’s are in the model)

So, since ABSMAG is the last variable entered into the model,

SSR(ABSMAG| REDSHIFT LINEFLUX LUMINOSITY AB1450)=**SSR1(ABSMAG)=SSR3(AMSMAG)=3**

**The t-statistics given in the coefficients table are based on Type III SS!**

T(ABSMAG) = 0.113 and (0.113)2 = 0.0128 which is the F(ABSMAG) in the Analysis of Variance Table.

IMPORTANT TO REMEMBER: p-values in Coefficients table are based on Type III SS and p-values in Analysis of Variance Table are based on Type I SS.

-----------------------

SSR=Sum of Squares Regression for REDSHIFT = SSR(REDSHIFT)=274

> lmod2 <- lm(RFEWIDTH ~ REDSHIFT,data=ex496)

> anova(lmod2)

Analysis of Variance Table

Response: RFEWIDTH

Df Sum Sq Mean Sq F value Pr(>F)

REDSHIFT 1 274 273.53 0.1173 0.7351

Residuals 23 53642 2332.26

F\*=274/MSE = 274/250 = 1.0960 with 1 and 19 df p=0.308661

-----------------------

SSR(LINEFLUX|REDSHIFT is already in the model) = 2218 with 1 df

> lmod3 <- lm(RFEWIDTH ~ REDSHIFT+LINEFLUX,data=ex496)

> anova(lmod3)

Analysis of Variance Table

Response: RFEWIDTH

Df Sum Sq Mean Sq F value Pr(>F)

REDSHIFT 1 274 273.53 0.1170 0.7355

LINEFLUX 1 2218 2217.81 0.9488 0.3406

Residuals 22 51424 2337.46

F\*=2218/MSE = 2218/250 = 8.8720 with 1 and 19 df p=.007719

-----------------------

SSR(LUMINOSITY|REDSHIFT,LINEFLUX) are already in the model) = 2342 with 1 df

> lmod4 <- lm(RFEWIDTH ~ REDSHIFT+LINEFLUX+LUMINOSITY,data=ex496)

> anova(lmod4)

Analysis of Variance Table

Response: RFEWIDTH

Df Sum Sq Mean Sq F value Pr(>F)

REDSHIFT 1 274 273.53 0.1170 0.7357

LINEFLUX 1 2218 2217.81 0.9489 0.3411

LUMINOSITY 1 2342 2342.31 1.0022 0.3282

Residuals 21 49082 2337.23

F\*=2342/MSE = 2342/250 = 9.3680 with 1 and 19 df p=0.006429

-----------------------