Anthony Cunningham

STAT 3200

Due 4/21/2017

**Homework 8**

#1. > lm.out=lm(intensity~commerce + tradition + commerce:tradition + midpeasant + inequality)

> summary(lm.out)

Coefficients:

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

(Intercept) 12.8425935 10.9301504 1.175 0.2507

commerce -0.8726188 0.3644404 -2.394 0.0241 \*

tradition -0.1958929 0.1302200 -1.504 0.1446

midpeasant -0.0005318 0.0160383 -0.033 0.9738

inequality 2.5073698 2.6291334 0.954 0.3490

commerce:tradition 0.0112353 0.0042432 2.648 0.0136 \*

Residual standard error: 1.109 on 26 degrees of freedom

Multiple R-squared: 0.6721, Adjusted R-squared: 0.609

F-statistic: 10.66 on 5 and 26 DF, p-value: 1.191e-05

A. The interaction coefficient is positive, which goes against the researchers’ expectations, since the coefficients for commerce and tradition are both negative in the model. This implies that the intensity level of the rebellion actually decreases when commerce and tradition predictors increase separately and independent of each other while holding other factors constant. Also, the p-value is significant at alpha-level 0.05, which means that there is evidence that the linear relationship between intensity and commerce, tradition, midpeasant, and inequality are not similar for differing commerce and tradition values.

B. Midpeasant’s estimated coefficient is barely negative (i.e. very close to 0), which completely goes against the sociological theory that says that intensity will be high when middle peasantry is high, but the coefficient says otherwise. In fact, the midpeasant predictor is judged to be insignificant in this model after controlling for the other predictors due to large p-value, which means that midpeasant has no significant effect on intensity level, which, once again, goes against what the researchers expected.

C. Inequality’s estimated coefficient is positive, which goes along with what the researchers expected. However, this predictor also tests to be insignificant in the model after accounting for the other factors due to large p-value, which means that inequality has a negligible effect on intensity level after controlling the other variables in the model, which defies the researchers’ expectations since they thought that inequality would significantly affect intensity levels.

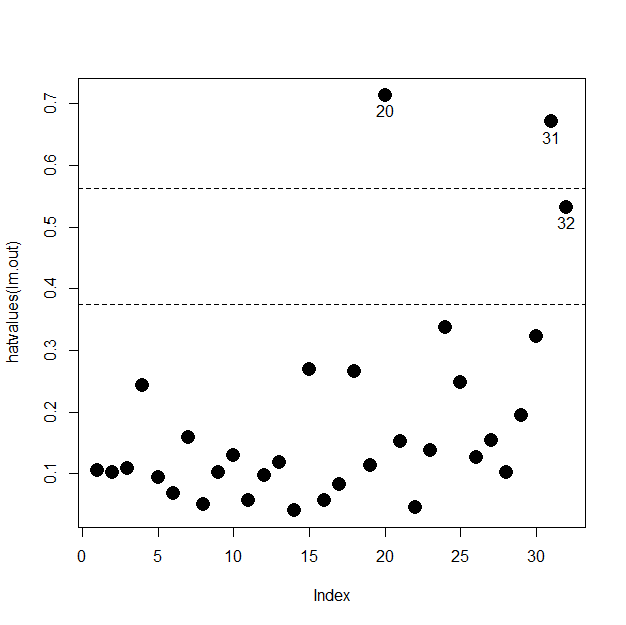
#2. > hatvalues(lm.out)

> plot(hatvalues(lm.out), pch=16, cex=2)

> abline(h=2\*6/32, lty=2)

> abline(h=3\*6/32, lty=2)

> identify(1:32, hatvalues(lm.out), row.names(Chirot))



A. The thresholds used to determine high and very high leverage are 2\*(average hat-value) and 3\*(average hat-value).

B. Observations w/ high leverage: 20 (h=0.714), 31 (h=0.671), 32 (h=0.532)

\* Observation #20 has the highest leverage in this data set.

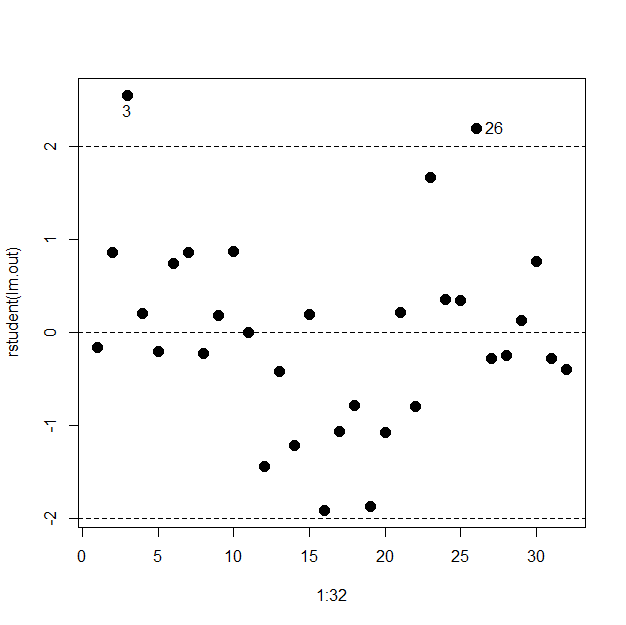
#3. > resids = data.frame(abs(rstudent(lm.out)), row.names(Chirot))

> resids[rev(order(resids[,1])), ]

> plot(1:32, rstudent(lm.out), pch=16, cex=1.5)

> abline(h=c(-2, 0, 2), lty=2)

> identify(1:32, rstudent(lm.out), row.names(Chirot))



A. Observations w/ large studentized residuals: 3 (r=2.553), 26 (r=2.195)

\* Observation #3 has the largest studentized residual.

B. > outlierTest(lm.out)

Largest |rstudent|: \* This studentized residual significantly

rstudent unadjusted p-value Bonferonni p differs from 0 due to large unadjusted

3 2.553413 0.017148 0.54874 p-value. However, the Bonferroni

p-value says that this residual is not

unusually large after comparing for

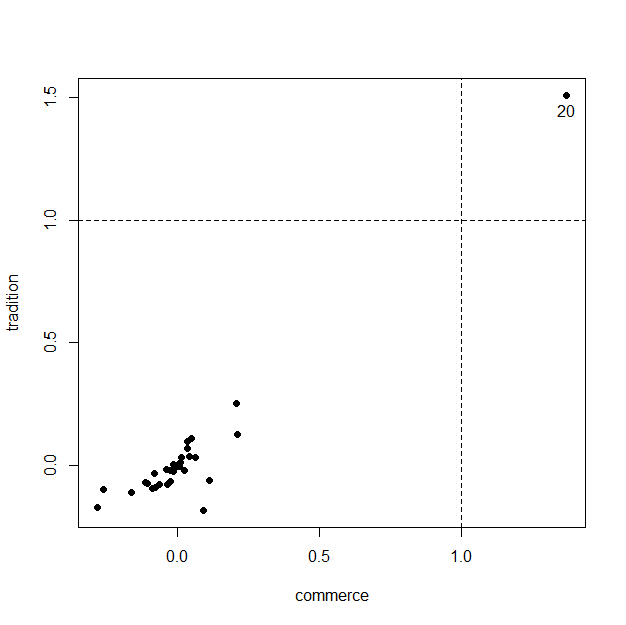
multiple comparisons.

#4. > plot(dfbetas(lm.out)[, c(2,3)], pch=16)

> abline(h=1, lty=2)

> abline(v=1, lty=2)

> identify(dfbetas(lm.out)[,2], dfbetas(lm.out)[,3], row.names(Chirot))



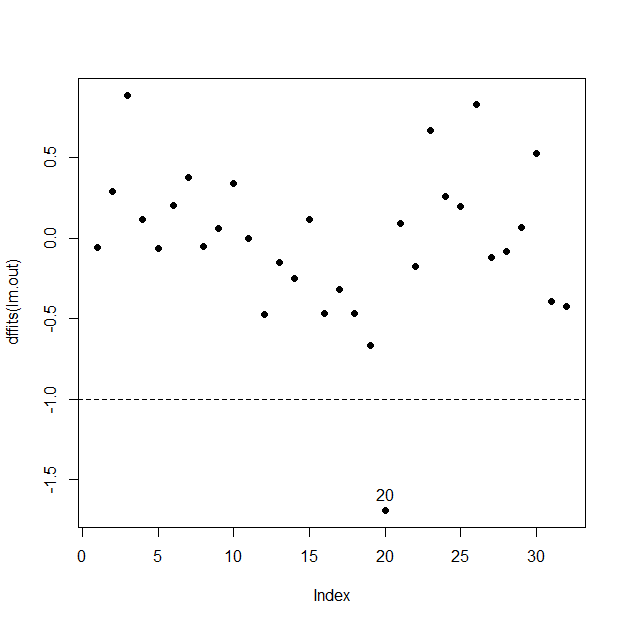
\* Observation #20 may have a large impact on the regression coefficients for both predictors tradition and commerce, after all other predictors are accounted for.

#5. > plot(dffits(lm.out), pch=16)

> abline(h=-1, lty=2)

> abline(h=1, lty=2)

> identify(1:32, dffits(lm.out), row.names(Chirot))

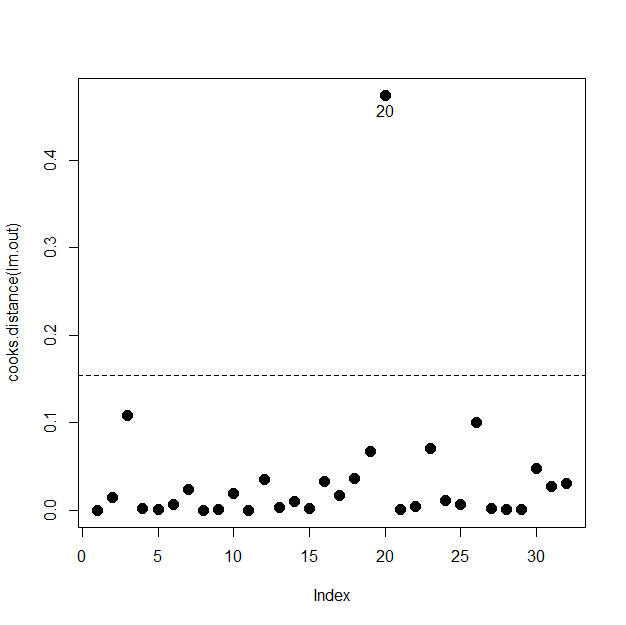


\* Observation #20 may have a large impact on its fitted value.

#6. > plot(cooks.distance(lm.out), pch=16,cex=1.5)

> abline(h=4/(32-5-1), lty=2)

> identify(1:32, cooks.distance(lm.out), row.names(Chirot))



A. Unusually influential cases occur when Cook’s D > 4\*(1/(n – k – 1)).

B. \*Observation #20 appears to be unusually influential.

Cook’sD = (ei\*2/(k+1))(hi/(1 – hi)) = ((1.071)2/(5+1)(0.714/(1 – 0.714))

= 0.191(2.497) = **0.477**

#7. > plot(hatvalues(lm.out), rstudent(lm.out), type="n")

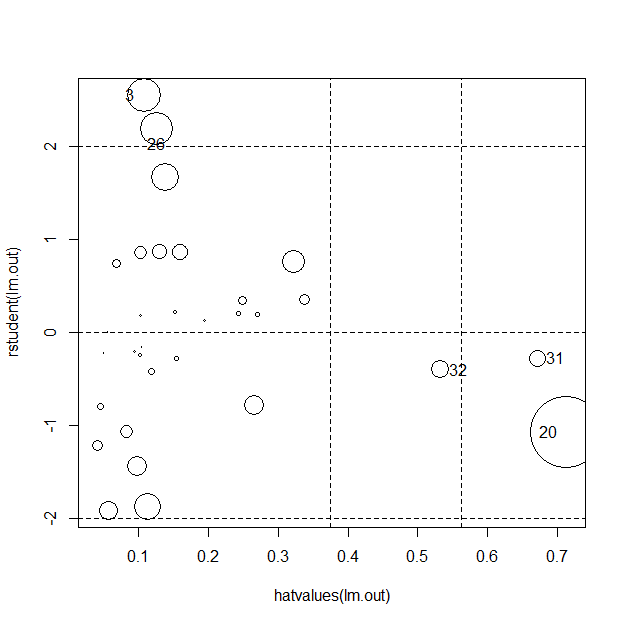
> cook = sqrt(cooks.distance(lm.out))

> points(hatvalues(lm.out), rstudent(lm.out), cex=10\*cook/max(cook))

> abline(h=c(-2,0,2), lty=2)

> abline(v=c(2,3)\*6/32, lty=2)

> identify(hatvalues(lm.out), rstudent(lm.out), row.names(Chirot))



A. Observations 20, 31, and 32 all have high leverage.

B. Observations 3 and 26 have large studentized residuals.

C. Observations with largest influence: 20, 26, 3

\* Observation #20 has the largest influence, mainly due to a large hat-value, although its studentized residual is decent-sized.

#8. > par(mfrow=c(3,2))

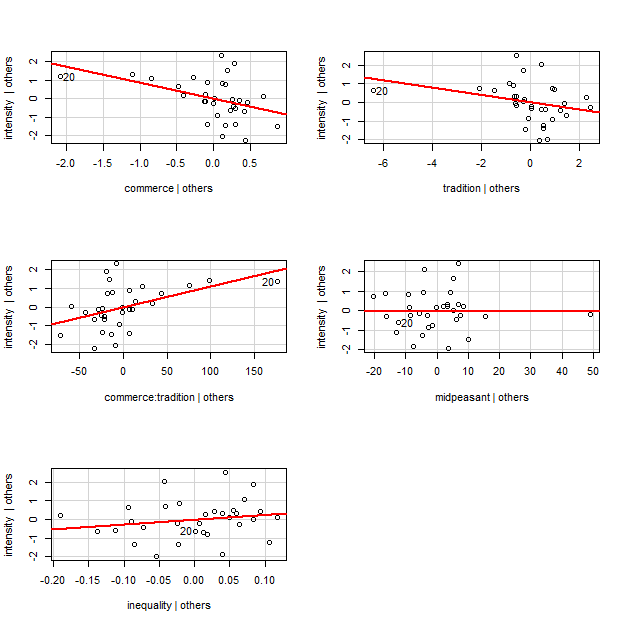
> avPlots(lm.out, "commerce", labels=row.names(Chirot),id.method=cooks.distance(lm.out), id.n=1)

> avPlots(lm.out, "tradition", labels=row.names(Chirot),id.method=cooks.distance(lm.out), id.n=1)

> avPlots(lm.out, "commerce:tradition", labels=row.names(Chirot),id.method=cooks.distance(lm.out), id.n=1)

> avPlots(lm.out, "midpeasant", labels=row.names(Chirot),id.method=cooks.distance(lm.out), id.n=1)

> avPlots(lm.out, "inequality", labels=row.names(Chirot),id.method=cooks.distance(lm.out), id.n=1)



\*Effect on estimated coefficients of predictors after removal of obs. #20 after accounting for other predictors:

(Commerce): estimated coefficient may be slightly larger (in negative direction)

(Tradition): estimated coefficient may be slightly larger (in negative direction)

(Commerce:Tradition): estimated coefficient may be slightly larger (in positive direction)

(Midpeasant): estimated coefficient may not be significantly affected

(Inequality): estimated coefficient may not be significantly affected

#9. > lm.out$coefficients

(Intercept) commerce tradition midpeasant

12.8425934747 -0.8726188387 -0.1958929071 -0.0005317925

Inequality commerce:tradition

2.5073697775 0.0112352820

> lm.out.no20 = lm(intensity[-20] ~ commerce[-20] + tradition[-20] + commerce[- 20]:tradition[-20] + midpeasant[-20] + inequality[-20])

> lm.out.no20$coefficients

(Intercept) commerce[-20]

29.787818597 -1.370165413

tradition[-20] midpeasant[-20]

-0.391682340 -0.006289326

inequality[-20] commerce[-20]:tradition[-20]

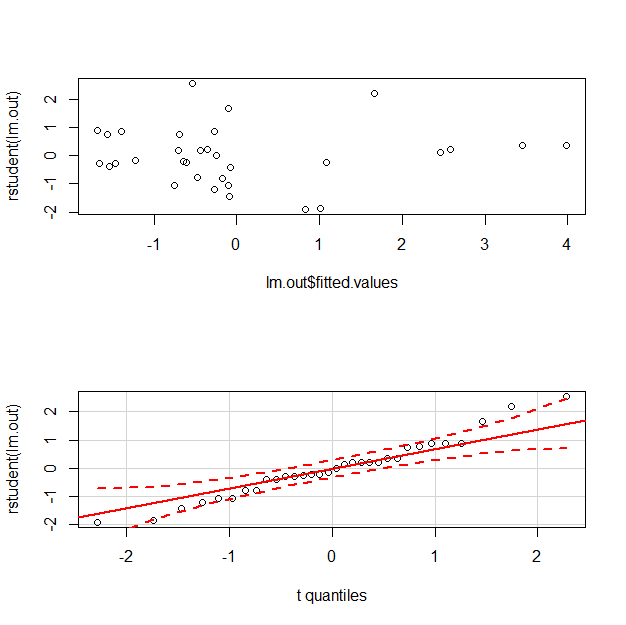
2.523594565 0.016957576

\*Yes, the results are pretty consistent with what I observed, although I may have overstated how much of an effect the removal of obs. #20 would have on “tradition” predictor.

#10. > par(mfrow=c(2,1))

> plot(lm.out$fitted.values, rstudent(lm.out))

> qqPlot(rstudent(lm.out), dist="t", df=32-5-2)



A. \*Constant Variance assumption does not appear to be violated, as the points on the studentized residual plot appear to be haphazard and random.

B. \*Normality assumption does not appear to be violated since nearly all studentized residuals for the data set are confined within t(35-5-2).

C. \*Linearity assumption appears to hold, as there is not obvious curvature in any of the Added-Variable Plots.