Assignment 7 Moeller

Moeller

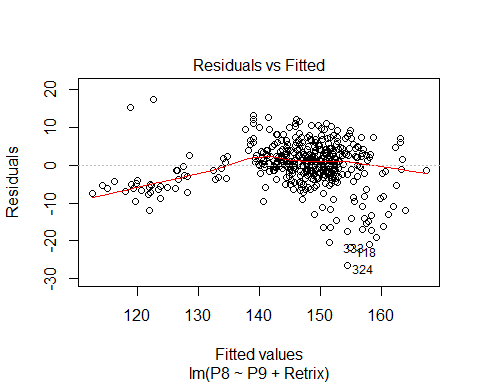
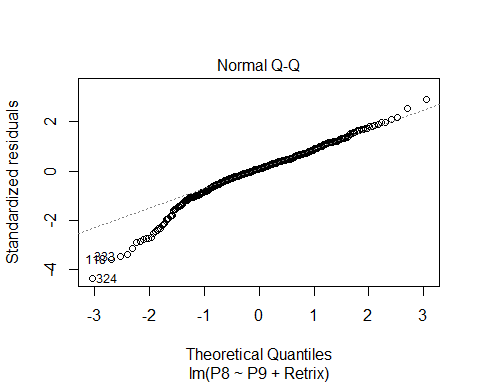
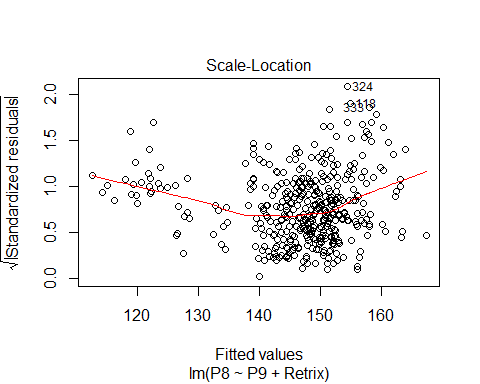
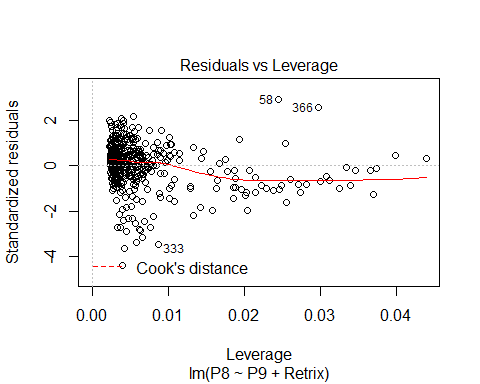
Sunday, October 18, 2015

Problem 1

# read in GLM data  
ptarm <- read.csv("C:/Users/anna.moeller/Documents/School/Research Design/R/ptarm.csv")  
  
# alternate models  
res1 <- lm(P8 ~ P9, data = ptarm)  
res2 <- lm(P8 ~ P9 + Retrix, data = ptarm)  
res3 <- lm(P8 ~ Sex + Site + Sex\*Site, data = ptarm)

Problem 2

# Pick res2 because it has the lowest AIC value. Do the data uphold the assumptions?   
plot(res2)

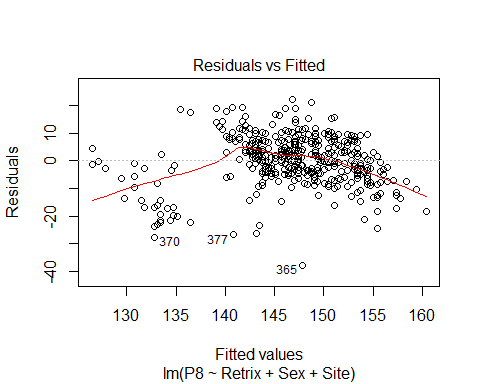
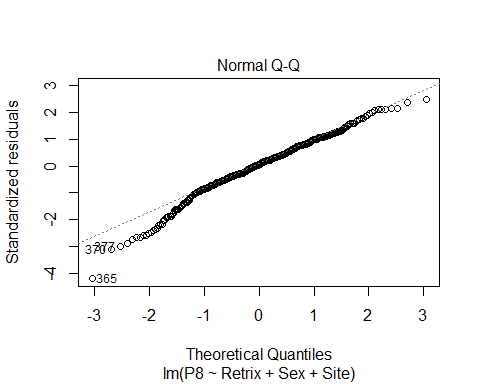
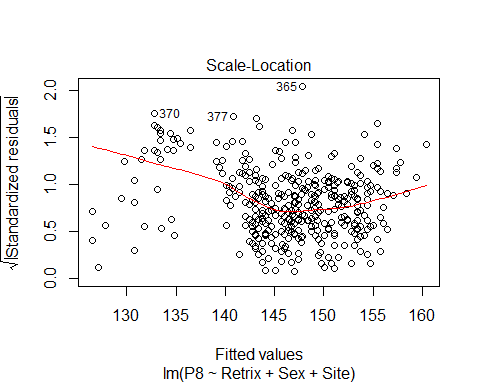
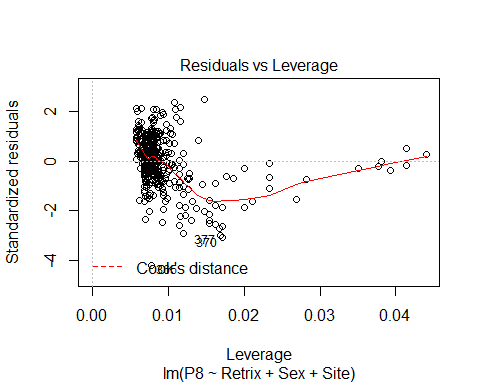
# first assumption: residuals are identically distributed  
# residuals vs. fitted:   
# data appear to have low variance at small values of y   
# and higher variance at larger values of y, but in general they are evenly  
# distributed around the line y = 0  
# assumption: residuals are normally distributed  
# normal Q-Q  
# data appear to be normal for the mid to high values of the theoretical quantiles but  
# non-normal for low values of the theoretical quantiles  
# scale-location  
# the line is curved slightly in the middle, which shows that the   
# standardized residuals are slightly lower for mid values of y than they are for  
# low or high values of y, but in general fit is okay.  
# residuals vs. leverage  
# the line is slightly curved outside the giant cloud of points, which shows that a few   
# data points are having more leverage than the cloud of points. However, no single point   
# seems to be driving the slope results.

Problem 3

res5 <- lm(P8 ~ Retrix + Sex + Site, data = ptarm)  
summary(res5)

##   
## Call:  
## lm(formula = P8 ~ Retrix + Sex + Site, data = ptarm)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -37.794 -4.758 0.579 6.255 22.210   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 184.9161 4.6166 40.054 < 2e-16 \*\*\*  
## Retrix -0.3322 0.0392 -8.472 3.96e-16 \*\*\*  
## SexMale 9.0388 0.9155 9.873 < 2e-16 \*\*\*  
## SiteKluane -3.6458 0.9140 -3.989 7.81e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 9.087 on 426 degrees of freedom  
## Multiple R-squared: 0.3113, Adjusted R-squared: 0.3065   
## F-statistic: 64.19 on 3 and 426 DF, p-value: < 2.2e-16

# Interpret the results  
# Multiple R-squared shows that this model explains 31% of the variance observed in the data.  
# The p-value for the F-statistic shows that it is unlikely to observe these data if   
# these covariates do not have a significant effect overall.   
# The p-values for each covariate show that it is unlikely to observe these data if these   
# covariates do not have a significant effect on P8  
# The estimates for Retrix and SiteKluane are negative, which shows that   
# birds with longer Retrices have shorter P8 than birds with short Retrices and  
# birds at Kluane have shorter P8 than birds at Chilkat  
# The estimate for SexMale is positive, which shoes that   
# male birds have longer P8 than female birds  
# The covariates are not scaled and centered, so we cannot interpret the relative magnitude  
# of the relationship for each one. If they were scaled and centered, I would compare  
# the Estimate value.   
# Does the model fit?  
plot(res5)

# residuals vs. fitted  
# the line is curved which indicates the fit is not great. The points on the left and right  
# side of the fitted values are all below the line y=0  
# normal Q-Q  
# this looks pretty good. The points below theoretical quantile -1 do not fall on the   
# line, but it is not as bad as other Q-Q plots we have seen  
# residuals vs. leverage  
# It appears that the points outside the cloud of points are driving the slope  
# quite a bit (because the line is not straight) but no points are outside the   
# Cook's distance, so the fit may be okay.   
# In general, fit doesn't appear great, but seems okay.