

Stat 404 Lab 11

SAS

1. Read in the file cuckoos.csv from the S: drive. The variables are Length (numerical) and Species (character).
2. Use an ODS HTML statement so that you can specify any style you want your report to look like (list of styles is on the website).
3. Run the following code: PROC GLM plot=(diagnostics residuals);
CLASS Species;
MODEL Length=Species;
MEANS Species;
RUN;
QUIT;
4. Close your ODS statement.

R

1. Read in the cuckoo.csv file (use `read.csv`).
2. Open a new R markdown document. Choose PDF document (if you have MikTeX or MacTeX installed), give it a title and specify your name if you want.
3. Run the following code in your document: `cuckoo.fit=lm(Length~Species)`

```
anova(cuckoo.fit)
# diagnostic plots
res=residuals(cuckoo.fit)
pred=fitted.values(cuckoo.fit)
par(mfrow=c(2,2))
# mean of residuals = 0
hist(res)
# homogenous (constant) variance
plot(pred,res,pch=17,main="Predicted vs. Residuals")
abline(0,0)
# independence of residuals
order=c(1:length(res))
plot(order,res,type='l')
abline(0,0)
# another way to check independence of residuals
# QQplot (normal probability plot)
qqnorm(res)
qqline(res)
par(mfrow=c(1,1))

# load the car package
library(car)

# look at hat values
plot(hatvalues(cuckoo.fit))
abline(h=2*(length(cuckoo.fit$coefficients))/length(cuckoo.fit$residuals), col=2)
# influence
# look at the distance between the regression coefficients with the ith observation
```

```
present or absent
plot(cooks.distance(cuckoo.fit))
abline(h=4/(length(cuckoo.fit$residuals)-length(cuckoo.fit$coefficients)), col=2)
bound=1.5*(2*(length(cuckoo.fit$coefficients))/length(cuckoo.fit$residuals))
# Influence plot in car-package combines the studentized residuals, hat values and Cook's
distances
# area of the circles correspond to Cook's distances
influencePlot(cuckoo.fit, xlim=c(0,bound), ylim=c(-5,5))
```