

Review

```
#USGS document
install.packages("NADA") #install NADA package
library(NADA) #load NADA package in R
data(package="NADA") #explore different datasets available in NADA
data(HgFish) # Mercury concentrations in fish across the United States
?HgFish
attach(HgFish) #Use attach when you work with only one dataset

#general look at the data
summary(HgFish)

#Explore distribution of response variable and transform it to make it normally
distributed if necessary
hist(Hg)
logHg <- log(Hg)
hist(logHg)
boxplot(Hg)
boxplot(logHg)
logHg_std <- (logHg-mean(logHg))/sd(logHg)
qqnorm(logHg_std)
abline(0,1)

shapiro.test(Hg)
shapiro.test(logHg)

Hg_Len <- Hg/Length
logHg_L <- log(Hg_Len)

shapiro.test(Hg_Len)
shapiro.test(logHg_L)

#Explore correlation and reasonable covariates
Hg_data <- data.matrix(HgFish) #Hg_data is all numerical not right

cor(Hg_data, use="pairwise.complete.obs")

cor.test(SedLOI, WatDOC, na.action= na.omit)

# subset data
Hg_AF <- subset(HgFish, LandUse == "AF")

#simple regression

# relation of Hg in Fish to Hg in water

plot(WatMeHg, logHg_L)
plot(log(WatMeHg), logHg_L)
reg2 <- lm(logHg_L ~ log(WatMeHg))
```

Review

```
summary(reg2)

abline(reg2)
fitted(reg2)
predict(reg2, se.fit=TRUE)
resid(reg2)

# residual analysis
plot(fitted(reg2),resid(reg2))
resid_reg2 <- resid(reg2)
resid_std <- (resid_reg2 - mean(resid_reg2)) / sd(resid_reg2)
qqnorm(resid_std)
abline(0,1)

coefficients(reg2)
anova(reg2)
df.residual(reg2)

#Only look at species of large mouth bass
logHg_L_Large <- logHg_L[Species=="LargemouthBass"]
logWatMeHg_L <- log(WatMeHg[Species=="LargemouthBass"])

plot(logWatMeHg_L, logHg_L_Large)
reg2 <- lm(logHg_L_Large ~ logWatMeHg_L)

sumy(reg2)
abline(reg2)

#Explore total mercury concentration in water
plot(WatTotHg, logHg_L)
plot(log(WatTotHg), logHg_L)

reg3 <- lm(logHg_L ~ log(WatTotHg))
summary(reg3)

abline(reg3)

# relation of Hg in Fish to percent of wetland
plot(PctWetland, logHg_L)
reg1 <- lm(logHg_L ~ PctWetland)

summary(reg1)
abline(reg1)

# multiple regression
SedAVS_c <- SedAVS ^ (1/3)
reg_all <- lm(logHg_L ~ PctWetland + log(WatMeHg) + log(WatTotHg) + SedMeHg +
```

Review

```
log(SedTotHg) + log(WatDOC) + SedLOI + SedAVS_c)

summary(reg_all)

plot(fitted(reg_all), resid(reg_all))
resid_reg <- resid(reg_all)
resid_std <- (resid_reg - mean(resid_reg)) / sd(resid_reg)
qqnorm(resid_std)
abline(0,1)

#Bring in land use
plot(LandUse, log(WatMeHg))

LU_Urb <- ifelse(LandUse=="Urb", 1, 0)
LU_Bkg <- ifelse(LandUse=="Bkg", 1, 0)
LU_Mine <- ifelse(LandUse=="Mine", 1, 0)
LU_Ag <- ifelse(LandUse=="Ag", 1, 0)
LU_AF <- ifelse(LandUse=="AF", 1, 0)

reg_all2 <- lm(logHg_L ~ LU_Urb+LU_Mine+LU_Ag+LU_AF+PctWetland + log(WatMeHg) +
log(WatTotHg) + SedMeHg + log(SedTotHg) + log(WatDOC) + SedLOI + SedAVS_c)
# it is the same as reg_all2 <- lm(logHg_L ~ factor(LandUse) + PctWetland +
log(WatMeHg) + log(WatTotHg) + SedMeHg + log(SedTotHg) + log(WatDOC) + SedLOI +
SedAVS_c)
summary(reg_all2)
plot(LandUse, PctWetland)

install.packages("car")
library(car)
vif(reg_all2) #Check multicollinearity

reduced1 <- lm(logHg_L ~ LU_Urb+LU_Mine+LU_Ag+LU_AF+PctWetland + log(WatMeHg) +
log(WatTotHg) + SedMeHg + log(WatDOC) + SedLOI + SedAVS_c)
summary(reduced1)
vif(reduced1)

reduced2 <- lm(logHg_L ~ LU_Urb+LU_Mine+LU_Ag+LU_AF+PctWetland + log(WatTotHg) +
SedMeHg + log(WatDOC) + SedLOI + SedAVS_c)
summary(reduced2)
vif(reduced2)
```