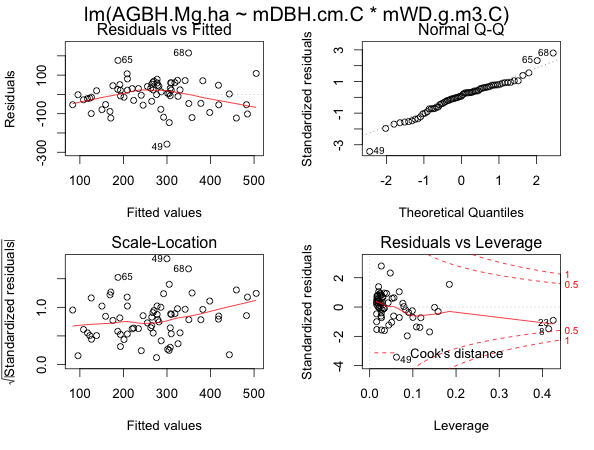
Problem 1

Null hypothesis: None of the variables in mean tree diameter, mean height, mean wood density, mean basal area, or presence of tree falls in the plot has influence on plot biomass;

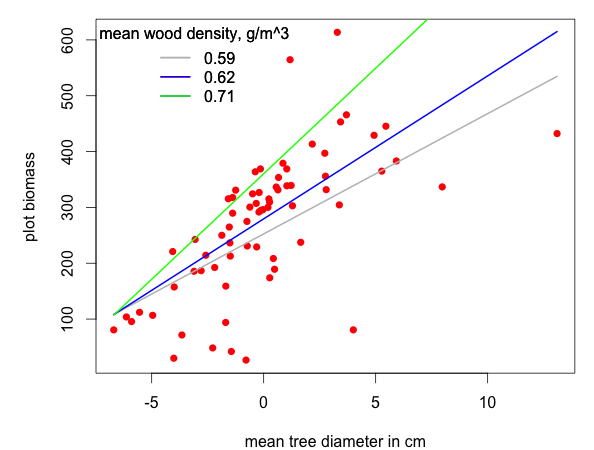
Alternative hypothesis: One of the variables in the mean tree diameter, mean height, mean wood density, mean basal area, and presence of tree falls in the plot has influence on plot biomass.

First I examined if the plot biomass meets the assumption of normality; then through pairwise plots I found that mDBH.cm, mH.m and mBA.cm2 are highly correlated. So I exclude mH.m and mBA.cm2 and build the initial model with mDBH.cm, mWD.g.m3, and factor Tree.Fall and all the interactions between these three.

After eliminating insignificant variables, excluding outlier observations, and centering the independent variables, I build the linear model: AGBH.Mg.ha = 270.482 + 24.251 \* mDBH.cm.C + 898.053 \* mWD.g.m3.C + 135.518 \* mDBH.cm.C \* mWD.g.m3.C. (Adjusted R-squared : 0.6067; F-statistic: 35.96 on 3 and 65 DF, p-value: 8.09e-14)



1 cm increase in mean tree diameter will result in more than 24.251Mg/ha increase in biomass; with greater mean wood density, the biomass will increase quicker.

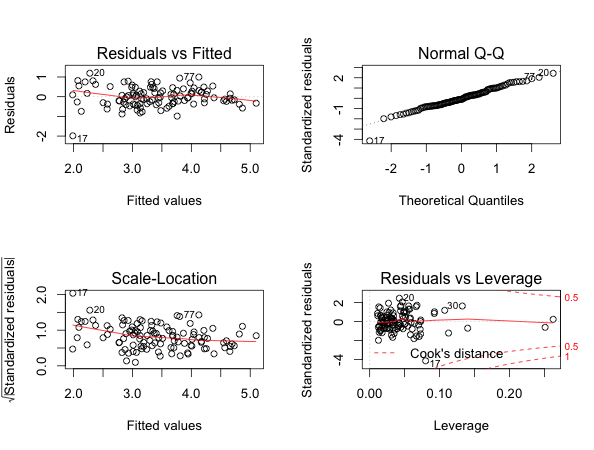


Problem 2

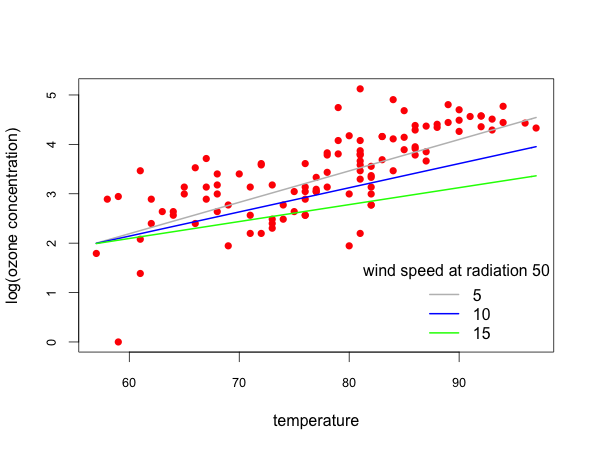
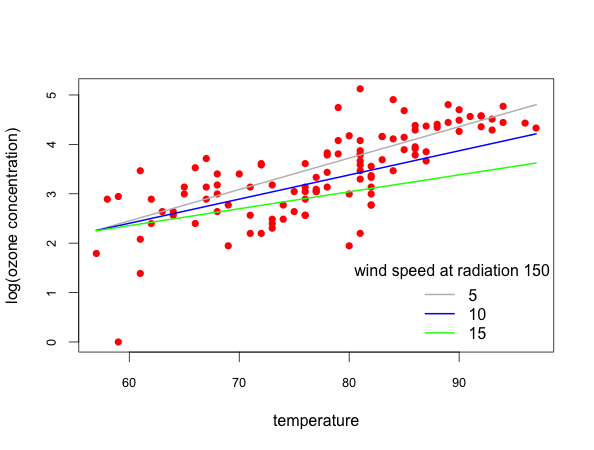
Null hypothesis: None of the variables in the temperature, wind speed or solar radiation has influence on ozone concentration;

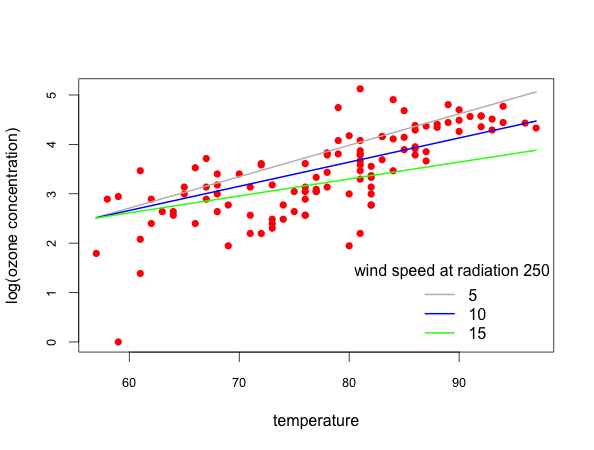
Alternative hypothesis: One of the variables in the temperature, wind speed and solar radiation has influence on ozone concentration.

I conducted logarithm transformation on ozone concentration so as to fit the normal assumption. Linear model I built: log(ozone) = -2.5845520 + 0.0782341 \*temp + 0.0025939 rad + 0.1661571 wind - 0.0029299 \*temp\*wind (Adjusted R-squared: 0.6669 ;F-statistic: 56.05 on 4 and 106 DF, p-value: < 2.2e-16)



The model shows 100 units increase in radiation will result in e^0.25939=1.30 or 30% increase in ozone concentration; at wind speed 10, 1 unit increase in the temperature will result in e^(0.0782341-0.0029299\*10) = 1.05 or 5% increase in ozone concentration. As shown in graph, the speed of the increase in the ozone concentration by 1 unit of the temperature will decrease as the wind speed increases.





Appendix

setwd("/Users/xuhong/Documents/Duke/Term 1 Courses/ENVIRON 710 Applied data analysis/Lab Github/ada/Lab 10")

source("pair.fun.R")

require("Sleuth3")

require("car")

# Problem 1

tdat <- read.csv("TreePlotsA.csv")

# exploratory analysis

pairs(tdat, lower.panel=panel.smooth, upper.panel=panel.cor,

diag.panel=panel.hist)]

cor(tdat[, -8])

# mDBH.cm, mH.m and mBA.cm2 show high correlation

df <- tdat[, c("mDBH.cm", "mH.m", "mBA.cm2")]

cor(df)

qqnorm(tdat$AGBH.Mg.ha)

qqline(tdat$AGBH.Mg.ha)

# build the full model. Drop highly correlate terms

lm0 <- with(tdat, lm(AGBH.Mg.ha ~ mDBH.cm\*mWD.g.m3\*factor(Tree.Fall)))

summary(lm0)

lm.f <- step(lm0)

summary(lm.f)

# diagnosis

par(mfrow=c(2,2))

plot(lm.f)

# 47th observation is dubious

lm1 <- with(tdat[-47,], lm(AGBH.Mg.ha ~ mDBH.cm\*mWD.g.m3 + factor(Tree.Fall)))

summary(lm1)

lm.f1 <- step(lm1)

summary(lm.f1)

plot(lm.f1)

# Tree.Fall is not significant

lm.final <- with(tdat[-47,], lm(AGBH.Mg.ha ~ mDBH.cm\*mWD.g.m3))

summary(lm.final)

plot(lm.final)

avPlots(lm.final)

# center variables

tdat$mDBH.cm.C <- tdat$mDBH.cm - mean(tdat$mDBH.cm)

tdat$mWD.g.m3.C <- tdat$mWD.g.m3 - mean(tdat$mWD.g.m3)

# fit model, one more time

lm.final.C <- with(tdat[-47,], lm(AGBH.Mg.ha ~ mDBH.cm.C\*mWD.g.m3.C))

summary(lm.final.C)

par(mfrow=c(2,2))

# diagnosis

plot(lm.final.C)

avPlots(lm.final.C)

vif(lm.final.C)

# First, plot biomass (AGBH.Mg.ha) against mean tree diameter (mDBH.cm)

with(tdat, plot(mDBH.cm.C, AGBH.Mg.ha, xlab = "mean tree diameter in cm",

ylab = "plot biomass", col="red", pch=16))

# Get three levels of mean wood density (mWD.g.m3)

summary(tdat$mWD.g.m3.C)

vals <- c(-0.02, 0.01, 0.10)

# Generate a sequence of mean tree diameter (mDBH.cm)

x <- seq(min(tdat$mDBH.cm.C), max(tdat$mDBH.cm.C), length=100)

# Plot curves at three levels of mean wood density (mWD.g.m3)

coef <- lm.final.C$coefficients

curve(coef[1] + coef[2]\*x + coef[3]\*vals[1] + coef[4]\*x\*vals[1], col="grey", lwd=2, add=T)

curve(coef[1] + coef[2]\*x + coef[3]\*vals[2] + coef[4]\*x\*vals[2], col="blue", lwd=2, add=T)

curve(coef[1] + coef[2]\*x + coef[3]\*vals[3] + coef[4]\*x\*vals[3], col="green", lwd=2, add=T)

vals.original <- vals + mean(tdat$mWD.g.m3)

vals.original <- round(vals.original, digits = 2)

legend("topleft", legend=vals.original, lty = 1, col = c("grey", "blue", "green"), lwd = 2,

bty = "n", title="mean wood density, g/m^3")

# Problem 2

odat <- read.csv("ozone.data.csv")

# exploratory data analysis

pairs(odat, lower.panel=panel.smooth, upper.panel=panel.cor,

diag.panel=panel.hist)

# transform ozone so as to fit normality

hist(log(odat$ozone))

qqnorm(log(odat$ozone))

qqline(log(odat$ozone))

odat$l.ozone <- log(odat$ozone)

pairs(odat, lower.panel=panel.smooth, upper.panel=panel.cor,

diag.panel=panel.hist)

# no significant correlation between indenpendent variables

cor(odat)

# build initial model

lm.oz.0 <- with(odat, lm(log(ozone) ~ temp\*rad\*wind))

summary(lm.oz.0)

# update

lm.oz.1 <- update(lm.oz.0, ~.-temp:rad:wind)

summary(lm.oz.1)

lm.oz.2 <- update(lm.oz.1, ~.-temp:rad)

summary(lm.oz.2)

lm.oz.3 <- update(lm.oz.2, ~.-rad:wind)

summary(lm.oz.3)

lm.oz.4 <- update(lm.oz.3, ~.-wind)

summary(lm.oz.4)

anova(lm.oz.0, lm.oz.1, lm.oz.2, lm.oz.3, lm.oz.4)

AIC(lm.oz.0, lm.oz.1, lm.oz.2, lm.oz.3, lm.oz.4)

# we end at lm.oz.3

par(mfrow=c(2,2))

plot(lm.oz.3)

# plot

summary(odat$wind)

summary(odat$rad)

wval <- c(5, 10, 15)

rval <- c(50, 150, 250)

par(mfrow=c(1,1))

x <- with(odat, seq(min(temp), max(temp), length=100))

coef <- lm.oz.3$coefficients

with(odat, plot(temp, l.ozone, xlab="temperature", ylab="log(ozone concentration)", pch=16,

col="red", cex.axis=0.8))

curve(coef[1] + coef[2]\*x + coef[3]\*rval[1] + coef[4]\*wval[1] + coef[5]\*x\*wval[1],

add = T, col = "grey", lwd = 2)

curve(coef[1] + coef[2]\*x + coef[3]\*rval[1] + coef[4]\*wval[2] + coef[5]\*x\*wval[2],

add = T, col = "blue", lwd = 2)

curve(coef[1] + coef[2]\*x + coef[3]\*rval[1] + coef[4]\*wval[3] + coef[5]\*x\*wval[3],

add = T, col = "green", lwd = 2)

legend("bottomright", legend=wval, lty = 1, col = c("grey", "blue", "green"), lwd = 2,

bty = "n", title="wind speed at radiation 50")

with(odat, plot(temp, l.ozone, xlab="temperature", ylab="log(ozone concentration)", pch=16,

col="red", cex.axis=0.8))

curve(coef[1] + coef[2]\*x + coef[3]\*rval[2] + coef[4]\*wval[1] + coef[5]\*x\*wval[1],

add = T, col = "grey", lwd = 2)

curve(coef[1] + coef[2]\*x + coef[3]\*rval[2] + coef[4]\*wval[2] + coef[5]\*x\*wval[2],

add = T, col = "blue", lwd = 2)

curve(coef[1] + coef[2]\*x + coef[3]\*rval[2] + coef[4]\*wval[3] + coef[5]\*x\*wval[3],

add = T, col = "green", lwd = 2)

legend("bottomright", legend=wval, lty = 1, col = c("grey", "blue", "green"), lwd = 2,

bty = "n", title="wind speed at radiation 150")

with(odat, plot(temp, l.ozone, xlab="temperature", ylab="log(ozone concentration)", pch=16,

col="red", cex.axis=0.8))

curve(coef[1] + coef[2]\*x + coef[3]\*rval[3] + coef[4]\*wval[1] + coef[5]\*x\*wval[1],

add = T, col = "grey", lwd = 2)

curve(coef[1] + coef[2]\*x + coef[3]\*rval[3] + coef[4]\*wval[2] + coef[5]\*x\*wval[2],

add = T, col = "blue", lwd = 2)

curve(coef[1] + coef[2]\*x + coef[3]\*rval[3] + coef[4]\*wval[3] + coef[5]\*x\*wval[3],

add = T, col = "green", lwd = 2)

legend("bottomright", legend=wval, lty = 1, col = c("grey", "blue", "green"), lwd = 2,

bty = "n", title="wind speed at radiation 250")