Math 327 Chapter 4 Homework

### \_\_ Michael Streyle \_\_

# Code for 4.3b  
# Open the data file, CH01PR2.txt  
mydata <- read.table(file.choose(),header=F,col.names=c("Y","X"))  
  
xname = "Service Time (minutes)"  
yname = "Copiers Serviced (#)"  
  
attach(mydata)  
  
myfit <- lm (Y ~ X)  
myfit

##   
## Call:  
## lm(formula = Y ~ X)  
##   
## Coefficients:  
## (Intercept) X   
## -0.5802 15.0352

summary(myfit)

##   
## Call:  
## lm(formula = Y ~ X)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -22.7723 -3.7371 0.3334 6.3334 15.4039   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -0.5802 2.8039 -0.207 0.837   
## X 15.0352 0.4831 31.123 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 8.914 on 43 degrees of freedom  
## Multiple R-squared: 0.9575, Adjusted R-squared: 0.9565   
## F-statistic: 968.7 on 1 and 43 DF, p-value: < 2.2e-16

qt(.9875, 43)

## [1] 2.322618

confint(myfit, level=.975)

## 1.25 % 98.75 %  
## (Intercept) -7.092642 5.932329  
## X 13.913221 16.157275

# Fit a regression through the origin, for 4.16a  
int0fit = lm (Y ~ 0 + X)  
int0fit

##   
## Call:  
## lm(formula = Y ~ 0 + X)  
##   
## Coefficients:  
## X   
## 14.95

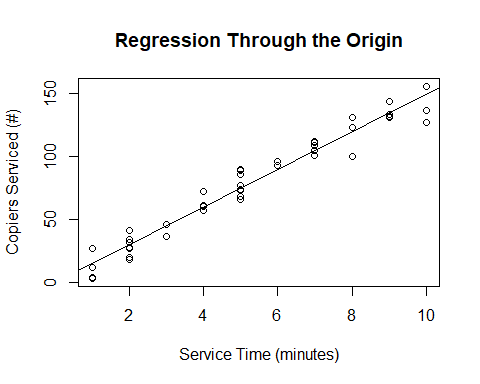
confint(int0fit, level=0.90)

## 5 % 95 %  
## X 14.56678 15.32767

predict(int0fit, data.frame(X=c(6)), interval="prediction", level=.90)

## fit lwr upr  
## 1 89.68338 74.69559 104.6712

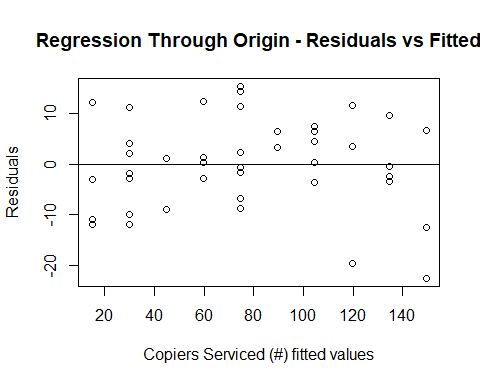
# Plot the data  
plot.new()  
plot(X, Y, xlab=xname, ylab=yname, main="Regression Through the Origin")  
abline(myfit)



# save the residuals  
int0resid = int0fit$residuals  
sum (int0resid)

## [1] -5.862797

# Plot residuals vs fitted  
plot.new()  
plot (int0fit$fitted.values, int0resid, xlab=paste(yname, "fitted values"), ylab="Residuals", main="Regression Through Origin - Residuals vs Fitted")  
abline(h=0)



# Lack of fit test  
full = lm (Y ~ 0 + as.factor(X))  
anova(int0fit, full)

## Analysis of Variance Table  
##   
## Model 1: Y ~ 0 + X  
## Model 2: Y ~ 0 + as.factor(X)  
## Res.Df RSS Df Sum of Sq F Pr(>F)  
## 1 44 3419.8   
## 2 35 2797.7 9 622.12 0.8648 0.5644

qf(.99, 9, 35)

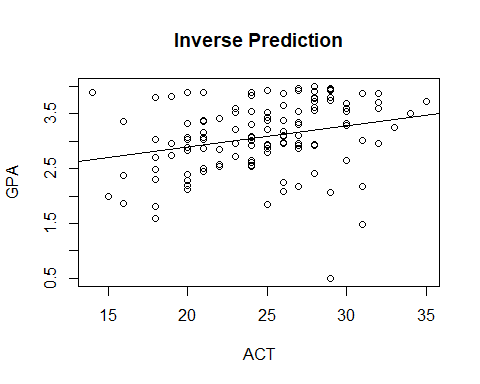
## [1] 2.963012

4.17c. Lack of fit test. State hypotheses, decision rule (can be in terms of p-value), and conclusion.

# inverse prediction using the GPA data.  
# Need to read in the GPA data, and change variable names.  
mydata2 <- read.table(file.choose(),header=F,col.names=c("Y","X"))  
xname = "ACT"  
yname = "GPA"  
attach(mydata2)

## The following objects are masked from mydata:  
##   
## X, Y

myfit = lm (Y ~ X)  
plot (X, Y, xlab=xname, ylab=yname, main="Inverse Prediction")  
abline (myfit)



b0 = myfit$coeff[1]  
b1 = myfit$coeff[2]  
xnew = (3.4 - b0)/b1  
summ = summary (myfit)  
mse = summ$sigma^2  
n = length(Y)  
numer = (xnew - mean(X))^2  
denom = sum ((X - mean(X))^2)  
s.predx.sq = (mse/b1^2)\*(1 + 1/n + numer/denom)  
xnew.lower = xnew - qt(0.90, n-2)\*sqrt(s.predx.sq) # Change confidence level, as needed  
xnew.upper = xnew + qt(0.90, n-2)\*sqrt(s.predx.sq) # Change confidence level, as needed  
xnew.lower

## (Intercept)   
## 12.0481

xnew.upper

## (Intercept)   
## 54.1917

# Prediction and Confidence interval for predicted X value at Y=16  
data.frame(Xnew = c(xnew), Lower = c(xnew.lower), Upper = c(xnew.upper), row.names=c("Prediction"))

## Xnew Lower Upper  
## Prediction 33.1199 12.0481 54.1917

# Equation to check the reasonableness of the interval, equation 4.33, p. 170 should be approx. < 0.1  
eqn4.33 = (qt(0.90, n-2)^2)\*mse/(b1^2 \* denom)  
eqn4.33

## X   
## 0.1797486