Math 327 Chapter 4 Homework

\*I did not like the formatting of knitting my written out answers, so I just wrote out my answers after the knitted section (pg 7-8)

### \_\_ 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

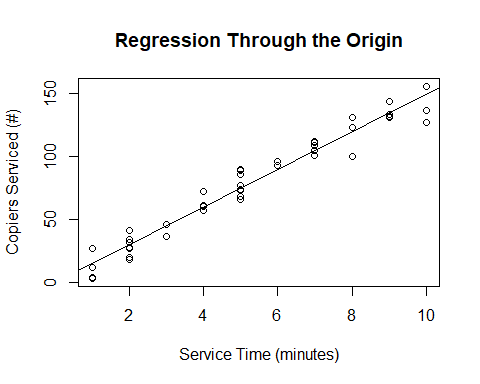
confint(int0fit, level=0.90)

## 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

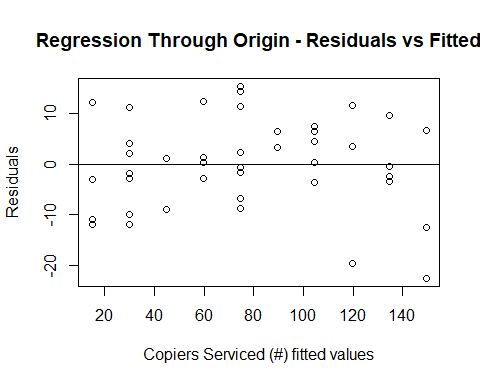
# Plot the data  
plot.new()  
plot(X, Y, xlab=xname, ylab=yname, main="Regression Through the Origin")  
abline(myfit)



# 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)

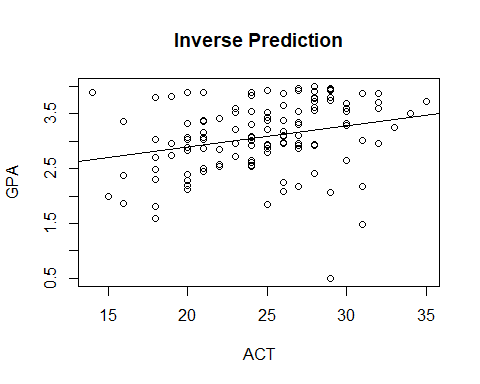
## [1] 2.963012

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

Chapter 4 Written Out Answers

3) a) b1 and b0 err in opposite directions because if b1 is overestimated, then it will decrease the y-intercept which decreases b0. Same concept if b1 is underestimated, it will tilt the line and increase b0.

b) Since the question asks for the Bonferroni joint confidence intervals for β0 and β1, using a 95 percent family confidence interval, that means the level of significance for each β0 and β1 should be divided by two: 1 - .95 = .05 then .05/2 = 0.025 = α for family confidence interval. Then using the qt function, we obtain B: qt(1-0.025/2, 45-2) = qt(.9875, 43) = 2.322618. From the lm function we get a b0 = -0.5802 and b1 = 15.0352. Then using the summary table, we find the standard errors and use them (2.8039 and 0.4831 respectively) to calculate the Bonferroni joint confidence intervals. β1 = 15.0352 ± 2.322618(0.4831) 🡪 13.913 ≤ β1 ≤ 16.157 and β0 = −0.5802 ± 2.32262(2.8039) 🡪 −7.093 ≤ β0 ≤ 5.932. This result could also be achieved easier by using the confint(myfit, level=.975) function as shown in the knit section – notice that the level = .975 not .95 because it is a family confidence interval (1 – α/2), but when in qt function, the (1 – α/2) still needs to be divided by 2.

c) Does my Bonferroni joint confidence intervals for β0 and β1 support the suggestion that β0 = 0 and β1 = 14.0? Since 0 and 14.0 are both in the respective joint confidence intervals from part (a) with a 95 percent family confidence interval, YES, I support the consultant’s suggestion.

16) a) Assuming linear regression through the origin is appropriate, the estimated regression function is Ŷ = 14.95X.

b) The 90 percent confidence interval for β1 is 14.56678 ≤ β1 ≥ 15.32767. This means that with 90 percent confidence, when the regression model goes through the origin, β1 falls between 14.56678 and 15.32767.

c) When a new call is placed in which 6 copiers are to be serviced, the predicted time is 89.683 minutes. With a 90 percent prediction interval, the interval, in minutes, to service 6 copiers is 74.696 to 104.671.

17) a) Yes, a fitted regression function that goes through the origin does seem to be a good fit here. It also makes sense that should take 0 minutes to service 0 copiers.

b) The residuals do not sum to 0. They sum to -5.862797. See the knitted section for the plot of the residuals against the fitted values of Ŷ*i*. The plot shows fairly constant variance in the residuals meaning the fitted regression line seems to be a good fit.

c) The alternatives for the Lack of Fit test are H0: E{Y } = β1X and Ha: E{Y } ≠ β1X. The sum of squares = 622.12 and the SSPE = 2797.66 both from the ANOVA table. F∗ = (622.12/9) / (2797.66/35) = 0.86478 which is equal to the F\* value from the ANOVA table in the knitted section. qf(.99; 9, 35) = 2.96301. The decision rule states that ff F∗ ≤ 2.96301 conclude H0, otherwise Ha. Since 0.86478 ≤ 2.96301, I conclude H0. The P-value of this test is 0.564.

19)

a) The 90 percent confidence interval for the student’s ACT test score is between 12.0481 and 54.1917, with a fitted value of 33.1199. This means that a new student that earned a GPA of 3.4 freshman year got an ACT score between 12 and 54, with 90 percent confidence. Since the ACT score can actually only be between 1 and 36 with extreme scores quite rare, the 12 to 54 interval is not particularly useful in this example.

b) The criterion for 4.33 is not met here because the quantity that is supposed to be less than 0.1 (from page 170 in the book where 4.33 is explained) is 0.1797486 which is > 0.1.