Math 327 Chapter 4 Homework

Ahmad M. Osman

10/01/2017

Chapter 4 – Problems 4.3, 4.16, 4.17, and 4.19

**Includes answers to:**

* 4.3: a, b, and c
* 4.16: a, b, and c
* 4.17: a, b, and c
* 4.19: a and b

I preferred writing my interpretations in Microsoft Word. To keep the formatting intact I did not write my comments in the knitted part but rather wrote it at the end of the word document file. Please look at page seven forward for my written answers and interpretations.

### \_\_ Ahmad M. Osman \_\_

# 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

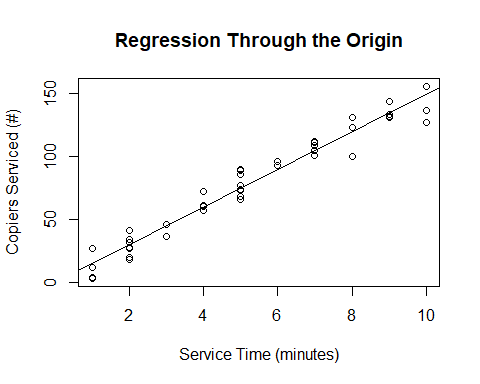
confint(int0fit, level=0.90)

## 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

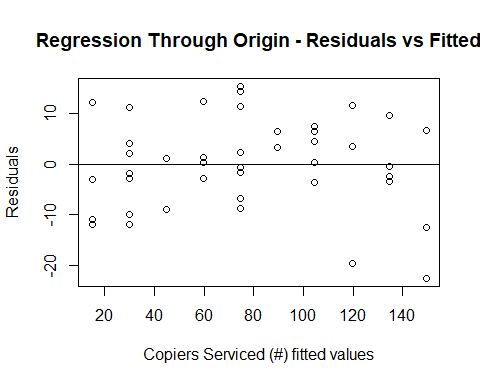
# Plot the data  
plot.new()  
plot(X, Y, xlab=xname, ylab=yname, main="Regression Through the Origin")  
abline(myfit)



# 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

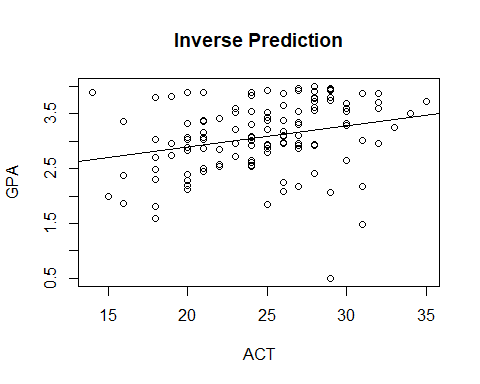
qf(.99, 9, 35)

## [1] 2.963012

# 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

**Answers Explained:**

**4.3**

**a)** b1 and b0 err are in opposite directions. The reason is that if b1 is overestimated, then it will decrease the y-intercept which decreases b0. Same applies for b1: if b1 is underestimated, the line will be tilt and increases b0.

**b)** As the questions asks for the Bonferroni joint confidence intervals for β0 and β1 using a 95% family confidence interval, the level of significance for each β0 and β1 should be divided by 2: 1 - .95 = .05 🡪 .05/2 = 0.025 = α for a family confidence interval. Afterward we use the qt function in R and obtain B as following: qt(1-0.025/2, 45-2) = qt(.9875, 43) = 2.322618. Using the lm function in R we get a b0 = -0.5802 and b1 = 15.0352. From here, we use the summary table and we find the standard errors (2.8039 and 0.4831 respectively) which we will use 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 in confint function is = .975 not .95 because it is a family confidence interval (1 – α/2), but when we use the qt function, the (1 – α/2) still needs to be divided by 2.

**c)** Whether or not the Bonferroni joint confidence intervals for β0 and β1 support the suggestion that β0 is equal to 0 and β1 is equal to 14? Since 0 and 14 are both in the respective joint confidence intervals from part (a) with a 95 percent family confidence interval, therefore the consultant’s suggestion is valid.

**Answers Explained:**

**4.16**

**a)** Assuming the appropriateness of linear regression through the origin, the estimated regression function is as following: Ŷ = 14.95X.

**b)** The 90% confidence interval for β1 is: 14.56678 ≤ β1 ≥ 15.32767. Meaning that with 90 percentage confidence, when the regression model goes through the origin, β1 is between 14.56678 and 15.32767.

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

**Answers Explained:**

**4.17**

**a)** The answer is yes. A fitted regression function in which goes through the origin indeed seems to be a good fit. Commonsense-wise, it should take 0 minutes to service 0 copiers.

**b)** The residuals do not sum to 0. Please see above for the knitted section of the plot of the residuals against the fitted values of Ŷ*i*. From the plot, a fairly constant variance in the residuals is shown. This means that the fitted regression line seems to be a good fit.

**c)** For the Lack of Fit test, the alternatives here are as following:

H0: E{Y} = β1X and Ha: E{Y} ≠ β1X.

The sum of squares is equal to 622.12 and the SSPE is equal to 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 above. Using the qf function: qf(.99; 9, 35) is equal to 2.96301. The decision rule here states that if F∗ ≤ 2.96301 conclude H0, otherwise conclude Ha. Since 0.86478 ≤ 2.96301, we conclude H0. The P-value for this test is 0.564.

**Answers Explained:**

**4.19**

**a)** The 90% confidence interval for the student’s ACT test score is between 12.0481 and 54.1917, with the fitted value equal to 33.1199. Which means that, with 90% confidence, a new student that earned a 3.4 GPA in his/her freshman year got a score between 12 and 54 on their ACT. As the ACT score can only be between 1 and 36 and that the extreme scores are rare, the 12 to 54 interval is not really useful in our case.

**b)** Since the quantity that is supposed to be less than 0.1 (explained on page 170 of the book as 4.33 is being demonstrated) is 0.1797486 which is > 0.1, the criterion for 4.33 is not met.