2. Do problem 3.13 on page 150.

Problem 3.13: Refer to Copier maintenance problem 1.20

A. What are the alternative conclusions when testing for lack of fit of a linear regression function. Note that the dependent variable (Y) is Service Time and the independent variable (X) is the number of copiers serviced.

The F statistic can be used to assess the lack of fit for a linear model which provides a formal way to test the following:

$H_o : E\{Y\} = \beta_0 + \beta_1 X$ * Concludes that the regression function is linear (a good fit)

$H_a : E\{Y\} \neq \beta_0 + \beta_1 X$ * Concludes that there is a lack of linear fit (lack of fit)

B. Perform the test indicated in part (a). Control the risk of Type 1 error at 0.05. State the decision rule and conclusion.

```r
df <- read.csv("data/CH. 1, PR 20.csv")
names(df) = c("ServiceTime", "NumCopiers")
result <- lm(ServiceTime ~ factor(NumCopiers), data=df)
F_crit <- qf(0.95, df1=1, df2=43)

result1 <- lm(ServiceTime ~ NumCopiers, data=df)
result2 <- lm(ServiceTime ~ factor(NumCopiers), data=df)

n <- dim(df)[1]
distinct_vals <- length(unique(df$NumCopiers))

# lack of fit degrees of freedom: number of distinct values - 2
lof_degf <- distinct_vals - 2
tot_degf <- n-1

# pure error degrees of freedom: n - lack of fit degrees of freedom
pe_degf <- n - distinct_vals

# error degrees of freedom
err_degf <- n - 2

SSR <- anova(result1)$"Sum Sq"[1]
SSE <- anova(result1)$"Sum Sq"[2]
SSPE <- anova(result2)$"Sum Sq"[2]
SSLF <- SSE - SSPE
SST <- SSR + SSE
MSR <- anova(result1)$"Mean Sq"[1]
MSE <- anova(result1)$"Mean Sq"[2]
MSPE <- anova(result2)$"Mean Sq"[2]
MSLF <- SSLF / (distinct_vals - 2)
F_mod <- anova(result1)$"F value"[1]
F_lof <- MSLF / MSPE

Source = c("Regression",
           "Residual Error",
           "Lack of Fit Error",
           "Pure Error",
```

```
            "Total")
DF <- c(1,err_degf,
        lof_degf,
        pe_degf,
        tot_degf)
SS <- c(SSR,
        SSE,
        SSLF,
        SSPE,
        SST)
MS <- c(as.character(MSR),
        as.character(MSE),
        as.character(MSLF),
        as.character(MSPE),"")
F_value <- c(as.character(F_mod),
             "",
             as.character(F_lof),
             "",
             "" )
result_df <- data.frame(Source, DF, SS, MS, F_value)

library(knitr)
kable(result_df)
```

Source	DF	SS	MS	F_value
Regression	1	76960.4230	76960.4229765013	968.657195979066
Residual Error	43	3416.3770	79.4506284534581	
Lack of Fit Error	8	618.7187	77.3398362706696	0.967557130626541
Pure Error	35	2797.6583	79.9330952380954	
Total	44	80376.8000		

```
txt = paste("F*: ", F_lof, "\nF: ", F_crit, "\n", sep="")
cat(txt)
```

```
## F*: 0.967557130626541
## F: 4.06704742642635
```

Let $\alpha = 0.05$. Since n = 45, F(0.95; 1,43) = 4.0670474. The decision rule is as follows:

- If $F* \leq 4.0670474$, conclude H_o
- If $F* > 4.0670474$, conclude H_a

Conclusion: there is a linear association between number of copiers and service time. So, conclude H_o:.

C. Does the test in part (b) detect other departures from regression model (2.1), such as lack of constant variance or lack of normality in the error terms? Could the results of the test of lack of fit be affected by such departures?

No, there is not a lack of fit. As shown below the data is clearly linear.

```r
with(df,plot(NumCopiers, ServiceTime))
```