HW8

2023-10-28

8.19. Refer to Copier maintenance Problems 1.20 and 8.15.

1.20. Copier maintenance. The Tri-City Office Equipment Corporation sells an imported copier on a franchise basis and performs preventive maintenance and repair service on this copier. The data below have been collected from 45 recent calls on users to perform routine preventive maintenance service; for each call, X is the number of copiers serviced and Y is the total number of minutes spent by the service person. Assume that first-order regression model (1.1) is appropriate. The data is in the file CH01PR20.txt.

8.15. (Copier maintenance Problem 1.20) The users of the copiers are either training institutions that use a small model, or business firms that use a large, commercial model. An analyst at Tri-City wishes to fit a regression model including both number of copiers serviced (X1) and type of copier (X2) as predictor variables and estimate the effect of copier model (S-small, L-large) on number of minutes spent on the service call. The models serviced in the 45 calls were in the data file CH08PR15.txt. Assume that regression model ($Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$) is appropriate. Let X2 = 1 if small model, and 0 if large, commercial model.

a. Fit regression model (8.49) and state the estimated regression function.

```
> # ---- Centering, m3.c
> # Center the continuous variable to alleviate multicollinearity
> X1.c <- X1 - mean(X1, na.rm=T)
> # Fit the interaction model again with the centered variable
> m3.c <- lm(Y ~ X1.c + X2.factor + X2.factor:X1.c, data=mydata)
> vif(m3.c)
there are higher-order terms (interactions) in this model
consider setting type = 'predictor'; see ?vif
         X1.c      X2.factor X1.c:X2.factor
     1.671186       1.008089       1.670894
> summary(m3.c) # Y.hat = 76.1034 + 14.3394(X1.c) + 0.9432(X2.factorSmall) + 1.7774(X1.c)(X2.factorSmall)

Call:
lm(formula = Y ~ X1.c + X2.factor + X2.factor:X1.c, data = mydata)

Residuals:
     Min       1Q   Median       3Q      Max
-19.2072  -6.7887  -0.1708   7.1504  14.7441

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          76.1034     1.6611  45.815   <2e-16 ***
X1.c                 14.3394     0.6146  23.333   <2e-16 ***
X2.factorSmall        0.9432     2.7078   0.348   0.7294
X1.c:X2.factorSmall   1.7774     0.9746   1.824   0.0755 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.771 on 41 degrees of freedom
Multiple R-squared:  0.9608,    Adjusted R-squared:  0.9579
F-statistic: 334.6 on 3 and 41 DF,  p-value: < 2.2e-16
```

Code:

```
# Re-name columns
# Y = total number of minutes spent by the service person
# X1 = number of copiers served
# X2 = type of copier (Small or Large), 1=small, 0=large commercial
colnames(mydata) <- c("Y", "X1", "X2")
```

```r
# Center the continuous variable to alleviate multicollinearity
x1.c <- x1 - mean(x1, na.rm=T)
# Fit the interaction model again with the centered variable
m3.c <- lm(Y ~ x1.c + X2.factor + X2.factor:x1.c, data=mydata)
vif(m3.c)
summary(m3.c) # Y.hat = 76.1034 + 14.3394(X1) + 0.9432(X2.factorSmall) + 1.7774(X1)(X2.factorSmall)
```

Y.hat = 76.1034 + 14.3394(X1.c) + 0.9432(X2.factorSmall) + 1.7774(X1.c)(X2.factorSmall)

① Fitted multiple regression model:

$$\widehat{(\text{total serviced time})} = 76.1034 + 14.3394(\text{No. copiers served.C}) + 0.9432(\text{Small}) + 1.7774(\text{No. copier served.C})(\text{Small})$$

centered, where $X_2 = \begin{cases} 0 & \text{Large} \\ 1 & \text{Small} \end{cases}$

② Desired model for "Large, commercial" copier only:

$$\hat{Y} = 76.1034 + 14.3394(X1.C) + 0.9432(0) + 1.7774(X1.C)(0)$$
$$= 76.1034 + 14.3394(X1.C)$$

③ Desired model for "small" copier only:

$$\hat{Y} = 76.1034 + 14.3394(X1.C) + 0.9434(1) + 1.7774(X1.C)(1)$$
$$= 77.0468 + 16.1168(X1.C)$$

Both intercept and slops show some difference between large and small copiers.

Interpretation

- Intercept (76.1034): The base service time for a large, commercial model copier when no copiers are served, is 76.1034 minutes.
- Coefficient for X1.C: For each additional "Large" copier served, the expected service time increases by 14.3394 min, holding the copier type constant.
- Coefficient for X2.factor small: Servicing a small model copier increases the expected service time by 0.9432 mins as compare to a large model, holding the No. of copier served constant.
- Coefficient for the interaction term X1C * X2.Factorsmall: For each additional small copier served, the time increses by additional 1.7774 mins, compared to serving an additional large copier.
  (⇒ Effect of serving an additional copier on the service time is 1.7774 mins greater for small model compare to the large model.
- For the "small" copier, the time increses by 16.1168 mins for each additional copier serviced.

b. Test whether the interaction term can be dropped from the model; control the α risk at 0.10. State the alternatives, decision rule, and conclusion. What is the P-value of the test? If the interaction term cannot be dropped from the model, describe the nature of the interaction effect. For the hypothesis testing, clearly state the five steps.

Step 1: Diagnostic (see R codes below)

(1) Linear relationship between Y & X_1. X_2 is categorical … no need to consider ($Y \sim X_1$)
(2) Outliers: No extreme outlier exist. (rstudent ~ $\hat{Y}$)
(3) Check for heteroskedasticity: Vertical spread is constant throughout $\hat{Y}$ (residual ~ $\hat{Y}$)
(4) Normal residual: (qqplot) : reasonably normal
(5) Independent observation: Yes.

Step 2: Hypothesis.

$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X1.X_2$

Additive model (M2.C) : $(\widehat{\text{Tot.serv.tm}}) = 76.1034 + 14.3394(X1.c) + 0.9432(X2.\text{factorsmall})$

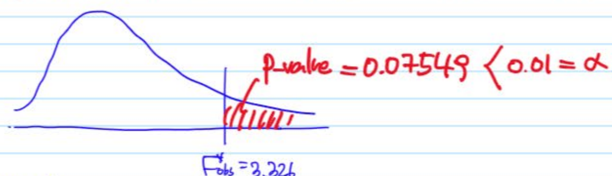
Interactive model (M3.C) : $(\widehat{\text{tot.serv.time}}) =$ " + " + " $+ 1.7774(X1.c)(X2.\text{factorsmall})$

H_0: Additive model. ($\beta_3 = 0$, Interaction term can be dropped)

H_1: Interaction model ($\beta_3 \neq 0$, " should be retained. The interaction term does have effect on Y.)

Step 3. Test statistic.

```
> anova(m3.c, m2.c)
Analysis of Variance Table

Model 1: Y ~ X1.c + X2.factor + X2.factor:X1.c
Model 2: Y ~ X1.c + X2.factor
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1     41 3154.4
2     42 3410.3 -1   -255.89 3.326 0.07549 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



Step 4. P-value
= 0.07549

Step 5. conclusion.

With $P = 0.07549$ which is $< 0.1 = \alpha$,
we reject H_0 and take H_1.
Thus there is enough evidence supporting "the interaction term does have effect on Y
with 1.7774 mins greater effect for small model compared to large model.

Nature of interaction effect :

The effect of X1.C (Number of copier serviced) on Y (total service time by a service person) differ depending on X2.factor (the copier types; small vs large model).
Specifically, for each additional unit of X1.C, the difference in Y between X2.factor.small (small model) increases by 1.7774 mins, with the significance level $\alpha < 0.1$.

Practical significance :

```
> summary(m2.c)$r.squared
[1] 0.9575707
> summary(m3.c)$r.squared
[1] 0.9607544
> # install.packages("rsq")
> library(rsq)
> rsq.partial(objF=m3.c, objR=m2.c)
$adjustment
[1] FALSE

$variables.full
[1] "X1.c"          "X2.factor"      "X1.c:X2.factor"

$variables.reduced
[1] "X1.c"      "X2.factor"

$partial.rsq
[1] 0.07503474
```

About 95.8% of observation can be explained by the reduced model (= additive ").

About 96% of observation observation can be explained by the Full model (= Interactive model.

The partial-R^2 = 0.075, pretty small.
Also, looking at the plot, noise is quite small.

In this case, we would not need interaction term for this data. We can take the smaller reduced model (additive model, M2.C), instead M3.C, because it is a more parsimonious model.

R code

Import and inspect data

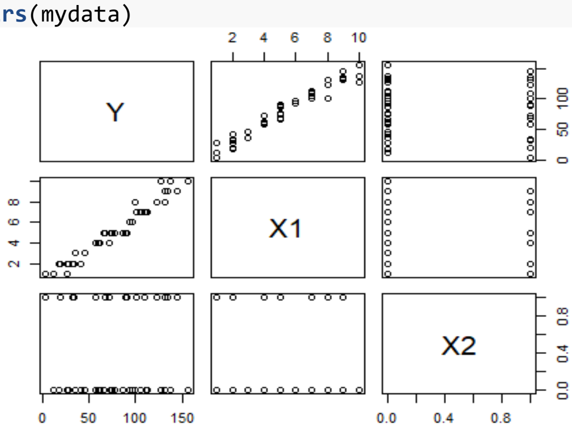
```
# ---------------------------------------------------------------------------

# Clean up the workspace for the new analysis
rm(list=ls())

# Load the libraries needed for this analysis
library(Hmisc)

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##     format.pval, units

library(Rmisc)

## Loading required package: lattice

## Loading required package: plyr

##
## Attaching package: 'plyr'

## The following objects are masked from 'package:Hmisc':
##
##     is.discrete, summarize

library(ggplot2)

# Set the following to your own folder
setwd("C:/Users/jyang/OneDrive - Arizona State University/10
Classes_OneDrive/2023_STP530_Regression")

# Import the data
mydata.1 <- read.table("CH01PR20.txt")
new.column <- read.table("CH08PR15.txt")
mydata <- cbind(mydata.1, new.column)
```

) combined

```
# Inspect the data
head(mydata)

##    V1 V2 V1
## 1  20  2  1
## 2  60  4  0
## 3  46  3  0
## 4  41  2  0
## 5  12  1  0
## 6 137 10  0

str(mydata)

## 'data.frame':    45 obs. of  3 variables:
##  $ V1: int  20 60 46 41 12 137 68 89 4 32 ...
##  $ V2: int  2 4 3 2 1 10 5 5 1 2 ...
##  $ V1: int  1 0 0 0 0 0 1 1 1 1 ...

Hmisc::describe(mydata)
```

```
## mydata
##
##  3  Variables      45  Observations
## ---------------------------------------------------------------------------
## V1
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##       45        0       40        1    76.27    49.78     13.2     20.0
##      .25      .50      .75      .90      .95
##     36.0     74.0    111.0    131.6    136.4
##
## lowest :   3   4  12  18  20, highest: 132 134 137 144 156
## ---------------------------------------------------------------------------
## V2
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##       45        0       10    0.983    5.111    3.212      1.0      2.0
##      .25      .50      .75      .90      .95
##      2.0      5.0      7.0      9.0      9.8
##
## Value          1     2     3     4     5     6     7     8     9    10
## Frequency      4     8     2     5     8     2     6     3     4     3
## Proportion 0.089 0.178 0.044 0.111 0.178 0.044 0.133 0.067 0.089 0.067
##
## For the frequency table, variable is rounded to the nearest 0
## ---------------------------------------------------------------------------
## V1
##        n  missing distinct     Info      Sum     Mean      Gmd
##       45        0        2    0.706       17   0.3778   0.4808
##
## ---------------------------------------------------------------------------
```

```r
# Re-name columns
# Y = total number of minutes spent by the service person
# X1 = number of copiers served
# X2 = type of copier (Small or Large), 1=small, 0=large commercial
colnames(mydata) <- c("Y", "X1", "X2")
pairs(mydata)
```



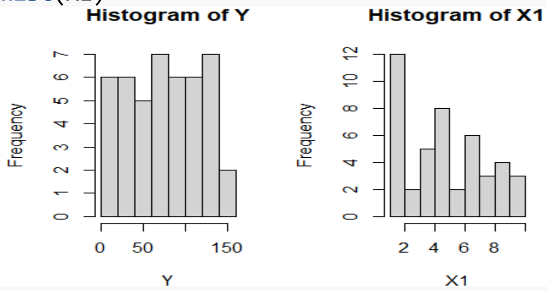

```r
# Dependent variable Y: total number of minutes spent by the service person
attach(mydata)
summary(Y)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    3.00   36.00   74.00   76.27  111.00  156.00
```

```r
sd(Y, na.rm=T)
```

```
## [1] 42.74044
```

```r
par(mfrow=c(1, 2))
hist(Y)

# X1: number of copiers served
summary(X1)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   2.000   5.000   5.111   7.000  10.000

hist(X1)
```




```r
# X2: type of copier (Small or Large), 1=small, 0=large commercial
table(X2, useNA="ifany")

## X2
##  0  1
## 28 17

X2.factor<- factor(X2, levels=c(0, 1), labels=c("Large", "Small"))
table(X2.factor)

## X2.factor
## Large Small
##    28    17
```

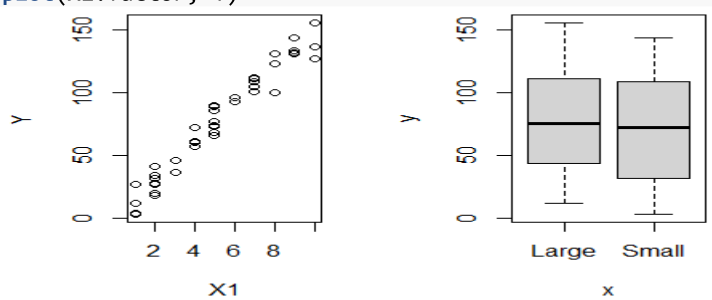
fit models: m1, m2, m3, m4 and diagnostics

```r
# ---------------------------------------------------------------------
# Regression
# Research Q1: Does the Y(service time) differ between X2.factor (machine type)
m1 <- lm(Y ~ X2.factor, data=mydata)
summary(m1) # not significantlly different

##
## Call:
## lm(formula = Y ~ X2.factor, data = mydata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -69.412 -38.412  -1.607  33.393  77.393
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      78.607      8.150   9.645 2.55e-12 ***
## X2.factorSmall   -6.195     13.260  -0.467    0.643
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 43.13 on 43 degrees of freedom
## Multiple R-squared:  0.005051,	Adjusted R-squared:  -0.01809
## F-statistic: 0.2183 on 1 and 43 DF,  p-value: 0.6427
```

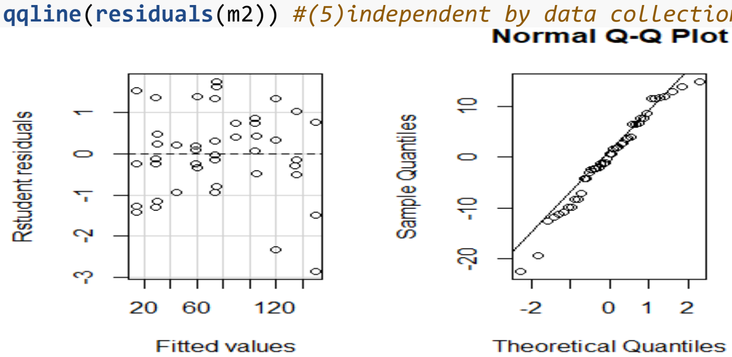
```
# ----Additive, categorical
# Research Q2: Does the Y(service time) still differ between copier types (S/L),
#  after controlling for the X1 (# of serviced copiers)
m2 <- lm(Y ~ X1 + X2.factor, data=mydata)
library(car)

## Loading required package: carData

vif(m2)

##        X1 X2.factor
##  1.006681  1.006681

summary(m2)

##
## Call:
## lm(formula = Y ~ X1 + X2.factor, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.5390  -4.2515   0.5995   6.5995  14.9330
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -0.9225     3.0997  -0.298    0.767
## X1              15.0461     0.4900  30.706   <2e-16 ***
## X2.factorSmall   0.7587     2.7799   0.273    0.786
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.011 on 42 degrees of freedom
## Multiple R-squared:  0.9576, Adjusted R-squared:  0.9556
## F-statistic: 473.9 on 2 and 42 DF,  p-value: < 2.2e-16

model.matrix(m2) #Show the dummy code system
##    (Intercept) X1 X2.factorSmall
## 1            1  2              1
## 2            1  4              0
## 3            1  3              0
## 4            1  2              0
## 5            1  1              0
## 6            1 10              0
## 7            1  5              1
## 8            1  5              1
## 9            1  1              1
## 10           1  2              1
## 11           1  9              1
## 12           1 10              0
## 13           1  6              0
## 14           1  3              0
## 15           1  4              1
## 16           1  8              0
## 17           1  7              0
## 18           1  8              0
## 19           1 10              0
## 20           1  4              0
## 21           1  5              0
## 22           1  7              1
## 23           1  7              1
## 24           1  5              0
```

```
## 25            1  9               1
## 26            1  7               0
## 27            1  2               0
## 28            1  5               0
## 29            1  7               0
## 30            1  6               0
## 31            1  8               1
## 32            1  5               1
## 33            1  2               1
## 34            1  2               0
## 35            1  1               1
## 36            1  4               1
## 37            1  5               0
## 38            1  9               0
## 39            1  7               0
## 40            1  1               0
## 41            1  9               1
## 42            1  2               1
## 43            1  2               0
## 44            1  4               0
## 45            1  5               0
## attr(,"assign")
## [1] 0 1 2
## attr(,"contrasts")
## attr(,"contrasts")$X2.factor
## [1] "contr.treatment"
```

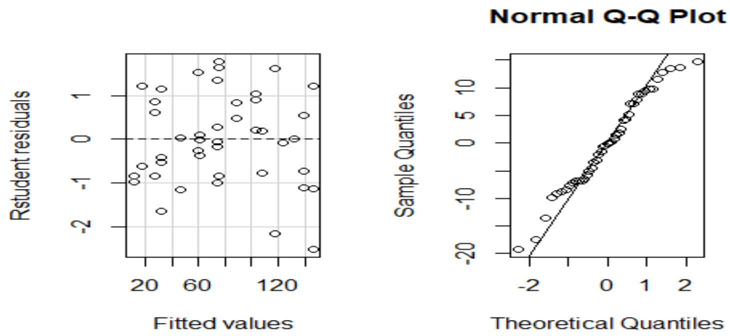
```
# Diagnostics of m2
par(mfrow=c(1, 2))
plot(X1, Y) #(1)linear relationship
plot(X2.factor, Y)
```




```
par(mfrow=c(1, 2))
library(car)
residualPlot(m2, type="rstudent", quadratic=F) #(2)Outliers, (3)Homoskedasticity
qqnorm(residuals(m2)) #(4)normal
qqline(residuals(m2)) #(5)independent by data collection,yes
```

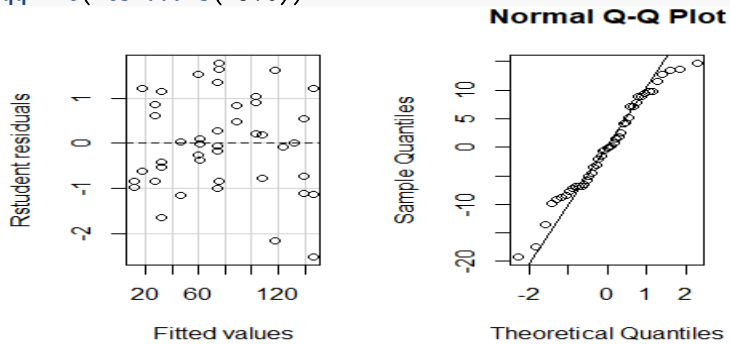


```r
# ----Interaction term, m3
# Research Q3: Does the relationship btw X1 and Y differ for the copier types?
# Is the effect of X1 on Y is differ depending on the X2.factor (copier types)?

m3 <- lm(Y ~ X1 + X2.factor + X2.factor:X1, data=mydata)
vif(m3) # multicollinearlity --> need to fix

## there are higher-order terms (interactions) in this model
## consider setting type = 'predictor'; see ?vif

##           X1    X2.factor X1:X2.factor
##     1.671186     4.280903     4.698387

summary(m3) # Y.hat = 2.8131 + 14.3394(X1) -8.1412(X2.factorSmall) +
1.7774(X1)(X2.factorSmall)

##
## Call:
## lm(formula = Y ~ X1 + X2.factor + X2.factor:X1, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.2072  -6.7887  -0.1708   7.1504  14.7441
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)           2.8131     3.6468   0.771   0.4449
## X1                   14.3394     0.6146  23.333   <2e-16 ***
## X2.factorSmall       -8.1412     5.5801  -1.459   0.1522
## X1:X2.factorSmall     1.7774     0.9746   1.824   0.0755 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.771 on 41 degrees of freedom
## Multiple R-squared:  0.9608, Adjusted R-squared:  0.9579
## F-statistic: 334.6 on 3 and 41 DF,  p-value: < 2.2e-16

anova(m3)

## Analysis of Variance Table
##
## Response: Y
##              Df Sum Sq Mean Sq   F value  Pr(>F)
## X1            1  76960   76960 1000.2987 < 2e-16 ***
## X2.factor     1      6       6    0.0786 0.78059
## X1:X2.factor  1    256     256    3.3260 0.07549 .
## Residuals    41   3154      77
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Diagnostics of m3
residualPlot(m3, type="rstudent", quadratic=F) #(2)Outliers, (3)Homoskedasticity
qqnorm(residuals(m3)) #(4)normal
qqline(residuals(m3))
```

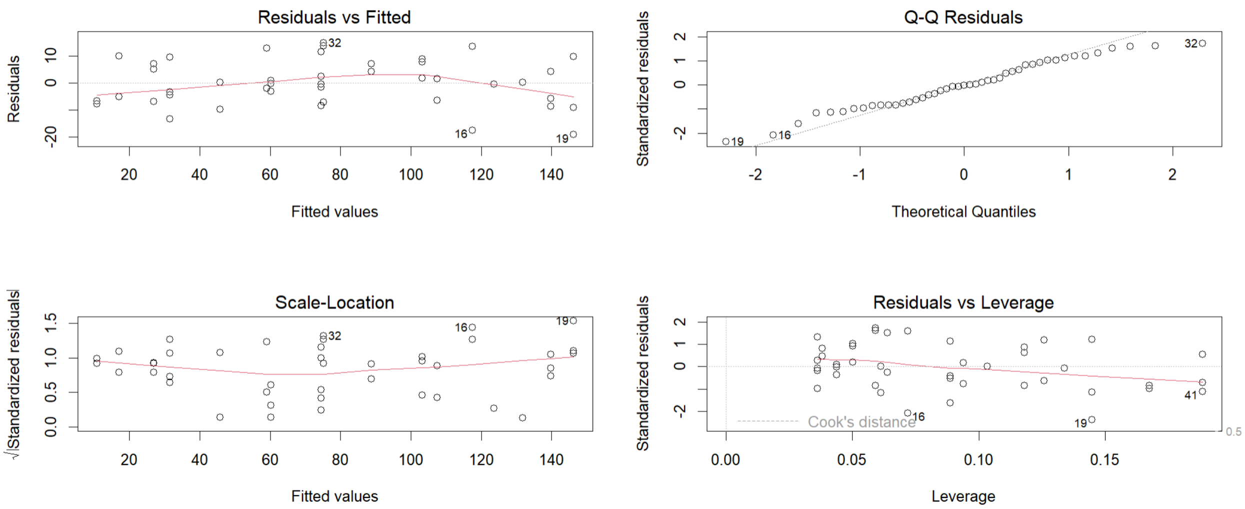



```r
# ---- Centering, m3.c
# Center the continuous variable to alleviate multicollinearity
X1.c <- X1 - mean(X1, na.rm=T)
# Fit the interaction model again with the centered variable
m3.c <- lm(Y ~ X1.c + X2.factor + X2.factor:X1.c, data=mydata)
vif(m3.c)

## there are higher-order terms (interactions) in this model
## consider setting type = 'predictor'; see ?vif

##           X1.c      X2.factor X1.c:X2.factor
##       1.671186       1.008089       1.670894

summary(m3.c) # Y.hat = 76.1034 + 14.3394(X1) + 0.9432(X2.factorSmall) +
1.7774(X1)(X2.factorSmall)

##
## Call:
## lm(formula = Y ~ X1.c + X2.factor + X2.factor:X1.c, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.2072  -6.7887  -0.1708   7.1504  14.7441
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)            76.1034     1.6611  45.815   <2e-16 ***
## X1.c                   14.3394     0.6146  23.333   <2e-16 ***
## X2.factorSmall          0.9432     2.7078   0.348   0.7294
## X1.c:X2.factorSmall     1.7774     0.9746   1.824   0.0755 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.771 on 41 degrees of freedom
## Multiple R-squared:  0.9608, Adjusted R-squared:  0.9579
## F-statistic: 334.6 on 3 and 41 DF,  p-value: < 2.2e-16

anova(m3.c)

## Analysis of Variance Table
##
## Response: Y
##                Df Sum Sq Mean Sq   F value   Pr(>F)
## X1.c            1  76960   76960 1000.2987 < 2e-16 ***
## X2.factor       1      6       6    0.0786 0.78059
## X1.c:X2.factor  1    256     256    3.3260 0.07549 .
## Residuals      41   3154      77
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Diagnostics of m4
residualPlot(m3.c, type="rstudent", quadratic=F) #(2)Outliers, (3)Homoskedasticity
qqnorm(residuals(m3.c)) #(4)normal
qqline(residuals(m3.c))
```




```r
# Residual plots for diagnostic purposes before statistical tests
par(mfrow=c(2, 2))
plot(m3.c)
```



Compare models: if the interaction term can be dropped.

```r
# Compare models using the general linear test approach
# Fit the model again
m2.c <- lm(Y ~ X1.c + X2.factor, data=mydata)
m3.c <- lm(Y ~ X1.c + X2.factor + X2.factor:X1.c, data=mydata)
summary(m2.c)

##
## Call:
## lm(formula = Y ~ X1.c + X2.factor, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.5390  -4.2515   0.5995   6.5995  14.9330
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       75.9800     1.7051  44.561   <2e-16 ***
## X1.c              15.0461     0.4900  30.706   <2e-16 ***
## X2.factorSmall     0.7587     2.7799   0.273    0.786
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.011 on 42 degrees of freedom
## Multiple R-squared:  0.9576, Adjusted R-squared:  0.9556
## F-statistic: 473.9 on 2 and 42 DF,  p-value: < 2.2e-16

summary(m3.c)

##
## Call:
## lm(formula = Y ~ X1.c + X2.factor + X2.factor:X1.c, data = mydata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -19.2072  -6.7887  -0.1708   7.1504  14.7441
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)             76.1034     1.6611  45.815   <2e-16 ***
## X1.c                    14.3394     0.6146  23.333   <2e-16 ***
## X2.factorSmall           0.9432     2.7078   0.348   0.7294
## X1.c:X2.factorSmall      1.7774     0.9746   1.824   0.0755 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.771 on 41 degrees of freedom
## Multiple R-squared:  0.9608, Adjusted R-squared:  0.9579
## F-statistic: 334.6 on 3 and 41 DF,  p-value: < 2.2e-16

# F-test of general linear testing approach
# anova(m4.lw, m2.lw) #marginally significant, sample size 45
anova(m3.c, m2.c)

## Analysis of Variance Table
##
## Model 1: Y ~ X1.c + X2.factor + X2.factor:X1.c
## Model 2: Y ~ X1.c + X2.factor
##   Res.Df    RSS Df Sum of Sq     F  Pr(>F)
## 1     41 3154.4
## 2     42 3410.3 -1   -255.89 3.326 0.07549 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# R^2, to look at practical significance by reducing the model
# summary(m2.lw)$r.squared
# summary(m4.lw)$r.squared
summary(m2.c)$r.squared

## [1] 0.9575707

summary(m3.c)$r.squared

## [1] 0.9607544

library(rsq)
rsq.partial(objF=m3.c, objR=m2.c)

## $adjustment
## [1] FALSE
##
## $variables.full
## [1] "X1.c"           "X2.factor"      "X1.c:X2.factor"
```

```
##
## $variables.reduced
## [1] "X1.c"      "X2.factor"
##
## $partial.rsq
## [1] 0.07503474
```

Graph

```
#-------------------------------------------------------------------------------
# Graph. Plot using the original data with occasional missing values.
par(mfrow=c(1, 2))
my.pred <- predict(m2.c, newdata=data.frame(X1, X2.factor))
plot(X1[X2.factor == "Large"], Y[X2.factor == "Large"],
     col="blue", xlab="The number of copiers served",
     ylab="Total time spent by the service person", pch=3)
points(X1[X2.factor == "Small"], Y[X2.factor == "Small"],
     col="orange", pch=5)
lines(X1[X2.factor == "Large"], my.pred[X2.factor == "Large"],
      col="blue", pch=3, lwd=3)
lines(X1[X2.factor == "Small"], my.pred[X2.factor == "Small"],
      col="orange", pch=3, lwd=3)
legend("bottomright", legend=c("m2.c.Large", "m2.c.Small"),
       col=c("blue", "orange"), pch=c(3,5))

my.pred <- predict(m3.c, newdata=data.frame(X1, X2.factor))
plot(X1[X2.factor == "Large"], Y[X2.factor == "Large"],
     col="forestgreen", xlab="The number of copiers served",
     ylab="Total time spent by the service person", pch=3)
points(X1[X2.factor == "Small"], Y[X2.factor == "Small"],
       col="magenta", pch=5)
lines(X1[X2.factor == "Large"], my.pred[X2.factor == "Large"],
      col="forestgreen", pch=3, lwd=3)
lines(X1[X2.factor == "Small"], my.pred[X2.factor == "Small"],
      col="magenta", pch=3, lwd=3)
legend("bottomright", legend=c("m4.c.Large", "m4.c.Small"),
       col=c("forestgreen", "magenta"), pch=c(3,5))
```