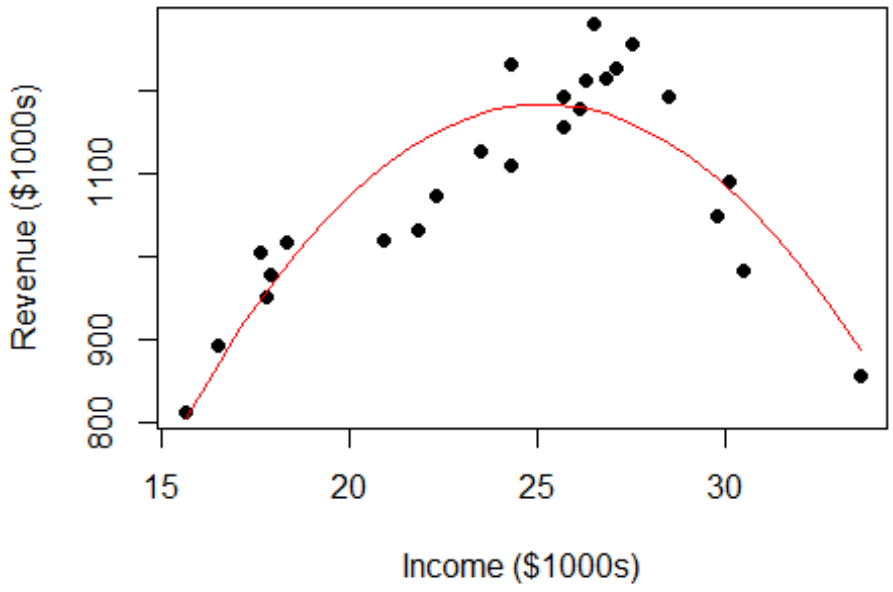
Regression_Extensions

```
library(readxl)

# Exercise 8.1 Fast Food
food <- read_excel("FastFood.xlsx", col_names = TRUE)
attach(food)
  # Model fit - 2 forms leading to same result
linefitQ <- lm(Revenue ~ Income + IncomeSqd)
summary(linefitQ)
## Call:
## lm(formula = Revenue ~ Income + IncomeSqd)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -105.452  -44.967    8.613   41.906  104.164
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1454.521    279.993  -5.195 3.29e-05 ***
## Income        209.815     24.084   8.712 1.39e-08 ***
## IncomeSqd      -4.170      0.504  -8.275 3.36e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 60.31 on 22 degrees of freedom
## Multiple R-squared:  0.8029, Adjusted R-squared:  0.785
## F-statistic: 44.82 on 2 and 22 DF,  p-value: 1.74e-08

linefitQa <- lm(Revenue ~ Income + I(Income^2))
summary(linefitQa)
## Call:
## lm(formula = Revenue ~ Income + I(Income^2))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -105.452  -44.967    8.613   41.906  104.164
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1454.521    279.993  -5.195 3.29e-05 ***
## Income        209.815     24.084   8.712 1.39e-08 ***
## I(Income^2)    -4.170      0.504  -8.275 3.36e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 60.31 on 22 degrees of freedom
## Multiple R-squared:  0.8029, Adjusted R-squared:  0.785
## F-statistic: 44.82 on 2 and 22 DF,  p-value: 1.74e-08
```

```r
  # scatter plot using alternate form Qa
newdata <- data.frame(Income = seq(min(Income), max(Income), length.out = 25)
)
newdata$pred <- predict(linefitQa, newdata, interval = "none")
plot(Income, Revenue, pch = 16, xlab = "Income ($1000s)", ylab = "Revenue ($1
000s)")
with(newdata, lines(x = Income, y = pred, col="red"))
```




```r
detach(food)
```

```r
# Exercise 8.2 Polishing
polish <- read_excel("Polishing.xlsx", col_names = TRUE)
  # convert Type to a factor variable TypeF, i.e., nominal
polish$typeF<-factor(polish$type)
    # attach names to the levels of the factor variable
levels(polish$typeF) <- c("bowl", "cass", "dish", "tray", "plate")
attach(polish)

  # fit the first-order linear model
fitPf <- lm(time ~ diam + typeF)
summary(fitPf)
```

```
##
## Call:
## lm(formula = time ~ diam + typeF)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -24.084  -6.949  -1.967   6.149  31.635
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.1880     5.0170  -0.237 0.813728
## diam          2.9992     0.4636   6.469 3.22e-08 ***
## typeFcass    16.9825     4.6756   3.632 0.000634 ***
## typeFdish     9.5477     5.0851   1.878 0.065949 .
## typeFtray     7.7619     4.9898   1.556 0.125763
## typeFplate   -7.1759     4.6722  -1.536 0.130520
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.77 on 53 degrees of freedom
## Multiple R-squared:  0.6502, Adjusted R-squared:  0.6172
## F-statistic: 19.71 on 5 and 53 DF,  p-value: 4.828e-11
```

```r
  # reorder the data frame with typeF = 5 (type = plate) as reference
detach(polish)
polish <- within(polish, typeF <- relevel(typeF, ref = "plate"))
attach(polish)
  # fit the first-order linear model in two ways
fitPf <- lm(time ~ diam + typeF)
summary(fitPf)
## Call:
## lm(formula = time ~ diam + typeF)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -24.084  -6.949  -1.967   6.149  31.635
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -8.3639     6.3805  -1.311  0.19556
## diam          2.9992     0.4636   6.469 3.22e-08 ***
## typeFbowl     7.1759     4.6722   1.536  0.13052
## typeFcass    24.1585     5.4586   4.426 4.82e-05 ***
## typeFdish    16.7236     5.9991   2.788  0.00735 **
## typeFtray    14.9378     5.6341   2.651  0.01055 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.77 on 53 degrees of freedom
## Multiple R-squared:  0.6502, Adjusted R-squared:  0.6172
## F-statistic: 19.71 on 5 and 53 DF,  p-value: 4.828e-11

fitP <- lm(time ~ diam + bowl + cass + dish + tray)
summary(fitP)
## Call:
## lm(formula = time ~ diam + bowl + cass + dish + tray)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -24.084  -6.949  -1.967   6.149  31.635
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -8.3639     6.3805  -1.311  0.19556
## diam          2.9992     0.4636   6.469 3.22e-08 ***
## bowl          7.1759     4.6722   1.536  0.13052
## cass         24.1585     5.4586   4.426 4.82e-05 ***
## dish         16.7236     5.9991   2.788  0.00735 **
## tray         14.9378     5.6341   2.651  0.01055 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.77 on 53 degrees of freedom
## Multiple R-squared:  0.6502, Adjusted R-squared:  0.6172
## F-statistic: 19.71 on 5 and 53 DF,  p-value: 4.828e-11
```

```r
  # fit the model with diam as sole predictor
fitSLR <- lm(time ~ diam)
summary(fitSLR)
```

```
##
## Call:
## lm(formula = time ~ diam)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -28.037  -8.287  -2.705   8.315  43.438
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.9547     5.4020  -0.362    0.719
## diam          3.4567     0.4667   7.407 6.67e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.69 on 57 degrees of freedom
## Multiple R-squared:  0.4905, Adjusted R-squared:  0.4815
## F-statistic: 54.86 on 1 and 57 DF,  p-value: 6.67e-10
```

```r
  # fit the model with type as predictor in two ways
    # using lm
fitT <- lm (time ~ typeF)
summary(fitT)
```

```
##
## Call:
## lm(formula = time ~ typeF)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -28.573  -9.367  -3.344   7.712  59.977
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   24.194      5.198   4.655 2.15e-05 ***
## typeFbowl      2.928      6.131   0.478 0.634894
## typeFcass     29.061      7.164   4.056 0.000162 ***
## typeFdish     10.816      7.858   1.376 0.174382
## typeFtray     25.209      7.164   3.519 0.000889 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.59 on 54 degrees of freedom
## Multiple R-squared:  0.3741, Adjusted R-squared:  0.3277
## F-statistic: 8.067 on 4 and 54 DF,  p-value: 3.562e-05
```

```r
anova(fitT)
```

```
## Analysis of Variance Table
##
## Response: time
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## typeF      4  7845.7 1961.44  8.0675 3.562e-05 ***
## Residuals 54 13128.9  243.13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
    # using anova
aovT <- aov(time ~ typeF)
summary(aovT)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## typeF        4   7846  1961.4   8.067 3.56e-05 ***
## Residuals   54  13129   243.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
TukeyHSD(aovT, conf.level = .90)
```

```
##   Tukey multiple comparisons of means
##     90% family-wise confidence level
##
## Fit: aov(formula = time ~ typeF)
##
## $typeF
##                   diff        lwr       upr     p adj
## bowl-plate    2.927729 -12.554072 18.409531 0.9890860
## cass-plate   29.060556  10.968529 47.152582 0.0014642
## dish-plate   10.815556  -9.028087 30.659198 0.6451429
## tray-plate   25.208556   7.116529 43.300582 0.0076063
## cass-bowl    26.132826  11.217775 41.047877 0.0004397
## dish-bowl     7.887826  -9.109452 24.885104 0.7671569
## tray-bowl    22.280826   7.365775 37.195877 0.0035580
## dish-cass   -18.245000 -37.649708  1.159708 0.1381402
## tray-cass    -3.852000 -21.461486 13.757486 0.9811826
## tray-dish    14.393000  -5.011708 33.797708 0.3439841
```

```r
      # R^2
(anova(aovT)[["Sum Sq"]][1])/((anova(aovT)[["Sum Sq"]][1])+(anova(aovT)[["Sum
Sq"]][2]))
```

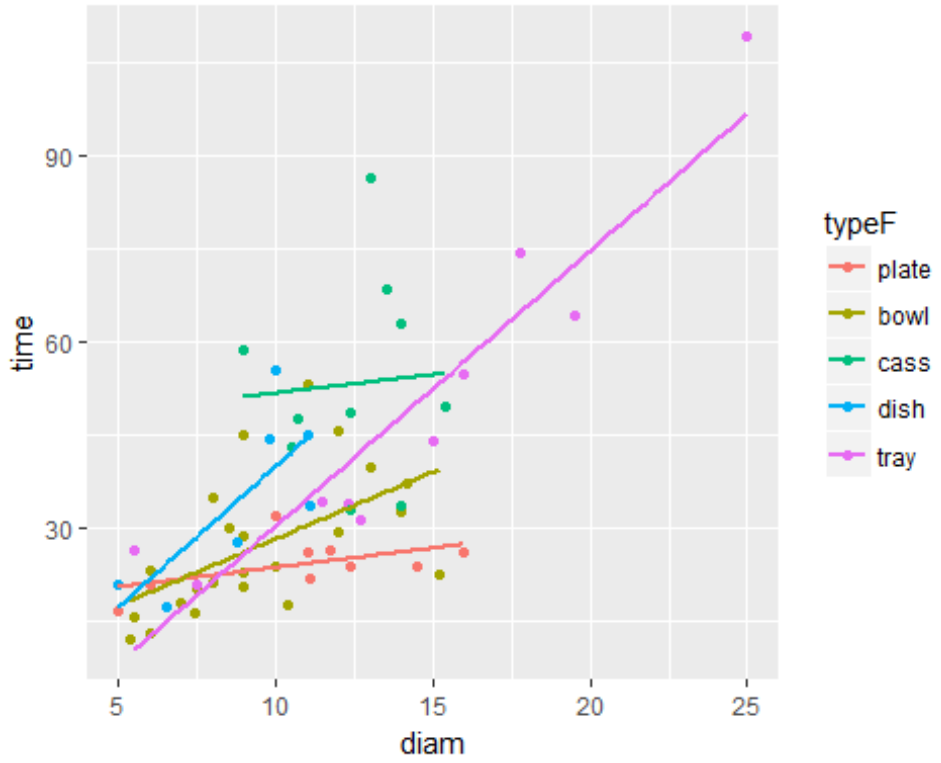
```
## [1] 0.3740579
```

```r
      # s = sqrt(MSE)
sqrt(anova(aovT)[["Sum Sq"]][2]/aovT$df.residual)
```

```
## [1] 15.59258
```

```r
  # fit the second-order model with interaction in two ways
fitPfi <- lm(time ~  diam * typeF)
summary(fitPfi)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -20.453  -5.953  -2.592   5.009  32.862
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     17.41075   12.21709   1.425  0.16046
## diam             0.62491    1.07514   0.581  0.56375
## typeFbowl      -10.62474   14.56155  -0.730  0.46908
## typeFcass       28.41439   26.54669   1.070  0.28970
## typeFdish      -22.84067   21.29184  -1.073  0.28864
## typeFtray      -31.39773   15.55788  -2.018  0.04907 *
## diam:typeFbowl   1.52955    1.34287   1.139  0.26024
## diam:typeFcass  -0.03004    2.15439  -0.014  0.98893
## diam:typeFdish   3.92621    2.18973   1.793  0.07914 .
## diam:typeFtray   3.81417    1.24636   3.060  0.00358 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Residual standard error: 10.83 on 49 degrees of freedom
## Multiple R-squared:  0.7258, Adjusted R-squared:  0.6755
## F-statistic: 14.41 on 9 and 49 DF,  p-value: 4.96e-11



fitPi <- lm(time ~  diam * (bowl + cass + dish + tray))
summary(fitPi)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -20.453  -5.953  -2.592   5.009  32.862
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  17.41075   12.21709   1.425  0.16046
## diam          0.62491    1.07514   0.581  0.56375
## bowl        -10.62474   14.56155  -0.730  0.46908
## cass         28.41439   26.54669   1.070  0.28970
## dish        -22.84067   21.29184  -1.073  0.28864
## tray        -31.39773   15.55788  -2.018  0.04907 *
## diam:bowl     1.52955    1.34287   1.139  0.26024
## diam:cass    -0.03004    2.15439  -0.014  0.98893
## diam:dish     3.92621    2.18973   1.793  0.07914 .
## diam:tray     3.81417    1.24636   3.060  0.00358 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Residual standard error: 10.83 on 49 degrees of freedom
## Multiple R-squared:  0.7258, Adjusted R-squared:  0.6755
## F-statistic: 14.41 on 9 and 49 DF,  p-value: 4.96e-11
```

```
    # line chart
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.4.4

ggplot(polish, aes(x = diam, y = time, color = typeF)) + geom_point() + geom_
smooth(method = "lm", se=F)
```




```
detach(polish)
```