14.3 Multiple Regression

Michael Sullivan

2023-08-29

Base R

Enter the data from Table 4 into R.

```
Age <- c(25, 25, 28, 32, 32, 32, 38, 42, 48, 51, 51, 58, 62, 65)
Fat <- c(19,28,19,16,24,20,31,20,26,24,32,21,21,30)
Cholesterol <- c(180, 195, 186, 180, 210, 197, 239, 183, 204, 221, 243, 208,
228, 269)
Table4 <- data.frame('Age'=Age, 'Fat'=Fat,'Cholesterol'=Cholesterol)
head(Table4)
```

```
##   Age Fat Cholesterol
## 1  25  19         180
## 2  25  28         195
## 3  28  19         186
## 4  32  16         180
## 5  32  24         210
## 6  32  20         197
```

To find a multiple regression model, use the lm() command. To include multiple explanatory variables, use + following ~ in the lm() command.

The summary() of the linear model object will include the least-squares regression model, R^2 and adjusted R^2, results of the F-test for lack of fit, and tests on the individual regression coefficients.

```
lm_object <- lm(Cholesterol ~ Age + Fat, data=Table4)
summary(lm_object)
```

```
##
## Call:
## lm(formula = Cholesterol ~ Age + Fat, data = Table4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.874  -8.192   3.479   8.151  14.907
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  90.8415    15.9887   5.682 0.000142 ***
## Age           1.0142     0.2427   4.179 0.001540 **
## Fat           3.2443     0.6632   4.892 0.000478 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.42 on 11 degrees of freedom
## Multiple R-squared:  0.8473, Adjusted R-squared:  0.8196
## F-statistic: 30.53 on 2 and 11 DF,  p-value: 3.239e-05
```

The regression model is $\hat{y} = 90.8415 + 1.0142Age + 3.2443Fat$. $R^2 = 84.7\%$ and $R^2_{adj} = 82.0\%$.

The F-test statistic is 30.53 with a P-value of 3.2e-05 = $3.2 \times 10^{-5} = 0.000032$.

The test statistic for the slope of Age is 4.179 with a P-value of 0.00154 and the test statistic for the slope of Fat is 4.892 with a P-value of 0.0048.

Residual Plots Using Base R

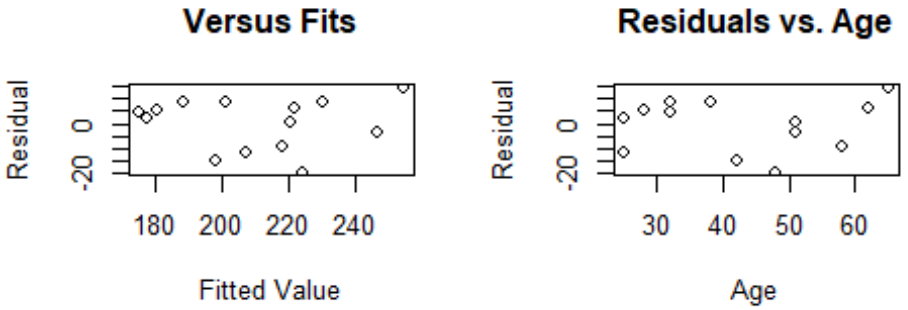
To perform a residual analysis on multiple regression residuals, you will draw 4 or more residual plots. The number of residual plots will always be the number of explanatory variables plus two (for the plot of residuals against the predicted values and the boxplot of residuals).

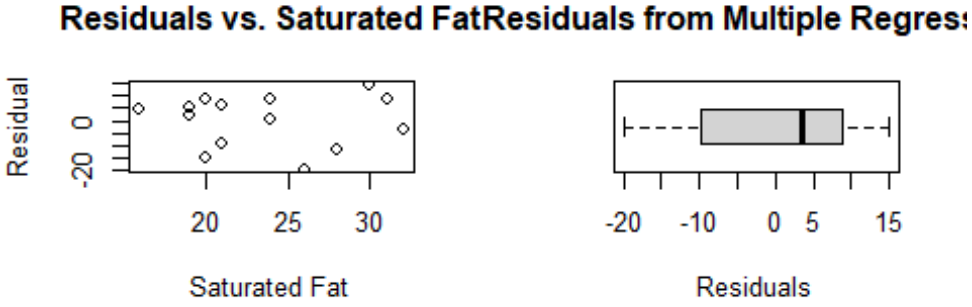
Use the following commands to plot residuals vs. fitted values, residuals vs. Age values, residuals vs. Saturated fat values, and a boxplot of residual values.

```
par(mfrow=c(2,2))
plot(model$fitted.values, model$residuals)
plot(Table4$Age, model$residuals)
plot(Table4$Fat, model$residuals)
boxplot(model$residuals, horizontal = TRUE)

## Note:** *par(mfrow=c(2,2))* splits the plotting area into a 2x2 area to
display 4 different plots at the same time. Use *par(mfrow=c(3,2))* if you
have 4 explanatory variables and *par(mfrow=c(4,2))* if you have 6
explanatory variables.

par(mfrow=c(2,2))
plot(lm_object$fitted.values, lm_object$residuals, main = "Versus Fits", xlab
= "Fitted Value", ylab = "Residual")
plot(Table4$Age, lm_object$residuals, main = "Residuals vs. Age", xlab =
"Age", ylab = "Residual")
plot(Table4$Fat, lm_object$residuals, main = "Residuals vs. Saturated Fat",
xlab = "Saturated Fat", ylab = "Residual")
boxplot(lm_object$residuals, horizontal = TRUE, main = "Residuals from
Multiple Regression", xlab = "Residuals")
```





Follow the procedure from 14.1.3 to draw a QQ Plot and determine the correlation of the QQ Plot.