# MULTIPLE REGRESSION

# Example: OmniFoods

- You are the marketing manager for OmniFoods, a large food products company. The company is planning a nationwide introduction of OmniPower, a new highenergy bar. Originally marketed to runners, mountain climbers, and other athletes, high-energy bars are now popular with the general public. OmniFoods is anxious to capture a share of this thriving market.
- Because the marketplace already contains several successful energy bars, you need to develop an effective marketing strategy. In particular, you need to determine the effect that price and in-store promotions will have on sales of OmniPower. Before marketing the bar nationwide, you plan to conduct a test-market study of OmniPower sales, using a sample of 34 stores in a supermarket chain.
- How can you develop a model to predict monthly sales volume per store of OmniPower bars and to determine what variables influence sales?

# Data

| Store | Sales Volumn | Price (cents) | Promotion ($) | Store | Sales Volumn | Price (cents) | Promotion ($) |
|---|---|---|---|---|---|---|---|
| 1 | 4141 | 59 | 200 | 18 | 2730 | 79 | 400 |
| 2 | 3842 | 59 | 200 | 19 | 2618 | 79 | 400 |
| 3 | 3056 | 59 | 200 | 20 | 4421 | 79 | 400 |
| 4 | 3519 | 59 | 200 | 21 | 4113 | 79 | 600 |
| 5 | 4226 | 59 | 400 | 22 | 3746 | 79 | 600 |
| 6 | 4630 | 59 | 400 | 23 | 3532 | 79 | 600 |
| 7 | 3507 | 59 | 400 | 24 | 3825 | 79 | 600 |
| 8 | 3754 | 59 | 400 | 25 | 1096 | 99 | 200 |
| 9 | 5000 | 59 | 600 | 26 | 761 | 99 | 200 |
| 10 | 5120 | 59 | 600 | 27 | 2088 | 99 | 200 |
| 11 | 4011 | 59 | 600 | 28 | 820 | 99 | 200 |
| 12 | 5015 | 59 | 600 | 29 | 2114 | 99 | 400 |
| 13 | 1916 | 79 | 200 | 30 | 1882 | 99 | 400 |
| 14 | 675 | 79 | 200 | 31 | 2159 | 99 | 400 |
| 15 | 3636 | 79 | 200 | 32 | 1602 | 99 | 400 |
| 16 | 3224 | 79 | 200 | 33 | 3354 | 99 | 600 |
| 17 | 2295 | 79 | 400 | 34 | 2927 | 99 | 600 |

```
# import data
OmniPower <- read.csv(file="OmniPower.csv")
attach(OmniPower)
```

# Model Fitting

```
# simple linear regression
OmniPowerFit <- lm(Sales ~ Price + Promotion)
```

# Model Fit and Significance

```
# model fit and significance
summary(OmniPowerFit)
```

```
Call:
lm(formula = Sales ~ Price + Promotion)

Residuals:
     Min       1Q   Median       3Q      Max
-1680.96  -406.40    53.45   297.48  1342.43

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 5837.5208   628.1502   9.293 1.79e-10 ***
Price        -53.2173     6.8522  -7.766 9.20e-09 ***
Promotion      3.6131     0.6852   5.273 9.82e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 638.1 on 31 degrees of freedom
Multiple R-squared: 0.7577, Adjusted R-squared: 0.7421
F-statistic: 48.48 on 2 and 31 DF,  p-value: 2.863e-10
```

# Residual Analysis

```
# observed sales value
Sales
```

```
 [1] 4141 3842 3056 3519 4226 4630 3507 3754 5000 5120 4011 5015 1916  675
[15] 3636 3224 2295 2730 2618 4421 4113 3746 3532 3825 1096  761 2088  820
[29] 2114 1882 2159 1602 3354 2927
```

```
# predicted sales value
fitted(OmniPowerFit)
```

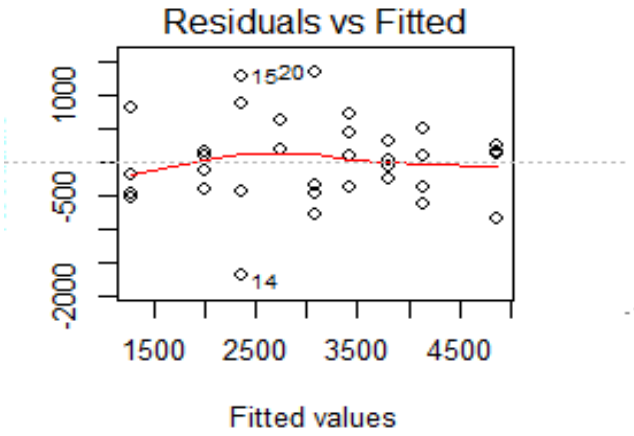
```
       1        2        3        4        5        6        7        8
3420.310 3420.310 3420.310 3420.310 4142.921 4142.921 4142.921 4142.921
       9       10       11       12       13       14       15       16
4865.533 4865.533 4865.533 4865.533 2355.963 2355.963 2355.963 2355.963
      17       18       19       20       21       22       23       24
3078.574 3078.574 3078.574 3078.574 3801.186 3801.186 3801.186 3801.186
      25       26       27       28       29       30       31       32
1291.616 1291.616 1291.616 1291.616 2014.228 2014.228 2014.228 2014.228
      33       34
2736.839 2736.839
```

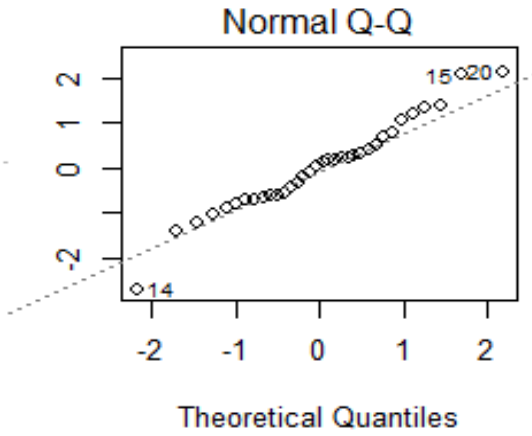
```
# residuals
residuals(OmniPowerFit)
```

```
       1        2        3        4        5        6
 720.69048  421.69048 -364.30952   98.69048   83.07887  487.07887
       7        8        9       10       11       12
-635.92113 -388.92113  134.46726  254.46726 -854.53274  149.46726
      13       14       15       16       17       18
-439.96280 -1680.96280 1280.03720  868.03720 -783.57440 -348.57440
      19       20       21       22       23       24
-460.57440 1342.42560  311.81399  -55.18601 -269.18601   23.81399
      25       26       27       28       29       30
-195.61607 -530.61607  796.38393 -471.61607   99.77232 -132.22768
      31       32       33       34
 144.77232 -412.22768  617.16071  190.16071
```

# Residual Analysis

```
# residual analyses
plot(OmniPowerFit)
```




```
> shapiro.test(res)

 Shapiro-Wilk normality test

data:  res
W = 0.975, p-value = 0.6116
```

Are the following assumptions adequately met?

1. **L**inearity
2. **I**ndependence of the residuals
3. **N**ormality of residual distribution
4. **E**qual variance of residuals

# Prediction

- You can use the multiple regression equation to predict values of the dependent variable. For example, what are the predicted sales for a store charging 79 cents during a month in which promotional expenditures are $400?

```
# store new values into a data frame
newdata <- data.frame(Price = 79, Promotion = 400)


# prediction
predict(OmniPowerFit, newdata)
```

```
       1
3078.574
```

This is from the model 5837.521 - 53.217 * 79 + 3.613 * 400

```
predict.lm(OmniPowerFit, newdata,
           interval = "predict")
       fit      lwr      upr
3078.574 1758.008 4399.14
```