Michelin

Chris Parrish

2017-04-12

Contents

Michelin

reference:
- Sheather, A Modern Approach to Regression with R, chapter 8, pp.263-268

Load package.

```
library(ggplot2)
library(boot)      # for inv.logit()
```

Import the data.

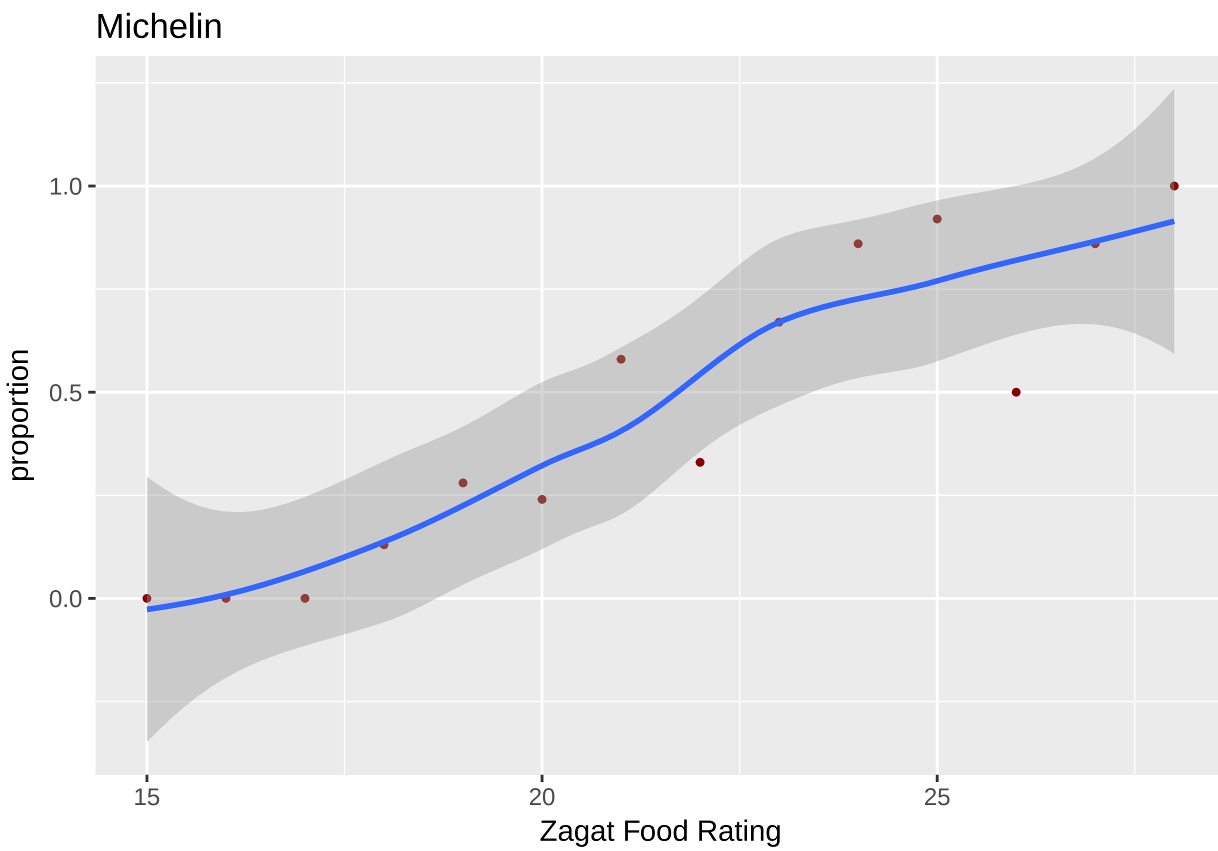
```
data <- read.delim("MichelinFood.txt")
data
```

```
##    Food InMichelin NotInMichelin mi proportion
## 1    15          0             1  1       0.00
## 2    16          0             1  1       0.00
## 3    17          0             8  8       0.00
## 4    18          2            13 15       0.13
## 5    19          5            13 18       0.28
## 6    20          8            25 33       0.24
## 7    21         15            11 26       0.58
## 8    22          4             8 12       0.33
## 9    23         12             6 18       0.67
## 10   24          6             1  7       0.86
## 11   25         11             1 12       0.92
## 12   26          1             1  2       0.50
## 13   27          6             1  7       0.86
## 14   28          4             0  4       1.00
```

Scatterplot.

```
ggplot(data, aes(Food, proportion)) +
  geom_point(shape = 20, color = "darkred") +
  geom_smooth(method = "loess") +
```

```r
  labs(x = "Zagat Food Rating",
       title = "Michelin")
```



logistic regression

logistic regression model

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x$$

```r
Michelin.glm <- glm(cbind(InMichelin, NotInMichelin) ~ Food, data = data, family = "binomial")
options(show.signif.stars = FALSE)
summary(Michelin.glm)
```

```
##
## Call:
## glm(formula = cbind(InMichelin, NotInMichelin) ~ Food, family = "binomial",
##     data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4850  -0.7987  -0.1679   0.5913   1.5889
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -10.84154    1.86236  -5.821 5.84e-09
```

```
## Food          0.50124    0.08768   5.717 1.08e-08
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 61.427  on 13  degrees of freedom
## Residual deviance: 11.368  on 12  degrees of freedom
## AIC: 41.491
##
## Number of Fisher Scoring iterations: 4
```

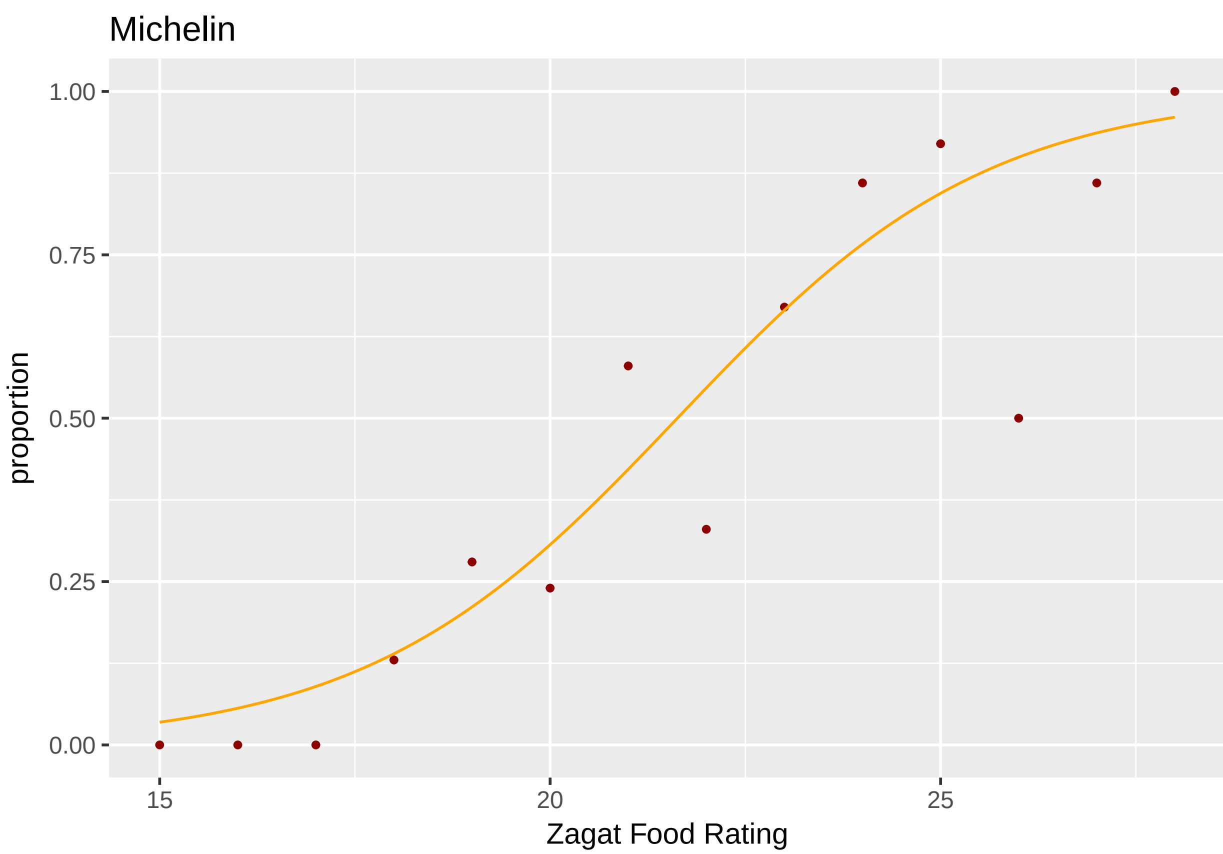
fit

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1 x = -10.8415389 + 0.5012367x$$

```r
summary(Michelin.glm)$coefficients
```

```
##                Estimate Std. Error   z value     Pr(>|z|)
## (Intercept) -10.8415389 1.86235775 -5.821405 5.835494e-09
## Food          0.5012367 0.08767565  5.716943 1.084576e-08
```

```r
ggplot(data, aes(x = Food, y = proportion)) +
  geom_point(shape = 20, color = "darkred") +
  stat_function(fun = f, color = "orange") +
  labs(x = "Zagat Food Rating",
       title = "Michelin")
```



95% CI for β_1

$$b_1 \pm z^* SE_{b_1}$$

```r
# b1
b1 <- summary(Michelin.glm)$coefficients[2, 1]
b1
```

```
## [1] 0.5012367
```

```r
#se.b1
se.b1 <- summary(Michelin.glm)$coefficients[2, 2]
se.b1
```

```
## [1] 0.08767565
```

```r
# z.star
alpha <- 0.05
z.star <- qnorm(1 - alpha/2)
z.star
```

```
## [1] 1.959964
```

```r
# ci
ci <- b1 + z.star * se.b1 * c(-1, 1)
ci
```

```
## [1] 0.3293956 0.6730778
```

Wald test for β_1

$$H_0 : \beta_1 = 0$$
$$H_a : \beta_1 \neq 0$$

```r
# beta1
beta1 <- 0                 # from H_0
# test statistic z
z <- (b1 - beta1) / se.b1
z
```

```
## [1] 5.716943
```

```r
# p.value
p.value <- 2 * (1 - pnorm(z))
p.value
```

```
## [1] 1.084576e-08
```

estimated odds for inclusion in the Michelin guide

$$estimated.odds = e^{b_1}$$

```r
estimated.odds <- exp(b1)
estimated.odds
```

```
## [1] 1.650761
```

95% CI for estimated.odds

$$e^{b_1} \pm e^{z^* SE_{b_1}}$$

```
ci.for.estimated.odds <- exp(ci)
ci.for.estimated.odds
```

```
## [1] 1.390128 1.960261
```

interpretation of the estimated odds

If the Zagat food score increases by 1, then the log odds of inclusion in the Michelin guide increases by b_1,

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1(x+1) = b_0 + b_1 x + b_1$$

and the odds increase by a factor of $e^{b_1} = 1.65$,

$$\frac{p}{1-p} = e^{b_1} * e^{b_0 + b_1 x}$$