To determine the price of used cars, a critical factor is how far the car has been driven. To examine this issue, a used-car dealer randomly selected 100 3-year old Toyota Camrys. The price ($1,000) and the number of miles (thousands) on the odometer were recorded (file `Xm16-02.csv`)
It is believed that the color of a car is a factor in determining its price and assumed that the colors that are most popular, white and silver, are likely to lead to different prices than other colors. Accordingly, a code of 1 was assigned to white cars, a code of 2 to silver cars, and a code of 3 to all other colors.
Fit a multiple regression model to predict the price of a used-car based on the odometer reading and the car's color. Define cars with other colors than white and silver, to be the base model.

```
setwd("C:/Users/USC Guest/Downloads2")
d0 = read.csv("Xm16-02.csv", header = T)
head(d0)
#  Price Odometer Color
#1  14.6     37.4     1
#2  14.1     44.8     1
#3  14.0     45.8     3
#4  15.6     30.9     3
#5  15.6     31.7     2
#6  14.7     34.0     2
str(d0)
#'data.frame':   100 obs. of  3 variables:
# $ Price   : num  14.6 14.1 14 15.6 15.6 14.7 14.5 15.7 15.1 14.8 ...
# $ Odometer: num  37.4 44.8 45.8 30.9 31.7 34 45.9 19.1 40.1 40.2 ...
# $ Color   : int  1 1 3 3 2 2 1 3 1 1 ...
d0$Color = as.factor(d0$Color)
str(d0)
#'data.frame':   100 obs. of  3 variables:
# $ Price   : num  14.6 14.1 14 15.6 15.6 14.7 14.5 15.7 15.1 14.8 ...
# $ Odometer: num  37.4 44.8 45.8 30.9 31.7 34 45.9 19.1 40.1 40.2 ...
# $ Color   : Factor w/ 3 levels "1","2","3": 1 1 3 3 2 2 1 3 1 1 ...

# 1 is white
# 2 is silver
# 3 is other color

levels(d0$Color)
# [1] "1" "2" "3"

# level 1 is base model (default)

# set "3" the base level
d0$Color = relevel(d0$Color,ref="3")
levels(d0$Color)
# "3" "1" "2"

# model
#-----------------------------------------------------------------------
m1 = lm(Price~.,d0)
coef(m1)
# (Intercept)    Odometer      Color1      Color2
# 16.83724755 -0.05912294  0.09113080  0.33036793

# Fitted equation for base model
# E[Price] = 16.83724755 - 0.05912294*(Odometer reading)

# Fitted equation for white cars (level 1)
```

```
# E[Price] = (16.83724755+0.09113080) - 0.05912294*(Odometer reading)

# Fitted equation for  silver cars (level 2)
# E[Price] = (16.83724755+0.33036793) - 0.05912294*(Odometer reading)

# Test coefficients
summary(m1)
#Coefficients:
#             Estimate Std. Error t value Pr(>|t|)
#(Intercept) 16.837248   0.197105  85.423  < 2e-16 ***
#Odometer    -0.059123   0.005065 -11.672  < 2e-16 ***
#Color1       0.091131   0.072892   1.250 0.214257
#Color2       0.330368   0.081650   4.046 0.000105 ***

#Residual standard error: 0.3043 on 96 degrees of freedom
#Multiple R-squared:  0.7008,    Adjusted R-squared:  0.6914
#F-statistic: 74.95 on 3 and 96 DF,  p-value: < 2.2e-16

# do not reject Ho: beta(Color1) = 0
# intercept for cars, white color, not different from base model

int = coef(m1)[1]
slope = coef(m1)[2]
b2 = coef(m1)[3]     # additional avg price of white cars
b3 = coef(m1)[4]

# plot
label = c("other colors","white","silver")
xb = c(0,50)
yb = c(12,18)
plot(Price~Odometer,d0,col=Color,pch=19,cex=0.6,xlim=xb,ylim=yb)
grid()
abline(int,slope,col=1)
abline(int+b2,slope,col=2)
abline(int+b3,slope,col=3)
legend("topright",legend=label,lty=c(1,1,1),col=c(1,2,3),cex=0.8)


# Combine levels
levels(d0$Color)
# [1] "3" "1" "2"
levels(d0$Color)=c("N","N","S")
levels(d0$Color)
# [1] "N" "S"
head(d0)
#   Price Odometer Color
# 1  14.6     37.4     N
```

```
# 2  14.1     44.8     N
# 3  14.0     45.8     N
# 4  15.6     30.9     N
# 5  15.6     31.7     S
# 6  14.7     34.0     S
str(d0)
#'data.frame':   100 obs. of  3 variables:
# $ Price   : num  14.6 14.1 14 15.6 15.6 14.7 14.5 15.7 15.1 14.8 ...
# $ Odometer: num  37.4 44.8 45.8 30.9 31.7 34 45.9 19.1 40.1 40.2 ...
# $ Color   : Factor w/ 2 levels "N","S": 1 1 1 1 2 2 1 1 1 1 ...

# model - simplified
#------------------------------------------------------------------------
m2 = lm(Price~.,d0)
coef(m2)
# (Intercept)    Odometer      ColorS
# 16.87739482 -0.05891037  0.28342190
summary(m2)
#Coefficients:
#             Estimate Std. Error t value Pr(>|t|)
#(Intercept) 16.877395   0.195036  86.535  < 2e-16 ***
#Odometer    -0.058910   0.005077 -11.603  < 2e-16 ***
#ColorS       0.283422   0.072713   3.898 0.000179 ***

#Residual standard error: 0.3051 on 97 degrees of freedom
#Multiple R-squared:  0.6959,    Adjusted R-squared:  0.6897
#F-statistic:   111 on 2 and 97 DF,  p-value: < 2.2e-16

# Fitted equation base model (non-silver color cars)
# E[Price] = 16.83724755 - 0.05891037*(Odometer reading)

# Fitted equation silver color cars
# E[Price] = (16.83724755+0.28342190) - 0.05891037*(Odometer reading)

# Silver cars cost 283.42190  more, on average
```

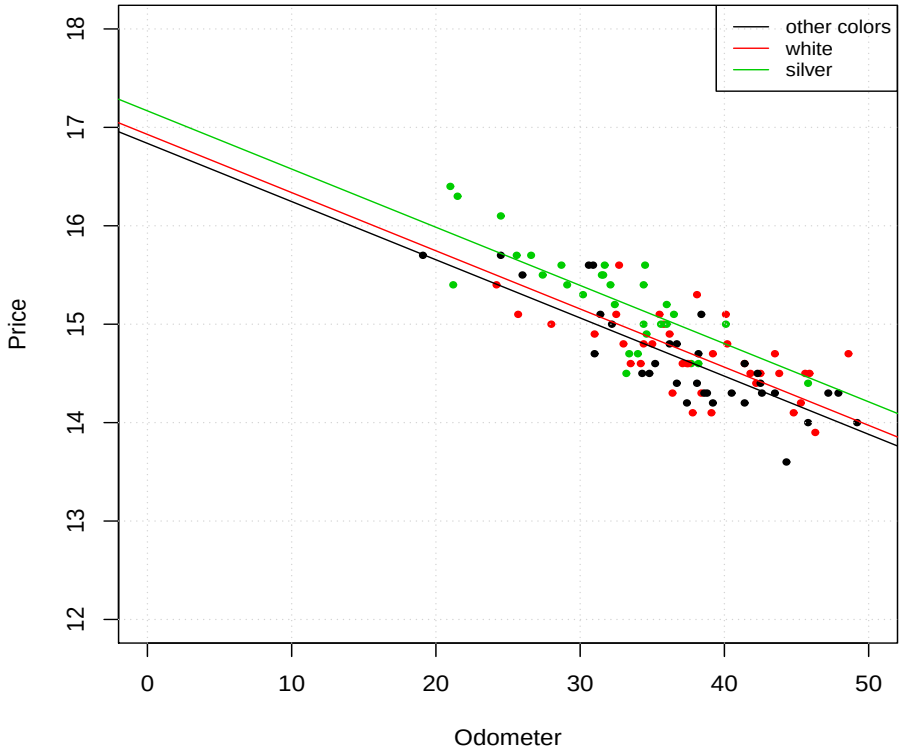

Figure 1: Scatterplot with fitted equations (p740)