Car dealers across North America use the so-called Blue Book to help them determine the value of used cars that their customers trade in when purchasing new cars. However, the Blue Book does not indicate the value determined by the odometer reading, despite the fact that a critical factor for used-car buyers is how far the car has been driven. To examine this issue, a used-car dealer randomly selected 100 3-year old Toyota Camrys that were sold at auction during the past month. The dealer recorded the price ($1,000) and the number of miles (thousands) on the odometer. The data is found in file `Xm16-02.csv`

1. Fit a regression model to predict a car's price using the odometer reading. Write the fitted equation $\hat{y} = b_0 + b_1\,x$. Interpret the slope b_1. Construct a scatterplot of car prices against odometer readings. Add the fitted equation to the scatterplot. Interpret the standard error of the estimate.
2. Test if there is a linear relationship between the auction price and the odometer reading for all 3-year-old Toyota Camrys. Find a 99% confidence interval for the slope β_1.
3. Find the coefficient of determination R^2. Interpret its numerical value. Verify that $R^2 = r^2$, where r is correlation between price and odometer reading.
4. Prediction
   a) Predict the average price of all 3-year-old Toyota Camry cars with 40,000 miles on the odometer. Find a 96% CI on this mean price.
   b) Predict the selling price of a 3-year-old Toyota Camry car with 40,000 miles on the odometer. Find a 95% PI on this price.
5. Verify the regression assumptions
6. Plot the fitted line with CIs and PIs

```
# slr.r

setwd("C:/Users/USC Guest/Downloads2")
d0 = read.csv("Xm16-02.csv", header = T)
head(d0)

# model
#-----------------------------------------------------------------------
m1 = lm(Price~Odometer,d0)

coef(m1)
# (Intercept)    Odometer
# 17.24872734 -0.06686089

# Fitted equation
# E[Price] = 17.24872734 - 0.06686089*(Odometer reading)

# scatterplot
plot(Price~Odometer,d0,pch=19,cex=0.4)
grid()
abline(m1,col="red")

# interpret slope b1
# For each additional mile on the odometer reading
# the car's price decreases by 0.067 dollars, on average

# Standard Error of the estimate S
#-----------------------------------------------------------------------
anova(m1)
# Analysis of Variance Table
# Response: Price
#           Df Sum Sq Mean Sq F value    Pr(>F)
# Odometer   1 19.256 19.2556  180.64 < 2.2e-16 ***
# Residuals 98 10.446  0.1066

MSE = 0.1066
S = sqrt(MSE)                 # [1] 0.3264966

# interpret S and MSE
# the average distance to the fitted line is 0.3264966 dollars.
# MSE is an estimate of the variance sigma^2
```

```
# test linear relationship Y vs X
#-----------------------------------------------------------------------
summary(m1)

#Coefficients:
#             Estimate Std. Error t value Pr(>|t|)
#(Intercept) 17.248727   0.182093   94.72   <2e-16 ***
#Odometer    -0.066861   0.004975  -13.44   <2e-16 ***

#Residual standard error: 0.3265 on 98 degrees of freedom
#Multiple R-squared:  0.6483,    Adjusted R-squared:  0.6447
#F-statistic: 180.6 on 1 and 98 DF,  p-value: < 2.2e-16

# T test for Ho: beta1=0
# second row of coeff table shows p-value <2e-16
# reject Ho, conclude that
# there is a relationship between average price and odometer reading

confint(m1,level=0.99)
#                   0.5 %      99.5 %
# (Intercept) 16.77038270 17.72707199
# Odometer    -0.07992892 -0.05379285

# slope beta1 lies between -0.07992892 and -0.05379285

# On average, for each additional mile
# the car's price decreases an amount between 0.07992892 and 0.05379285 dollars


# Coeff of determination R-squared and square of correlation r
#-----------------------------------------------------------------------

# from coeff table, R2 = 0.6483

r = cor(d0$Price,d0$Odometer)
# -0.805168
R2 = r^2
R2
# 0.6482955 agrees with Multiple R-squared from coeff table

# interpret R2
# 64.83% of variation of selling prices is explained by variation in odometer readings
```

```
# prediction
#-----------------------------------------------------------------------

newval = data.frame(Odometer=40)
predict(m1,newval)
# 14.57429   000s dollars

predict(m1,newval,interval = "conf", level=0.96)
#        fit      lwr      upr
# 1 14.57429 14.49477 14.65382

predict(m1,newval,interval = "pred")
#        fit      lwr      upr
# 1 14.57429 13.92196 15.22662

# prediction intervals wider than CIs, always

# residuals plots
#-----------------------------------------------------------------------
library(PASWR2)
checking.plots(m1)

# test for normality
# Ho: observations are normal

sres = rstandard(m1)
shapiro.test(sres)
# Shapiro-Wilk normality test
# W = 0.9848, p-value = 0.307

# do not reject Ho

# assumptions (normality and equal variance) hold
```

```
# plot with conf bands
#-----------------------------------------------------------------------

xaxis = 19:50
newval = data.frame(Odometer=xaxis)
d2=predict(m1,newval,interval = "conf")
head(d2)
#       fit      lwr      upr
#1 15.91151 15.74069 16.08233
#2 15.84465 15.68292 16.00638
#3 15.77779 15.62505 15.93053
#4 15.71093 15.56707 15.85479
#5 15.64407 15.50895 15.77919
#6 15.57721 15.45066 15.70375
d3=predict(m1,newval,interval = "pred")

# plot
plot(Price~Odometer,d0,pch=19,cex=0.4)
grid()
abline(m1,col="red")
# CIs
lines(xaxis,d2[,2],lty=2)
lines(xaxis,d2[,3],lty=2)
# PIs
lines(xaxis,d3[,2],lty=3)
lines(xaxis,d3[,3],lty=3)

# intervals at 40k miles
abline(v=40,lty=2,col=4)
```

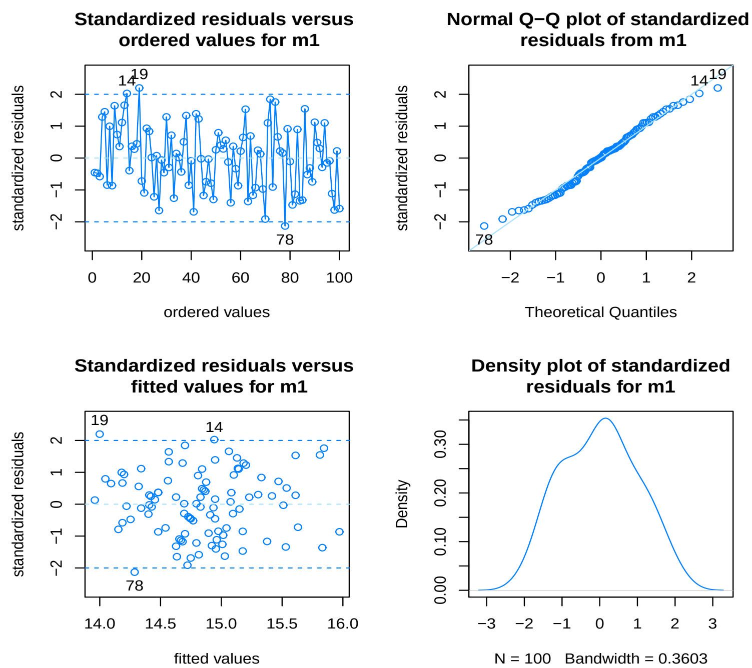

Figure 1: Residual plots

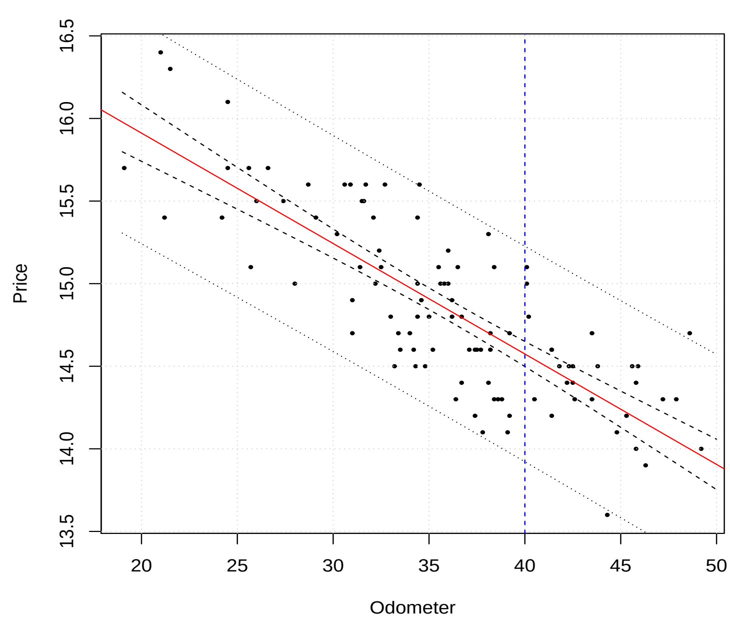

Figure 2: Fitted line with statistical intervals