Solutions to Assignment 3

Rongfei Jin

January 31, 2025

Conceptual 1

Solution.

The null hypothesis is the following: $H_0 : \beta_0 = 0$ (TV) $H_0 : \beta_1 = 0$ (Radio) $H_0 : \beta_2 = 0$ (Newspaper)

Assume we pick $\alpha = 0.05$ as the significance level. Then, the p-value of TV and Radio are both smaller than 0.05, so we reject the null hypothesis and conclude that TV and Radio have significant impact on sales. However, the p-value of Newspaper is larger than 0.05, so we fail to reject the null hypothesis and conclude that Newspaper does not have significant impact on sales.

■

Conceptual 3

Solution.

(a) Let the GPA and IQ be $c_1, c_2 \in \mathbb{Z}^*$ The model then can be written as:

$$\hat{y} = 50 + 20c_1 + 0.07c_2 + 35x_3 + 0.01c_1c_2 - 10x_3c_1 + 0.01x_3c_2$$

- (i) wrong, let $c_1 = 4$ then $x_3 = 1$ will result in a smaller predicted value than when $x_3 = 0$
- (ii) wrong, let $c_1 = 1$, then $x_3 = 1$ will result in a greater predicted value than when $x_3 = 0$
- (iii) right, let $c_1 = 4$ then $x_3 = 1$ will result in a smaller predicted value than when $x_3 = 0$
- (iv) wrong since it is the opposite of (iii)

(b) substitute the values into the expression above, $\hat{y} = 137.1$

(c) small interaction does not necessary means the interaction is not significant. The significance of the interaction term should be determined by the p-value and the p-value is determined by standard error of the parameter and the parameter itself.

■

Conceptual 6

Solution. We can substitute the value of $\bar{x}$ into (3.4) to get

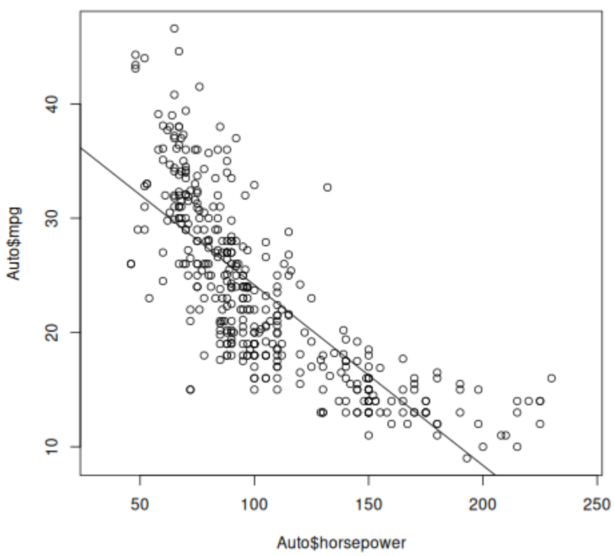
$$\begin{aligned} \hat{y} &= \hat{\beta}_0 + \hat{\beta}_1\bar{x} \\ &= \bar{y} - \hat{\beta}_1\bar{x} + \hat{\beta}_1\bar{x} \\ &= \bar{y} \end{aligned}$$

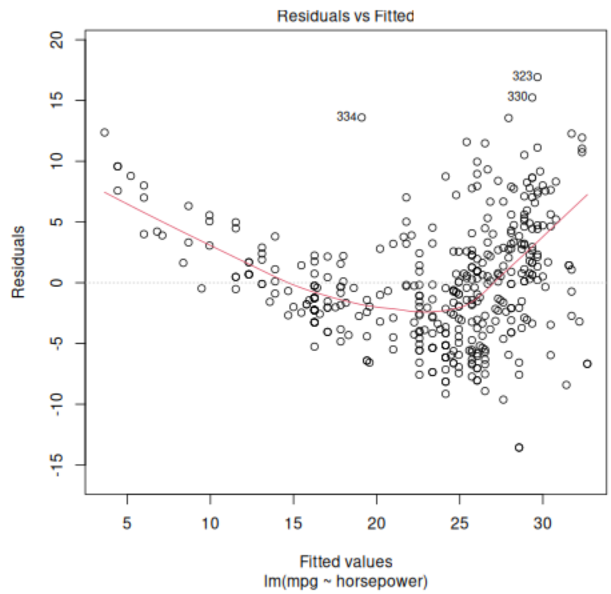
■

Applied 8

```
library(ISLR2)
model <- lm(mpg ~ horsepower, data = Auto)
summary(model)

# set plot ratio
plot(Auto$horsepower, Auto$mpg, asp = 5)
abline(model)

plot(model, which=1)
```

[lineos,breaklines] A linear relationship exists between mpg and horsepower. The p-value of horsepower is less than 0.05, so we reject the null hypothesis and conclude that horsepower has significant impact on mpg. This impact is negative since the coefficient of horsepower is negative. The R-squared is 0.6011, which means 60.11% of the variation in mpg can be explained by the linear relationship between mpg and horsepower. The strength of the linear relationship is moderate.

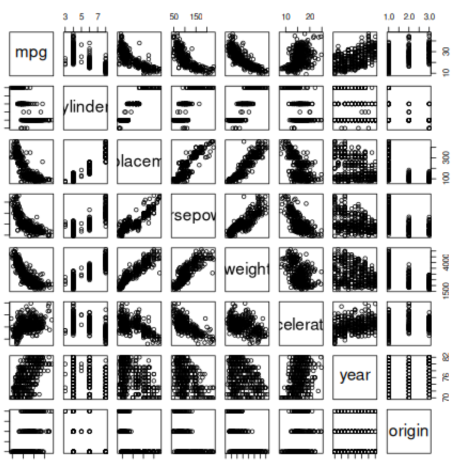




The Residual plot at fitted values between 5-20 shows consistent residual variations. However, above 20, the greater variations appear, suggesting that the linear model may not be appropriate for the data above 20.

Applied 9

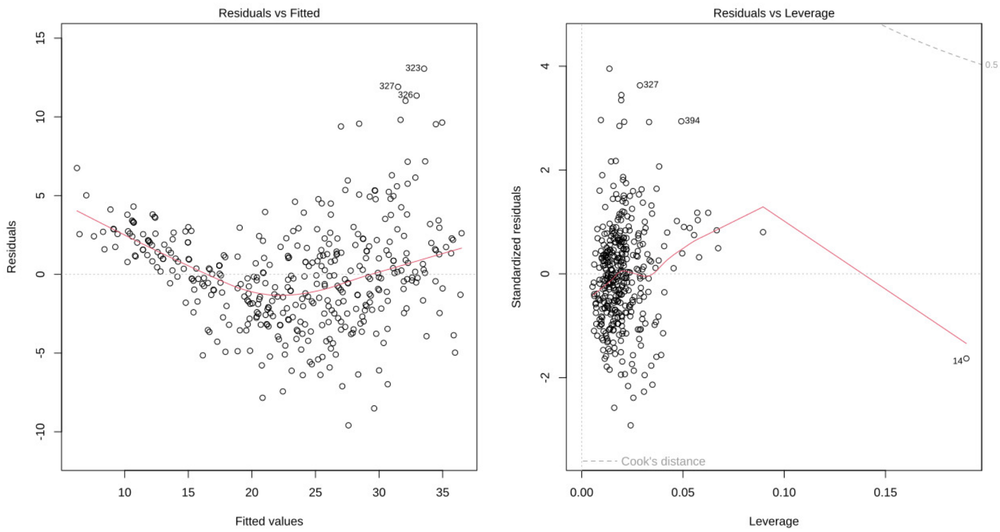
```r
library(ISLR2)
# exclude the name column
Auto <- Auto[, -9]
str(Auto)

# a
pairs(Auto)

# b
cor(Auto)

# c
model <- lm(mpg ~ ., data = Auto)
summary(model)
par(mfrow = c(1, 2))
plot(model, which=1)
plot(model, which=5)

model_2 <- lm(mpg ~ . + cylinders:horsepower, data = Auto)
summary(model_2)

model_2 <- lm(mpg ~ . + acceleration:origin, data = Auto)
summary(model_2)


model_3 <- lm (mpg ~ cylinders+I(displacement)^2 + log(weight)+ weight + year + acceleration + origin + horse
summary(model_3)
```



(a)

There exists a relationship between mpg and other predictors. Cylinder, displacement, weights, year and origin are significant predictors. The coefficient of year is poisitive 0.75, indicating that the mpg increases by 0.75 for each additional year.

(b) (See code above)



(c)

The residual plots shows that the distribution of residual is not consistently. So the linear model may not be appropriate for the data.

The leverage plot shows that there exists 1 extreme outliers and some less extreme outliers

(d) The interactions of some predictors, such as cylinders times horsepower, are significant.

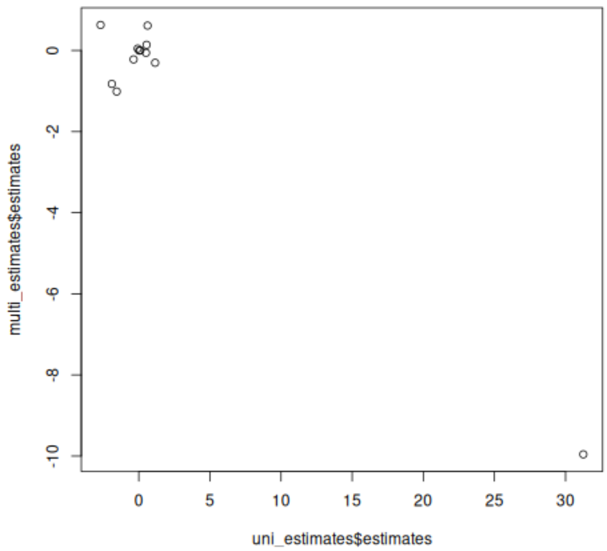
(e) The transformation fo the response further improves the model by lowering the residual standard errors and increasing the R-squared.

Applied 15

```r
library(ISLR2)

predictors = subset(Boston, select = -crim)
models = fits <- lapply(predictors, function(x) lm(Boston$crim ~ x))


# put p values in a dataframe
p_values = data.frame(p = sapply(models, function(x) summary(x)$coefficients[2, 4]))

print(p_values)

multiple_model = lm(crim ~ ., data = Boston)
summary(multiple_model)


uni_estimates = data.frame(estimates = sapply(models, function(x) summary(x)$coefficients[2, 1]))
print(uni_estimates)

multi_estimates = data.frame(estimates = summary(multiple_model)$coefficients[-1, 1])

multi_estimates

# plot (multi_estimates vs uni_estimates)
par(mfrow = c(1, 1))
plot(uni_estimates$estimates, multi_estimates$estimates)

cubic_models <- lapply(predictors, function(x) lm(Boston$crim ~ x + I(x^2) + I(x^3)))

for (name in names(cubic_models)) {
  cat(name, "\n")
  print(summary(cubic_models[[name]])$coefficients[,4])
    cat("\n")
}
```

(a) All the predictors except chaos are significant

(b) The multiple linear regression model yields a model with 0.45 R squared. Predictor zn, dis, rad and medv are significant predictors to reject null hypothesis. The coefficient of dis is negative, indicating that the crime rate increases as the distance to employment centers decreases. The coefficient of rad is positive, indicating that the crime rate increases as the accessibility to radial highways increases. The coefficient of medv is negative, indicating that the crime rate decreases as the median value of owner-occupied homes increases.



(c) The estimates from the multiple linear regression model and the stepwise regression model are different especially for the estimator of nox

(d) Some coefficients in polynomial model are significant such as x^2, x^3 of nox, indus and medv, suggesting that the data may be non-linear

Additional 1

$$\begin{aligned}
\nabla f(\boldsymbol{x}) &= \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\partial}{\partial x_1}\boldsymbol{u}^T\boldsymbol{x} \\ \vdots \\ \frac{\partial}{\partial x_n}\boldsymbol{u}^T\boldsymbol{x} \end{bmatrix} \\
&= \begin{bmatrix} \frac{\partial}{\partial x_1}\sum u_i x_i \\ \vdots \\ \frac{\partial}{\partial x_n}\sum u_i x_i \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{u}_1 \\ \vdots \\ \boldsymbol{u}_n \end{bmatrix}
\end{aligned}$$

Additional 2

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \\ \vdots \\ \frac{\partial}{\partial x_n}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \end{bmatrix}$$

We take x_1 and the rest will follow

$$\begin{aligned}
\frac{\partial f}{\partial x_1} &= \frac{\partial}{\partial x_1}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \\
&= \frac{\partial}{\partial x_1}\sum^{j}\sum^{i} a_{ij}x_i x_j \\
&= \frac{\partial}{\partial x_1}(\sum_{j=2}^{n} a_{1j}x_1 x_j + \sum_{i=2}^{n} a_{i1}x_i x_1 + x_1^2 a_{11}) + 0 \\
&= \sum_{j=2}^{n} a_{1j}x_j + \sum_{i=2}^{n} a_{i1}x_i + 2a_{11}x_1 \\
&= \sum_{j=1}^{n} a_{1j}x_j + \sum_{i=1}^{n} a_{i1}x_i \\
&= 2\sum_{j=1}^{n} a_{1j}x_j && \text{symmetric property of } A \\
&= 2(\boldsymbol{A}_1.\boldsymbol{x})
\end{aligned}$$

We can write the rest in a similar way and get the final result

$$\nabla f(\boldsymbol{x}) = 2\boldsymbol{A}\boldsymbol{x}$$

Additional 3

From Exercise 2, we know

$$
\begin{aligned}
\frac{\partial f}{\partial x_1} &= 2\sum_{j=1}^{n} a_{1j}x_j \\
\frac{\partial^2 f}{\partial^2 x_1} &= 2\sum_{j=1}^{n} a_{1j}
\end{aligned}
$$