production

Chris Parrish

December 8, 2016

Contents

production

reference:
- Sheather, A Modern Approach to Regression with R, chapter 2, pp.15-30; chapter 3, p.70

Load package.

```
library(ggplot2)
```

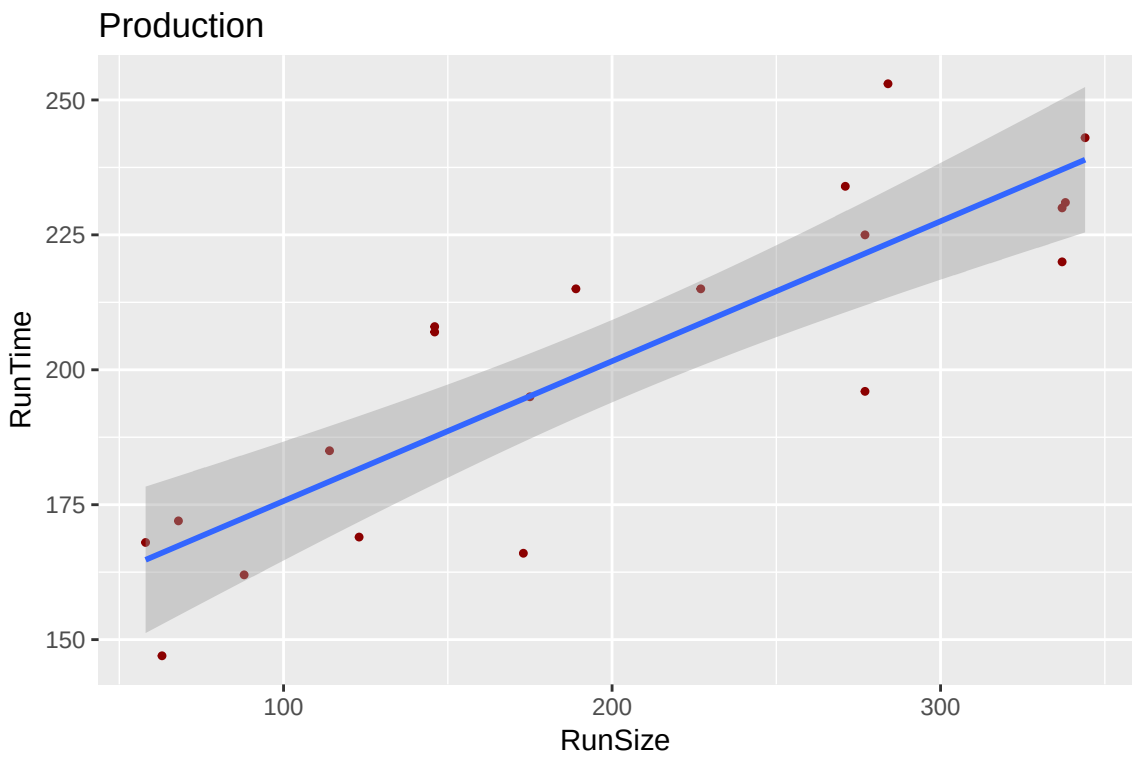
Import the data.

```
data <- read.delim("production.txt", header=TRUE)
head(data)
```

```
##   Case RunTime RunSize
## 1    1     195     175
## 2    2     215     189
## 3    3     243     344
## 4    4     162      88
## 5    5     185     114
## 6    6     231     338
```

linear model

```
ggplot(data, aes(RunSize, RunTime)) +
  geom_point(shape = 20, color = "darkred") +
  geom_smooth(method = "lm") +
  ggtitle("Production")
```

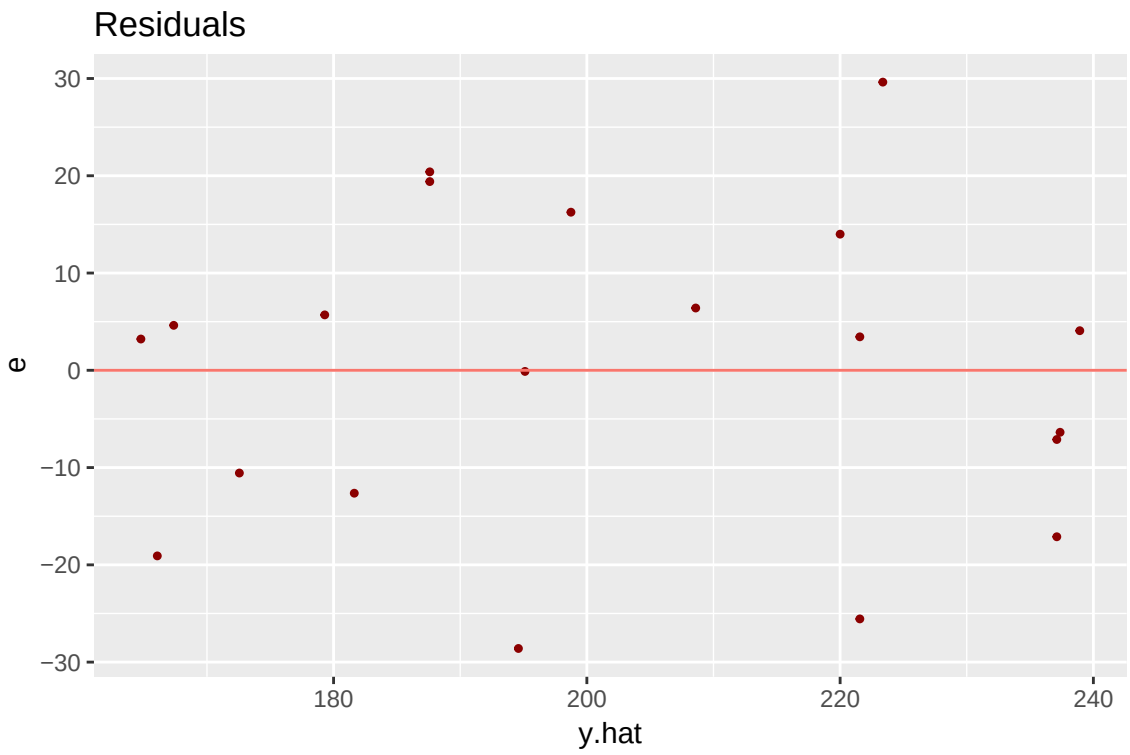



```
production.lm <- lm(RunTime ~ RunSize, data=data)
summary(production.lm)
```

```
##
## Call:
## lm(formula = RunTime ~ RunSize, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -28.597 -11.079   3.329   8.302  29.627
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 149.74770    8.32815   17.98 6.00e-13
## RunSize       0.25924    0.03714    6.98 1.61e-06
##
## Residual standard error: 16.25 on 18 degrees of freedom
## Multiple R-squared:  0.7302, Adjusted R-squared:  0.7152
## F-statistic: 48.72 on 1 and 18 DF,  p-value: 1.615e-06
```

residuals

```
data.lm <- data.frame(y.hat = predict(production.lm),
                      e = production.lm$residuals)
ggplot(data.lm, aes(y.hat, e)) +
  geom_point(shape = 20, color = "darkred") +
  geom_hline(aes(yintercept = 0, color = "orange")) +
  theme(legend.position="none") +
  ggtitle("Residuals")
```



CI for β_1

```
coef <- coefficients(summary(production.lm))
coef
```

```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 149.7477030 8.32815141 17.980905 5.996749e-13
## RunSize       0.2592431 0.03714198  6.979788 1.614863e-06
```

```
b1 <- coef[2, 1]
se.b1 <- coef[2, 2]
alpha <- 0.05
n <- nrow(data)
t.star <- qt(1 - alpha/2, df=n - 2)
t.star
```

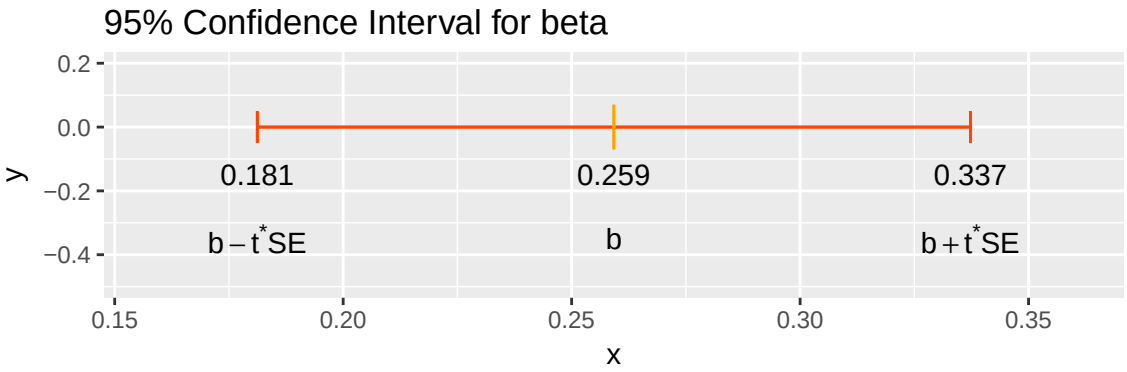
```
## [1] 2.100922
```

```
ci <- b1 + t.star * se.b1 * c(-1, 1)
ci
```

```
## [1] 0.1812107 0.3372755
```

View the CI.

```
center <- b1
title <- "95% Confidence Interval for beta"
gg.draw.ci.mean.sigma.unknown(center, ci, title)
```



Check.

```
confint(production.lm)
```

```
##                   2.5 %      97.5 %
## (Intercept) 132.2509062 167.2444999
## RunSize       0.1812107   0.3372755
```

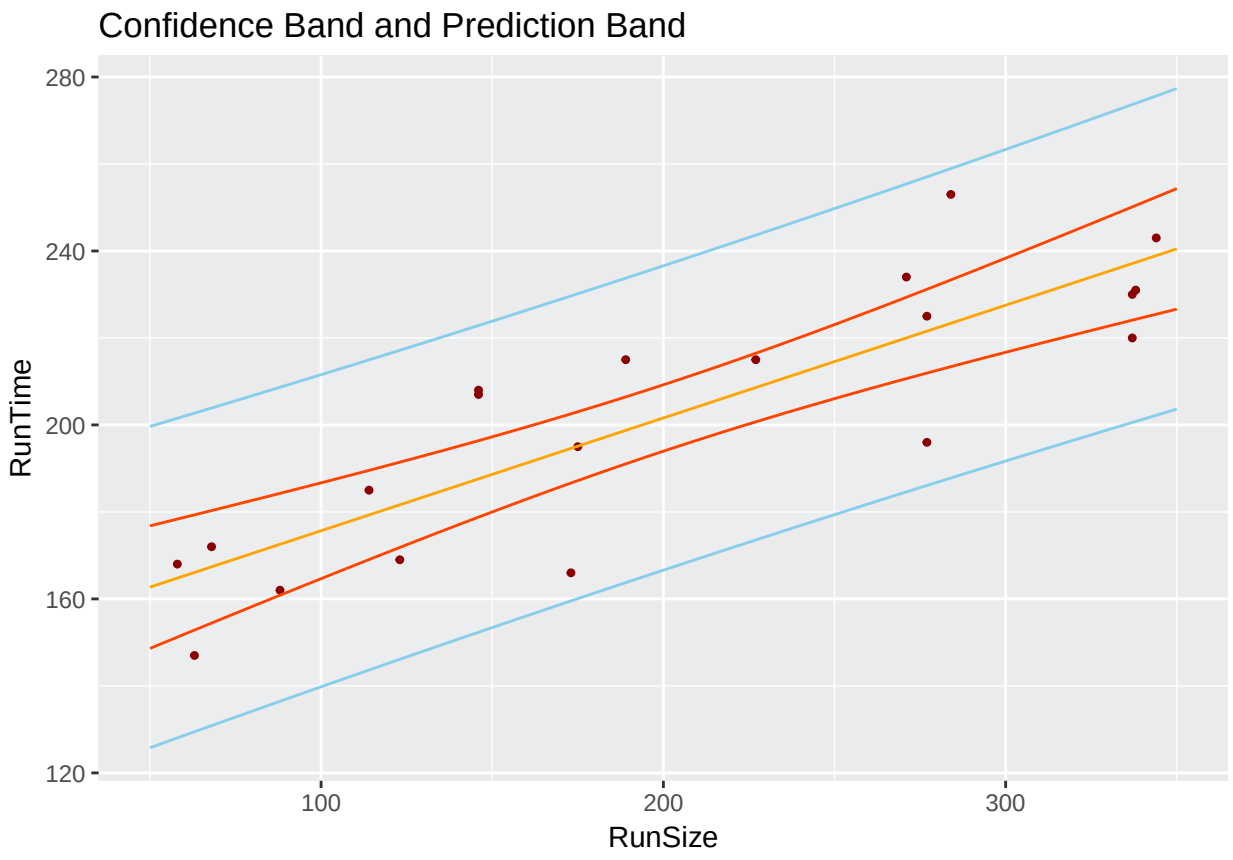
confidence band and prediction band

Illustration of CI for μ_y given x and PI for y given x

Notice that the confidence band and prediction band are narrowest near the point $(\bar{x}, \bar{y})$, but then both bands widen as the distance to that point gets larger.

```
xs <- seq(from=50, to=350, by=0.02)
new.data <- data.frame(RunSize=xs)
CI <- predict(production.lm, new.data, interval="confidence")
CI.df <- as.data.frame(CI)
CI.df$RunSize <- xs
PI <- predict(production.lm, new.data, interval="prediction")
PI.df <- as.data.frame(PI)
PI.df$RunSize <- xs
PI.df$RunTime <- PI.df$fit
```

```
ggplot(data, aes(RunSize, RunTime)) +
  geom_ribbon(data = PI.df,
              aes(x = RunSize, ymin = lwr, ymax = upr),
              fill = "aliceblue", alpha = 0.2) +
  geom_point(shape = 20, color = "darkred") +
  geom_line(data = PI.df, aes(RunSize, fit), color = "orange") +
  geom_line(data = PI.df, aes(RunSize, upr), color = "skyblue") +
  geom_line(data = PI.df, aes(RunSize, lwr), color = "skyblue") +
  geom_line(data = CI.df, aes(RunSize, upr), color = "orangered") +
  geom_line(data = CI.df, aes(RunSize, lwr), color = "orangered") +
  ggtitle("Confidence Band and Prediction Band")
```

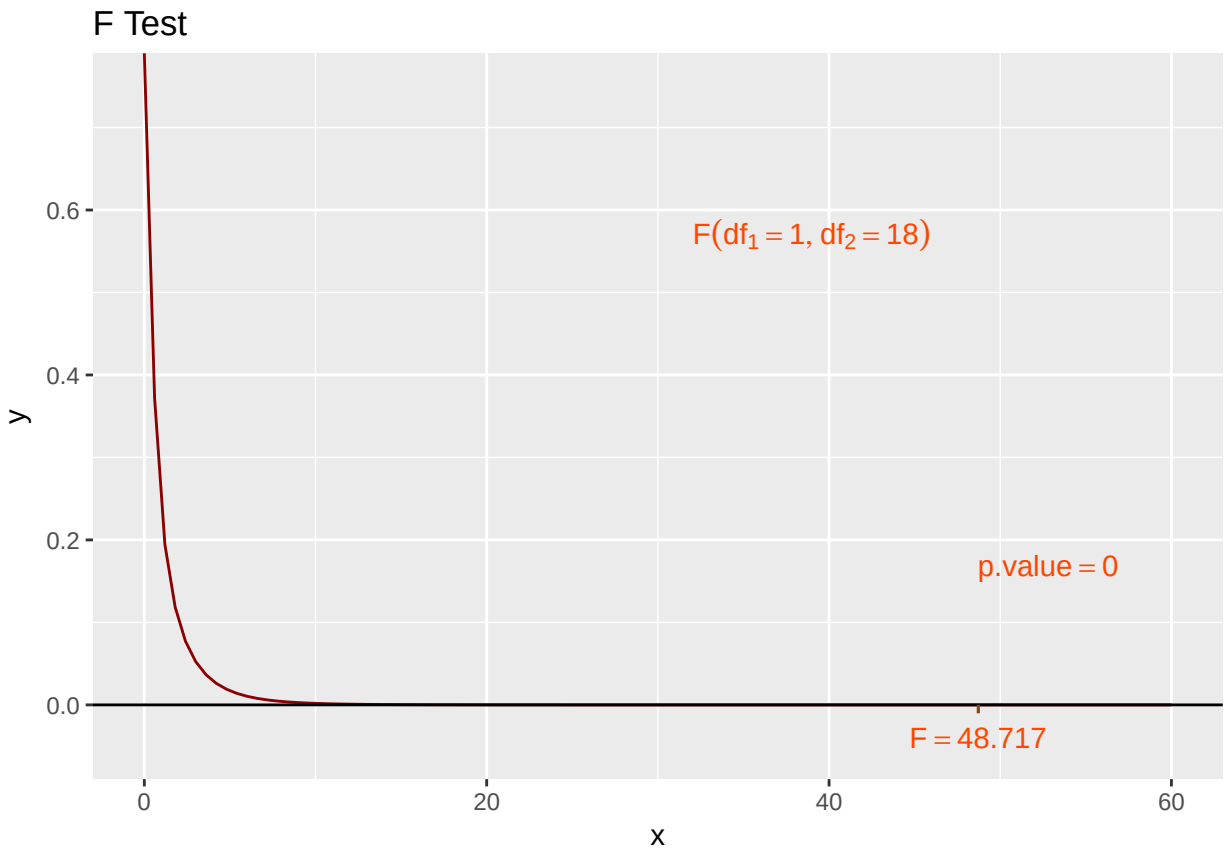


ANOVA F test for model utility.

```
anova(production.lm)
```

```
## Analysis of Variance Table
##
## Response: RunTime
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## RunSize    1 12868.4 12868.4  48.717 1.615e-06
## Residuals 18  4754.6   264.1
```

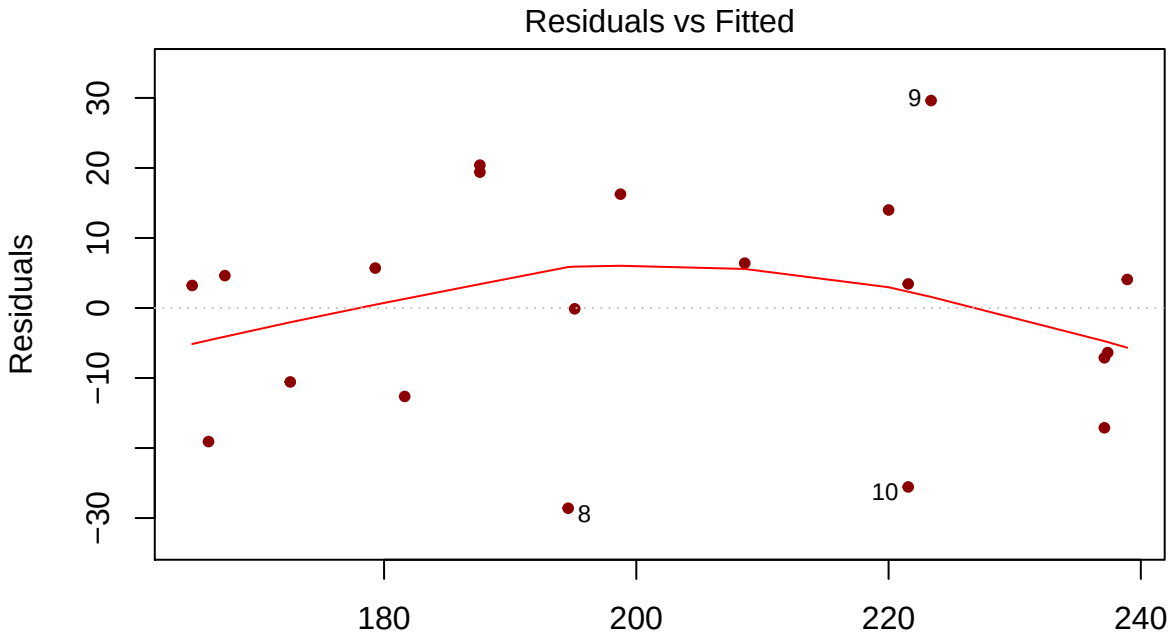
```
n <- 20   #  n observations
g <- 2    #  g parameters: beta0 and beta1

x.max <- 60
y.max <- 0.75
f.val <- 48.717
f.df1 <- g - 1
f.df2 <- n - g
f.p.value <- round(1 - pf(f.val, df1=f.df1, df2=f.df2), 3)
title <- "F Test"

gg.draw.F(x.max, y.max, f.val, f.df1, f.df2, f.p.value, title)
```

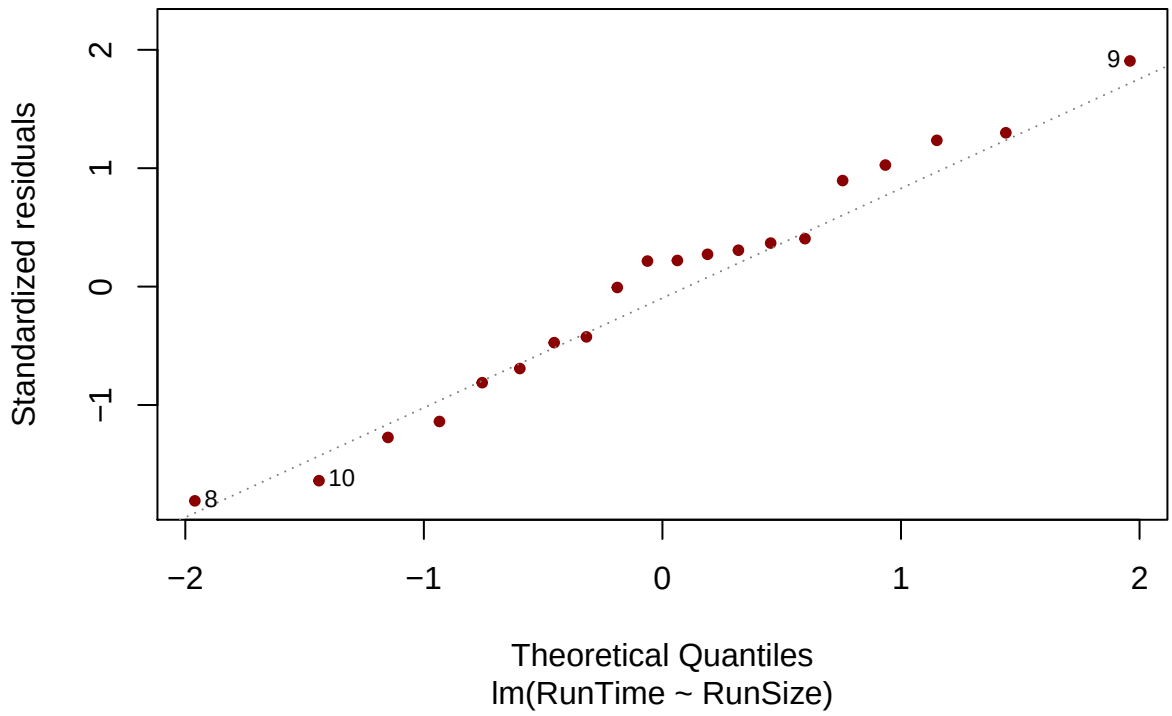


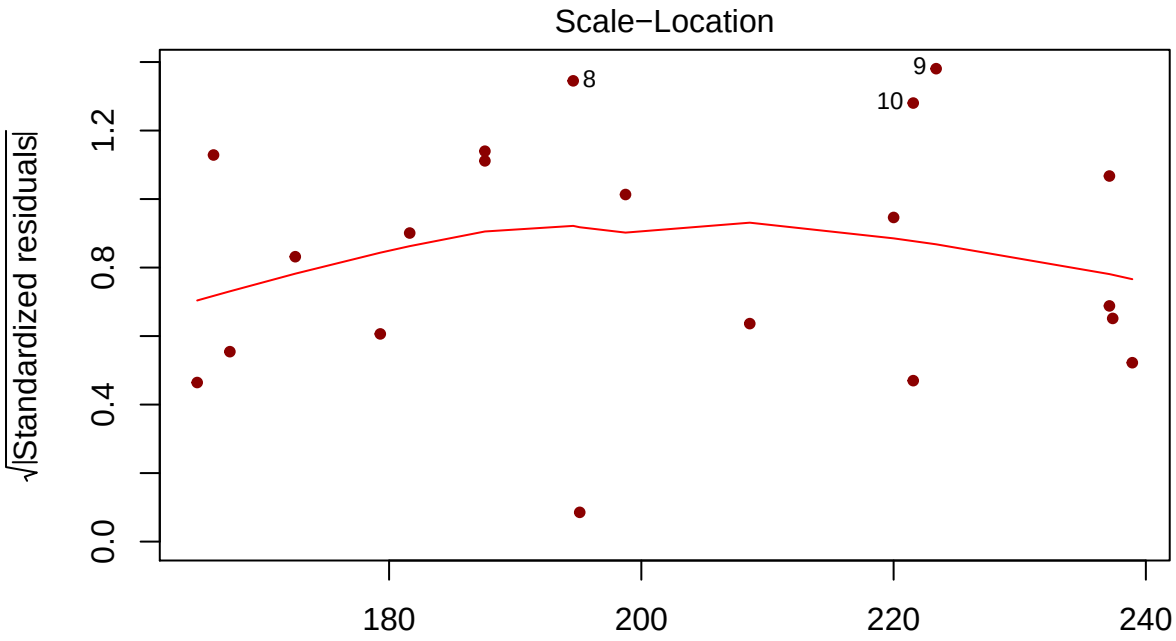
regression diagnostics

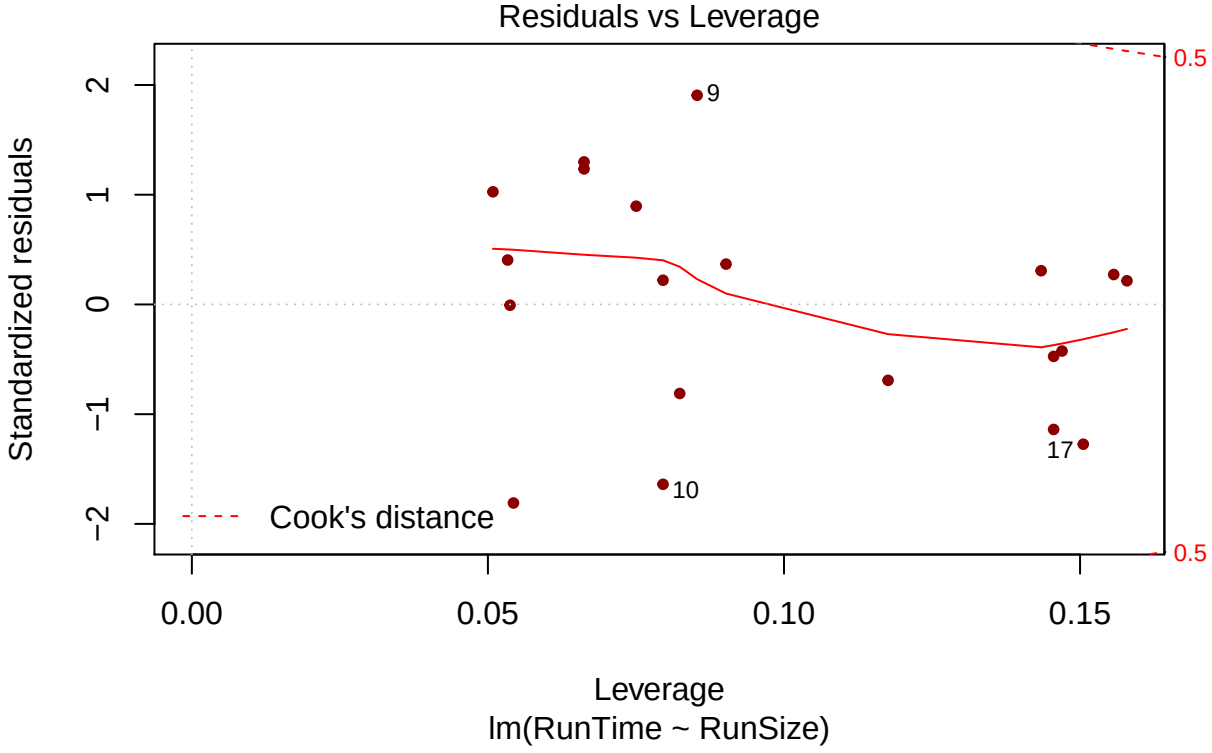
reference:
- Sheather, A Modern Approach to Regression with R, chapter 3, p.70

```
plot(production.lm, pch=20, col="darkred")
```






Scale−Location
√|Standardized residuals|
0.0
0.4
0.8
1.2
180
200
220
240
8
9
10
Fitted values
lm(RunTime ~ RunSize)


Residuals vs Leverage
Standardized residuals
−2
−1
0
1
2
0.00
0.05
0.10
0.15
0.5
0.5
9
10
17
Cook's distance
Leverage
lm(RunTime ~ RunSize)