



hw5_yw3204

Yuhao Wang

10/27/2018

1.

Assume $X \sim MVN(\mu, \Sigma)$.

Then, we have $\mathbb{E}((X - \mu)(X - \mu)^T) = \Sigma$.

For AX , we know its mean is $\mathbb{E}(AX) = A\mu$.

By definition,

$$\begin{aligned}
& Cov(AX) \\
&= \mathbb{E}((AX - A\mu)(AX - A\mu)^T) \\
&= A * \mathbb{E}((X - \mu)(X - \mu)^T) * A^T \\
&= A\Sigma A^T
\end{aligned}$$

2.

Denote the response vector as Y and the original design matrix as $X = (1, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)$.

Assume X 's column vectors are linearly independent. Then by Gram-Schmidt orthogonalization, we can turn them into a set of orthogonal vectors with length 1.

Denote the new design matrix as $\tilde{X} = (\tilde{\mathbf{x}}_0, \tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_p)$.

Then, we rewrite the model as $y = \gamma_0 * \tilde{x}_0 + \gamma_1 * \tilde{x}_1 + \dots + \gamma_p * \tilde{x}_p$.

The hypothesis is thus equivalent to

$$H_0 : \gamma_1 = 0 \text{ and } H_1 : \gamma_1 \neq 0.$$

The estimator of γ is

$$\hat{\gamma} = \tilde{X}^T Y \text{ or } \hat{\gamma}_i = \tilde{x}_i^T * Y.$$

Due to the orthogonal property of \tilde{X}^T , the covariance matrix of $\hat{\gamma}$ is $\sigma^2 I_{p+1}$, which means $\hat{\gamma}_i$ are mutually independent.

The estimator of σ^2 is

$$\frac{(Y - \tilde{X}\tilde{X}^T Y)^T (Y - \tilde{X}\tilde{X}^T Y)}{n - p - 1}, \text{ which has a chi-square distribution with } df = n - p - 1.$$

We can prove it by expanding the column space of \tilde{X} . But here, we just take it for granted.

The t-statistic is thus

$$t = \frac{\hat{\gamma}_1}{\sqrt{\hat{\sigma}^2}}$$

And the F-statistic is

$$F = \frac{SSE(\tilde{x}_0, \tilde{x}_2, \dots, \tilde{x}_p) - SSE(\tilde{x}_0, \tilde{x}_1, \tilde{x}_p)}{SSE(\tilde{x}_0, \tilde{x}_1, \tilde{x}_p) / (n - p - 1)}$$

The denominator is exactly $\hat{\sigma}^2$.

For the numerator, it can be written as

$$\begin{aligned}
& SSE(\tilde{x}_0, \tilde{x}_2, \dots, \tilde{x}_p) - SSE(\tilde{x}_0, \tilde{x}_1, \dots, \tilde{x}_p) \\
&= z^T z - (z - \hat{\gamma}_1 * \tilde{\mathbf{x}}_1)^T (z - \hat{\gamma}_1 * \tilde{\mathbf{x}}_1), \text{ in which } z = Y - \sum_{i \neq 1} \hat{\gamma}_i \tilde{\mathbf{x}}_i. \\
&= 2 * \hat{\gamma}_1 * z^T * \tilde{\mathbf{x}}_1 - \hat{\gamma}_1^2 \\
&= 2 * \hat{\gamma}_1 * Y^T * \tilde{\mathbf{x}}_1 - \hat{\gamma}_1^2 \text{ (because of the orthogonality)} \\
&= 2 * \hat{\gamma}_1 * \hat{\gamma}_1 - \hat{\gamma}_1^2 \\
&= \hat{\gamma}_1^2
\end{aligned}$$

Hence, the F-statistic can be written as

$$F = \frac{\hat{\gamma}_1^2}{\sigma^2}$$

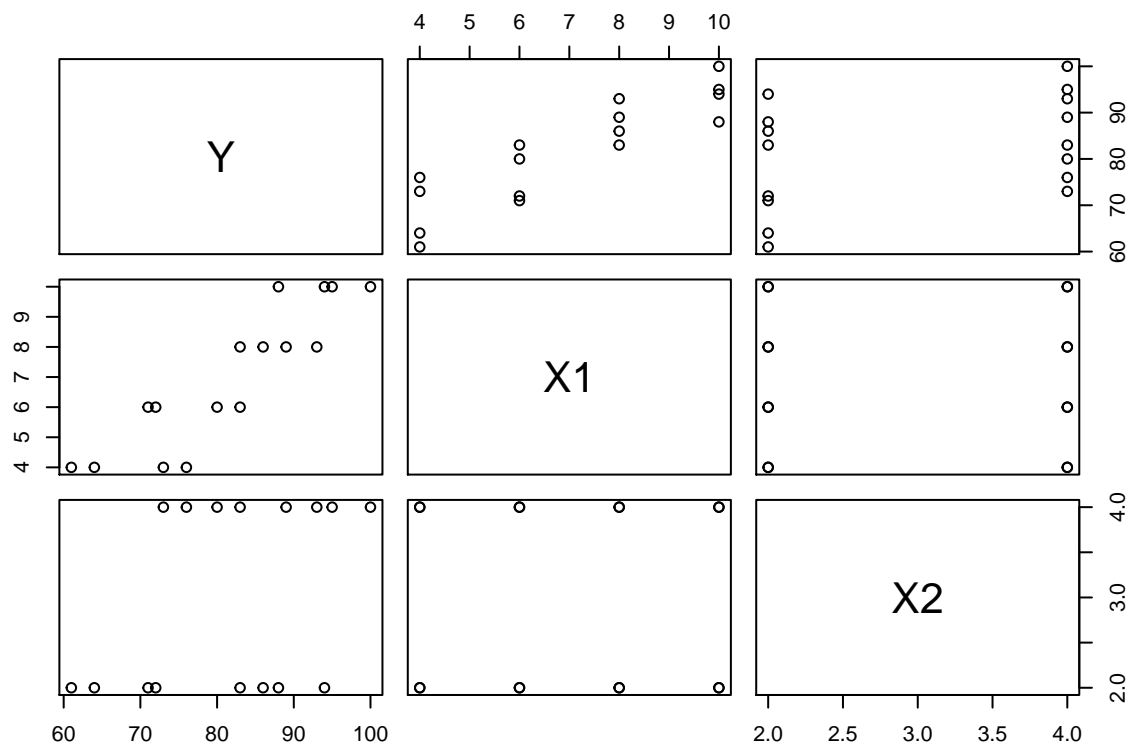
It is also the square of the t-statistic. Thus, we proved the equivalence between t-test and F-test.

6.5

a.

```
brand <- read.table("CH06PR05.txt")
names(brand) <- c("Y", "X1", "X2")

pairs(brand)
```



```
cor(brand)
```

```
##           Y           X1           X2
## Y  1.000000  0.8923929  0.3945807
## X1  0.8923929  1.0000000  0.0000000
## X2  0.3945807  0.0000000  1.0000000
```

Observe the scatter plot matrix and the correlation matrix, we find there is a strong linear relationship between Y and X_1 . And X_2 can be regarded as a categorical variable.

b.

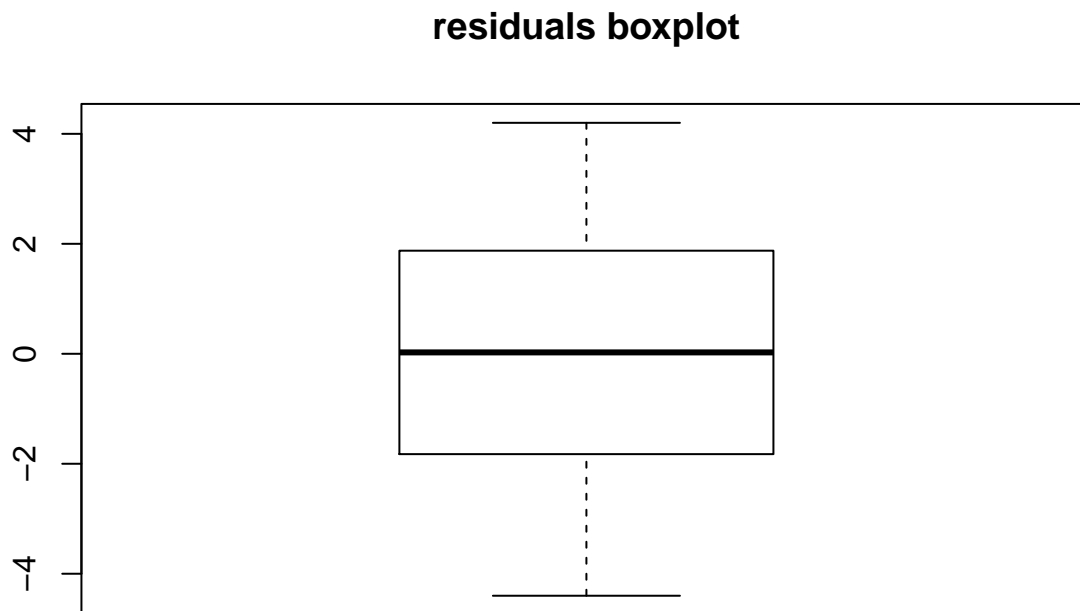
```
lm1 <- lm(Y ~ X1+X2, brand)
lm1

##
## Call:
## lm(formula = Y ~ X1 + X2, data = brand)
##
## Coefficients:
## (Intercept)          X1          X2
##      37.650       4.425       4.375
```

The fitted line is $Y = 37.650 + 4.425X_1 + 4.375 * X_2$. Here, b_1 means holding other variables fixed, the response Y changes 4.425 units when X_1 changes 1 unit.

c.

```
boxplot(resid(lm1), main = "residuals boxplot")
```



The plot seems quite symmetric with mean around 0 which matches with the equal variance and zero mean assumption in the model.

6.7

a.

```
summary(lm1)$r.squared
```

```
## [1] 0.952059
```

The coefficient of multiple determination is 0.95 here which is incredibly high.

b.

```
Y <- brand$Y
Y_hat <- predict(lm1, brand[, c(2, 3)])
summary(lm(Y~Y_hat))$r.squared
```

```
## [1] 0.952059
```

Yes, they are equal.

6.8

a.

```
predict(lm1, newdata = data.frame(X1 = 5, X2 = 4), interval = "confidence", level = 0.99)
```

```
##      fit      lwr      upr
## 1 77.275 73.88111 80.66889
```

The interval is [73.88, 80.67] and it means the true mean at $X_1 = 5$ and $X_2 = 4$ falls into it with probability 99%.

b.

```
predict(lm1, newdata = data.frame(X1 = 5, X2 = 4), interval = "prediction", level = 0.99)
```

```
##      fit      lwr      upr
## 1 77.275 68.48077 86.06923
```

The interval is [68.48, 86.07] and it means the new observation at $X_1 = 5$ and $X_2 = 4$ falls into it with probability 99%.

6.25

We can subtract $\beta_2 * X_{i2}$ from Y_i . That's to say, we regress $(Y_i - \beta_2 * X_{i2})$ on the remaining variables.