Chapter 1 Homework

1678226 Chanmi Yoo (유찬미)

Regression Analysis
20657-02
Jae Keun Yoo
Sep 23, 2019

Q1.2

Let’s denote Y = yearly cost for memership, X = the number of visits.

Then, the relation between X and Y is:

Y = 300 + 2X, which is statistical relation.

Q1.7

(a)

The regression model would be Y = 20X + 100, so the point estimate for X=5 is:

```
100 + 20*5
```

```
## [1] 200
```

However, in order to state the probability, we should know the form of distribution of the error terms (and hence of the Yi). So, with these limited information, we cannot solve (a) question.

(b)

In (b) question, the normality assumption for the error terms is given. Thus, we can now state the exact probability. With the given information, the point estimate for X=5 is 200 and sigma is 5. Denote the probability as Z. Then we should find Z that satisfies the equation 200 - Z*5 = 195 and 200 + Z*5 = 205. So the Z is

```
(200 - 195) / 5  # Z-score
```

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

```
(1 - pnorm(1)) *2  # probability
```

```
## [1] 0.3173105
```

Thus, the probability would be 0.32.

Q1.10

Yes, it would increase until 47 and then decrease.

Q1.18

True. When the regression line is fitted, the sum of residuals is 0 and the sum of error term would be 0.

Q1.21

(a)

```
CH0121 <- read.table("Data Sets/Chapter  1 Data Sets/CH01PR21.txt"); names(CH0121) <- c("Y", "X"); attach(CH0121)

x.bar <- mean(X)
y.bar <- mean(Y)
sxy <- sum((X-x.bar)*(Y-y.bar))
sxx <- sum((X-x.bar)^2)
b1 <- sxy/sxx; print(b1)
```

```
## [1] 4
```

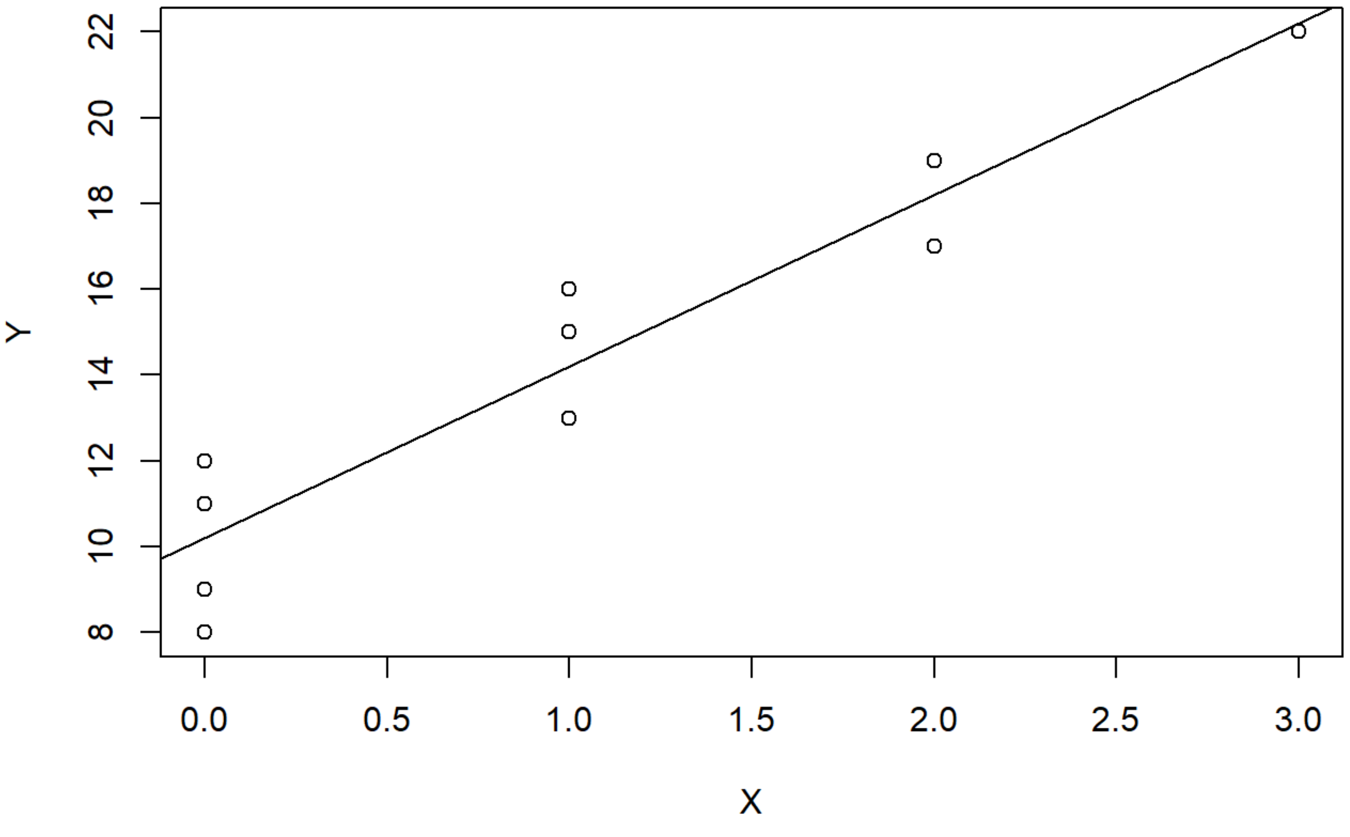
```
b0 <- y.bar - b1*x.bar; print(b0)
```

```
## [1] 10.2
```

Thus, the estimated regression function is Y = 10.2 + 4.0*X.

The plot of the estimated regression function is like this:

```
plot(X, Y) # scatter plot
abline(b0, b1) # adding the regression line
```



The linear regression function appears to give a good fit, since the regression line has little distance from the scattering of points. The regression line explains the relationship between the number of aircraft-shipment transfer (X) and the number of broken amplues (Y) well.

(b)

The point estimate of the expected number of broken ampules when X = 1 is:

```
10.2 + 4.0*1
```

```
## [1] 14.2
```

(c)

The slope is:

```
b1
```

```
## [1] 4
```

Therefore, broken ampules are expected to increase by 4, as the transfer increases by 1.

(d)

```
10.2 + 4.0*x.bar == y.bar
```

```
## [1] TRUE
```

Since y.bar equals to the estimated regression function of x.bar, the point (x.bar, y.bar) goes through the fitted regression line.

Q1.25

(a)

```
fitted <- b0 + b1*X
resid <- Y - fitted
resid[1]  # the residual of the first case
```

```
## [1] 1.8
```

The difference between error and residual is that error is yield from the true regression line (which is unknown) but residuals are calculated from the estimated line (which is made from observed values). So we do not know errors, but know residuals. As the estimated line goes close to the true regression line, the residual becomes better estimate of the error.

(b)

is:

```
SSE <- sum(resid^2); print(SSE)  # the sum of squared residuals
```

```
## [1] 17.6
```

```
MSE <- SSE / (nrow(CH0121) - 2); print(MSE)  # MSE
```

```
## [1] 2.2
```

The sqaure root of MSE is the point estimate of sigma, the standard deviation of the probability distribution of Y for any X. In this case, 1.48 transfers is the standard deviation of this distribution.