hw_06

LeoFontenot

12/9/2021

1

a

```
x = c(110.5, 105.4, 118.1, 104.5, 93.6, 84.1, 77.8, 75.6)
y = c(5.755, 5.939, 6.010, 6.545, 6.730, 6.750, 6.899, 7.862)
my_lm = lm(y ~ x)
summary(my_lm)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.34626 -0.27605 -0.09448  0.27023  0.53495
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.137455   0.842265  12.036    2e-05 ***
## x           -0.037175   0.008653  -4.296  0.00512 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3624 on 6 degrees of freedom
## Multiple R-squared:  0.7547, Adjusted R-squared:  0.7138
## F-statistic: 18.46 on 1 and 6 DF,  p-value: 0.005116
```

The estimate of B1 is 10.1374553, -0.0371747

b

```
anova(my_lm)
```

```
## Analysis of Variance Table
##
```

```
## Response: y
##           Df  Sum Sq Mean Sq F value   Pr(>F)
## x          1 2.42357 2.42357  18.455 0.005116 **
## Residuals  6 0.78794 0.13132
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(my_lm)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.34626 -0.27605 -0.09448  0.27023  0.53495
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.137455   0.842265  12.036    2e-05 ***
## x           -0.037175   0.008653  -4.296  0.00512 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3624 on 6 degrees of freedom
## Multiple R-squared:  0.7547, Adjusted R-squared:  0.7138
## F-statistic: 18.46 on 1 and 6 DF,  p-value: 0.005116
```

Since the F statistic is larger than the critical value, we can reject the null hyphothesis

c

d

```
resid(my_lm)
```

```
##          1          2          3          4          5          6          7
## -0.2746519 -0.2802428  0.2628757  0.2922999  0.0720958 -0.2610638 -0.3462643
##          8
##  0.5349514
```

Fitted regression line is Y= 10.137455x -0.037175

e

```
summary(my_lm)$sigma^2
```

```
## [1] 0.1313228
```

f

```
predict(my_lm, newdata = data.frame(x=100), interval = "confidence")
```

```
##        fit      lwr      upr
## 1 6.419986 6.096321 6.743651
```

g

```
predict(my_lm, newdata = data.frame(x=100), interval = "prediction")
```

```
##        fit      lwr      upr
## 1 6.419986 5.476038 7.363934
```

The prediction interval is wider than the confidence interval

h

```
x= summary(my_lm)
x$r.squared
```

```
## [1] 0.7546518
```

The coefficient of determination is the proportion of the variation in the dependent variable that is predictable from the independent variables. It provides a measure of how well observations are replicated by the model. This coefficient is 0.75 meaning that 75% of the time, predictions match observations.

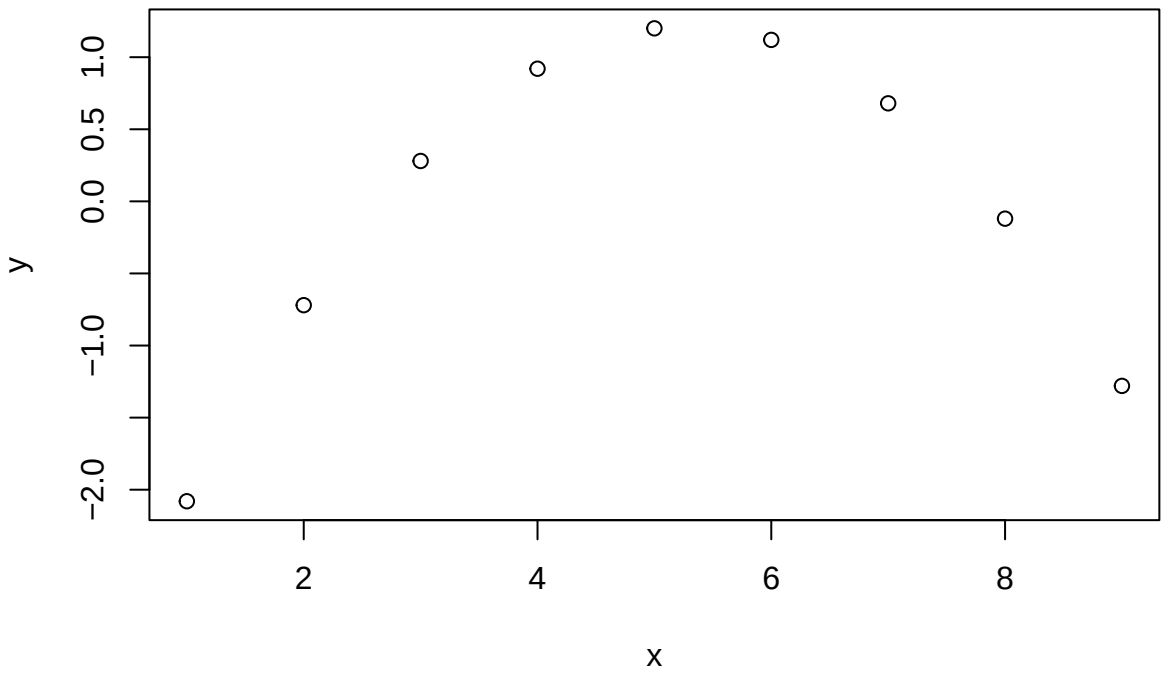
2

```
x = c(1, 2, 3, 4, 5, 6, 7, 8, 9)
y = c(-2.08, -0.72, 0.28, 0.92, 1.20, 1.12, 0.68, -0.12, -1.28)
my_lm2 = lm(y~x)
```

a

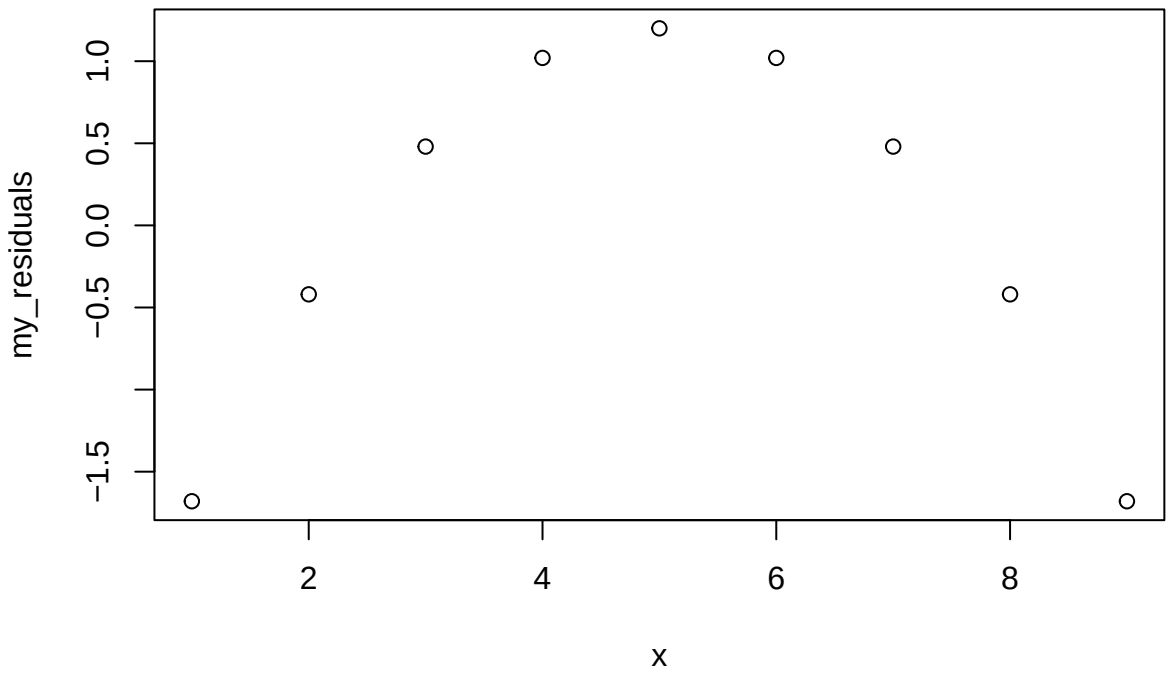
Plot of y ~ x is below

```
plot(x, y)
```



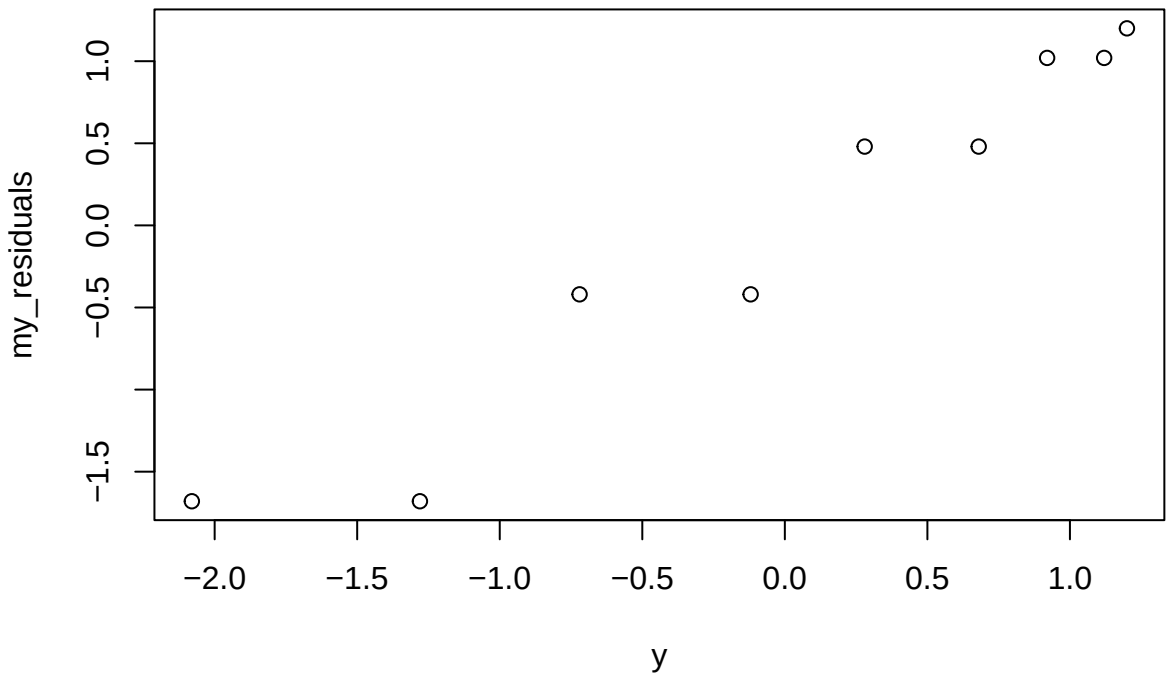
b

```
my_residuals= resid(my_lm2)
plot(x = x, y = my_residuals)
```



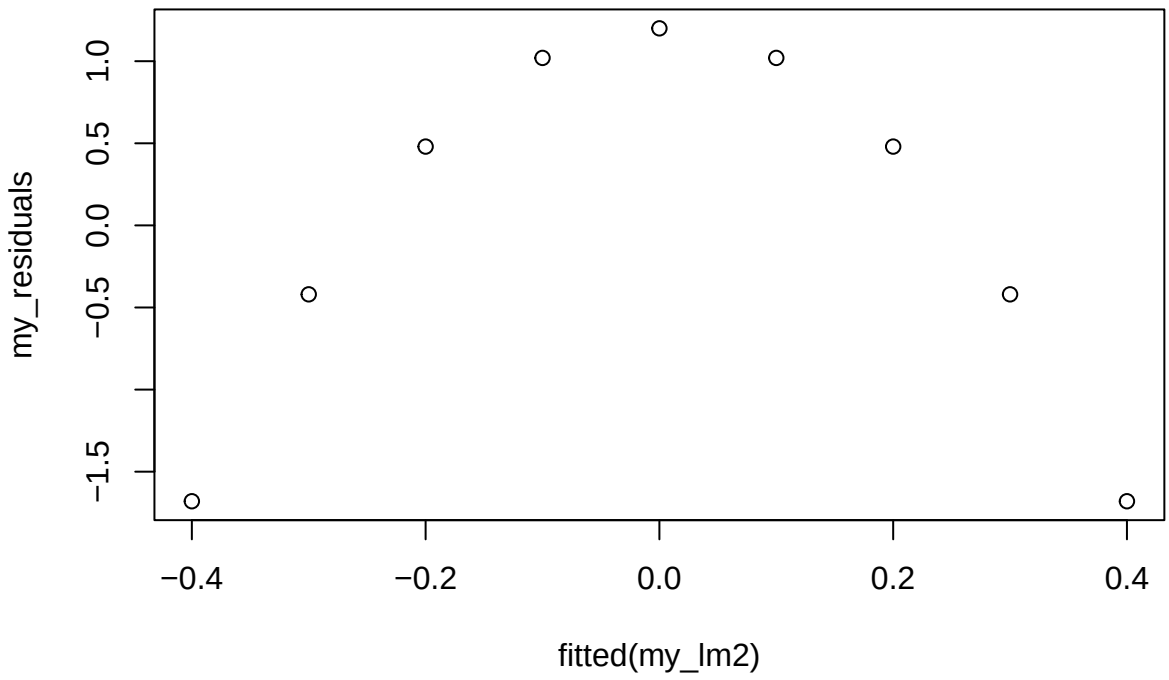
c

```
my_residuals= resid(my_lm2)
plot(x = y, y = my_residuals)
```



d

```
plot(x = fitted(my_lm2), y = my_residuals)
```



e

The plot against y and y-hat are different.When the residuals are plotted against y does not show that the linear regression model is not good.It is better to plot against x or y-hat.