How well the line fit the data?

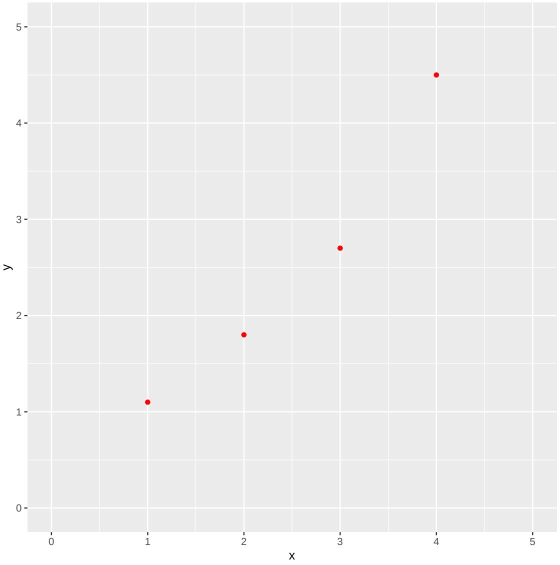
- Not every data can be fitted by a linear model
- We can measure a goodness of the fit in three ways
- Coefficient of Determination or R^2
- F-test
- t-test

Coefficient of Determination

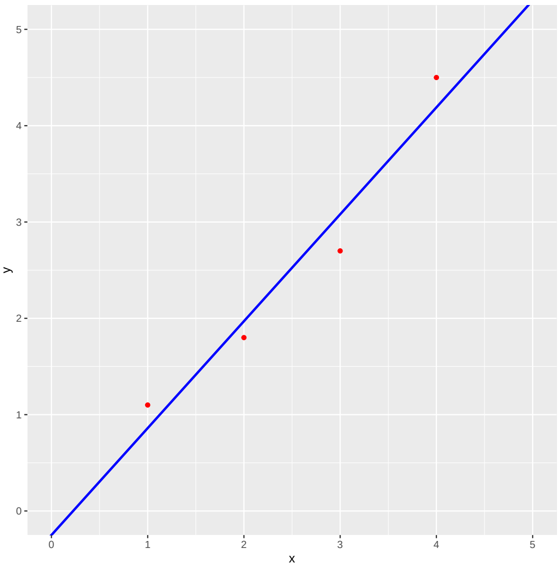
$$R^2 = \frac{RegSS}{TSS} = 1 - \frac{RSS}{TSS}$$

- RSS: Residual Sums Squares
- TSS: Total Sums Squares

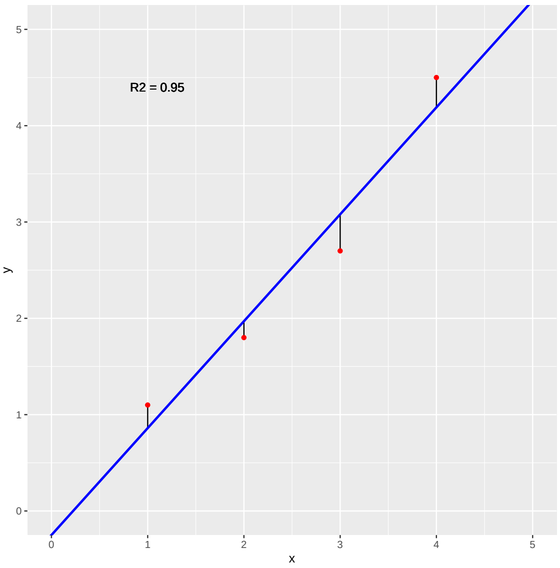
- If $RSS = 0$, $R^2 = 1$. This is a perfect fit.
- If $RSS = TSS$, $R^2 = 0$ means. This is the lowest R^2 could be.
- The closer R^2 to 1, the better the fit.

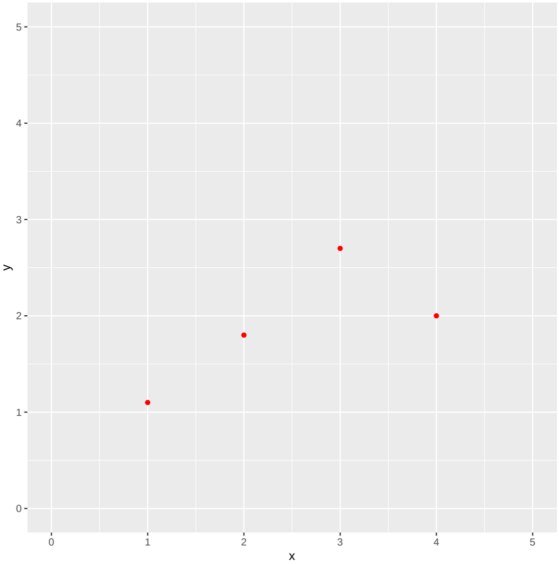


y
5
4
3
2
1
0
0
1
2
3
4
5
x

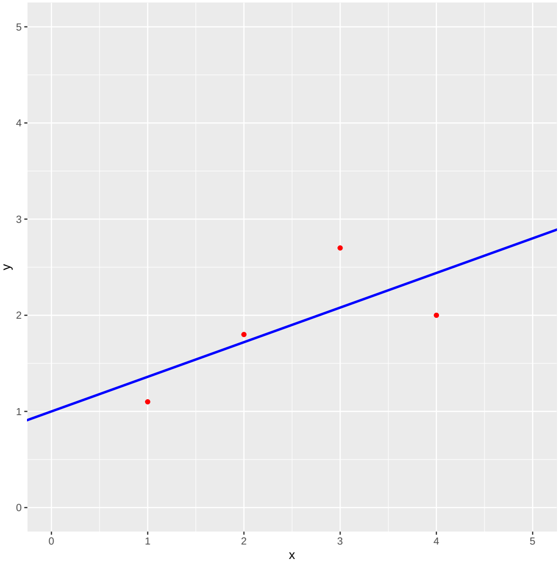


5
4
3
2
1
0
y
0
1
2
3
4
5
x

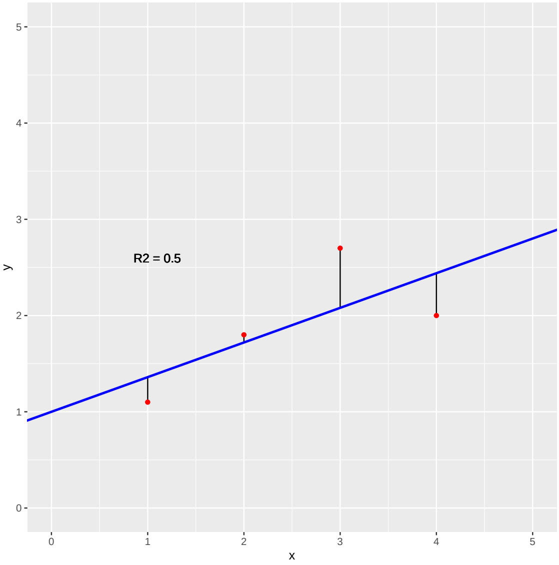

R2 = 0.95
y
x
0
1
2
3
4
5
0
1
2
3
4
5



5
4
3
2
1
0
y
0
1
2
3
4
5
x



5
4
3
2
1
0
y
0
1
2
3
4
5
x


R2 = 0.5
y
x
0
1
2
3
4
5
0
1
2
3
4
5

- Some examples of R^2

F-test

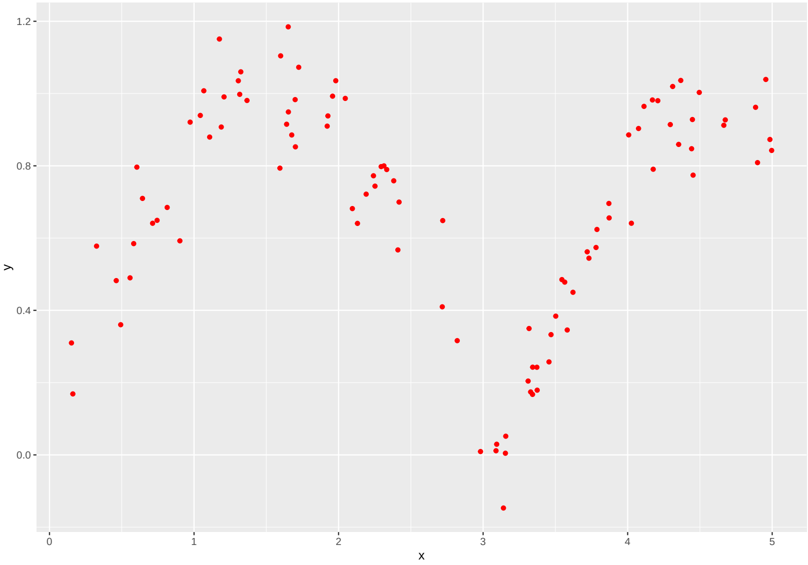
$H_0 : \beta_1 = 0$
$H_0 : y = \beta_0 + \epsilon$
H_0 : The linear model is not a good fit

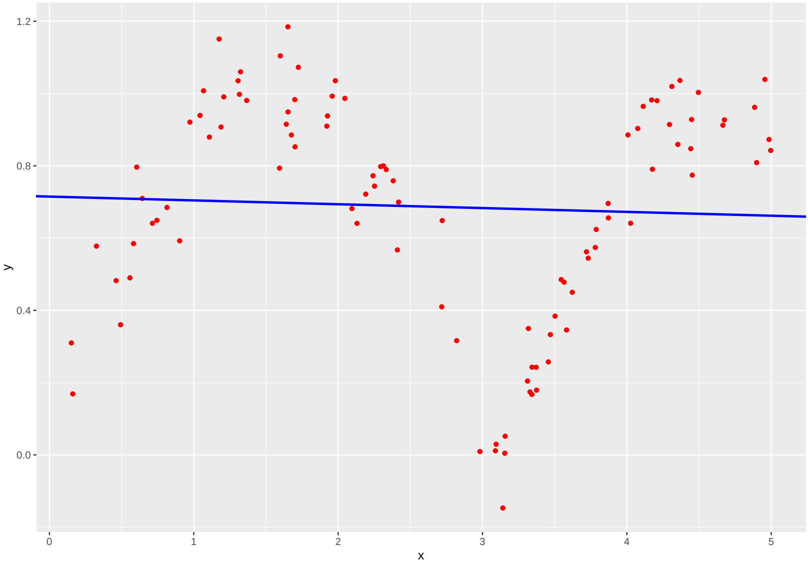
$H_\alpha : \beta_1 \neq 0$
$H_\alpha : y = \beta_0 + \beta_1 x + \epsilon$
H_α : The linear model is a good fit

$$F = \frac{RegSS}{RSS/(n-2)}$$

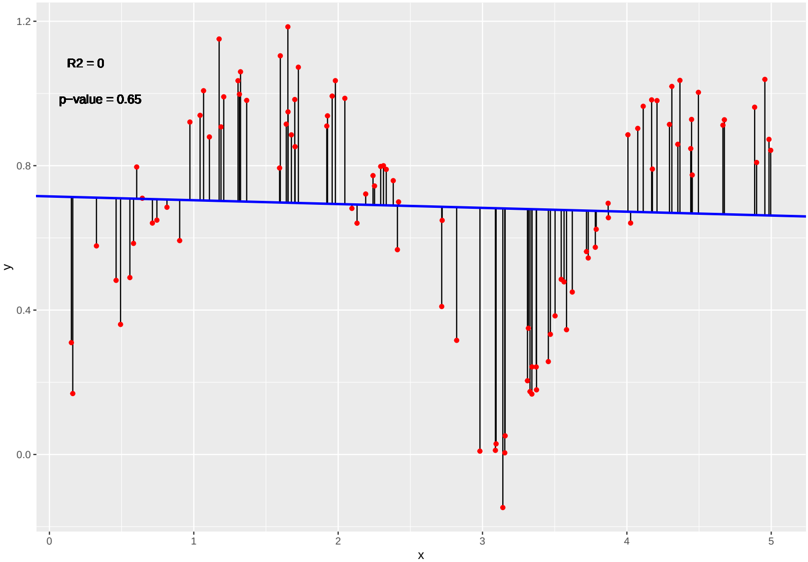
These are equivalent

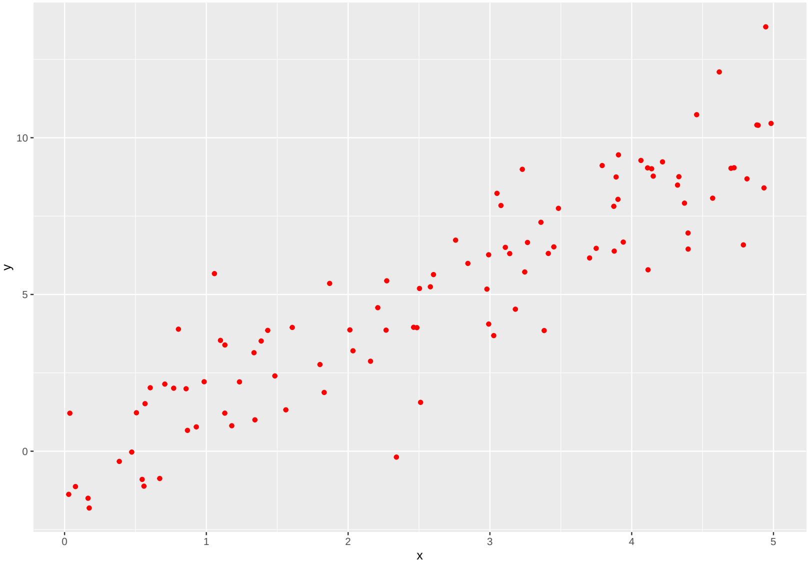
- Larger F
- Larger Reg SS
- Smaller Residual SS (RSS)
- Smaller p-value
- More support for H_α or More useful the model


1.2
0.8
0.4
0.0
y
0
1
2
3
4
5
x

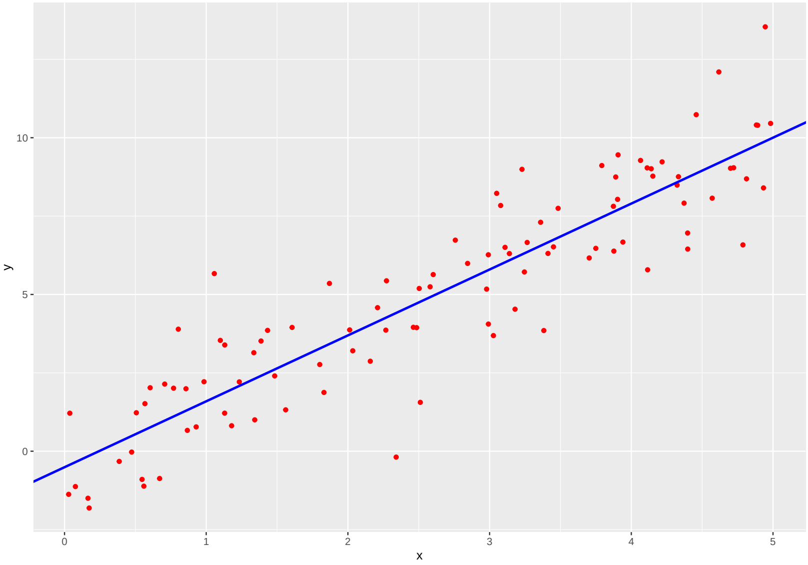


1.2
0.8
0.4
0.0
y
0
1
2
3
4
5
x

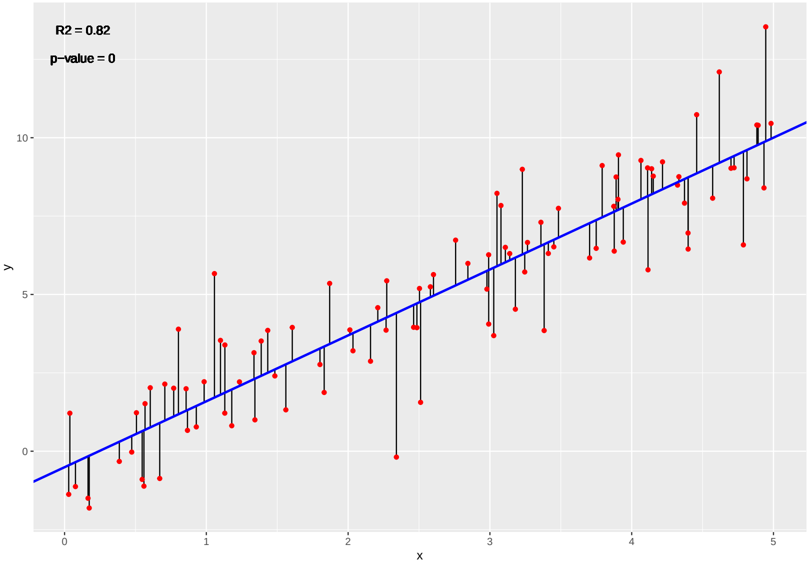

R2 = 0
p−value = 0.65
1.2
0.8
0.4
0.0
y
0
1
2
3
4
5
x



10
5
0
y
0
1
2
3
4
5
x



10
5
0
y
0
1
2
3
4
5
x


R2 = 0.82
p−value = 0
10
5
0
y
0
1
2
3
4
5
x

t-test

$H_0 : \beta_1 = 0$
$H_0 : y = \beta_0 + \epsilon$
H_0 : The linear model is not a good fit

$H_\alpha : \beta_1 \neq 0$
$H_\alpha : y = \beta_0 + \beta_1 x + \epsilon$
H_α : The linear model is a good fit

$$t = \frac{\hat{\beta}_1}{sd(\hat{\beta}_1)}$$

Relation between F-test, t-test and and R2

- We have

$$F = t^2 = \frac{(n-2)R^2}{(1-R^2)}$$

$$\implies \frac{1}{F} = \left(\frac{1}{R^2} - 1\right) \cdot \frac{1}{n-2}$$

- This means when $R^2 \nearrow$, $\frac{1}{R^2} \searrow$, $\frac{1}{F} \searrow$ and $F \nearrow$.
- The p-values of F-test and t-test are the same.