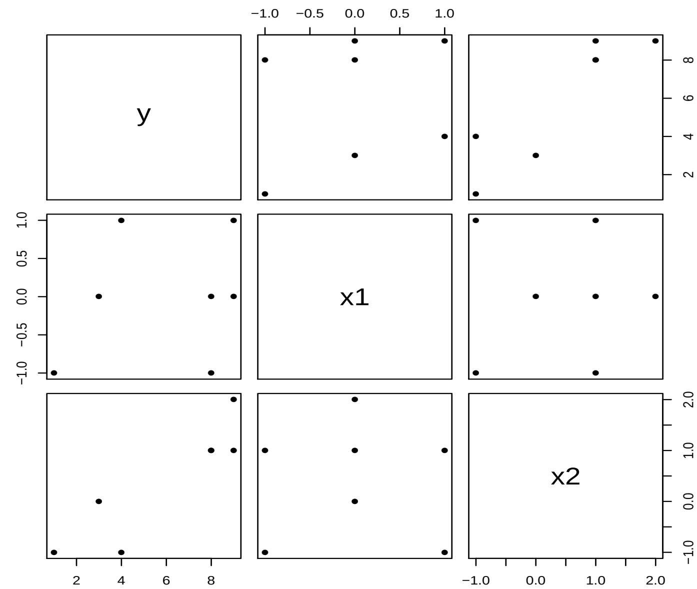

- Fit a linear regression model with terms x_1, x_2 and x_1^2.
- Find the residual sum of squares of the fitted model.
- Test for the significance of the regression.
- Calculate the R^2 and the R^2_{adj} for this model.
- Assess the importance of adding x_1^2 to a model that already contains x_1 and x_2.

FM: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \varepsilon$

```
> f1=lm(y~x1+x2+I(x1^2));
> model.matrix(f1)

  (Intercept) x1 x2 I(x1^2)
1           1 -1 -1       1
2           1  1 -1       1
3           1 -1  1       1
4           1  1  1       1
5           1  0  0       0
6           1  0  1       0
7           1  0  2       0
attr(,"assign")
[1] 0 1 2 3
```

```
Call:
lm(formula = y ~ x1 + x2 + I(x1^2))

Residuals:
      1       2       3       4       5       6       7
-0.5000  0.5000  0.5000 -0.5000 -0.6667  1.3333 -0.6667

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.6667     0.7817   4.690  0.01832
x1            1.0000     0.5528   1.809  0.16815
x2            3.0000     0.4513   6.647  0.00694
I(x1^2)       1.8333     0.9574   1.915  0.15140

Residual standard error: 1.106 on 3 degrees of freedom
Multiple R-squared: 0.9427,        Adjusted R-squared: 0.8854
F-statistic: 16.45 on 3 and 3 DF,  p-value: 0.02288

Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value   Pr(>F)
x1         1  4.000   4.000  3.2727 0.168147
x2         1 51.852  51.852 42.4242 0.007351
I(x1^2)    1  4.481   4.481  3.6667 0.151401
Residuals  3  3.667   1.222
```

- Test $H_0 : \beta_{11} = 0$
- Test $H_0 : \beta_{11} = 0, \beta_1 = \beta_2 = \beta$

$H_0 : \beta_{11} = 0$ **RM:** $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$

```
Call:
lm(formula = y ~ x1 + x2)

Residuals:
      1       2       3       4       5       6       7
-0.2963  0.7037  1.5185  0.5185 -1.8889  0.5185 -1.0741

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)   4.8889     0.5827   8.390  0.00110
x1            1.0000     0.7136   1.401  0.23374
x2            2.5926     0.5139   5.045  0.00726

Residual standard error: 1.427 on 4 degrees of freedom
Multiple R-squared: 0.8727,       Adjusted R-squared: 0.809
F-statistic: 13.71 on 2 and 4 DF,  p-value: 0.01621

Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value   Pr(>F)
x1         1  4.000   4.000  1.9636 0.233742
x2         1 51.852  51.852 25.4545 0.007255
Residuals  4  8.148   2.037
```

$H_0 : \beta_{11} = 0$ **RM:** $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$

The partial F-test:

SS due to the hypothesis. SSres.f2-SSres.f1=8.1481-3.6667=4.4815

$\frac{\text{SSres.f2-SSres.f1}}{\text{df.f2-df.f1}} = \frac{8.1481-3.6667}{4-3} = 4.4815/1 = 4.4815$

The appropriate test statistic for H_0 is

$$\frac{(\text{SSres.f2-SSres.f1})/(\text{df.f2-df.f1})}{\text{SSres.f1/df.f1}} = \frac{4.4815}{1.2222} = 3.6667$$

Since $F(0.05, 1, 3) = 10.13$, we *do not reject* H_0 which implies $\beta_{11} = 0$, a more plausible model would be $E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$.

$H_0 : \beta_{11} = 0$ **RM:** $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$

```
> anova(f2,f1)

Analysis of Variance Table

Model 1: y ~ x1 + x2
Model 2: y ~ x1 + x2 + I(x1^2)
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1      4 8.1481
2      3 3.6667  1    4.4815 3.6667 0.1514
```

$H_0 : \beta_{11} = 0, \beta_1 = \beta_2 = \beta$ **RM:** $y = \beta_0 + \beta(x_1 + x_2) + \varepsilon$

```
> z=x1+x2
> f4=lm(y~z)
> summary(f4)
```

```
Call:
lm(formula = y ~ z)

Residuals:
       1        2        3        4        5        6        7 
-0.02439 -1.12195  2.87805 -0.21951 -2.12195  0.82927 -0.21951 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.1220     0.6857   7.470 0.000679
z             2.0488     0.5032   4.072 0.009618

Residual standard error: 1.722 on 5 degrees of freedom
Multiple R-squared: 0.7683,       Adjusted R-squared: 0.722
F-statistic: 16.58 on 1 and 5 DF,  p-value: 0.009618

> anova(f4)

Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value   Pr(>F)
z          1 49.171  49.171  16.579 0.009618
Residuals  5 14.829   2.966
```

$H_0 : \beta_{11} = 0, \beta_1 = \beta_2 = \beta$ **RM:** $y = \beta_0 + \beta(x_1 + x_2) + \varepsilon$

The partial F-test:

SS due to the hypothesis. SSres.f4-SSres.f1=14.8293-3.6667=11.1626

$\frac{\text{SSres.f4-SSres.f1}}{\text{df.f4-df.f1}} = \frac{14.8293-3.6667}{5-3} = 11.1626/2 = 5.5813$

The appropriate test statistic for H_0 is

$$\frac{(\text{SSres.f4-SSres.f1})/(\text{df.f4-df.f1})}{\text{SSres.f1/df.f1}} = \frac{5.5813}{1.2222} = 4.5665$$

Since $F(0.05, 2, 3) = 9.55$, we *do not reject* H_0 which implies $\beta_{11} = 0, \beta_1 = \beta_2 = \beta$, a more plausible model would be $E(y) = \beta_0 + \beta(x_1 + x_2)$.

```
> anova(f4, f1)

Analysis of Variance Table

Model 1: y ~ z
Model 2: y ~ x1 + x2 + I(x1^2)
```

```
  Res.Df     RSS Df Sum of Sq      F Pr(>F)
1      5 14.8293
2      3  3.6667  2   11.1626 4.5665 0.1229
```

```
> anova(f4, f2)
```

```
Analysis of Variance Table

Model 1: y ~ z
Model 2: y ~ x1 + x2
  Res.Df     RSS Df Sum of Sq      F Pr(>F)
1      5 14.8293
2      4  8.1481  1    6.6811 3.2798 0.1444
```