Linear Models: R Examples — The full rank model: inference

Model adequacy

Why give up a good running example? We continue with the clover model. We first test $H_0 : \boldsymbol{\beta} = \mathbf{0}$.

```
> SSReg <- t(y) %*% y - SSRes
> SSReg

         [,1]
[1,] 358.9289

> Fstat <- (SSReg/p)/(SSRes/(n - p))
> Fstat

         [,1]
[1,] 551.1428

> pf(Fstat, p, n - p, lower.tail = FALSE)

             [,1]
[1,] 5.361324e-78
```

The R way

```
> basemodel <- lm(area ~ 0, data = clover)
> anova(basemodel, model)

Analysis of Variance Table

Model 1: area ~ 0
Model 2: area ~ midrib + estim
  Res.Df    RSS  Df Sum of Sq      F    Pr(>F)
1    145 389.75
2    142  30.83   3    358.93 551.14 < 2.2e-16 ***
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ
```

$H_0 : \beta_0 = 0$

```
> X2 <- X[, -1]
> Rg2 <- t(y) %*% X2 %*% inv(t(X2) %*% X2) %*% t(X2) %*% y
> Rg2

         [,1]
[1,] 357.5173

> Rg1g2 <- SSReg - Rg2
> Rg1g2

         [,1]
[1,] 1.411603

> r <- 1
```

$H_0 : \beta_0 = 0$

```
> Fstat <- (Rg1g2/r)/(SSRes/(n - p))
> Fstat
         [,1]
[1,] 6.502639
> pf(Fstat, r, n - p, lower.tail = FALSE)
           [,1]
[1,] 0.01183194
```

$H_0 : \beta_0 = 0$ (R)

```
> basemodel <- lm(area ~ 0 + midrib + estim, data = clover)
> anova(basemodel, model)

Analysis of Variance Table

Model 1: area ~ 0 + midrib + estim
Model 2: area ~ midrib + estim
  Res.Df    RSS  Df Sum of Sq      F  Pr(>F)
1    143 32.237
2    142 30.826   1     1.412 6.5026 0.01183 *
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ
```

$H_0 : \beta_1 = 0$

```
> X2 <- X[, -2]
> Rg2 <- t(y) %*% X2 %*% inv(t(X2) %*% X2) %*% t(X2) %*% y
> Rg2

         [,1]
[1,] 357.7500

> Rg1g2 <- SSReg - Rg2
> Rg1g2

         [,1]
[1,] 1.178859

> r <- 1
```

$H_0 : \beta_1 = 0$

```
> Fstat <- (Rg1g2/r)/(SSRes/(n - p))
> Fstat
         [,1]
[1,] 5.43049
> pf(Fstat, r, n - p, lower.tail = FALSE)
           [,1]
[1,] 0.02119664
```

$H_0 : \beta_1 = 0$ (R)

```
> basemodel <- lm(area ~ estim, data = clover)
> anova(basemodel, model)

Analysis of Variance Table

Model 1: area ~ estim
Model 2: area ~ midrib + estim
  Res.Df    RSS  Df Sum of Sq      F  Pr(>F)
1    143 32.004
2    142 30.826   1     1.179 5.4305 0.02120 *
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ
```

$H_0 : \beta_2 = 0$

```
> X2 <- X[, -3]
> Rg2 <- t(y) %*% X2 %*% inv(t(X2) %*% X2) %*% t(X2) %*% y
> Rg2
         [,1]
[1,] 350.2031
> Rg1g2 <- SSReg - Rg2
> Rg1g2
         [,1]
[1,] 8.725833
> r <- 1
```

$H_0 : \beta_2 = 0$

```
> Fstat <- (Rg1g2/r)/(SSRes/(n - p))
> Fstat
         [,1]
[1,] 40.1961
> pf(Fstat, r, n - p, lower.tail = FALSE)
             [,1]
[1,] 2.86636e-09
```

$H_0 : \beta_2 = 0$ (R)

```
> basemodel <- lm(area ~ midrib, data = clover)
> anova(basemodel, model)

Analysis of Variance Table

Model 1: area ~ midrib
Model 2: area ~ midrib + estim
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    143 39.551
2    142 30.826  1     8.726 40.196 2.866e-09 ***
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ
```

```
> summary(model)

Call:
lm(formula = area ~ midrib + estim, data = clover)

Residuals:
     Min       1Q   Median       3Q      Max
-2.31730 -0.07022  0.08005  0.18787  1.14160

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.1741     0.4604   -2.55   0.0118 *
midrib        0.5240     0.2248    2.33   0.0212 *
estim         0.7338     0.1157    6.34 2.87e-09 ***
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ.âĂŹ 0.

Residual standard error: 0.4659 on 142 degrees of freedom
Multiple R-squared: 0.7078, Adjusted R-squared: 0.7036
F-statistic:   172 on 2 and 142 DF,  p-value: < 2.2e-16
```

Corrected sum of squares — $H_0 : \beta_1 = \beta_2 = 0$

```
> X2 <- X[, 1]
> Rg2 <- t(y) %*% X2 %*% inv(t(X2) %*% X2) %*% t(X2) %*% y
> Rg2

         [,1]
[1,] 284.2732

> Rg1g2 <- SSReg - Rg2
> Rg1g2

         [,1]
[1,] 74.65567

> r <- 2
```

$H_0 : \beta_1 = \beta_2 = 0$

```
> Fstat <- (Rg1g2/r)/(SSRes/(n - p))
> Fstat
        [,1]
[1,] 171.953
> pf(Fstat, r, n - p, lower.tail = FALSE)
             [,1]
[1,] 1.167508e-38
```

$H_0 : \beta_1 = \beta_2 = 0$ (R)

```
> basemodel <- lm(area ~ 1, data = clover)
> anova(basemodel, model)

Analysis of Variance Table

Model 1: area ~ 1
Model 2: area ~ midrib + estim
  Res.Df     RSS Df Sum of Sq      F    Pr(>F)
1    144 105.481
2    142  30.826  2    74.656 171.95 < 2.2e-16 ***
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.05 âĂŸ
```

Partial vs. sequential tests

All the previous tests were conducted with the assumption that all other variables are included in the model (i.e. partial tests). We now try some sequential tests. We test in the order $\beta_0 \rightarrow \beta_1 \rightarrow \beta_2$.

```
> X0 <- X[, 1]
> X1 <- X[, c(1, 2)]
> R0 <- t(y) %*% X0 %*% inv(t(X0) %*% X0) %*% t(X0) %*% y
> R0

         [,1]
[1,] 284.2732

> R1 <- t(y) %*% X1 %*% inv(t(X1) %*% X1) %*% t(X1) %*% y
> R1

         [,1]
[1,] 350.2031
```

$\beta_0 \longrightarrow \beta_1 \longrightarrow \beta_2$

```
> R10 <- R1 - R0
> R10
         [,1]
[1,] 65.92984
> R21 <- SSReg - R1
> R21
         [,1]
[1,] 8.725833
> R0 + R10 + R21 - SSReg
     [,1]
[1,]    0
```

$\beta_0 \longrightarrow \beta_1 \longrightarrow \beta_2$

```
> R0/(SSRes/(n - p))
         [,1]
[1,] 1309.522
> R10/(SSRes/(n - p))
         [,1]
[1,] 303.7099
> R21/(SSRes/(n - p))
        [,1]
[1,] 40.1961
> qf(0.95, 1, n - p)
[1] 3.907782
```

$\beta_0 \longrightarrow \beta_1 \longrightarrow \beta_2$ (R)

```
> bm1 <- lm(area ~ 0, data = clover)
> bm2 <- lm(area ~ 1, data = clover)
> bm3 <- lm(area ~ midrib, data = clover)
```

$\beta_0 \to \beta_1 \to \beta_2$ (R)

```
> anova(bm1, bm2, bm3, model)
Analysis of Variance Table

Model 1: area ~ 0
Model 2: area ~ 1
Model 3: area ~ midrib
Model 4: area ~ midrib + estim
  Res.Df    RSS  Df Sum of Sq        F    Pr(>F)
1    145 389.75
2    144 105.48   1    284.27 1309.522 < 2.2e-16 ***
3    143  39.55   1     65.93  303.710 < 2.2e-16 ***
4    142  30.83   1      8.73   40.196 2.866e-09 ***
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.0
```

$H_0 : (\beta_0, \beta_1, \beta_2) = (-1, 0.5, 1)$

```
> bst <- as.vector(c(-1, 0.5, 1))
> Fstat <- ((t(b - bst) %*% t(X) %*% X %*% (b - bst))/p)/(SSRes/
+     p))
> Fstat

         [,1]
[1,] 69.17173

> pf(Fstat, p, n - p, lower.tail = FALSE)

             [,1]
[1,] 1.246212e-27
```

$H_0 : (\beta_0, \beta_1, \beta_2) = (-1, 0.5, 1)$ (R)

```
> h0 <- X %*% bst
> basemodel <- lm(area ~ 0, data = clover, offset = h0)
> anova(basemodel, model)

Analysis of Variance Table

Model 1: area ~ 0
Model 2: area ~ midrib + estim
  Res.Df    RSS  Df Sum of Sq      F    Pr(>F)
1    145 75.873
2    142 30.826   3    45.048 69.172 < 2.2e-16 ***
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.0
```

$H_0 : (\beta_0, \beta_1, \beta_2) = (-1.1, 0.5, 0.7)$

```
> bst <- as.vector(c(-1.1, 0.5, 0.7))
> Fstat <- ((t(b - bst) %*% t(X) %*% X %*% (b - bst))/p)/(SSRes/
+     p))
> Fstat

          [,1]
[1,] 0.6726326

> pf(Fstat, p, n - p, lower.tail = FALSE)

          [,1]
[1,] 0.5701846
```

$H_0 : (\beta_0, \beta_1, \beta_2) = (-1.1, 0.5, 0.7)$ (R)

```
> h0 <- X %*% bst
> basemodel <- lm(area ~ 0, data = clover, offset = h0)
> anova(basemodel, model)

Analysis of Variance Table

Model 1: area ~ 0
Model 2: area ~ midrib + estim
  Res.Df    RSS  Df Sum of Sq      F Pr(>F)
1    145 31.264
2    142 30.826   3     0.438 0.6726 0.5702
```

$H_0 : \beta_0 = -1, \beta_1 = \beta_2$

```
> C <- matrix(c(1, 0, 0, 1, 0, -1), 2, 3)
> C

     [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    1   -1

> dst <- as.vector(c(-1, 0))
> Fstat <- (t(C %*% b - dst) %*% inv(C %*% inv(t(X) %*% X) %*%
+     t(C)) %*% (C %*% b - dst)/r)/(SSRes/(n - p))
> Fstat

         [,1]
[1,] 4.185408

> pf(Fstat, 2, n - p, lower.tail = FALSE)

           [,1]
[1,] 0.01713404
```

$H_0 : \beta_0 = -1, \beta_1 = \beta_2$ (R)

```
> linear.hypothesis(model, C, dst)
Linear hypothesis test

Hypothesis:
(Intercept) = -1
midrib - estim = 0

Model 1: area ~ midrib + estim
Model 2: restricted model

  Res.Df    RSS  Df Sum of Sq      F  Pr(>F)
1    142 30.826
2    144 32.643  -2    -1.817 4.1854 0.01713 *
---
Signif. codes:  0 âĂŸ***âĂŹ 0.001 âĂŸ**âĂŹ 0.01 âĂŸ*âĂŹ 0.0
```