

Teste F-parcial

INGREDIENTES

A hipótese nula, H_0 , define o modelo restrito. A hipótese alternativa, H_a : H_0 é falsa, define o modelo irrestrito.

- SQR_r : soma de quadrado dos resíduos associada à estimação dos parâmetros do modelo restrito (modelo definido sob H_0);
- SQR_{ir} : soma de quadrado dos resíduos associada ao modelo irrestrito;
- g : número de restrições a serem testadas, sob H_0 ;

O teste F-parcial para H_0 é definido por

$$F = \frac{\frac{SQR_r - SQR_{ir}}{g}}{\frac{SQR_{ir}}{n-p}}$$

onde p é o número de coeficientes na regressão irrestrita (incluindo, potencialmente, o intercepto/constante).

DISTRIBUIÇÃO F

Sob a hipótese nula e, ainda, admitindo a validade das suposições MLR.1 a MLR.6, a estatística F acima segue uma distribuição F-de-Snedecor:

$$F_{g;n-p}$$

com g e $n - p$ graus de liberdade.

O teste F-parcial pode ser utilizado como:

- Forma de verificar a contribuição de uma ou mais variáveis explicativas (regressoras/independentes) como se estas fossem as últimas variáveis que entram no modelo;
- Critério de seleção da melhor equação de regressão.

EXEMPLO `temco.txt`

O gerente de uma empresa terceirizada, responsável pelo recrutamento e seleção de novos funcionários para a empresa TEMCO, acredita que os salários dos funcionários da TEMCO sofrem um acréscimo médio de 700,00 dólares, por ano a mais na empresa, e que a experiência prévia na função não tem impacto no salário, uma vez que a TEMCO mantém uma política de contratar recém-formados e trabalhadores sem experiência, pois prefere fornecer um treinamento customizado aos recém-contratados, *ceteris paribus*.

Para tanto, a análise inferencial deve ser feita a partir da estimação dos parâmetros de um modelo de regressão linear múltipla que apresenta **educ**, **anosemp** e **exprev** como regressores e **salario** como regressando.

Adotando um nível de significância de 5%, a desconfiança do gerente procede ou não.

MODELOS

Modelo irrestrito:

$$\text{salario}_i = \beta_0 + \beta_1 \text{educ}_i + \beta_2 \text{anosemp}_i + \beta_3 \text{expprev}_i + \varepsilon_i$$

Hipóteses de interesse:

$$H_0 : (\beta_2, \beta_3) = (700, 0) \quad \text{versus} \quad H_a : (\beta_2, \beta_3) \neq (700, 0)$$

Modelo restrito:

$$\text{salario}_i = \beta_0 + \beta_1 \text{educ}_i + 700 \text{anosemp}_i + \varepsilon_i$$

ou

$$(\text{salario}_i - 700 \text{anosemp}_i) = \beta_0 + \beta_1 \text{educ}_i + \varepsilon_i$$

```

data = read.table("temco.txt",header=TRUE)

attach(data)

n=nrow(data)

# Modelo irrestrito
reg.irr = lm(SALARIO~EDUC+ANOSEMP+EXPPREV)
summary(reg.irr)
anova(reg.irr)
SQR.irr = sum(reg.irr$res^2)

# Modelo restrito
y = SALARIO-700*ANOSEMP
reg.r = lm(y~EDUC)
summary(reg.r)
anova(reg.r)
SQR.r = sum(reg.r$res^2)

# F test
num   = (SQR.r-SQR.irr)/2
den   = SQR.irr/(n-4)
Ftest = num/den

```


MODELO IRRESTRITO

Call:

```
lm(formula = SALARIO ~ EDUC + ANOSEMP + EXPPREV)
```

Residuals:

Min	1Q	Median	3Q	Max
-17741.8	-1975.2	429.7	2123.3	11959.0

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	23480.46	2027.70	11.580	1.18e-14	***
EDUC	1925.88	384.44	5.010	1.03e-05	***
ANOSEMP	671.33	143.21	4.688	2.92e-05	***
EXPPREV	-73.83	232.78	-0.317	0.753	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5799 on 42 degrees of freedom

Multiple R-squared: 0.7405, Adjusted R-squared: 0.722

F-statistic: 39.96 on 3 and 42 DF, p-value: 2.262e-12

MODELO IRRESTRITO

Analysis of Variance Table

Response: SALARIO

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
EDUC	1	3286777227	3286777227	97.7293	1.579e-12	***
ANOSEMP	1	741570903	741570903	22.0499	2.841e-05	***
EXPPREV	1	3382779	3382779	0.1006	0.7527	
Residuals	42	1412520494	33631440			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

MODELO RESTRITO

Call:

```
lm(formula = y ~ EDUC)
```

Residuals:

Min	1Q	Median	3Q	Max
-17543.0	-1574.8	570.4	2002.9	12011.5

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	23120	1726	13.398	< 2e-16 ***
EDUC	1872	298	6.281	1.3e-07 ***

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 5675 on 44 degrees of freedom

Multiple R-squared: 0.4728, Adjusted R-squared: 0.4608

F-statistic: 39.45 on 1 and 44 DF, p-value: 1.302e-07

MODELO RESTRITO

Analysis of Variance Table

Response: y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
EDUC	1	1270701489	1270701489	39.453	1.302e-07	***
Residuals	44	1417159792	32208177			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Como $g = 2$, $n - p = 46 - 4 = 42$,

$$SQR_{ir} = 1412520494$$

$$SQR_r = 1417159792$$

$$F_{test} = 0.06897264$$

Rejeita-se H_0 se o teste F parcial, isto é, $F = 0.06897264$ for maior que o valor crítico da distribuição $F_{2,42}$ ao nível de significância α .

Se $\alpha = 0.05$, é fácil¹ verificar que $F_{2,42}^{0.05} = 3.219942$.

Se $\alpha = 0.10$, é fácil verificar que $F_{2,42}^{0.10} = 2.433564$.

Como F é menor que $F_{2,42}^{0.05}$ (ou $F_{2,42}^{0.10}$), não temos evidência para rejeitar a hipótese nula ao nível de 5% (ou 10%) de significância.

¹Em R usa-se a função `qf`: `qf(0.95, 2, 42)`

VAMOS FAZER JUNTOS?

O sindicato, ao qual pertencem os funcionários da empresa TEMCO, afirma ao diretor que deve haver um acréscimo médio anual de U\$ 2.700,00 quando aumenta-se conjuntamente 1 ano no tempo de empresa e 1 ano de estudo após o 2o grau, mantendo-se o tempo de experiência prévia fixo. Conclua se a empresa segue a norma com 95% de confiança.

MODELOS

Modelo irrestrito:

$$\text{salario}_i = \beta_0 + \beta_1 \text{educ}_i + \beta_2 \text{anosemp}_i + \beta_3 \text{expprev}_i + \varepsilon_i$$

Hipóteses de interesse:

$$H_0 : \beta_1 + \beta_2 = 2700 \quad \text{versus} \quad H_a : \beta_1 + \beta_2 \neq 2700$$

Modelo restrito:

$$\text{salario}_i = \beta_0 + (2700 - \beta_2) \text{educ}_i + \beta_2 \text{anosemp}_i + \beta_3 \text{expprev}_i + \varepsilon_i$$

ou

$$(\text{salario}_i - 2700 \text{educ}_i) = \beta_0 + \beta_2 (\text{anosemp} - \text{educ}_i) + \beta_3 \text{expprev}_i + \varepsilon_i$$

```
# Modelo irrestrito
reg.irr = lm(SALARIO~EDUC+ANOSEMP+EXPPREV)
summary(reg.irr)
anova(reg.irr)
SQR.irr = sum(reg.irr$res^2)

# Modelo restrito
y = SALARIO-2700*EDUC
x = ANOSEMP-EDUC
reg.r = lm(y~x+EXPPREV)
summary(reg.irr)
anova(reg.irr)
SQR.r = sum(reg.r$res^2)

# F test
num = (SQR.r-SQR.irr)/1
den = SQR.irr/(n-4)
Ftest = num/den
Ftest
[1] 0.1041627
```