F-test exercises

ST552 Winter 2015

January 23, 2015

In a study of cheddar cheese from the LaTrobe Valley of Victoria, Australia, samples of cheese were analyzed for their chemical composition and were subjected to taste tests. Overall taste scores were obtained by combining the scores from several tasters.

`cheddar` is a data frame with 30 observations on the following 4 variables:

taste, a subjective taste score
Acetic, concentration of acetic acid (log scale)
H2S, concentration of hydrogen sulfice (log scale)
Lactic, concentration of lactic acid

The following model:

$$\textbf{Full Model:} \quad \text{taste}_i = \beta_0 + \beta_1 \text{Acetic}_i + \beta_2 \text{H2S}_i + \beta_3 \text{Lactic}_i + \epsilon_i$$

was fit in R and the output is shown below.

```
data(cheddar, package = "faraway")
# full model, aka Model 6
fit <- lm(taste ~ . , data = cheddar)
summary(fit)
```

```
##
## Call:
## lm(formula = taste ~ ., data = cheddar)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -17.39  -6.61  -1.01   4.91  25.45
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -28.877     19.735   -1.46   0.1554
## Acetic         0.328      4.460    0.07   0.9420
## H2S            3.912      1.248    3.13   0.0042 **
## Lactic        19.671      8.629    2.28   0.0311 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.1 on 26 degrees of freedom
## Multiple R-squared:  0.652,  Adjusted R-squared:  0.612
## F-statistic: 16.2 on 3 and 26 DF,  p-value: 3.81e-06
```

#2

#1

These alternative models were also fit:

Model 1:	$\text{taste}_i = \beta_0$	RSS $= 7662.9$
Model 2:	$\text{taste}_i = \beta_0 + \beta_1 \text{Acetic}_i + \epsilon_i$	RSS $= 5348.7$
Model 3:	$\text{taste}_i = \beta_1 \text{H2S}_i + \beta_2 \text{Lactic}_i + \epsilon_i$	RSS $= 3601.8$
Model 4:	$\text{taste}_i = \beta_0 + \beta_1 \text{H2S}_i + \beta_2 \text{Lactic}_i + \epsilon_i$	RSS $= 2669$
Model 5:	$\text{taste}_i = \beta_0 + \beta_1 (\text{Acetic}_i + \text{H2S}_i) + \beta_2 \text{Lactic}_i + \epsilon_i$	RSS $= 2719.9$
Model 6:	$\text{taste}_i = \beta_0 + \beta_1 \text{Acetic}_i + \beta_2 \text{H2S}_i + \beta_3 \text{Lactic}_i + \epsilon_i$	RSS $= 2668.4$

1. Find the overall regression F-statistic. Where is this reported in the R output?
2. Find the F-statistic for testing the null hypothesis that $\beta_1 = 0$ in the full model. What distribution should this statistic be compared to? Identify the equivalent t-test statistic and p-value in the `lm` output.
3. Find the F-statistic for testing the null hypothesis that $\beta_1 = 0$ in **Model 2**. The p-value for this test is 0.0017, why is this conclusion different to the one above?
4. Find the F-statistic for testing the null hypothesis that $\beta_0 = \beta_1 = 0$ in the full model. Can you predict the conclusion from the R output?
5. Find the F-statistic for testing the null hypothesis that ~~$\beta_1 = \beta_2$~~ $\beta_1 = \beta_2$ in the full model. What distribution should this F-statistic be compared to?

(1) Null: $\beta_1 = \beta_2 = \beta_3 = 0$ Model 1

$$\frac{(RSS_{M1} - RSS_{M6})/(4-1)}{RSS_{M6}/(30-4)} = \frac{(7662.9 - 2668.4)/3}{2668.4/26}$$

$$= \frac{1664.83}{102.63}$$

$$= 16.22$$

p-value < 0.001

There is very strong evidence that the mean taste score is associated with the concentration of at least one of acetic acid, lactic acid or hydrogen sulfide.

(2) $\beta_1 = 0$ in $taste_i = \beta_0 + \beta_1 Acetic_i + \beta_2 H2S_i + \beta_3 lactic_i + \varepsilon_i$

Null corresponds to M4

$$F = \frac{(RSS_{M4} - RSS_{M6})/(4-3)}{RSS_{M6}/26}$$

$$= \frac{2669 - 2668.4}{102.63}$$

$$= 0.0058 \quad \text{compare to } F_{1,26}$$

$$t = \sqrt{F} = 0.076 \qquad \text{p-value} = 0.9420$$

from Acetic line in output.

There is no evidence the mean taste score is associated with the concentration of Acetic acid, after accounting for the concentration of lactic acid and hydrogen sulfice.

(3) Models 1 & 2

$$F = \frac{(7662.9 - 5348.7)/1}{5348.7/(30-2)} = \frac{2314.2}{191.03}$$

$$= 12.11$$

This test asks "Is mean taste score associated with Acetic acid?" and ignores the varying lactic & H2S

Considering Acetic alone we see a relationship with taste, but after accounting for the effects of Lactic & H_2S we do not see a relationship.

(4) Model 3 & Model 6

$$F = \frac{(3601.8 - 2668.4)/2}{102.63}$$

$$= \frac{466.7}{102.63} = 4.55 \quad \text{compare to } F_{2,26}$$

In R:

1 - pf(4.55, 2, 26)

= 0.02

No, you cannot predict the result of testing multiple parameters from their individual t-tests!

(5)

$\beta_1 = \beta_2$ i $\quad taste_i = \beta_0 + \beta_1 Acetic + \beta_2 H_2S + \beta_3 Lactic + \varepsilon_i$

$$\Rightarrow taste_i = \beta_0 + \beta_1 Acetic + \underline{\beta_1} H_2S + \beta_3 lactic + \varepsilon_i$$

$$= \beta_0 + \beta_1 (Acetic + H_2S) + \beta_3 lactic + \varepsilon_i$$

Model 5

$$F = \frac{(2719.9 - 2668)/1}{102.63} = 0.51 \quad \text{compare to } F_{1,26}$$