F-test exercises

ST552 Winter 2019

2019-01-25

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:

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

was fit in R and the output is shown below.

```
##
## Call:
## lm(formula = taste ~ ., data = cheddar)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.390  -6.612  -1.009   4.908  25.449
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -28.8768    19.7354  -1.463  0.15540
## Acetic        0.3277     4.4598   0.073  0.94198
## H2S           3.9118     1.2484   3.133  0.00425 **
## Lactic       19.6705     8.6291   2.280  0.03108 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.13 on 26 degrees of freedom
## Multiple R-squared:  0.6518, Adjusted R-squared:  0.6116
## F-statistic: 16.22 on 3 and 26 DF,  p-value: 3.81e-06
```

```r
# M1
fit_mean_only <- lm(taste ~ 1, data = cheddar)
# M2
fit_acetic <- lm(taste ~ Acetic, data = cheddar)
# M3
fit_LH <- lm(taste ~ Lactic + H2S - 1, data = cheddar)
# M4
fit_notacetic <- lm(taste ~ Lactic + H2S, data = cheddar)
# M5
fit_subspace <- lm(taste ~ I(Acetic + H2S) + Lactic, data = cheddar)
```

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?

```r
anova(fit_mean_only, fit)
```

```
## Analysis of Variance Table
##
## Model 1: taste ~ 1
## Model 2: taste ~ Acetic + H2S + Lactic
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     29 7662.9
## 2     26 2668.4  3    4994.5 16.221 3.81e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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.

```r
anova(fit_notacetic, fit)
```

```
## Analysis of Variance Table
##
```

```
## Model 1: taste ~ Lactic + H2S
## Model 2: taste ~ Acetic + H2S + Lactic
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     27 2669.0
## 2     26 2668.4  1   0.55427 0.0054  0.942
```

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?

```
## Analysis of Variance Table
##
## Model 1: taste ~ 1
## Model 2: taste ~ Acetic
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     29 7662.9
## 2     28 5348.7  1    2314.1 12.114 0.001658 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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?

```
anova(fit_LH, fit)
```

```
## Analysis of Variance Table
##
## Model 1: taste ~ Lactic + H2S - 1
## Model 2: taste ~ Acetic + H2S + Lactic
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     28 3601.8
## 2     26 2668.4  2     933.4 4.5474 0.02025 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

5. Find the F-statistic for testing the null hypothesis that $\beta_1 = \beta_2$ in the full model. What distribution should this F-statistic be compared to?

```
anova(fit_subspace, fit)
```

```
## Analysis of Variance Table
##
## Model 1: taste ~ I(Acetic + H2S) + Lactic
## Model 2: taste ~ Acetic + H2S + Lactic
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     27 2719.9
## 2     26 2668.4  1    51.446 0.5013 0.4852
```