# Inference in regression: F-test

ST552 Lecture 8

Charlotte Wickham

January 22, 2016

# HW#1 Q1 - Code

Most people lost $\frac{1}{2}$ - 1 pt for being lazy to check the type of an object.

80% of unexpected R errors I see are people giving a function an object of the wrong type.

str is your friend!

x*5    package lintr

# Exercises

*ST552 Winter 2016*

*January 20, 2016*

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")
fit <- lm(taste ~ . , data = cheddar)
summary(fit)


##
## 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
```

$\hat{\sigma}$

1. Write down the form of the $\beta$ vector and the $\hat{\beta}$ vector.

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} \qquad \hat{\beta} = \begin{pmatrix} -28.8 \\ 0.33 \\ 3.91 \\ 19.7 \end{pmatrix}$$

2. What are the values of $\hat{\sigma}$, $n$, $\hat{\sigma}^2$, $\sum_{i=1}^{n} e_i^2$, $\sigma$?

$n = 30$ $\qquad\qquad$ $\sigma$ = unknown!

$\hat{\sigma}^2 = (10.13)^2$

$\sum_{i=1}^{n} e_i^2 = 26\,(10.13)^2$ $\qquad\qquad$ $\hat{\varepsilon} = e$

↑ residual

$$
\begin{array}{cc}
 & \begin{array}{cccc} \hat{\beta}_0 & \hat{\beta}_1 & \hat{\beta}_2 & \hat{\beta}_3 \end{array} \\
(X^TX)^{-1} = & \begin{pmatrix} 3.80 & -0.76 & 0.09 & -0.07 \\ -0.76 & \boxed{0.19} & \boxed{-0.02} & -0.13 \\ 0.09 & -0.02 & 0.02 & -0.05 \\ -0.07 & -0.13 & -0.05 & 0.73 \end{pmatrix}
\end{array}
$$

3. Verify the reported value for $SE(\hat{\beta}_1)$.

$$SE(\hat{\beta}_1) = \hat{\sigma}\sqrt{(X^TX)^{-1}_{(1,1)}}$$

start conting at zero

$$= 10.13 \times \sqrt{0.19}$$

4. What is the value of $\hat{Cov}(\hat{\beta}_1, \hat{\beta}_2)$?

$$= 10.13^2 \times -0.02$$

# F-test

design matrix

Let $\Omega$ denote a larger model of interest with $p$ parameters $X_{n \times p}$
and $\omega$ a smaller model that represents some simplification of $\Omega$ with $q$ parameters. $X_{0_{n \times q}}$ $q < p$

**Intuition:** If both models "fit" as well as each other, we should prefer the simpler model, $\omega$. If $\Omega$ shows substantially better fit than $\omega$, that suggests the simplification is not justified.

How do we measure fit? What is substantially better fit?

We need $\omega$ to be "nested" within $\Omega$

The colspace$(X_0) \subset$ colspace$(X)$

In practice : - some $\beta$ in $\Omega$ are set to zero $\beta_1 = 0$
- linear constraint on parameters, $\beta_1 = \beta_2$

# F-statistic

"badness of fit"
measure of fit of small model

always smaller than $\mathrm{RSS}_\omega$

$$F = \frac{(\mathrm{RSS}_\omega - \mathrm{RSS}_\Omega)/(p-q)}{\mathrm{RSS}_\Omega/(n-p)}$$

RSS = sum of squared residuals

$\to \hat{\sigma}^2_\Omega$

Null hypothesis: the simplification to $\Omega$ implied by the simpler model, $\omega$.

Under the null hypothesis, the F-statistic has an F-distribution with $p-q$ and $n-p$ degrees of freedom.

Leads to tests of the form: reject $H_0$ for $F > F^{(\alpha)}_{p-q,n-p}$.

Deriving this fact is beyond this class (take Linear Models), but let's try to appreciate its pieces.

$$Z \sim F_{\nu_1 \nu_2}$$

if

$$Z = \frac{X / \nu_1}{Y / \nu_2} \qquad X \sim \chi^2_{\nu_1} \qquad Y \sim \chi^2_{\nu_2}$$

In F-stat: $\quad RSS_{\Omega} \sim \chi^2_{n-p} \qquad E \sum_{i=1}^{n} e_i^2 = n-p$

# Example: Overall regression F-test

The overall regression F-test asks if any predictors are related to the response.

**Full model:** $Y = X\beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I)$
**Reduced model:** $Y = \beta_0 + \epsilon$

**Null hypothesis:** $H_0 : \beta_1 = \beta_2 = \ldots = \beta_{p-1} = 0$ except $\beta_0$
All the parameters (other than the intercept) are zero.

**Alternative hypothesis:** At least one parameter is non-zero.

**Exercise**: question #1 on handout

# Other examples

- More than one parameter
- A subspace of the parameter space

**Exercise:** questions #4 & #5 on handout

# Example: One predictor

**Null hypothesis**: $\beta_j = 0$

Equivalent to the t-test, reject null if

$$|t_j| = \left| \frac{\hat{\beta}_j}{\mathsf{SE}\left(\hat{\beta}_j\right)} \right| > t_{n-p}^{\alpha/2}$$

In fact, in this case, $F = t_j^2$.

**Exercise**: questions #2 & #3 on handout

If there is **evidence against the null hypothesis**:

- The null is not true, or
- the null is true but we got unlucky, or
- the full model isn't true and the F-test is meaningless.

Control this Probability of this with Type I error

If there is **no evidence against the null hypothesis**:

- The null is true, or
- the null is false but we didn't gather enough evidence to reject it, or
- the full model isn't true and the F-test is meaningless.

power

# We can't do F-tests when

- we want to test non-linear hypotheses, e.g. $H_0 : \beta_j \beta_k = 1$ (we might be able to make use of the Delta method, though)
- we want to compare non-nested models (find an example on the handout)
- the models fitted use different data