Tutorial 11 Board Solutions

Adam Shen

April 1, 2019

Question 10.1.10, Page 420

Please remember to go over this proof. The solution can be found in the file `ANOVA and Regression Proofs.pdf` in the Winter 2018 folder. This provides the ground work as to why we can do what we are doing.

Question 10.1.6, Page 419

```
library(knitr)
library(kableExtra)
i1 <- c(20.5, 28.1, 27.8, 27.0, 28.0, 25.2, 25.3, 27.1, 20.5, 31.3)
i2 <- c(26.3, 24.0, 26.2, 20.2, 23.7, 34.0, 17.1, 26.8, 23.7, 24.9)
i3 <- c(29.5, 34.0, 27.5, 29.4, 27.9, 26.2, 29.9, 29.5, 30.0, 35.6)
i4 <- c(36.5, 44.2, 34.1, 30.3, 31.4, 33.1, 34.1, 32.9, 36.3, 25.5)
dt <- rbind(i1,i2,i3,i4)
```

Display the data

```
kable(dt, booktabs=TRUE) %>%
  add_header_above(c(" " = 1, "Fe Amount" = 10)) %>%
  kable_styling(latex_options="striped")
```

	Fe Amount									
i1	20.5	28.1	27.8	27.0	28.0	25.2	25.3	27.1	20.5	31.3
i2	26.3	24.0	26.2	20.2	23.7	34.0	17.1	26.8	23.7	24.9
i3	29.5	34.0	27.5	29.4	27.9	26.2	29.9	29.5	30.0	35.6
i4	36.5	44.2	34.1	30.3	31.4	33.1	34.1	32.9	36.3	25.5

We have four treatment groups and ten observations per group. Then:

- $I = 4$
- $J_i = 10$ for $i = 1, 2, 3, 4$

Compute treatment sums and grand mean

```
sapply(list(i1,i2,i3,i4), sum)
```

```
## [1] 260.8 246.9 299.5 338.4
```

This tells us that:

$$\sum_{j=1}^{10} x_{1j} = 260.8 \quad \sum_{j=1}^{10} x_{2j} = 246.9 \quad \sum_{j=1}^{10} x_{3j} = 299.5 \quad \sum_{j=1}^{10} x_{4j} = 338.4$$

$$
\begin{aligned}
\overline{x}_{\cdot\cdot} &= \frac{\sum_{i=1}^{I}\sum_{j=1}^{J_i} x_{ij}}{\sum_{i=1}^{I} J_i} \\
&= \frac{\sum_{i=1}^{I}\sum_{j=1}^{J} x_{ij}}{IJ} \quad \text{(when } J_i \text{ is constant)} \\
&= \frac{\sum_{i=1}^{4}\left(\sum_{j=1}^{10} x_{ij}\right)}{40} \\
&= \frac{260.8 + 246.9 + 299.5 + 338.4}{40} \\
&= 28.64
\end{aligned}
$$

Compute treatment sample mean and variances

The formulas for the sample means and sample variances are exactly the same (just with different notation). So we can compute them the regular way in `R`.

```r
sapply(list(i1,i2,i3,i4), mean)
```

```
## [1] 26.08 24.69 29.95 33.84
```

This gives us:

$$\overline{x}_{1\cdot} = 26.08 \quad \overline{x}_{2\cdot} = 24.69 \quad \overline{x}_{3\cdot} = 29.95 \quad \overline{x}_{4\cdot} = 33.84$$

Similarly:

```r
sapply(list(i1,i2,i3,i4), var)
```

```
## [1] 11.50178 19.58322  8.14500 23.34044
```

$$s_1^2 = 11.50 \quad s_2^2 = 19.58 \quad s_3^2 = 8.15 \quad s_4^2 = 23.34$$

Compute SSTr

$$
\begin{aligned}
\text{SSTr} &= \sum_{i=1}^{I} J_i(\overline{x}_{i\cdot} - \overline{x}_{\cdot\cdot})^2 \\
&= J\sum_{i=1}^{I}(\overline{x}_{i\cdot} - \overline{x}_{\cdot\cdot})^2 \quad \text{(for } J_i \text{ constant)} \\
&= 10\left((26.08 - 28.64)^2 + (24.69 - 28.64)^2 + (29.95 - 28.64)^2 + (33.84 - 28.64)^2\right) \\
&= 509.122
\end{aligned}
$$

Compute SSE

$$\text{SSE} = \sum_{i=1}^{I}(J_i - 1)s_i^2$$

$$= (J-1)\sum_{i=1}^{I} s_i^2 \quad \text{(for } J_i \text{ constant)}$$

$$= 9\left((11.50 + 19.58 + 8.15 + 23.34)\right)$$

$$= 563.134$$

Compute SST

SST = SSTr + SSE, so:

$$\text{SSTr} = 509.122 + 563.134 = 1072.256$$

Create ANOVA table

Source	df	SS	MS	F-value
Treatment	a	d	h	F
Error	b	e	k	
Total	c	g		

$a = I - 1 = 4 - 1 = 3$

$b = I(J-1) = 4(10-1) = 4 \times 9 = 36$

$c = IJ - 1 = a + b = 3 + 36 = 39$

$d = \text{SSTr} = 509.122$

$e = \text{SSE} = 563.134$

$g = \text{SST} = 1072.256$

$h = \frac{d}{a} = \frac{509.122}{3} = 169.707$

$k = \frac{e}{b} = \frac{563.134}{36} = 15.643$

$F = \frac{h}{k} = \frac{169.707}{15.643} = 10.849$

Filling in the table:

Source	df	SS	MS	F-value
Treatment	3	509.122	169.707	10.849
Error	36	563.134	15.643	
Total	39	1072.256		

Perform the hypothesis test

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$

H_A : At least one pair unequal

This will always be an upper tailed test by the information given in Lecture 7 Slide 8. We reject the null hypothesis if $F > F_{I-1,I(J-1),1-\alpha}$, where $1-\alpha$ is the area to the left. (α was given as 0.01).

$$F_{I-1,I(J-1),1-\alpha} = F_{3,36,0.99}$$

```r
qf(0.99, df1=3, df2=36)
```

```
## [1] 4.377096
```

Since $10.849 > 4.377$, we reject the null hypothesis at the 1% significance level. In other words, the data suggests that the treatment means are not equal for at least one pair. In the context of this question, we conclude that mean iron content is not the same among the four iron formations.

NOTE: Table A.9 in the textbook only gives $F_{3,30,0.01} = 4.51$ and $F_{3,40,0.01} = 4.31$ (0.01 is the area to the right). On an assignment, you should be using `R` to get an exact quantile value. In a test setting, even if you don't have the exact value of $F_{3,36,0.01}$, you could still infer that this value is somewhere between 4.31 and 4.51. Since our observed test value is 10.849, which is greater than both of these bounds, we can still make a conclusion despite not knowing the exact quantile value.

The p-value of this test can be calculated as:

```r
pf(10.849, df1=3, df2=36, lower.tail=FALSE)
```

```
## [1] 3.199146e-05
```

Since this value is less than 0.01, using the p-value method we would also reject the null hypothesis at the 1% significance level. (Remember that we want area to the right of F since this is an upper-tailed test so we need `lower.tail=FALSE`).

Question 10.1.7, Page 419

Fill in the ANOVA table

Source	df	SS	MS	F-value
Mixture	a	d	g	F
Error	b	e	13.929	
Total	c	5664.415		

We are told that there are six different mixtures and 26 measurements per mixture. Then we have:

- $I = 6$
- $J_i = 26$ for $i = 1, 2, ..., 6$.

$a = I - 1 = 5$

$b = I(J-1) = 6(26-1) = 150$

$c = a + b = 5 + 150 = 155$

$\frac{e}{b} = \frac{e}{150} = 13.929 \Longrightarrow e = 13.929 \times 150 = 2089.35$

$d + e = d + 2089.35 = 5664.415 \Longrightarrow d = 5664.415 - 2089.35 = 3575.065$

$g = \frac{d}{a} = \frac{3575.065}{5} = 715.013$

$F = \frac{g}{13.929} = \frac{715.013}{13.929} = 51.33$

Source	df	SS	MS	F-value
Mixture	5	3575.065	715.013	51.33
Error	150	2089.35	13.929	
Total	155	5664.415		

Perform the hypothesis test

$H_0 : \mu_1 = \mu_2 = ... = \mu_6$

H_A : At least one pair unequal

Take $\alpha = 0.05$. We reject the null hypothesis if $F > F_{I-1,I(J-1),1-\alpha}$, where $1 - \alpha$ is the area to the left.

$$F_{I-1,I(J-1),1-\alpha} = F_{5,150,0.95}$$

```
qf(0.95, df1=5, df2=150)
```

```
## [1] 2.274491
```

Since $51.33 > 2.27$, we can reject the null hypothesis at the 5% significance level. We conclude that there is evidence against the assumption of equal mean electrical resistivity among mixtures.

Using the p-value method:

```
pf(51.33, df1=5, df2=150, lower.tail=FALSE)
```

```
## [1] 8.32117e-31
```

Since this value is less than 0.05, we would reject the null hypothesis, as expected.