Math 170S

Lecture Notes Section 9.3 [*†]

One-factor analysis of variance

Instructor: Swee Hong Chan

NOTE: The notes is a summary for materials discussed in the class and is not supposed to substitute the textbook. Please refer back to the textbook when studying for exams; materials that appear in the textbook but do not appear in the lecture notes might still be tested. Please send me an email if you find typos.

1 Motivating example

Example 1. We would like to estimate the likability index of four different instructors, Instructor 1, Instructor 2, Instructor 3, Instructor 4, by checking at their reviews from three students. Let X_i (for $i \in \{1, 2, 3, 4\}$) by the (random) likability index of Instructor i, which is a normal random variable with mean μ_i and variance σ^2 (all instructors share the same variance). The review scores from the three students are given by

	Student A	Student B	Student C
Instructor 1	13	8	9
Instructor 2	15	11	13
Instructor 3	8	12	7
Instructor 4	11	15	10

The hypothesis are:

[*]Version date: Saturday 6$^{\text{th}}$ June, 2020, 14:02.

[†]This notes is based on Hanbaek Lyu's and Liza Rebrova's notes from the previous quarter, and I would like to thank them for their generosity. "*Nanos gigantum humeris insidentes* (I am but a dwarf standing on the shoulders of giants)".

- The null hypotheses H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$;
- The alternative hypotheses H_1: $\mu_1 \neq \mu_2$, or $\mu_1 \neq \mu_3$, or $\mu_1 \neq \mu_4$.

Can we reject H_0 with significance level $\alpha = 0.05$? $\triangle$

2 One-factor analysis of variance

Our problem is of the following form.

- **Object:**
  - $X_1, X_2, \ldots, X_m$ are **independent** random variables with **unknown** mean μ_1, μ_2,... μ_m and **unknown** variance σ^2.
- **Hypotheses:**
  - **Null Hypothesis** H_0: $\mu_1 = \mu_2 = \ldots = \mu_m$.
  - **Alternative Hypothesis** H_1: $\mu_1 \neq \mu_2$, or $\mu_1 \neq \mu_3$, ..., or $\mu_1 \neq \mu_m$.
- **Input:** Significance level α, and n_1 many random samples for X_1, n_2 many random samples for X_2, ..., n_m many random samples for X_m, i.e.,

Samples for X_1	X_{11}	X_{12}	...	X_{1n_1}
Samples for X_2	X_{21}	X_{22}	...	X_{2n_2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Samples for X_m	X_{m1}	X_{m2}	...	X_{mn_m}

- **Methodology:**
  - Compute $n = n_1 + n_2 + \ldots + n_m$;
  - For each $i \in \{1, 2, \ldots, m\}$, compute

$$\overline{\mathrm{X}}_i \;:=\; \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij},$$

and

$$\overline{\mathrm{X}} \;=\; \frac{1}{n} \sum_{i=1}^{m} n_i \, \overline{\mathrm{X}}_i \,.$$

  - Compute SS(TO), SS(T), SS(E).

– Reject H_0 if

$$\frac{\mathrm{SS(T)}/(m-1)}{\mathrm{SS(E)}/(n-m)} \;\geq\; F_\alpha(m-1, n-m),$$

where $F_\alpha(m-1, n-m)$ can be computed from Table VII in the textbook.

Definition 2. The *total sum of squares* is

$$\mathrm{SS(TO)} \;:=\; \sum_{i=1}^{m}\sum_{j=1}^{n_i} \left(X_{ij} \;-\; \overline{\mathrm{X}}\right)^2.$$

The *error sum of squares* is

$$\mathrm{SS(E)} \;:=\; \sum_{i=1}^{m}\sum_{j=1}^{n_i} \left(X_{ij} \;-\; \overline{\mathrm{X}}_i\right)^2.$$

The *between-tratment sum of squares* is

$$\mathrm{SS(T)} \;:=\; \sum_{i=1}^{m} n_i \left(\; \overline{\mathrm{X}}_i \;-\; \overline{\mathrm{X}}\right)^2.$$

△

Note that there are formulas equivalent to Definition 2 that sometimes are simpler to compute:

$$\begin{aligned}
\mathrm{SS(TO)} \;&=\; \left(\sum_{i=1}^{m}\sum_{j=1}^{n_i} X_{ij}^2\right) \;-\; n(\overline{\mathrm{X}})^2;\\
\mathrm{SS(T)} \;&=\; \left(\sum_{i=1}^{m} n_i(\overline{\mathrm{X}}_i)^2\right) \;-\; n\,(\overline{\mathrm{X}})^2;\\
\mathrm{SS(E)} \;&=\; \mathrm{SS(TO)} \;-\; \mathrm{SS(T)}.
\end{aligned}$$

Answer to Example 1. From the sample data, we have

$$\begin{aligned}
n \;&= n_1 + n_2 + n_3 + n_4 \;=\; 12;\\
\overline{\mathrm{X}}_1 \;&= \frac{13+8+9}{3} \;=\; 10;\\
\overline{\mathrm{X}}_2 \;&= \frac{15+11+13}{3} \;=\; 13;\\
\overline{\mathrm{X}}_3 \;&= \frac{8+12+7}{3} \;=\; 9;\\
\overline{\mathrm{X}}_4 \;&= \frac{11+15+10}{3} \;=\; 12;\\
\overline{\mathrm{X}} \;&= \frac{(3)(10)+(3)(13)+(3)(9)+(3)(12)}{12} \;=\; 11.
\end{aligned}$$

which gives us

$$\begin{aligned}\mathrm{SS(TO)} &= (13-11)^2 + (8-11)^2 + (9-11)^2 + (15-11)^2 + (11-11)^2 + (13-11)^2 \\ &\quad + (8-11)^2 + (12-11)^2 + (7-11)^2 + (11-11)^2 + (15-11)^2 + (10-11)^2 \\ &= 80,\end{aligned}$$

and

$$\mathrm{SS(T)} = (3)(10-11)^2 + (3)(13-11)^2 + (3)(9-11)^2 + (3)(12-11)^2 = 30,$$

and

$$\mathrm{SS(E)} = \mathrm{SS(TO)} - \mathrm{SS(T)} = 80 - 30 = 50.$$

This gives us

$$\frac{\mathrm{SS(T)}/(m-1)}{\mathrm{SS(E)}/(n-m)} = \frac{30/3}{50/8} = 1.6.$$

On the other hand, the value for $F_\alpha(m-1, n-m)$ is

$$F_\alpha(m-1, n-m) = F_{0.05}(3, 8) = 4.07.$$

Since the former is smaller than the latter, we conclude that the test is inconclusive. □

Remark 3. The notation $\overline{\mathrm{X}}_i$ here is written as $\overline{\mathrm{X}}_{i\cdot}$ in the textbook. The notation $\overline{\mathrm{X}}$ here is written as $\overline{\mathrm{X}}_{\cdot\cdot}$ in the textbook. △

Remark 4. We can drop the assumption that $X_1, \ldots X_m$ are normal random variables, assuming that $n_1, n_2, \ldots, n_m$ are large enough. △

3 ANOVA table

Definition 5. The *analysis-of-variance table* (ANOVA table) is

Source	Sum of squares (SS)	Degrees of freedom	Mean square (MS)	F ratio
Treatment	SS(T)	$m-1$	MS(T) $= \frac{\text{SS(T)}}{m-1}$	$\frac{\text{MS(T)}}{\text{MS(E)}}$
Error	SS(E)	$n-m$	MS(E) $= \frac{\text{SS(E)}}{n-m}$	
Total	SS(TO)	$n-1$		

$\triangle$

For example, the ANOVA table for Example 1 is

Source	Sum of squares (SS)	Degrees of freedom	Mean square (MS)	F ratio
Treatment	30	3	10	1.6
Error	50	8	6.25	
Total	80	11		