MEM-264 Applied Statistics

Department of Mathematics and Applied Mathematics, University of Crete

Costas Smaragdakis (kesmarag@uoc.gr)

18th lecture - 09-04-2021

- In the one-way ANOVA, we classified populations according to a single factor (categorical variable).
- In the two-way ANOVA, we will extent the model to take into account two factors.

Example : Student's Grades

columns.

Department	Male (j=1)	Female (j=2)
Mathematics (i=1)	310 (mean: 6.38)	400 (mean: 6.52)
Physics (i=2)	510 (mean: 6.41)	330 (mean: 6.56)
Computer Science (i=3)	190 (mean: 6.44)	210 (mean: 6.41)

Rows

(i,j)

$n_{2,1} = 510$ $i=2,\ j=1,\ k=1,\dots,510$

$310 + 400 = 710$

mean: $\frac{310 \cdot 6.38 + 400 \cdot 6.52}{710}$

- **With this analysis we can investigate interactions between factors!**

(page 6)   overall sample mean $\bar{Y} = \frac{310 \cdot 6.38 + 400 \cdot 6.52 + \dots 210 \cdot 6.41}{310 + 510 + \dots + 210}$

One-way: $Y_{ik} = \mu + \alpha_i + \varepsilon_{ik}$, $\varepsilon_{ik} \sim N(0,\sigma^2)$ $\quad \sum \alpha_i = 0$

individual 1st cat. var. 2nd Cat. var. interactions Random error

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}, \quad \epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$$

for all j=1,...,J $\quad$ for all i=1,...,I

$$\sum_{i=1}^{I} \alpha_i = 0, \quad \sum_{j=1}^{J} \beta_j = 0, \quad \sum_{i=1}^{I} \gamma_{ij} = 0, \quad \sum_{j=1}^{J} \gamma_{ij} = 0$$

$$\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_I \end{bmatrix} \quad \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_J \end{bmatrix} \quad \Gamma = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1J} \\ \vdots & & \\ \gamma_{I1} & & \gamma_{IJ} \end{bmatrix}$$

- PAIR (i=1, j=1) : $\{Y_{1,1,1}, \ldots, Y_{1,1,n_{1,1}}\}$ i.i.d sample
- PAIR (i=1, j=2) : $\{Y_{1,2,1}, \ldots, Y_{1,2,n_{1,2}}\}$ i.i.d sample

  ⋮

- ~~GROUP~~ PAIR (i=I,j=J) : $\{Y_{I,J,1}, \ldots, Y_{I,J,n_{I,J}}\}$ i.i.d sample

IJ-i.i.d samples

- There are three null hypotheses (Rows, Columns, interactions)

The three null hypotheses of the model

$$H_{I0} : \alpha_1 = \alpha_2 = \cdots = \alpha_I$$

$$H_{J0} : \beta_1 = \beta_2 = \cdots = \beta_J$$

$$H_{IJ0} : \gamma_{11} = \cdots = \gamma_{IJ}$$

Sample mean for the estimation of the mean μ_{ij}

$$\bar{Y}_{ij,n_{i,j}} = \frac{1}{n_{i,j}} \sum_{k=1}^{n_{i,j}} Y_{ijk}$$

In our example
$\bar{Y}_{21,510} = 6.41$ (see the table)

Sample variance for the estimation of the variance σ^2

$$S^2_{ij,n_{i,j}} = \frac{1}{n_{i,j}-1} \sum_{k=1}^{n_{i,j}} e^2_{ijk}, \quad e_{ijk} = Y_{ijk} - \bar{Y}_{ij,n_{i,j}}$$

Pooled sample variance

$$S_p^2 = \frac{\sum_{i,j}(n_{i,j}-1) S^2_{ij,n_{i,j}}}{\sum_{i,j}(n_{i,j}-1) = N - IJ}$$

this is a better approximation of σ^2

$$N = \sum_{i=1}^{I} \sum_{j=1}^{J} n_{i,j}$$

Overall sample mean

$$\bar{Y} = \frac{1}{\sum_{i,j} n_{i,j}} \sum_{i,j,k} Y_{ijk}$$

Sample variance between groups

$$S_b^2 = \frac{\mathsf{SSM}}{\mathsf{DF}} = \frac{1}{IJ - 1} \sum_{i,j} (\bar{Y}_{ij,n_{ij}} - \bar{Y})^2$$

In two-way ANOVA, the terms SSM and DF can be decomposed as follows:

SSIJ = SSM − (SSI + SSJ)

$$\mathrm{SSM} = \mathrm{SSI} + \mathrm{SSJ} + \mathrm{SSIJ}$$

$$\mathrm{DF} = \mathrm{DFI} + \mathrm{DFJ} + \mathrm{DFIJ}$$

Two-way ANOVA table

		DF	F (numerator)	F (denominator)		
F_I	Rows	I-1	SSI/(I-1)	S_p^2	DF = N-IJ	$F_I = \frac{\frac{SSI}{I-1}}{S_p^2}$
F_J	Columns	J-1	SSJ/(J-1)	S_p^2	DF = N-IJ	
F_{IJ}.	Interaction	(I-1)(J-1)	SSIJ/[(I-1)(J-1)]	S_p^2	DF = N-IJ	

P-value_I

P_value_J

P-value -IJ

$\alpha = 0.05$

After the analysis we will obtain three p-values

p-value for I (or/and J)

- If this p-value is smaller than the considered α, this means that the first (or/and second) factor have a statistically significant effect on the values.

p-value for IJ

- If this p-value is smaller than the considered α, this means that there is a statistically significant interaction effect beetween the factors.

Back to the student's grades

P-value-II $< \alpha$

Department	Male (j=1)	Female (j=2)	Mean
Mathematics (i=1)	310 (mean: 6.38)	400 (mean: 6.52)	
Physics (i=2)	510 (mean: 6.41)	330 (mean: 6.56)	
Computer Science (i=3)	190 (mean: 6.44)	210 (mean: 6.41)	
Mean			