HMMA 307 : Advanced Linear Modeling

Chapter 4 : ANOVA 2 Factors

KHALIFI OUMAYMA KANDOUCI WALID SAHBANE ABDESSTAR

https://github.com/WalidKandouci/HMMA307_Modeles_Lineaires_Avances_Cours_5

Université de Montpellier

Summary

Introduction

We discussed in the previous paragraph the one-way ANOVA and its uses.
In this paragraph, we will be looking at two-way ANOVA, an extension of the one-way ANOVA that examines the influence of two different categorical independent variables on one continuous dependent variable.
The two-way ANOVA not only aims at assessing the main effect of each independent variable but also if there is any interaction between them.

introductory example:

Let's suppose that we have two judges who are tasting 2 different wines, that we will call Wine 1 and Wine 2 such as:

- Judge 1 tasts Wine 1, 3 times and Wine 2, 4 times.
- Judge 2 tasts Wine 1, 3 times and Wine 2, 1 time.

introductory example:

We summarize the example in the following table:

	Wine 1	Wine 2
Judge 1	[6,7,8]	[1,2,3,5]
Judge 2	[3,8,4]	[1]

If factor 1 is Judge 2 and factor 2 is Wine 1, then we have :

$$y_{211} = 3, y_{212} = 8, y_{213} = 4$$

Here, we have:

$$n = n_{11} + n_{12} + n_{21} + n_{22} = 3 + 4 + 3 + 1 = 11$$

we must adapt the table so as to have: $n_{ij} = cte \ \forall i, j \in [\![1, 2]\!]$.
This is done either by eliminating or adding elements.

Two factors :

- Factor 1 : I levels / I classes.
- Factor 2 : J levels / I classes.

n_{ij} : number of repetitions / observations of factor 1 in the class i and to factor 2 in j class.

We obtain the following constraints :

$$n = \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij}$$

Model equation

Model:

$$y_{i,j,k} \overset{iid}{\sim} \mathcal{N}\left(\mu_{ij}, \sigma^2\right), \quad \forall i \in [\![1, I]\!], \forall j \in [\![1, J]\!], \forall k \in [\![1, n_{ij}]\!]$$

$$y_{i,j,k} = \mu + \alpha_i + \beta_j + \varepsilon_{i,j,k}$$

- $\mathrm{Cov}\left(\varepsilon_{i,j,k}, \varepsilon_{i',j',k'}\right) = \sigma^2 \delta_{i,i'} \delta_{j,j'} \delta_{k,k'}$
- $\mu \in \mathbb{R}$: the average effect.
- α_i: the specific effect of level i for the first factor.
- β_j:the specific effect of level j for the second factor.

Note:

If the design of the experiment is not balanced (i.e., the n_{ij} are different), the mathematical analysis is difficult.

Therefore we will assume in order to facilitate the analysis that:

$$\forall i \in [\![1, I]\!], \quad \forall j \in [\![1, J]\!], \quad n_{ij} = K$$

Finaly we get: $n = IJK$ observations.

We can write the matrix form of model just by following a usual approach that is "least squares":

$$X = \begin{bmatrix} \mathbb{1}_n & \mathbb{1}_{C_1} & \dots & \mathbb{1}_{C_I} & \mathbb{1}_{D_1} & \dots & \mathbb{1}_{D_J} \end{bmatrix} \in \mathbb{R}^{n \times (1+I+J)}$$

Where:

$$\mathrm{rang}(X) = I + J + 1 - 2 = I + J - 1 \text{ et } \mathbb{1}_n = (1, \dots, 1)^\top$$

Design matrix

$$X = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \end{bmatrix} \qquad y = \begin{bmatrix} 6 \\ 7 \\ 8 \\ 1 \\ 2 \\ 3 \\ 1 \\ . \\ . \\ . \end{bmatrix}$$

- 3*4=12 rows
- Columns of X: (J_1, J_2, V_1, V_2)

Definition

$$(\widehat{\mu}, \widehat{\alpha}, \widehat{\beta}) \in \underset{(\mu,\alpha,\beta)\in\mathbb{R}\times\mathbb{R}^I\times\mathbb{R}^J}{\arg\min} \frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (y_{i,j,k} - \mu - \alpha_i - \beta_j)^2$$

$$\text{s.c.} \quad \sum_{i=1}^{I} \alpha_i = 0$$

$$\sum_{j=1}^{J} \beta_j = 0$$

For this problem we get the following Lagrangian:

$$\mathcal{L}\left(\mu,\alpha,\beta,\lambda_\alpha,\lambda_\beta\right)=\frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(y_{i,j,k}-\mu-\alpha_i-\beta_j\right)^2+\lambda_\alpha\left(\sum_{i=1}^{I}\alpha_i\right)+\lambda$$

We should solve the following system:

$$
\begin{cases}
\frac{\partial \mathcal{L}}{\partial \mu} = 0 \\
\frac{\partial \mathcal{L}}{\partial \alpha} = 0 \\
\frac{\partial \mathcal{L}}{\partial \beta} = 0 \\
\frac{\partial \mathcal{L}}{\partial \lambda_\alpha} = 0 \\
\frac{\partial \mathcal{L}}{\partial \lambda_\beta} = 0
\end{cases}
$$

We get as results:

$$\frac{\partial \mathcal{L}}{\partial \mu} = 0 \Longrightarrow n\widehat{\mu} = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} y_{i,j,k} \Longrightarrow \widehat{\mu} = \bar{y}_n$$

$$\frac{\partial \mathcal{L}}{\partial \alpha} = 0 \Longrightarrow \forall i \in [1, I], \quad \widehat{\alpha}_i = \underbrace{\bar{y}_{i,:,}}_{=\frac{1}{JK}\sum_{j=1}^{N}\sum_{k=1}^{K} y_{i,j,k}} - \widehat{\mu}$$

$$\frac{\partial \mathcal{L}}{\partial \beta} = 0 \Longrightarrow \forall j \in [1, J], \widehat{\beta}_j = \underbrace{\bar{y}_{:,j,:}}_{=\frac{1}{IK}} \sum_{i=1}^{I}\sum_{k=1}^{K} y_{i,j,k} - \widehat{\mu}$$

Associated predictor:

$$\widehat{y_{ij}} = \widehat{\mu} + \widehat{\alpha_i} + \widehat{\beta}_j$$

$$= \bar{y}_{i,:,:} + \bar{y}_{:,j,:} - \widehat{\mu}$$

$$= \bar{y}_{i,:,:} + \bar{y}_{:,j,:} - \bar{y}_n$$

The affected unbiased estimator is:

$$\widehat{\sigma^2} = \frac{1}{n - (I + J - 1)} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \left(\widehat{y_{ij}} - y_{i,j,k}\right)^2$$

The estimator σ^2 is:

$$\mathbb{E}\left(\widehat{\sigma^2}\right) = \sigma^2$$

Global test:

$$\textbf{Facteur} \quad 1{:}H_0 : \text{"}_{\alpha_1} = \alpha_2 = \cdots = \alpha_I\text{"}$$

$$\textbf{Facteur} \quad 2{:}H_0 : \text{"}\beta_1 = \beta_2 = \cdots = \beta_J\text{"}$$

For factor 1:

$$F_{obs} = \frac{1}{\widehat{\sigma^2}} \frac{KJ}{(I-1)} \sum_{i=1}^{I} (\bar{y}_{i,;;:} - \bar{y}_n)^2 \sim \mathcal{F}^{I-1}_{n-(I+J-1)}$$

Test: $F_{\text{obs}} > \mathcal{F}^{I-1}_{n-(I+J-1)}(1-\alpha)$