MULTIVARIATE ANALYSIS OF VARIANCE (MANOVA)

06/04

(p. 273)

(univariate : ANOVA)

We have g samples coming from g different populations:

$$\left.\begin{array}{l} \underline{X}_{11}, \dots, \underline{X}_{1n_1} \overset{iid}{\sim} N_p(\mu_1, \underline{\Sigma}) \\ \underline{X}_{21}, \dots, \underline{X}_{2n_2} \overset{iid}{\sim} N_p(\mu_2, \underline{\Sigma}) \\ \vdots \\ \underline{X}_{g1}, \dots, \underline{X}_{gn_g} \overset{iid}{\sim} N_p(\mu_g, \underline{\Sigma}) \end{array}\right\} \text{ independent } (\perp\!\!\!\perp)$$

they must be all the same (we have to check the equivalence of the estimates of the Σs)

Goal: make inference on the means $(\mu_1, \dots, \mu_g)$ (even if it's called "Analysis of variance")

Why "Analysis of VARIANCE"? We want to see if there is enough variability (enough difference) among the estimators of these means to guarantee that they're different ↪ unknown

Basic idea: compare variability "**between**" with variability "**within**"

between the means

within each population

Case: $p \geq 1$ and $g = 2$:

$$\left.\begin{array}{l} \underline{X}_{11}, \dots, \underline{X}_{1n_1} \overset{iid}{\sim} N_p(\mu_1, \Sigma) \\ \underline{X}_{21}, \dots, \underline{X}_{2n_2} \overset{iid}{\sim} N_p(\mu_2, \Sigma) \end{array}\right\} \perp\!\!\!\perp$$

Goal: Inference on $\boxed{\mu_1 - \mu_2}$ (**Ex.** patients with and without drugs)

$\bar{\underline{X}}_1 = \frac{1}{n_1} \sum_{j=1}^{n_1} X_{1j}$ estimator for μ_1

$\bar{\underline{X}}_2 = \frac{1}{n_2} \sum_{j=1}^{n_2} X_{ij}$ estimator for μ_2

$$\left.\begin{array}{l} \bar{\underline{X}}_1 \sim N_p(\mu_1, \frac{1}{n_1}\Sigma) \\ \bar{\underline{X}}_2 \sim N_p(\mu_2, \frac{1}{n_2}\Sigma) \end{array}\right\} \perp\!\!\!\perp$$

$$\Longrightarrow \bar{\underline{X}}_1 - \bar{\underline{X}}_2 \sim N_p(\mu_1 - \mu_2, \frac{1}{n_1}\Sigma + \frac{1}{n_2}\Sigma)$$

$$\Longrightarrow \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-\frac{1}{2}} \left((\bar{\underline{X}}_1 - \bar{\underline{X}}_2) - (\mu_1 - \mu_2)\right) \sim N_p(\underline{0}, \Sigma)$$

That's not enough for inference: Σ?

$$S_1 = \frac{1}{n_1-1}\sum_{j=1}^{n}(\underline{X}_{1j}-\bar{\underline{X}}_1)(\underline{X}_{1j}-\bar{\underline{X}}_1)^T$$ estimate for Σ in group 1

$$S_2 = \frac{1}{n_2-1}\sum_{j=1}^{n}(\underline{X}_{2j}-\bar{\underline{X}}_2)(\underline{X}_{2j}-\bar{\underline{X}}_2)^T$$ " " in group 2

$$\left.\begin{aligned}(n-1)\,S_1 &\sim \text{Wish}(\Sigma, n-1)\\ (n-2)\,S_2 &\sim \text{Wish}(\Sigma, n-2)\end{aligned}\right\} \perp\!\!\!\perp$$ (since they're based on independent samples)

$$\Longrightarrow \boxed{(n_1-1)\,S_1 + (n_2-1)\,S_2 \sim \text{Wish}(\Sigma, n_1+n_2-2)}$$

$$\boxed{S_{pooled} = \frac{(n_1-1)\,S_1 + (n_2-1)\,S_2}{n_1+n_2-2}}$$

$$\Longrightarrow \boxed{(n_1+n_2-2)\,S_{pooled} \sim \text{Wish}(\Sigma, n_1+n_2-2)}$$

and is $\perp\!\!\!\perp$ of $\bar{\underline{X}}_1$ and $\bar{\underline{X}}_2$

Hotelling's theorem:

$$\Longrightarrow \boxed{\begin{aligned}&\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1}\left[(\bar{\underline{X}}_1-\bar{\underline{X}}_2)-(\underline{\mu}_1-\underline{\mu}_2)\right]^T S_{pooled}^{-1}\left[(\bar{\underline{X}}_1-\bar{\underline{X}}_2)-(\underline{\mu}_1-\underline{\mu}_2)\right] \sim \\ &\qquad \sim \frac{(n_1+n_2-2)\,p}{n_1+n_2-1-p}\,F(p, n_1+n_2-1-p)\end{aligned}}$$

pivotal statistics

Now that we have the pivotal stat "we're in business" for inference, for example:

$$\begin{cases} H_0: \underline{\mu}_1-\underline{\mu}_2 = \underline{\delta}_0 \\ H_1: \underline{\mu}_1-\underline{\mu}_2 \neq \underline{\delta}_0 \end{cases}$$

- Reject at level $\alpha \in (0,1)$ if:

$$\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1}\left((\bar{\underline{X}}_1-\bar{\underline{X}}_2)-\underline{\delta}_0\right)^{-1} S_{pooled}^{-1}\left((\bar{\underline{X}}_1-\bar{\underline{X}}_2)-\underline{\delta}_0\right) > \frac{(n_1+n_2-2)\,p}{n_1+n_2-1-p}\,F_\alpha(p, n_1+n_2-1-p)$$

- $CR_{1-\alpha}(\underline{\mu}_1-\underline{\mu}_2) = \Big\{\underline{\delta}\in\mathbb{R}^p$ s.t. :

$$\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{-1}\left[(\bar{\underline{X}}_1-\bar{\underline{X}}_2)-\underline{\delta}\right]^T S_p^{-1}\left[(\bar{\underline{X}}_1-\bar{\underline{X}}_2)-\underline{\delta}\right] \leq \frac{(n_1+n_2-2)\,p}{n_1+n_2-1-p}\,F_\alpha(p, n_1+n_2-1-p)\Big\}$$

- Sim Conf $I_{1-\alpha}$ for $\underline{a}^T(\underline{\mu}_1-\underline{\mu}_2)$
- Bonferroni's CI, ...

What if : $\underline{X}_{11}, \dots, \underline{X}_{1n_1} \overset{iid}{\sim} N_p(\mu_1, \Sigma_1)$

$\underline{X}_{21}, \dots, \underline{X}_{2n_2} \overset{iid}{\sim} N_p(\mu_2, \Sigma_2)$

but $\boxed{\Sigma_1 \neq \Sigma_2}$? (Behrens - Fisher problem)

Can we test $H_0 : \Sigma_1 = \Sigma_2$ vs. $H_1 : \Sigma_1 \neq \Sigma_2$? "Yes, maybe."

- Anderson, 2006 (assuming gaussianity) Extension of Levine test
- Non parametric test (use distances for pos. def. matrices and permutational test)

For large n_1 and n_2 : (we don't need gaussianity because we're using asymptotic theory) (we only need mean and variance)

$$\left.\begin{array}{l} \underline{X}_{11}, \dots, \underline{X}_{1n_1} \overset{iid}{\sim} (\mu_1, \Sigma_1) \\ \underline{X}_{21}, \dots, \underline{X}_{2n_2} \overset{iid}{\sim} (\mu_2, \Sigma_2) \end{array}\right\} \perp$$

$\overset{\text{by CLT}}{\Longrightarrow}$

$$\left.\begin{array}{l} \bar{X}_1 \sim N_p(\mu_1, \frac{1}{n_1}\Sigma_1) \\ \bar{X}_2 \sim N_p(\mu_2, \frac{1}{n_2}\Sigma_2) \end{array}\right\} \perp$$

$$\Longrightarrow \underline{\bar{X}}_1 - \underline{\bar{X}}_2 \quad N_p\left(\mu_1 - \mu_2, \frac{1}{n_1}\Sigma_1 + \frac{1}{n_2}\Sigma_2\right)$$

$$\Longrightarrow \left[(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)\right]^T \left(\frac{1}{n_1}\Sigma_1 + \frac{1}{n_2}\Sigma_2\right)^{-1} \left[(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)\right] \sim \chi^2(p)$$

we don't know Σ_1 and Σ_2, it is not a pivotal quantity yet:

$S_1 \xrightarrow{\mathbb{P}} \Sigma_1$ for $n_1 \to \infty$

$S_2 \xrightarrow{\mathbb{P}} \Sigma_2$ for $n_2 \to \infty$

$\overset{n_1 \text{ very large},\ n_2 \text{ very large}}{\Longrightarrow}$

$$\underbrace{\boxed{\left[(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)\right]^T \left(\frac{S_1}{n_1} + \frac{S_2}{n_2}\right)^{-1} \left[(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)\right] \sim \chi^2(p)}}_{\text{pivotal quantity}}$$

- $CR_{1-\alpha}(\mu_1 - \mu_2) = \{ \underline{\delta} \in \mathbb{R}^p$ s.t.

$$\left[(\bar{X}_1 - \bar{X}_2 - \underline{\delta})\right]^T \left(\frac{S_1}{n_1} + \frac{S_2}{n_2}\right)^{-1} \left[(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)\right] \leq \chi_\alpha(p) \}$$

- Test $H_0 : \mu_1 - \mu_2 = \underline{\delta}_0$ vs. $H_1 : \mu_1 - \mu_2 \neq \underline{\delta}_0$
- ...

Case: $p = 1$, $g \geq 1$ (ANOVA)

$$X_{11}, \dots, X_{1n_1} \overset{iid}{\sim} N_1(\mu_1, \sigma^2)$$
$$X_{21}, \dots, X_{2n_2} \overset{iid}{\sim} N_1(\mu_2, \sigma^2)$$
$$\vdots$$
$$X_{g1}, \dots, X_{gn_g} \overset{iid}{\sim} N_1(\mu_g, \sigma^2)$$

We can see it like: we have a treatment that we can apply in $g-1$ different ways. Does it change if we apply the treatment in way i or j?

$x_{11}, \dots, x_{1n_1} \to$ no treat
$x_{21}, \dots, x_{2n_2} \to$ treat at level 1
$\vdots$
$x_{g1}, \dots, x_{gn_g} \to$ treat at level $g-1$

Goal: 1. $\begin{cases} H_0 : \mu_1 = \mu_2 = \dots = \mu_g \\ H_1 : \exists i \neq j : \mu_i \neq \mu_j \end{cases}$ (:= treatment doesn't have effect)

2. If we reject H_0, estimate μ_i's

New parametrization:

$$\mu_i = \mu + \tau_i \qquad i = 1, \dots, g$$

μ: overall mean

τ_i: how different μ_i is from the overall mean (treatment effect)

We moved from g parameters $\longrightarrow$ $g+1$ parameters: $(\mu, \tau_1, \tau_2, \dots, \tau_g)$

We need some constraints on $\mu, \tau_1, \dots, \tau_g$. (since we add a parameter)

Model for X_{ij}'s:

$$X_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad \mu \in \mathbb{R} \quad \tau_i \in \mathbb{R} \quad i = 1, \dots, g \quad \varepsilon_{ij} ?$$

they're gaussian $\Downarrow$

$$\varepsilon_{ij} \overset{iid}{\sim} N_1(0, \sigma^2) \qquad i = 1, \dots, g \qquad j = 1, \dots, n_i$$

- Estimator for μ (overall mean): $(n_1 + n_2 + \dots + n_g = n)$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{g} \sum_{j=1}^{n_i} X_{ij}$$

$$\mathbb{E}[\bar{X}] = \frac{1}{n} \sum_{i=1}^{g} \sum_{j=1}^{n_i} \mathbb{E}[X_{ij}] = \frac{1}{n} \sum_{i=1}^{g} \sum_{j=1}^{n_i} (\mu + \tau_i)$$
$$= \frac{1}{n} \sum_{i=1}^{g} (n_i \mu + n_i \tau_i)$$
$$= \mu \frac{\sum^g n_i}{n} + \frac{\sum^g n_i \tau_i}{n}$$

$$\Longrightarrow \quad \mathbb{E}[\bar{x}] = \mu + \underbrace{\frac{1}{n}\sum_{i=1}^{g} n_i \tau_i}_{}$$

if this is $=0$ $\Longrightarrow$ $\bar{X}$ is unbiased $\Big\}$ $\leftarrow$ we have now our constraint : $\boxed{\sum_{i=1}^{g} n_i \tau_i = 0}$!

If we have balanced design : $n_1 = n_2 = \dots = n_g$ $\Longrightarrow$ $\boxed{\text{Constraint : } \sum \tau_i = 0}$

- **Estimator for τ_i**, $i = 1, \dots, g$:

$\bar{x}_i - \bar{x}$ where $\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$

$$\mathbb{E}[\bar{x}_i - \bar{x}] = \mathbb{E}[\bar{x}_i] - \mu = \mu + \tau_i - \mu = \tau_i \Longrightarrow \text{unbiased}$$

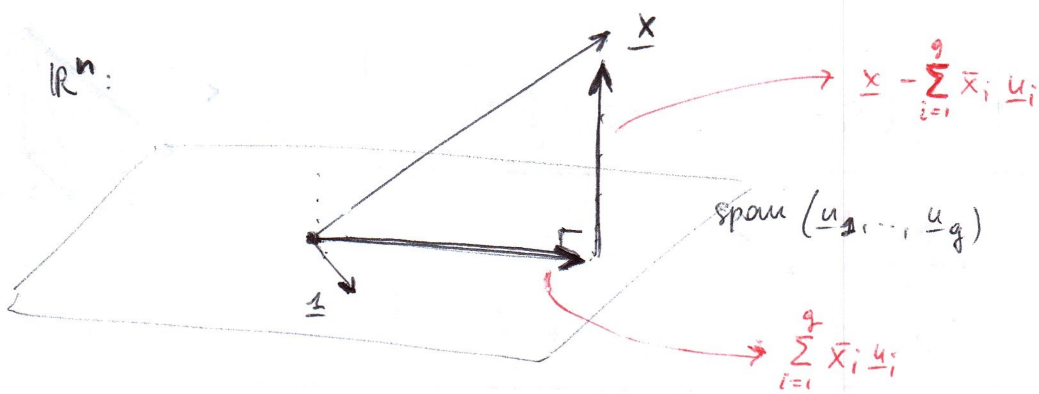
GEOMETRY OF THE ANALYSIS OF VARIANCE :

Consider :

$$\underline{u}_1 = \begin{bmatrix} 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \Big\} n_1 \qquad \underline{u}_2 = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \Big\} n_2 \qquad \dots \qquad \underline{u}_g = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \Big\} n_g$$

- $\underline{u}_1, \dots, \underline{u}_g$ are linearly independent
- $\underline{1} \in \text{span}(\underline{u}_1, \dots, \underline{u}_g)$ $\quad \left(\underline{1} = \sum_{i=1}^{g} \underline{u}_i \right)$
- $\underline{u}_1, \dots, \underline{u}_g$ are $\perp$

$\mathbb{R}^n$:



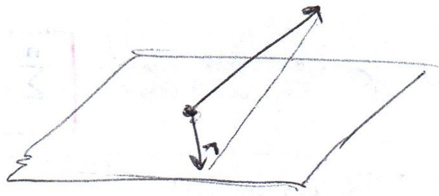
$\underline{X} = [x_{11}, \dots, x_{1n_1}, x_{21}, \dots, x_{2n_2}, \dots, x_{g1}, \dots, x_{gn_g}]^T$ = vector that collects all the variables $\in \mathbb{R}^n$

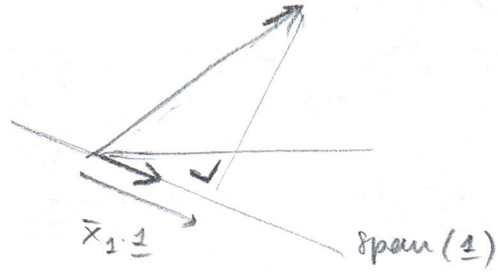
Orthogonal projection of $\underline{X}$ on $\text{span}(\underline{u}_1, \dots, \underline{u}_g)$:

$$\sum_{i=1}^{g} \frac{\underline{u}_i \underline{u}_i^T}{\underline{u}_i^T \underline{u}_i} \underline{X} \underset{\uparrow}{=} \sum_{i=1}^{g} \left(\frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \right) \underline{u}_i = \sum_{i=1}^{g} \bar{x}_i \underline{u}_i$$

$\underline{u}_i^T \underline{X}$ = summing the comp. of $\underline{X}$ corresponding for the pos. of group i

Now we want the orthogonal projection of $\underline{X}$ on span $(\underline{1})$:





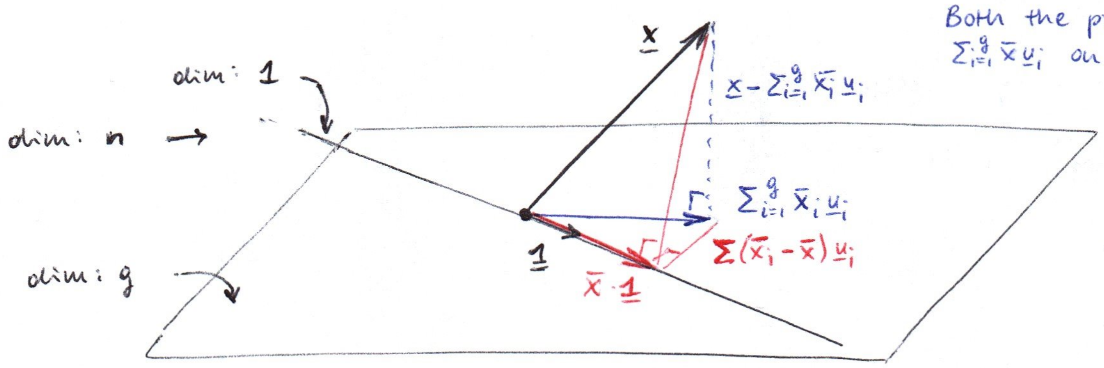
$$\frac{\underline{1}\,\underline{1}^T}{\underline{1}^T\underline{1}} \cdot \underline{X} = \left(\frac{1}{n} \sum_i \sum_j x_{ij}\right) \underline{1} = \bar{x} \cdot \underline{1}$$

Orthogonal projection of $\sum_{i=1}^{g} \bar{x}_i \, \underline{u}_i$ of span $(\underline{1})$:

$$\frac{\underline{1}\,\underline{1}^T}{\underline{1}^T\underline{1}} \left(\sum_{i=1}^{g} \bar{x}_i \, \underline{u}_i \right) = \sum_{i=1}^{g} \bar{x}_i \frac{\underline{1} \cdot \underline{1}^T}{\underline{1}^T\underline{1}} \underline{u}_i$$

$$= \left(\frac{1}{n} \sum_{i=1}^{g} n_i \bar{x}_i \right) \underline{1}$$

$$= \left(\frac{1}{n} \sum_{i=1}^{g} \sum_{j=1}^{n_i} x_{ij} \right) \underline{1} = \bar{x}\, \underline{1}$$

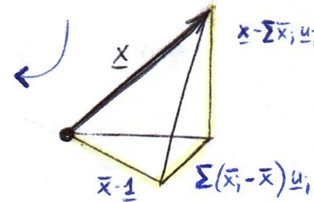


Both the projections of $\underline{x}$ and $\sum_{i=1}^{g} \bar{x}\, \underline{u}_i$ on $\mathcal{L}(\underline{1})$ are equal to $\bar{x} \cdot \underline{1}$

$\Downarrow$

we can now write $\underline{x}$ as the sum of 3 $\perp$ vectors :

- $\sum (\bar{x}_i - \bar{x}) \underline{u}_i$
- $\bar{x} \cdot \underline{1}$
- $\underline{x} - \sum_{i=1}^{g} \bar{x}_i \underline{u}_i$



Hence:
$$\boxed{\underline{X} = \underbrace{\bar{x} \cdot \underline{1}} + \underbrace{\sum_{i=1}^{g} (\bar{x}_i - \bar{x}) \underline{u}_i} + \underbrace{\left(\underline{x} - \sum_{i=1}^{g} \bar{x}_i \underline{u}_i \right)}}$$

these 3 are $\perp$

since they're $\perp$
$$\Longrightarrow \quad \| \underline{x} \|^2 = \| \bar{x} \cdot \underline{1} \|^2 + \left\| \sum_{i=1}^{g} (\bar{x}_i - \bar{x}) \underline{u}_i \right\|^2 + \left\| \underline{x} - \sum_{i=1}^{g} \bar{x}_i \underline{u} \right\|^2$$

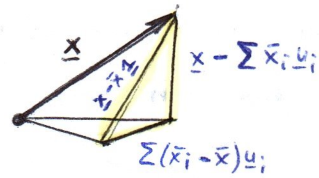
$$\boxed{\sum_{i=1}^{g} \sum_{j=1}^{n_i} x_{ij}^2 = n \bar{x}^2 + \sum_{i=1}^{g} (\bar{x}_i - \bar{x})^2 n_i + \sum_{i=1}^{g} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}$$

:= DECOMPOSITION OF VARIANCE :

$$\boxed{SS_{obs} = SS_{mean} + SS_{treat} + SS_{residuals}}$$

Also :

$$\| \underline{x} - \bar{x}\underline{1} \|^2 = \| \sum_{i=1}^{g} (\bar{x}_i - \bar{x}) \underline{u}_i \|^2 + \| \underline{x} - \sum_{i=1}^{g} \bar{x}_i \underline{u}_i \|^2$$



$$\boxed{\sum_{i=1}^{g} \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 = \sum_{i=1}^{g} (\bar{x}_i - \bar{x})^2 n_i + \sum_{i=1}^{g} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}$$

:= 2nd form of the **DECOMPOSITION OF VARIANCE** : (both are called variance decomp. formulas)

$$\boxed{SS_{centered} = SS_{treat} + SS_{res}}$$

$\Longrightarrow$ **Tot.** variability around mean = variability **between** groups + variability **within** groups

$\sum_{i=1}^{g} (\bar{x}_i - \bar{x})^2 n_i$ — we're checking the variability in the case that every unit is equal to its mean (inside the group they're the same, outside?)

$\sum_{i=1}^{g} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$ — we consider that all the groups have the same variability, how about inside each group? what is the variability?

Our goal was :

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \dots = \mu_g \\ H_1 : \exists i \neq j \text{ s.t. } \mu_i \neq \mu_j \end{cases} = \begin{cases} H_0 : \tau_1 = \tau_2 = \dots = \tau_g = 0 \\ H_1 : \exists i : \tau_i \neq 0 \end{cases}$$

If **H_0** is true : we expect (all the means are equal $\Rightarrow$ all the obs. are coming from the same population (which has only one mean))

$$\boxed{\sum_{i=1}^{g} (\bar{x}_i - \bar{x}) \underline{u}_i \quad \text{to be "small"} \quad \left(\text{w.r.t.} \quad \underline{x} - \sum_{i=1}^{g} \bar{x}_i \underline{u}_i \right)}$$

$\Rightarrow$ because we expect projecting on $\mathcal{L}(\underline{1})$ (where all the components have the same mean) or projecting on span$(\underline{u}_1, \dots, \underline{u}_g)$ (where we allow for different groups to have different means) be almost the same

$\Longrightarrow$ Reject H_0 if :

$$\frac{\| \sum_{i=1}^{g} (\bar{x}_i - \bar{x}) \underline{u}_i \|^2}{\| \underline{x} - \sum_{i=1}^{g} \bar{x}_i \underline{u}_i \|^2} \quad \text{is large}$$

or if :

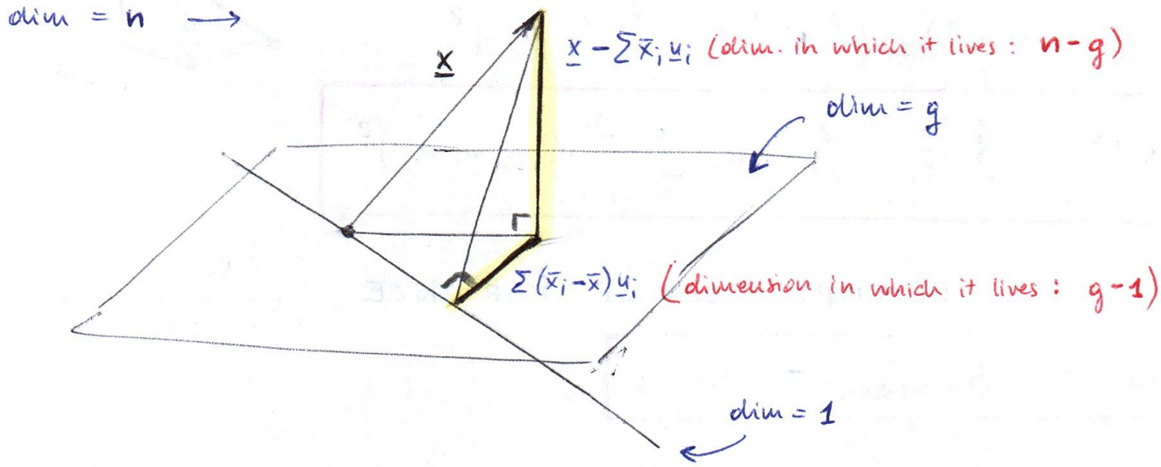
$$(*) \quad \frac{\sum_{i=1}^{g} (\bar{x}_i - \bar{x})^2 n_i}{\sum_{i=1}^{g} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2} \quad \text{is large}$$

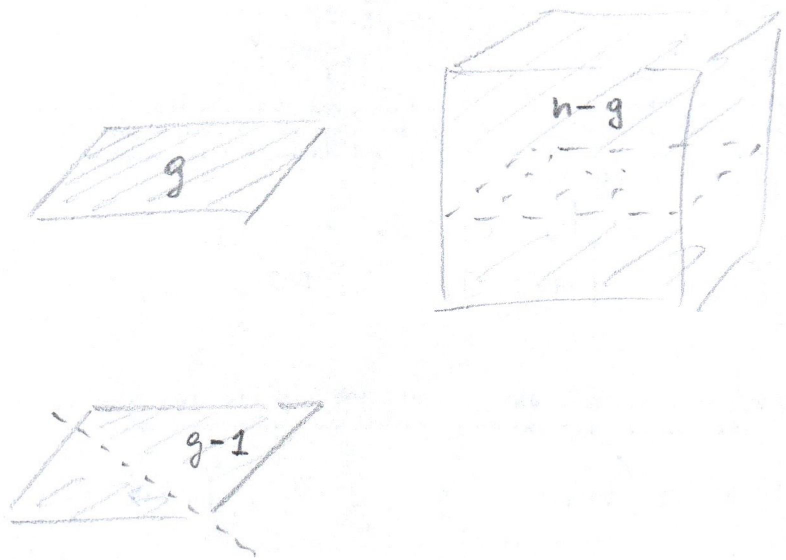
$\underline{x}$ is gaussian $\Rightarrow$ projecting orthogonally a gaussian vector we obtain gaussian vectors $\Rightarrow$ $\Sigma(\bar{x}_i - \bar{x})\underline{u}_i$, $\bar{x}\underline{1}$, $\underline{x} - \Sigma \bar{x}_i \underline{u}_i$ are gaussian (and orthogonal) $\Rightarrow$ they're $\perp\!\!\!\perp$ $\Rightarrow$ we have 3 $\perp\!\!\!\perp$ gaussian $\Rightarrow$ the square of the length of a gaussian vector means the square of its components, it's a χ^2 distribution $\Rightarrow$ it's not surprising that we end up with the ratio of two χ^2 distribution, meaning an F distrib.

How large?
We need to know the distribution of $(*)$ if H_0 is true.

$$\boxed{F_0 = \frac{\dfrac{\sum_{i=1}^{g} (\bar{x}_i - \bar{x})^2 n_i}{(g-1)}}{\dfrac{\sum_{i=1}^{g} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}{(n-g)}} \sim F(g-1, n-g)}$$

Why those degree of freedom?





ANOVA (flash RECAP)

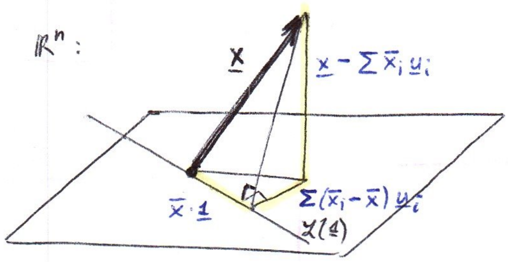
X_{ij} $\quad i = 1, \dots, g$ (groups) $\quad j = 1, \dots, n_i$ (observation per group)

$X_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad \mu \in \mathbb{R}, \quad \tau_i \in \mathbb{R} \text{ s.t. } \sum_{i=1}^{g} n_i \tau_i = 0$

$\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2) \qquad \sigma^2 > 0$

$$\underset{\text{vector of observations}}{\underline{X}} = \begin{bmatrix} X_{11} \\ X_{12} \\ \vdots \\ X_{g n_g} \end{bmatrix} \in \mathbb{R}^n \qquad n = \sum_{i=1}^{g} n_i$$

$$\underline{u}_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \Big\} n_i \qquad i = 1, \dots, g$$



$$\underline{X} = \underset{(1)}{\bar{X} \cdot \underline{1}} + \underset{(2)}{\sum_{i=1}^{g} (\bar{X}_i - \bar{X}) \underline{u}_i} + \underset{(3)}{\left(\underline{X} - \sum_{i=1}^{g} \bar{X}_i \underline{u}_i\right)}$$

- $\bar{X} \cdot \underline{1} \in \operatorname{span}(\underline{1})$
- $\sum_{i=1}^{g} (\bar{X}_i - \bar{X}) \underline{u}_i \in \left((\operatorname{span}(\underline{1}))^{\perp} \cap \operatorname{span}(\underline{u}_1, \dots, \underline{u}_g) \right)$
- $\underline{X} - \sum_{i=1}^{g} \bar{X}_i \underline{u}_i \in (\operatorname{span}(\underline{u}_1, \dots, \underline{u}_g))^{\perp}$

	degrees of freedom for :
dim = 1	mean
dim = g − 1	treatment effects
dim = n − g	residuals

Decomposition of variance:

$$\sum_{i=1}^{g} \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 = \sum_{i=1}^{g} n_i (\bar{x}_i - \bar{x})^2 + \sum_{i=1}^{g} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$

$$SS_{centered} = SS_{treat} + SS_{residuals}$$

Observations:

- $\underline{X}$ is gaussian $\Longrightarrow$ (1), (2), (3) are gaussian (since they're linear transf. of a gaussian vector)
- (1), (2), (3) are independent (because they are orthogonal projections of a gaussian vector whose covariance is $\sigma^2 I$.

Exercise: suppose $\underline{X} \sim N_p(\mu, \sigma^2 I)$

$P : \mathbb{R}^p \longrightarrow \mathbb{R}^p$ be an orthogonal projector

Prove that $P\underline{X}$ and $(I - P)\underline{X}$ are independent.

- SS_{treat} and $SS_{residuals}$ are independent (since (2) ⫫ (3))

Now we want to understand the distribution of (1), (2), (3).

Note that[1]: $SS_{res} = \sum_{i=1}^{g} \sum_{j=1}^{n_i} (\bar{x}_i - \bar{x})^2 = \sum_{i=1}^{g} (n_i - 1) S_i^2$

$$S_i^2 = \frac{1}{n-1} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 \quad \text{estimator of } \sigma^2 \text{ in group } i.$$

$$\begin{cases} S_i \sim \sigma^2 \chi^2(n_i - 1) \text{ distr.} \\ S_i\text{'s are independent} \end{cases}$$

(they reffer to samples (from group 1, group 2, ..) that are independent)

we're summing ⫫ χ^2

$$\Longrightarrow SS_{res} = \boxed{\sum_{i=1}^{g} (n_i - 1) S_i^2 \sim \sigma^2 \chi^2(n-g)}$$

Note that[2]: If $H_0 : \tau_1 = \tau_2 = .. = \tau_g = 0$ is true

$$\Longrightarrow \boxed{\sum_{i=1}^{g} \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 = (n-1) S^2 \sim \sigma^2 \chi^2(n-1)}$$

Hence, if H_0 is true :

$$\underset{(\sim \sigma^2 \chi^2(n-1))}{SS_{cent}} = \underset{?}{SS_{treat}} + \underset{(\sim \sigma^2 \chi^2(n-g))}{SS_{res}}$$

⫫

$SS_{treat} \sim$? We have to add something to a χ^2 to obtain an other χ^2 $\Rightarrow$ we add a χ^2

SS_{treat} must have $\sim \sigma^2 \chi^2(g-1)$: $\boxed{SS_{treat} \sim \sigma^2 \chi^2(g-1)}$

If H_0 is true : (we know the distr. of SS_{treat} and SS_{res}):

$$\boxed{F_0 = \frac{\frac{SS_{treatment}}{g-1}}{\frac{SS_{res}}{n-g}} \sim F(g-1, n-g)}$$

Therefore reject H_0 at level $\alpha \in (0,1)$ if $F_0 > F_\alpha(g-1, n-g)$ (p-value ..)

MANOVA : $p \geq 1, \ g \geq 2$

$\underline{X}_{ij} \in \mathbb{R}^p$: $\underline{X}_{ij} = \underline{\mu} + \underline{\tau}_i + \underline{\varepsilon}_{ij}$

$i = 1, .., g$
$j = 1, .., n_i$

$\underline{\varepsilon}_{ij} \overset{iid}{\sim} N(\underline{0}, \Sigma)$

$\sum_{i=1}^{g} n_i \underline{\tau}_i = \underline{0}$

We have: $\begin{cases} \underline{X}_{11}, \dots, \underline{X}_{1n_1} \sim N_p(\underline{\mu}_1, \Sigma) \\ \vdots \\ \underline{X}_{g1}, \dots, \underline{X}_{gn_g} \sim N_p(\underline{\mu}_g, \Sigma) \end{cases}$

Observation :

- for a fixed component k of $\underline{X}_{ij}$ $\Longrightarrow$ ANOVA :

$$X_{ijk} = \mu_k + \tau_{ik} + \varepsilon_{ijk} \qquad i = 1, .., g$$
$$j = 1, .., n_i$$
$$\varepsilon_{ijk} \overset{iid}{\sim} N(0, \sigma_{kk}) \qquad k = 1, .., p$$
$$\sum n_i \tau_{ik} = 0$$

!! So we're just doing p-ANOVA models simultaneously? NO. We assume that the statistical units are ⫫ and also the different groups are ⫫, we're not assuming that the components of the vector are ⫫.
But there might be dependence between component k and component l (of the same vector) : σ_{kl}

$\underline{\bar{X}}$ is the estimator for $\underline{\mu}$

$\underline{\bar{X}}_i$ is the estimator for $\underline{\mu}_i$

$\underline{\bar{X}}_i - \underline{\bar{X}}$ estimator for $\underline{\tau}_i$

unbiased if $\sum n_i \underline{\tau}_i = \underline{0}$

Goal : $H_0 : \underline{\tau}_1 = \underline{\tau}_2 = .. = \underline{\tau}_g = \underline{0}$
$H_1 : \exists\, \underline{\tau}_i \neq \underline{0}$

Covariance decomposition formula :

$$\boxed{\underbrace{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(\underline{X}_{ij}-\underline{\bar{X}})(\underline{X}_{ij}-\underline{\bar{X}})^T}_{\text{covariability that we would compute if all the groups hold the same mean}} = \sum_{i=1}^{g}(\underline{\bar{X}}_i-\underline{\bar{X}})(\underline{\bar{X}}_i-\underline{\bar{X}})^T n_i + \sum_{i=1}^{g}\sum_{j=1}^{n_i}(\underline{X}_{ij}-\underline{\bar{X}}_i)(\underline{X}_{ij}-\underline{\bar{X}}_i)^T}$$

proof.
(Exercise) (algebrically) ■

(Note that $\underline{X}_{ij} - \underline{\bar{X}} = (\underline{\bar{X}}_i - \underline{\bar{X}}) + (\underline{X}_{ij} - \underline{\bar{X}}_i)$ (an identity))

$$B := \sum_{i=1}^{g} n_i(\underline{\bar{X}}_i - \underline{\bar{X}})(\underline{\bar{X}}_i - \underline{\bar{X}})^T$$

$$W := \sum_{i=1}^{g}\sum_{j=1}^{n_i}(\underline{X}_{ij} - \underline{\bar{X}}_i)(\underline{X}_{ij} - \underline{\bar{X}}_i)^T$$

$$= \sum_{i=1}^{g}(n_i - 1)\, S_i$$

sample covariances that we would compute in group i

"B = between" : it's capturing the covariance between groups (as if in every group, every statistical unit was the same and equal to the sample mean): What would be the covariability in the dataset if every statistical unit would be the same? (no variability within)

"W = within": we place in group i, we look the covariability w.r.t. the mean of group i and then we change the group ($i \to j$). Then we sum up all these covariabilities

(We can say that this is a pooled covariability, averaging all the estimators for Σ that we make for every group)

By analogy with ANOVA:

we would like to consider " $\frac{B}{W}$ " — we always want to compare how big is the variability between groups against the variability within

Different proposals: (of test statistics for comparing those two covariabilities):

- **WILKS**:

$$\boxed{\Lambda_W = \frac{\mathrm{Det}(W)}{\mathrm{Det}(W+B)}}$$

Λ_W large ($\equiv$ gen. variance within the groups is not different from the overall gen. variance) $\Longrightarrow$ the treatment did not produce much effect (it was not able to increase the variability among groups)

Reject $H_0: \underline{\tau}_1 = \underline{\tau}_2 = \dots = \underline{\tau}_g = \underline{0}$ if Λ_W is **small**

= the treatment does not produce effects

- **LAWLEY - HOTELLING**:

$$\boxed{\Lambda_{LH} = \mathrm{Tr}(BW^{-1})}$$

(closest to ANOVA)
Λ_{LH} large $\Rightarrow$ B is adding a lot of variability w.r.t. the variability that is within the groups

Reject $H_0: \underline{\tau}_1 = \underline{\tau}_2 = \dots = \underline{\tau}_g = \underline{0}$ if Λ_{LH} is **large**

- **PILLAI**:

$$\boxed{\Lambda_P = \mathrm{Tr}(B \cdot (B+W)^{-1})}$$

if the variability introduced by the treatment is large we expect to have $B \cdot (B+W)^{-1}$ with large variability $\Rightarrow$ we reject if Λ_P is large
Λ_P large $\Rightarrow$ the treat. produced much effect

Reject $H_0: \underline{\tau}_1 = \underline{\tau}_2 = \dots = \underline{\tau}_g = \underline{0}$ if Λ_P is **large**

All of them are functions of $\lambda_1, \dots, \lambda_s$ eigenvalues of BW^{-1}
where $s = \min(g-1, p)$. $(s = \mathrm{rank}(B))$ $(B \in \mathbb{R}^{p \times p})$

$(\underline{\bar{X}}_i - \underline{\bar{X}})$ lives in $g-1$ dimension (vector that create B), so if $g-1 < p$ we know that the determinant is gonna be zero; there are only a number of eigenvalues less than p $\Rightarrow$ for this reason we take $s = \min(g-1, p)$

What is Λ large or small? We need to know the distributions of Λ:

Distribution of Λ's under H_0 are generally unknown

$\Longrightarrow$ we use simulations.

About Λ_W:

- if $p \geq 1$ and $g = 2, 3$ distribution under H_0 is known
- if $p = 2$ and $g \geq 1$ " "
- asymptotic distribution of Λ_W is known for $n \to \infty$ (all the n_i are big, not just one group)

$$-\left(n-1-\frac{p+g}{2}\right)\log\Lambda_W \sim \chi^2(p(g-1)) \qquad \text{(BARTLETT'S approx.)}$$

Reject at level $\alpha \in (0,1)$ if

$$-\left(n-1-\frac{p+g}{2}\right)\log\Lambda_W > \chi^2(p(g-1))$$

If H_0 has been rejected, then we want to estimate the effect of the treatment:

$\underline{\tau}_i$ compared with $\underline{\tau}_k$ component-wise — not only we want to see if the level i produced a difference w.r.t. level k, but also on what component

$\Longrightarrow$ BONFERRONI CI for $\tau_{il} - \tau_{kl}$

$k, i = 1, \dots, g$
$l = 1, \dots, p$

Estimator for: $\tau_{i\ell} - \tau_{k\ell}$

- $\boxed{\bar{x}_{i\ell} - \bar{x}_{k\ell}} \sim N\left(\tau_{i\ell} - \tau_{k\ell}, \frac{1}{n_i}\sigma_{\ell\ell} + \frac{1}{n_k}\sigma_{\ell\ell}\right)$

we need an estimator for $\sigma_{\ell\ell} \rightarrow W$

since we're considering a mean

- $\frac{1}{n-g} W$ is an estimator of $\Sigma \Longrightarrow \frac{w_{\ell\ell}}{n-g}$ is an estimator of $\sigma_{\ell\ell}$

- How many comparisons: $p\,\frac{g(g-1)}{2}$

p confidence intervals (one for each component) and for each component we are comparing 2 groups, so how many possibilities we have? We have g possibilities for first group and $g-1$ for the second, all divided by 2 because if we compare A and B, and B and A, it's the same

$\Downarrow$

so we're trying to compute $p \cdot \frac{g(g-1)}{2}$ confidence intervals simultaneously

Bonf Sim $CI_{1-\alpha}(\tau_{i\ell} - \tau_{k\ell}) =$

$$= \left[\bar{x}_{i\ell} - \bar{x}_{k\ell} \pm t_{\frac{\alpha}{2}\frac{2}{pg(g-1)}}(n-g)\sqrt{\frac{w_{\ell\ell}}{n-g}\left(\frac{1}{n_k} + \frac{1}{n_i}\right)}\right]$$

Exercise: Bonferroni CI for ANOVA $(p=1)$

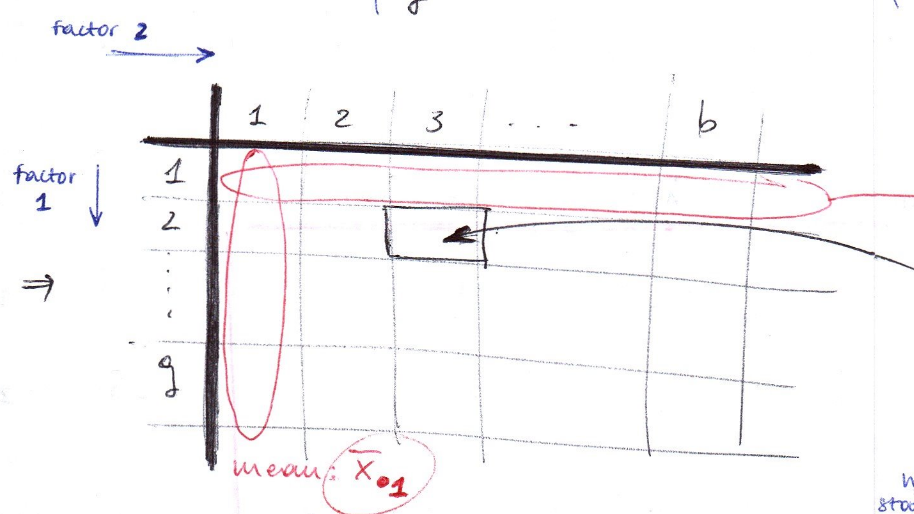
TWO-WAYS (M) ANOVA

(not one but two treatments simultaneously)

We have two factor generating the treatment:

factor 1 (treatment 1) — levels of factor 1: $1, 2, 3, \dots, g$

factor 2 (treatment 2) — levels of factor 2: $1, 2, \dots, b$



$\rightarrow$ mean: $\bar{X}_{1\bullet}$ (we're averaging along all the observations for which treat 1 is at level 1)

X_{ijk}

i = factor 1 (level of F1)

j = factor 2 (" F2)

$k = 1, \dots, n$

how many statistical units we have in this group characterized by these two levels of these two factors

(BALANCED: same sample size in each group)

$X_{ijk} \in \mathbb{R}$ $(p=1)$ (for semplicity)

Model: $X_{ijk} = \mu + \tau_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$

τ_i: effects of treat. 1; β_j: effects of treat. 2; γ_{ij}: interactions between treatment

$i = 1, \dots, g$

$j = 1, \dots, b$

$k = 1, \dots, n$

$$\sum_{i=1}^{g} \tau_i = 0 \qquad \sum_{i=1}^{g} \gamma_{ij} = \sum_{j=1}^{b} \gamma_{ij} = 0$$

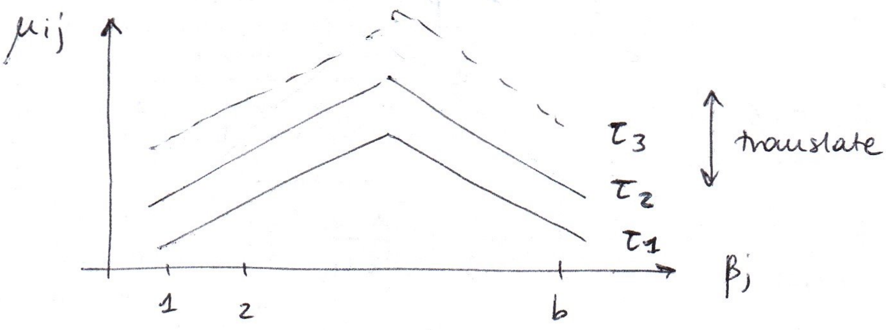
$$\sum_{j=1}^{b} \beta_j = 0$$

$\varepsilon_{ijk} \overset{iid}{\sim} N(0, \sigma^2) \qquad \sigma^2 > 0$

Do we want interactions?

Model <u>without</u> interactions (additive model)

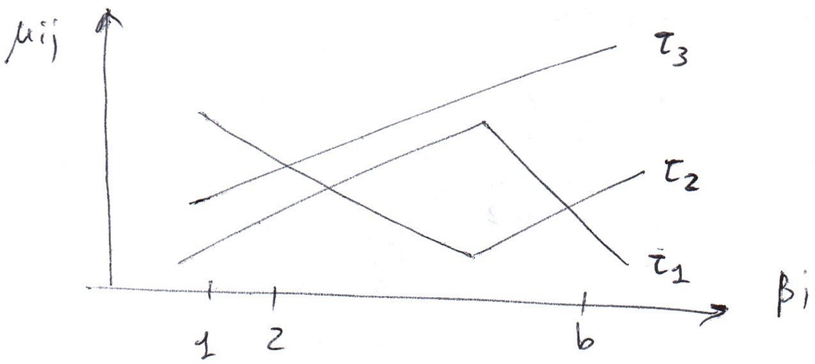
$\mu_{ij} = \mu + \tau_i + \beta_j$:= mean for group ij $(\gamma_{ij} = 0)$



Once we fix τ, then μ_{ij} only depends on β. If we change τ, it's the same μ_{ij}-profile but translated up or down.

Model <u>with</u> interactions (complete model)

$\mu_{ij} = \mu + \tau_i + \beta_j + \gamma_{ij}$



This model is richer than the additive model. Should we always use this? In this case we have many more parameters for the mean and then, therefore, less degrees of freedom for estimating σ^2 ⇒ we have a less bias for the mean but greater uncertainty about what we're doing because we don't know the meter with wich we've comparing the differencences (we don't have a way to estimate σ^2) ⇒ if we don't have much date this is a BAD CHOICE

Decomposition of variance:

b cells
n observations in each cell

$$\underbrace{\sum_{i=1}^{g}\sum_{j=1}^{b}\sum_{k=1}^{n}(x_{ijk}-\bar{x})^2}_{SS_{centered}} = \sum_{i=1}^{g} bn(\bar{x}_{i\bullet}-\bar{x})^2 + \qquad SS_{treat\ 1}$$

$$+ \sum_{j=1}^{b} gn(\bar{x}_{\bullet j}-\bar{x})^2 + \qquad SS_{treat\ 2}$$

$$+ \sum_{i=1}^{g}\sum_{j=1}^{b}(\bar{x}_{ij}-\bar{x}_{i\bullet}-\bar{x}_{\bullet j}+\bar{x})^2 n + \qquad SS_{interactions}$$

$$+ \sum_{i=1}^{g}\sum_{j=1}^{b}\sum_{k=1}^{n}(x_{ijk}-\bar{x}_{ij})^2 \qquad SS_{residuals}$$

this sum up in the ADDITIVE MODEL

	degrees of freedom (= spaces where we're projecting)
SS treat 1	$g-1$
SS treat 2	$b-1$
SS interact.	$(g-1)(b-1)$
SS residuals	$g\,b(n-1)$

	dof
$SS_{centered}$	1

interactions are introducing enough variability among these cells (of the design) so we can conclude that they exist?

$$\begin{cases} H_0: \gamma_{ij} = 0 \\ H_1: \exists\, \gamma_{ij} \neq 0 \end{cases}$$ $\equiv$ there are interactions vs. there are not

Reject at level $\alpha \in (0,1)$ if:

$$\boxed{\frac{\frac{SS_{interactions}}{(g-1)(b-1)}}{\frac{SS_{res}}{gb(n-1)}} > F_\alpha\left((g-1)(b-1),\, gb(n-1)\right)}$$

If we <u>do not</u> reject H_0 $\Longrightarrow$ additive model

and so:

$$\begin{cases} H_0: \tau_1 = \tau_2 = \ldots = \tau_g = 0 \\ H_1: \exists\, \tau_i \neq 0 \end{cases}$$

Reject at level $\alpha \in (0,1)$ if:

$$\boxed{\frac{\frac{SS_{treat\,1}}{g-1}}{\frac{SS_{residuals}}{gbn-g}} > F_\alpha\left(g-1,\, gbn-g\right)}$$