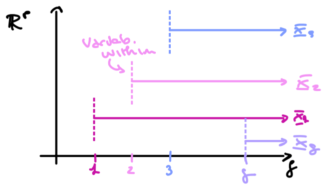
Multivariate Analysis of variance (MANOVA)

$$\mathcal{Y} = \begin{cases} \underline{x}_{11}, \dots, \underline{x}_{1n_1} & \text{iid} \sim N_p(\underline{\mu}_1, \Sigma) \\ \underline{x}_{21}, \dots, \underline{x}_{2n_2} & \text{iid} \sim N_p(\underline{\mu}_2, \Sigma) \\ \vdots & \\ \underline{x}_{g1}, \dots, \underline{x}_{gn_g} & \text{iid} \sim N_p(\underline{\mu}_g, \Sigma) \end{cases}$$



Variability between samples, to see if the different means are justified

Goal: inference on $\underline{\mu}_1, \dots, \underline{\mu}_g$

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

$$\mathcal{Y} = \begin{cases} \underline{x}_{11}, \dots, \underline{x}_{1n_1} & \text{iid} \sim N_p(\underline{\mu}_1, \Sigma) \\ \underline{x}_{21}, \dots, \underline{x}_{2n_2} & \text{iid} \sim N_p(\underline{\mu}_2, \Sigma) \end{cases}$$

Goal: Inference on $\underline{\mu}_1, \underline{\mu}_2$

$$\mathcal{Y} = \begin{cases} \bar{\underline{x}}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} \underline{x}_{i1} \sim N_p(\underline{\mu}_1, \frac{1}{n_1}\Sigma) \\ \bar{\underline{x}}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} \underline{x}_{i2} \sim N_p(\underline{\mu}_2, \frac{1}{n_2}\Sigma) \end{cases}$$

$$\bar{\underline{x}}_1 - \bar{\underline{x}}_2 \sim N_p\left(\underline{\mu}_1 - \underline{\mu}_2, \Sigma\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right) \quad \text{unbiased for } \underline{\mu}_1 - \underline{\mu}_2$$

$$\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1}\left((\bar{\underline{x}}_1 - \bar{\underline{x}}_2) - (\underline{\mu}_1 - \underline{\mu}_2)\right) \sim N_p(\underline{0}, \Sigma)$$

$$S'_1 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (\underline{x}_{i1} - \bar{\underline{x}}_1)(\underline{x}_{i1} - \bar{\underline{x}}_1)' \quad \text{est of } \Sigma \qquad (n_1 - 1) S'_1 \sim Wish(\Sigma, n_1 - 1)$$

$$S'_2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (\underline{x}_{i2} - \bar{\underline{x}}_2)(\underline{x}_{i2} - \bar{\underline{x}}_2)' \quad \text{est of } \Sigma \qquad (n_2 - 1) S'_2 \sim Wish(\Sigma, n_2 - 1)$$

unbiased

$$\Longrightarrow (n_1 - 1) S'_1 + (n_2 - 1) S'_2 \sim Wish(\Sigma, n_1 + n_2 - 2) \Longrightarrow (n_1 + n_2 - 2) S_{pooled} \sim Wish(\Sigma, n_1 + n_2 - 2)$$

$$S_{pooled} = \frac{(n_1 - 1) S'_1 + (n_2 - 1) S'_2}{n_1 + n_2 - 2}$$

Hotelling's thm: $\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1} \left[(\bar{\underline{x}}_1 - \bar{\underline{x}}_2) - (\underline{\mu}_1 - \underline{\mu}_2)\right]' S_{pooled}^{-1} \left[(\bar{\underline{x}}_1 - \bar{\underline{x}}_2) - (\underline{\mu}_1 - \underline{\mu}_2)\right] \sim \frac{(n_1 + n_2 - 2) p}{n_1 + n_2 - 1 - p} F(p, n_1 + n_2 - 1 - p)$ pivotale

$\alpha \in (0,1)$ $CR_{1-\alpha}(\underline{\mu}_1 - \underline{\mu}_2) = ?$

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

Reject H_0 at level α if: $\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1} \left[(\bar{\underline{x}}_1 - \bar{\underline{x}}_2) - \underline{\delta}_0\right]' S_{pooled}^{-1} \left[(\bar{\underline{x}}_1 - \bar{\underline{x}}_2) - \underline{\delta}_0\right] \geq \frac{(n_1 + n_2 - 2) p}{n_1 + n_2 - 1 - p} F_{1-\alpha}(p, n_1 + n_2 - 1 - p)$

$$CR_{1-\alpha}(\underline{\mu}_1 - \underline{\mu}_2) = \left\{ \underline{\varphi} = \underline{\mu}_1 - \underline{\mu}_2 \in \mathbb{R}^p : d^2(\bar{\underline{x}}_1 - \bar{\underline{x}}_2, \underline{\varphi}) \leq \frac{(n_1 + n_2 - 2) p}{n_1 + n_2 - 1 - p} F_{1-\alpha}(p, n_1 + n_2 - 1 - p) \right\}$$

Obs: $\underline{x}_{11}, \dots, \underline{x}_{1n_1}$ iid $\sim N_p(\underline{\mu}_1, \Sigma_1)$

$\underline{x}_{21}, \dots, \underline{x}_{2n_2}$ iid $\sim N_p(\underline{\mu}_2, \Sigma_2)$

$\longrightarrow$ Assumption: $\Sigma_1 = \Sigma_2$

Test: $H_0: \Sigma_1 = \Sigma_2$ vs $H_1: \Sigma_1 \neq \Sigma_2$

1) Extension of Levene's test $\longrightarrow$ Anderson (2006), assumption perfect gaussian distribution

n_1 & n_2 large

$\bar{\underline{x}}_1 \sim N_p(\underline{\mu}_1, \frac{1}{n_1}\Sigma_1)$

$\bar{\underline{x}}_2 \sim N_p(\underline{\mu}_2, \frac{1}{n_2}\Sigma_2)$

$\bar{\underline{x}}_1 - \bar{\underline{x}}_2 \sim N_p(\underline{\mu}_1 - \underline{\mu}_2, \frac{1}{n_1}\Sigma_1 + \frac{1}{n_2}\Sigma_2)$

$\left[(\bar{\underline{x}}_1 - \bar{\underline{x}}_2) - (\underline{\mu}_1 - \underline{\mu}_2)\right]' \left(\frac{1}{n_1} S_1 + \frac{1}{n_2} S_2\right) \left[(\bar{\underline{x}}_1 - \bar{\underline{x}}_2) - (\underline{\mu}_1 - \underline{\mu}_2)\right] \sim \chi^2(p)$ pivotale

$S'_1 \xrightarrow{P} \Sigma_1$ LLN

$S'_2 \xrightarrow{P} \Sigma_2$

ANOVA

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

$$\mathcal{Y} = \begin{cases} x_{11}, \dots, x_{1n_1} & \text{iid} \sim N_1(\mu_1, \sigma^2) \\ x_{21}, \dots, x_{2n_2} & \text{iid} \sim N_1(\mu_2, \sigma^2) \\ \vdots & \\ x_{g1}, \dots, x_{gn_g} & \text{iid} \sim N_1(\mu_g, \sigma^2) \end{cases}$$

Reparametrize the problem: $X_{ij} = \mu + \tau_i + \varepsilon_{ij}$ i.e. $\mu_i = \mu + \tau_i$

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

$\mu_1, \dots, \mu_g$ g param $\longrightarrow$ $\mu, \tau_1, \dots, \tau_g$ $\quad g+1$ param.

Needs a constraint on $\tau_1, \dots, \tau_g$: $\sum n_i \tau_i = 0$ $\quad$ if $n_1 = \dots = n_g \implies \sum \tau_i = 0$

Estimate for μ : $\bar{x} = \frac{1}{n} \sum_{i=1}^{g} \sum_{j=1}^{n_i} x_{ij}$ estimate for μ $\quad (n = n_1 + n_2 + \dots + n_g)$

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

Estimate for τ_i : $\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$ $\quad i = 1$

$\bar{x}_i - \bar{x}$ $\quad \mathbb{E}[\bar{x}_i - \bar{x}] = (\mu + \tau_i) - \mu = \tau_i$

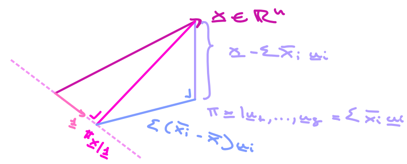
Goal: $H_0: \mu_1 = \mu_2 = \dots = \mu_g$ vs $H_1: H_0^c$

$H_0: \tau_1 = \tau_2 = \dots = \tau_g = 0$ vs $H_1: \exists \tau_i \neq 0$

Decomposition of variance: $\underline{x} = (x_{11}, \dots, x_{1n_1}, \dots, x_{2n_2}, \dots, x_{g1}, \dots, x_{gn_g})' \in \mathbb{R}^n$

$$\underline{w}_1 = \begin{pmatrix} 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}1\\ \vdots\\ 1\end{matrix}}\right\} n_1 \\ \\ \\ \end{matrix}, \quad \underline{w}_2 = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{matrix} \} n_1 \\ \} n_2 \\ \\ \end{matrix}, \quad \dots \quad \underline{w}_g = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \begin{matrix} \\ \\ \} n_g \end{matrix}$$

note that :
1) $\underline{w}_1, \dots, \underline{w}_g$ are lin. indep.
2) $\underline{w}_i' \underline{w}_j = 0$, $\underline{w}_1, \dots, \underline{w}_g$ orthogonal. indep.
3) Span $(\underline{w}_1, \dots, \underline{w}_g) \ni \underline{1}$



a. $\pi_{\underline{x} | \underline{w}_1, \dots, \underline{w}_g} = \sum_{i=1}^{g} \bar{x}_i \underline{w}_i$

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

c. $\pi_{\sum_{i=1}^{g} \bar{x}_i \underline{w}_i | \underline{1}} = \frac{\underline{1}\underline{1}'}{\underline{1}'\underline{1}} \sum \bar{x}_i \underline{w}_i = \sum_{i=1}^{g} \bar{x}_i \frac{\underline{1}\underline{1}'}{\underline{1}'\underline{1}} \underline{w}_i = \left(\frac{1}{n} \sum_{i=1}^{g} n_i \bar{x}_i \right) \underline{1} = \left(\frac{1}{n} \sum_{i=1}^{g} \sum_{j=1}^{n_i} x_{ij} \right) \underline{1} = \bar{x} \cdot \underline{1}$

$$\underline{x} = \bar{x} \cdot \underline{1} + \sum_{i=1}^{g} (\bar{x}_i - \bar{x}) \underline{w}_i + \left(\underline{x} - \sum_{i=1}^{g} \bar{x}_i \underline{w}_i \right)$$

orthogonal

$\| \underline{x} \|^2 = \| \bar{x} \cdot \underline{1} \|^2 + \| \sum (\bar{x}_i - \bar{x}) \underline{w}_i \|^2 + \| \underline{x} - \sum \bar{x}_i \underline{w}_i \|^2$

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

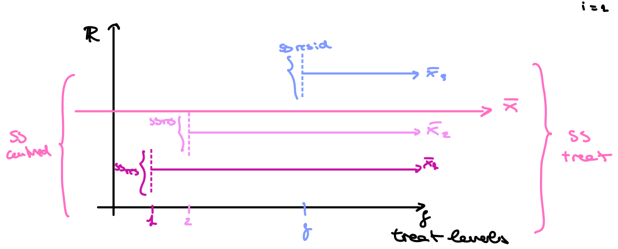
Decomposition of variance formula (vers 1):

$$\underbrace{\sum_{i=1}^{g} \sum_{j=1}^{n_i} x_{ij}^2}_{SS_{tot}} = \underbrace{n \bar{x}^2}_{SS_{mean}} + \underbrace{\sum_{i=1}^{g} n_i (\bar{x}_i - \bar{x})^2}_{SS_{treat}} + \underbrace{\sum_{i=1}^{g} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}_{SS \text{ residual}}$$

Decomposition of variance formula (vers 2):

$$\| \underline{x} - \bar{x} \cdot \underline{1} \|^2 = \| \sum (\bar{x}_i - \bar{x}) \underline{w}_i \|^2 + \| \underline{x} - \sum \bar{x}_i \underline{w}_i \|^2$$

$$\underbrace{\sum_{i=1}^{g} \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2}_{SS \text{ centred}} = \underbrace{\sum_{i=1}^{g} n_i (\bar{x}_i - \bar{x})^2}_{SS_{treat}} + \underbrace{\sum_{i=1}^{g} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}_{SS \text{ residual}}$$



$H_0: \tau_1 = \tau_2 = \dots = \tau_g = 0$ vs $H_1: \exists \tau_i \neq 0$

Idea: reject H_0 if SS_{treat} is large w.r.t. SS resid.

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

$S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$

Recall: $x_1, \dots, x_n$ iid $\sim N_1(\mu, \sigma^2)$

$S^2 = \frac{1}{n-1} \sum_{j=1}^{n} (x_j - \bar{x})^2$ $\qquad (n-1) S^2 \sim \sigma^2 \chi^2(n-1)$

Pooled est. of var: $\frac{1}{n-g} \sum (n_i - 1) S_i^2 = S_{pooled}$

If H_0 is true: $\mu_1 = \mu_2 = \dots = \mu_g$

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

If H_0 is true: SS centred $\sim \sigma^2 \chi^2(n-1)$
SS residual $\sim \sigma^2 \chi^2(n-g)$
SS treat $\sim$?

EXERCISE:

$X \sim N_n(\mu, \Sigma)$ $\quad P$ orth proj

$\Longrightarrow PX \perp\!\!\!\perp (I-P)X$

hint: $\begin{pmatrix} P \\ H-I \end{pmatrix} X$

$H_0: \tau_1 = \tau_2 = \dots = \tau_g = 0$ vs $H_1: \exists\, \tau_i \neq 0$

Idea: reject H_0 if $\frac{\text{SS treat}}{\text{SS res}}$ is big

$$\frac{\text{SS treat} / g-1}{\text{SS res} / n-g} \sim \frac{\sigma^2\chi^2(g-1)/(g-1)}{\sigma^2\chi^2(n-g)/(n-g)} = F(g-1, n-g)$$

$H_0: \mu_1 = \mu_2 = \dots = \mu_g$ vs $H_1: H_0^c$

$H_0: \tau_1 = \tau_2 = \dots = \tau_g$ vs $H_1: \exists\, \tau_i \neq 0$

Reject H_0 at level $\alpha \in (0,1)$: $\quad \frac{(\sum_i^g n_i(\bar{x}_i - \bar{x})^2)/(g-1)}{(\sum_i^g \sum_j^{n_i}(x_{ij} - \bar{x}_i)^2/(n-g)} > F_{1-\alpha}(g-1, n-g)$

MANOVA

Case: $p \geq 1$, $g \geq 1$

$$Y = \begin{cases} X_{11}, \dots, X_{1n_1} & \text{iid} \sim N_p(\mu_1, \Sigma) \\ X_{21}, \dots, X_{2n_2} & \text{iid} \sim N_p(\mu_2, \Sigma) \\ \vdots \\ X_{g1}, \dots, X_{gn_g} & \text{iid} \sim N_p(\mu_g, \Sigma) \end{cases}$$

Reparametrize: $\circledast \begin{cases} X_{ij} = \mu + \tau_i + \varepsilon_{ij} & i = 1, \dots, g \quad j = 1, \dots, n_i \\ \varepsilon_{ij} \text{ iid} \sim N_p(0, \Sigma) \end{cases}$ $\quad \mu, \tau_i \in \mathbb{R}^p$, $\sum_{i=1}^g n_i \tau_i = 0$

Obs: Fix component $k \in \{1, \dots, p\}$ $\circledast$ (component wise)

$x_{ij} = \mu_k + \tau_{ik} + \varepsilon_{ijk}$ $\quad \mu_k, \tau_{ik} \in \mathbb{R}$ $\quad$ (ANOVA)

ε_{ijk} iid $\sim N_1(0, \sigma_{kk})$ $\quad \sum_{i=1}^g n_i \tau_{ik} = 0$

We don't go trough this way, because we are missing correlation between the components

Decomposition of covariance: $\underbrace{\sum_{i=1}^g\sum_{j=1}^{n_i}(X_{ij} - \bar{X})(X_{ij} - \bar{X})'}_{\text{TOTAL COV}} = \underbrace{\sum_{i=1}^g n_i(\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})'}_{\text{COV. treat} = B} + \underbrace{\sum_{i=1}^g\sum_{j=1}^{n_i}(x_{ij} - \bar{X}_i)(x_{ij} - \bar{X}_i)'}_{\text{COV. residual} := W}$

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

$S_i = \sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)'$ est. of Σ in group i, $i = 1, \dots, g$

Proof: $X_{ij} - \bar{X} = (\bar{X}_i - \bar{X}) + (X_{ij} - \bar{X}_i)$ $\quad i = 1, \dots, g \quad j = 1, \dots, n_i$

$\Longrightarrow$ do the algebra

$H_0: \mu_1 = \mu_2 = \dots = \mu_g$ vs $H_1: H_0^c$

$H_0: \tau_1 = \tau_2 = \dots = \tau_g = 0$ vs $H_1: \exists\, \tau \neq 0$

Idea: reject H_0 if "B/W is large"

Test statistics:

- Wilks $\quad \Delta_W = \frac{\det(W)}{\det(B+W)}$ $\quad$ reject when it's small
- Pillai $\quad \Delta_P = tr(B \cdot (B+W)^{-1})$ $\quad$ reject when it's small
- Hotelling-Lawley $\quad \Delta_{LH} = tr(BW^{-1})$ $\quad$ reject when it's large

Obs: $\Delta_W, \Delta_P, \Delta_{LH}$ can be expressed in terms of eigenvalues of BW^{-1}

$\lambda_1, \dots, \lambda_s \quad s = \min(g-1, p)$

Case of Δ_ω

If H_0 is true, distribution of Δ_ω ?

Idea: reject H_0 if Δ_ω is small

Distribution of Δ_ω when H_0 is true:
- known for $p \geq 1$ and $g = 2, 3$;
- known for $g \geq 1$ and $p = 2$;
- unknown for general p, g.

Bartlet's asymptotic approx

H_0 is true: $\left(n - 1 - \frac{p+g}{2}\right) \log \Delta_\omega \sim \chi^2(p(g-1))$

For $\alpha \in (0,1)$ reject H_0 at level α if: $\left(n-1-\frac{p+g}{2}\right)\log \Delta_\omega > \chi_{1-\alpha}(p(g-1))$

If you reject H_0 $\Longrightarrow$ Confidence int for $\tau_{ie} - \tau_{ke}$ $\quad i,k = 1, \dots, g \quad e = 1, \dots, p$

$\alpha \in (0,1)$ $\quad$ Sim Bonf $CI_{1-\alpha}(\tau_{ie} - \tau_{ke})$ $\quad i,k = 1, \dots, g \quad e = 1, \dots, p$

Point estimat. for $\tau_{ie} - \tau_{ke}$: $\quad \bar{x}_i - \bar{x}$ estimator for τ_i

$(\bar{x}_{ie} - \bar{x}_e) - (\bar{x}_{ke} - \bar{x}_e) = \bar{x}_{ie} - \bar{x}_{ke}$

compl of $\bar{x}_i$ $\quad$ comp of $\bar{x}$

$\bar{x}_{ie} - \bar{x}_{ke} \sim N_1\left(\tau_{ie} - \tau_{ke},\ \sigma_{ee}\left(\frac{1}{n_i} + \frac{1}{n_k}\right)\right)$

σ_{ee} is the e-element on the diagonal of Σ

Est of σ_{ee}?

Est of Σ: $\quad W = \sum_{i=1}^{g}\sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)' = \sum_{i=1}^{g} (n_i-1) S_i$

$S_i = \frac{1}{n_i-1} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)' \qquad i = 1, \dots, g$

$S_{pooled} = \frac{1}{n-g} \sum_{i=1}^{g} (n_i - 1) S_i \qquad$ estim of Σ

Estimate for σ_{ee}: $\quad \frac{1}{n-g} w_{ee} = S_{pooled}(ee)$

$\text{Sim Bonf } CI_{1-\alpha}(\tau_{ie} - \tau_{ke}) = \left[\bar{x}_{ie} - \bar{x}_{ke} \pm t_{1-\frac{\alpha}{2}}(n-g)\sqrt{\frac{w_{ee}}{n-g}\left(\frac{1}{n_i} + \frac{1}{n_k}\right)}\right]$ $\qquad$ n° of CI's $= p\frac{(g-1)g}{2}$

$$\boxed{\text{Sim Bonf } CI_{1-\alpha}(\tau_{ie} - \tau_{ke}) = \left[\bar{x}_{ie} - \bar{x}_{ke} \pm t_{1-\frac{\alpha}{(g-1)gp}}(n-g)\sqrt{\frac{w_{ee}}{n-g}\left(\frac{1}{n_i} + \frac{1}{n_k}\right)}\right]}$$

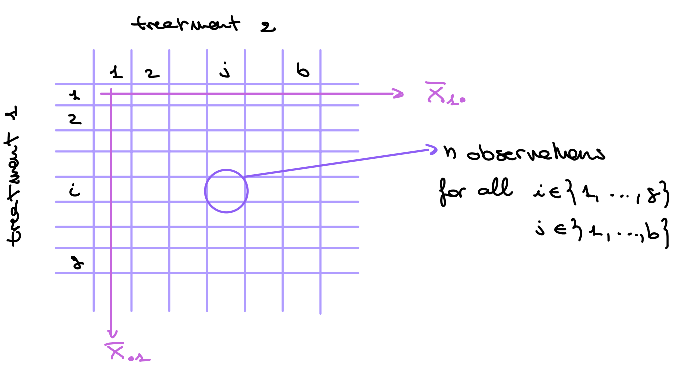
Short excursus on two-way ANOVA

extension $\begin{cases} \text{two-way MANOVA} \\ \text{three - four - } \dots \text{ way} \end{cases}$

Two treatments – two factors

$\rightarrow$ treat. 1 : g levels $\{1, \dots, g\}$

treat. 2 : b levels $\{1, \dots, b\}$



$\mathbb{R} \ni x_{ijk} = \mu_{ij} + \varepsilon_{ijk}$

$i = 1, \dots, g$

$j = 1, \dots, b$

$k = 1, \dots, n$ $\longrightarrow$ balanced exp

ε_{ijk} iid $\sim N_1(0, \sigma^2)$

$\bar{x}_{1\cdot} = \frac{1}{nb}\sum_{j=1}^{b}\sum_{k=1}^{n} x_{1jk}$ $\qquad$ $\bar{x}_{\cdot 1} = \frac{1}{ng}\sum_{i=1}^{g}\sum_{k=1}^{n} x_{i1k}$

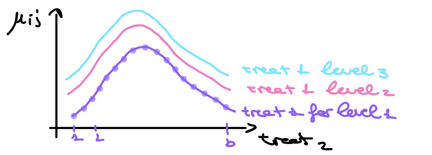
$\bar{x}_{i\cdot} = \frac{1}{nb}\sum_{j=1}^{b}\sum_{k=1}^{n} x_{ijk}$ $\qquad$ $\bar{x}_{ij} = \frac{1}{n}\sum_{k=1}^{n} x_{ijk}$ $\qquad$ $\bar{x} = \frac{1}{ngb}\sum_{j=1}^{b}\sum_{i=1}^{g}\sum_{k=1}^{n} x_{ijk}$

Model for μ_{ij}: 1. complete model: $\mu_{ij} = \mu + \tau_i + \beta_j + \gamma_{ij}$
2. Additive model: $\mu_{ij} = \mu + \tau_i + \beta_j$

Say $\gamma_{ij} = 0 \quad \forall ij$

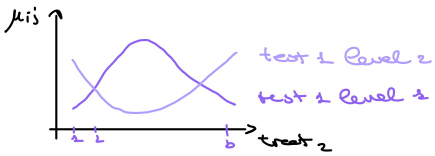
Fix level 1 for treat 1 $\quad \mu_{ij} = \mu + \tau_1 + \beta_j$

Fix level 2 for treat 1 $\quad \mu_{ij} = \mu + \tau_2 + \beta_j$



Fix level 1 for treat 1 $\quad \mu_{ij} = \mu + \tau_1 + \boxed{\beta_j + \gamma_{1j}}$

Fix level 2 for treat 1 $\quad \mu_{ij} = \mu + \tau_2 + \boxed{\beta_j + \gamma_{2j}}$



Estimates:

μ	$\bar{x}$
τ_i	$\bar{x}_{i\cdot} - \bar{x}$
β_j	$\bar{x}_{\cdot j} - \bar{x}$
γ_{ij}	$\bar{x}_{ij} - (\bar{x}_{i\cdot} - \bar{x}) - (\bar{x}_{\cdot j} - \bar{x}) - \bar{x} = \bar{x}_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x}$

Model: $\quad \mu_{ij} = \underset{1}{\mu} + \underset{g}{\tau_i} + \underset{b}{\beta_j} + \gamma_{ij} \qquad i = 1, \dots, g \qquad j = 1, \dots, b$

$$0 = \sum \tau_i = \sum \beta_j = \underbrace{\sum_i \gamma_{ij} = \sum_j \gamma_{ij}}_{g+b-1} \qquad \forall i, j$$

Decomposition of variance:

$$\sum_{i=1}^{g} \sum_{j=1}^{b} \sum_{k=1}^{n} (x_{ijk} - \bar{x})^2 = \sum_{i=1}^{g} (\bar{x}_{i\cdot} - \bar{x})^2 nb \quad \leftarrow SS_{treat.1} \quad (g-1)$$
$$+ \sum_{j=1}^{b} (\bar{x}_{\cdot j} - \bar{x})^2 ng \quad \leftarrow SS_{treat2} \quad (b-1)$$
$$+ \sum_i \sum_j (\bar{x}_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x})^2 n \quad \leftarrow SS_{interaction} \quad (g-1)(b-1)$$
$$+ \sum_i \sum_j \sum_k (x_{ijk} - \bar{x}_{jk}) \quad \leftarrow SS_{residuals} \quad gb(n-1)$$

$$\underline{x} = \begin{bmatrix} x_{111} \\ x_{112} \\ \vdots \\ x_{11n} \\ x_{121} \\ x_{122} \\ \vdots \end{bmatrix} \in \mathbb{R}^{gbn} \qquad \bar{x} \cdot \underline{1} = \frac{\underline{1}\,\underline{1}'}{\underline{1}'\underline{1}} \underline{x}$$

testing: $\quad H_0: \gamma_{ij} = 0 \quad \begin{matrix} i = 1, \dots, g \\ j = 1, \dots, b \end{matrix} \quad$ vs $\quad H_1: \exists\, \gamma_{ij} \neq 0$

$\alpha \in (0, 1)$

Reject H_0 at level α if: $\quad \dfrac{\dfrac{SS_{interaction}}{(g-1)(b-1)}}{\dfrac{SS_{residual}}{gb(n-1)}} > F_{1-\alpha}\big((g-1)(b-1), gb(n-1)\big)$

If you do not reject H_0 $\Longrightarrow$ move to the additive model $\quad \mu_{ij} = \mu + \tau_i + \beta_j$

$\Longrightarrow$ $SS_{res}(add) = SS_{inter}(comp) + SS_{res}(comp)$

$(g-1)(b-1) + gb(n-1)$