

Formulaire

1 Les moyennes

1.1 Moyenne arithmétique \bar{X}

Données non groupées

$$\frac{\sum_{i=1}^n X_i}{n}$$

Données groupées

$$\frac{\sum_{i=1}^n f_i * X_i}{\sum_{i=1}^n f_i}$$

1.2 Moyenne pondérée

$$\frac{\sum_{i=1}^n w_i * X_i}{\sum_{i=1}^n w_i}$$

1.3 Moyenne géométrique

$$MG = \sqrt[n]{\prod_{i=1}^n X_i}$$

1.4 Moyenne harmonique

$$\frac{1}{MH} = \frac{1}{n} * \sum_{i=1}^n \frac{1}{X_i}$$

2 Les mesures de dispersion

2.1 Variance S^2 des données non groupées

$$\frac{\sum_{i=1}^n x_i^2}{n-1} = \frac{\sum_{i=1}^n X_i^2 - \frac{\left(\sum_{i=1}^n X_i\right)^2}{n}}{n-1} \quad \text{avec} \quad x_i^2 = (X_i - \bar{X})^2$$

2.2 Variance S^2 des données groupées

$$\frac{\sum_{i=1}^n f_i * X_i^2 - \frac{\left(\sum_{i=1}^n f_i * X_i\right)^2}{\sum_{i=1}^n f_i}}{\sum_{i=1}^n f_i - 1}$$

2.3 Le coefficient de variation

$$\frac{S}{\bar{X}} * 100$$

2.4 L'erreur standard (Déviation standard de la moyenne)

$$S_{\bar{X}} = \sqrt{\frac{S^2}{n}}$$

3 Distributions discrètes

3.1 Binomiale

$$P(r) = B(r|n, p) = C_n^r * p^r * q^{n-r} \quad \text{avec} \quad p + q = 1$$

$$\text{où } C_n^r = \frac{n!}{r!(n-r)!}$$

3.2 Multinomiale

Pour le cas d'une distribution trinomiale:

$$p(k; l; m) = \frac{n!}{k!l!m!} * p^k * q^l * r^m \quad \text{avec} \quad p + q + r = 1$$

3.3 Binomiale négative

$$p(r) = C_{n-1}^{r-1} * p^r * q^{n-r}$$

3.4 Poisson

$$p(k) = \frac{e^{-\mu} * \mu^k}{k!}$$

3.5 Hypergéométrique

Dans un ensemble de taille N , constitué de deux sous-ensembles de tailles N_1 et $N_2 = N - N_1$, la probabilité de prélever $a_1 (< N_1)$ individus dans l'ensemble de taille N_1 quand on prélève un total de $a (< N)$ individus vaut:

$$P(a_1; N_1; a; N) = \frac{C_{N_1}^{a_1} * C_{N_2}^{a-a_1}}{C_N^a} \quad \text{avec} \quad a_1 + a_2 = a$$

4 Distributions continues

4.1 Normale

$$f(x) = \frac{1}{\sigma * \sqrt{2\pi}} * e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

4.2 Chi-carré

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - A_i)^2}{A_i}$$

où k est le nombre de catégories, O_i et A_i sont les nombres d'observés et d'attendus dans la catégorie i , respectivement.

5 Puissance

$$Puissance = Pr [Z \leq Z_1 (1 - \beta) \mid H_1]$$

où $Z_1 (1 - \beta)$ s'obtient par:

$$Z_1 (1 - \beta) = Z_0 (\alpha) + \frac{\sqrt{n}}{\sigma} * (\mu_0 - \mu_1)$$

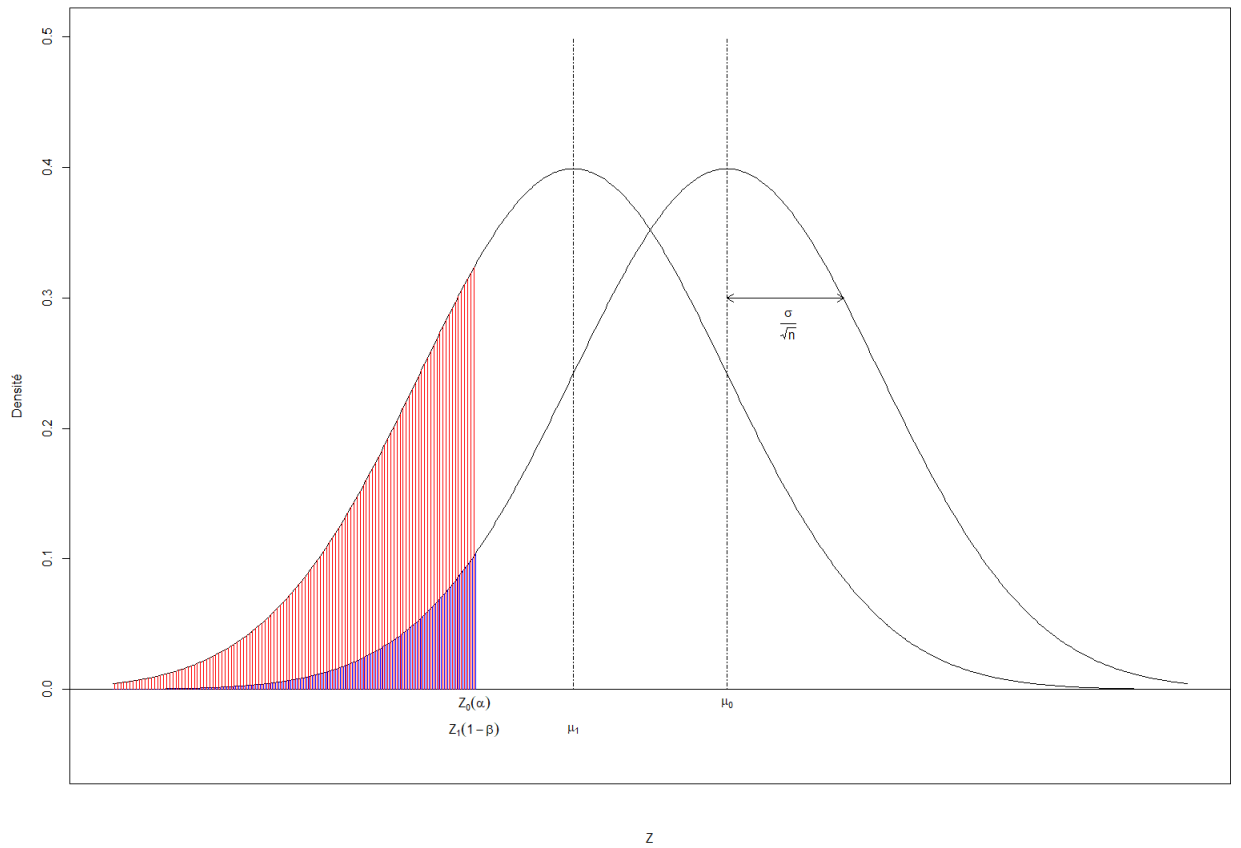


Figure 1: Représentation dans la puissance (surface rouge).

6 Comparaisons de moyennes

6.1 Un seul échantillon

6.1.1 Variance σ^2 connue - Test de Z

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

6.1.2 Variance σ^2 non connue - Test de t

$$t = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

6.2 Deux lots au hasard

6.2.1 Variance σ^2 connue - Test de Z

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

6.2.2 Variance σ^2 non connue - Test de t

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sum x_1^2 + \sum x_2^2}{n_1 + n_2 - 2} * (\frac{1}{n_1} + \frac{1}{n_2})}}$$

6.3 Données paarées

$$t = \frac{\bar{d} - \delta}{S_{\bar{d}}}$$

$$S_{\bar{d}} = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n_d}}{(n_d - 1) * n_d}}$$

7 Comparaisons de pourcentage

7.1 Comparaison d'un pourcentage obtenu \hat{p} à un pourcentage prévu p

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}}$$
$$Z = \frac{r - np}{\sqrt{npq}}$$

Remarque:

$$\chi^2 = \frac{(O_1 - A_1)^2}{A_1} + \frac{(O_2 - A_2)^2}{A_2} = Z^2$$

7.2 Comparaison de deux pourcentages observés

$$Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p} * \hat{q} * \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

où $p_1 = \frac{r_1}{n_1}$, $p_2 = \frac{r_2}{n_2}$ et $\hat{p} = \frac{r_1 + r_2}{n_1 + n_2}$

7.3 Comparaison de plus de deux pourcentages

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - A_i)^2}{A_i}$$

8 Intervalles de confiance

8.1 Moyenne (Variance connue)

$$P \left\{ \bar{x} - z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} \right\} = 1 - \alpha$$

où $z_{\frac{\alpha}{2}}$ est la valeur de z qui n'est dépassée qu'avec une probabilité de $\frac{\alpha}{2}$ dans la distribution de z . Par exemple, si $\alpha = 5\%$, $z_{\frac{\alpha}{2}} = 1.96$.

8.2 Moyenne (Variance non connue)

$$P \left\{ \bar{x} - t_{\frac{\alpha}{2}} * \frac{S}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\frac{\alpha}{2}} * \frac{S}{\sqrt{n}} \right\} = 1 - \alpha$$

8.3 Proportions

8.3.1 La limite supérieure p_s est telle que:

$$\sum_{k=0}^r C_n^k * p_s^k * (1 - p_s)^{n-k} = \frac{\alpha}{2}$$

8.3.2 La limite inférieure p_i est telle que:

$$\sum_{k=r}^n C_n^k * p_i^k * (1 - p_i)^{n-k} = \frac{\alpha}{2}$$

8.3.3 Approximation normale

$$P \left\{ \hat{p} - z_{\frac{\alpha}{2}} * \sqrt{\frac{\hat{p}\hat{q}}{n}} \leq p \leq \hat{p} + z_{\frac{\alpha}{2}} * \sqrt{\frac{\hat{p}\hat{q}}{n}} \right\} \approx 1 - \alpha$$

8.4 Variance

$$P \left\{ \frac{\sum x^2}{\chi_{1-\frac{\alpha}{2}}^2} \leq \sigma^2 \leq \frac{\sum x^2}{\chi_{\frac{\alpha}{2}}^2} \right\} = 1 - \alpha$$

8.5 Prédiction de X au prochain tirage

$$P \left\{ \bar{x} - t_{\frac{\alpha}{2}} * \sqrt{S^2 * \left(\frac{n+1}{n}\right)} \leq X \leq \bar{x} + t_{\frac{\alpha}{2}} * \sqrt{S^2 * \left(\frac{n+1}{n}\right)} \right\} = 1 - \alpha$$

9 ANOVA

9.1 ANOVA I

$$SCA = \sum_i \sum_j (\bar{Y}_{i.} - \bar{Y}_{..})^2$$

$$SCE = \sum_i \sum_j (Y_{ij} - \bar{Y}_{i.})^2$$

$$SCT = \sum_i \sum_j (Y_{ij} - \bar{Y}_{..})^2$$

$$F = \left(\frac{SCA}{g-1} \right) / \left(\frac{SCE}{n-g} \right)$$

où g est le nombre de groupes, n est le nombre de données et SCA , SCE et SCT font référence aux sommes de carrés du facteur A , de l'erreur et totale, respectivement.

9.1.1 *Least Significant Difference* (LSD) - Plus petite différence significative

$$LSD = t_{n-g}(\alpha/2) * \sqrt{\frac{SCE}{n-g} * \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

où $t_{n-g}(\alpha/2)$ est la valeur de t qu'on ne dépasse qu'avec une probabilité de $\alpha/2$ dans la distribution avec $(n-g)$ degrés de liberté, et n_1 et n_2 sont les tailles des deux groupes qu'on compare.

9.2 ANOVA II croisée

$$SCA = \sum_i \sum_j (\bar{Y}_{i.} - \bar{Y}_{..})^2$$

$$SCB = \sum_i \sum_j (\bar{Y}_{.j} - \bar{Y}_{..})^2$$

$$SCE = \sum_i \sum_j (Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})^2$$

$$SCT = \sum_i \sum_j (Y_{ij} - \bar{Y}_{..})^2$$

$$F_A = \left(\frac{SCA}{g_A - 1} \right) / \left(\frac{SCE}{n - g_A - g_B + 1} \right)$$

$$F_B = \left(\frac{SCB}{g_B - 1} \right) / \left(\frac{SCE}{n - g_A - g_B + 1} \right)$$

où g_A est le nombre de groupes pour le facteur A , g_B est le nombre de groupes pour le facteur B et SCA , SCB , SCE et SCT font référence aux sommes de carrés du facteur A , du facteur B , de l'erreur et totale, respectivement.

9.3 ANOVA II hiérarchique

$$SCA = \sum_i \sum_j \sum_k (\bar{Y}_{i..} - \bar{Y}_{...})^2$$

$$SCAB = \sum_i \sum_j \sum_k (\bar{Y}_{ij.} - \bar{Y}_{i..})^2$$

$$SCE = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij.})^2$$

$$SCT = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{...})^2$$

$$F_A = \left(\frac{SCA}{g_A - 1} \right) / \left(\frac{SCE}{n - g_{AB}} \right)$$

$$F_{AB} = \left(\frac{SCAB}{g_{AB} - g_A} \right) / \left(\frac{SCE}{n - g_{AB}} \right)$$

où g_{AB} est le nombre de combinaisons des facteurs A et B , et SCA , $SCAB$, SCE et SCT font référence aux sommes de carrés du facteur A , du facteur B niché dans le facteur A , de l'erreur et totale, respectivement.

9.4 ANOVA II avec interaction

$$SCA = \sum_i \sum_j \sum_k (\bar{Y}_{i..} - \bar{Y}_{...})^2$$

$$SCB = \sum_i \sum_j \sum_k (\bar{Y}_{.j.} - \bar{Y}_{...})^2$$

$$SCI = \sum_i \sum_j \sum_k (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2$$

$$SCE = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij.})^2$$

$$SCT = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{...})^2$$

$$F_A = \left(\frac{SCA}{g_A - 1} \right) / \left(\frac{SCE}{n - g_{AB}} \right)$$

$$F_B = \left(\frac{SCB}{g_B - 1} \right) / \left(\frac{SCE}{n - g_{AB}} \right)$$

$$F_I = \left(\frac{SCI}{g_{AB} - g_A - g_B + 1} \right) / \left(\frac{SCE}{n - g_{AB}} \right)$$

où g_{AB} est le nombre de combinaisons des facteurs A et B , et SCA , SCB , SCI , SCE et SCT font référence aux sommes de carrés du facteur A , du facteur B , de l'interaction entre les facteurs A et B , de l'erreur et totale, respectivement.

10 Régression

Le modèle est:

$$Y_i = \alpha + \beta * X_i + e_i$$

Les valeurs prédites par le modèle sont:

$$\hat{Y}_i = a + b * X_i$$

où:

$$a = \bar{Y} - b * \bar{X} \quad \text{et} \quad b = \frac{\sum x*y}{\sum x^2}$$

10.1 Test de $H_0 : \beta = 0$

$$t_{n-2} = \frac{b - \beta}{S_b}$$

ou, de manière équivalente:

$$P \{ b - t_{\frac{\alpha}{2}} * S_b \leq \beta \leq b + t_{\frac{\alpha}{2}} * S_b \} = 1 - \alpha$$

avec:

$$S_b^2 = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{(n-2) * \sum_{i=1}^n (X_i - \bar{X})^2}$$

ou encore, en passant par une analyse de variance de la régression:

$$F_{1,n-2} = \left(\frac{SCR}{1} \right) / \left(\frac{SCE}{n-2} \right) = t_{n-2}^2$$

où $SCR = b^2 * \sum_{i=1}^n x_i^2 = b * \sum_{i=1}^n x_i * y_i$ et $SCE = \sum_{i=1}^n (y_i - b * x_i)^2$.

10.2 Intervalles de confiance de la prédiction

10.2.1 Moyenne des Y en fonction de X ($x = X - \bar{X}$)

$$P \left\{ \hat{Y} - t_{\frac{\alpha}{2}} S_{yx} \sqrt{\frac{1}{n} + \frac{x^2}{\sum x^2}} \leq \mu_{yx} \leq \hat{Y} + t_{\frac{\alpha}{2}} S_{yx} \sqrt{\frac{1}{n} + \frac{x^2}{\sum x^2}} \right\} = 1 - \alpha$$

où S_{yx} est la déviation standard de l'erreur de la régression de Y sur X.

10.2.2 Valeur de Y en fonction de X

$$P \left\{ \hat{Y} - t_{\frac{\alpha}{2}} S_{yx} \sqrt{1 + \frac{1}{n} + \frac{x^2}{\sum x^2}} \leq \mu_{yx} \leq \hat{Y} + t_{\frac{\alpha}{2}} S_{yx} \sqrt{1 + \frac{1}{n} + \frac{x^2}{\sum x^2}} \right\} = 1 - \alpha$$

où S_{yx} est la déviation standard de l'erreur de la régression de Y sur X.

10.3 Equation de régression multiple

Le modèle est:

$$Y(i) = \beta_0 + \beta_1 * X_1(i) + \beta_2 * X_2(i) + e(i)$$

Les valeurs prédites par le modèle sont:

$$\hat{Y}(i) = a + b_1 * X_1(i) + b_2 * X_2(i)$$

10.3.1 Calcul de a, b₁ et b₂

$$a = \bar{Y} - b_1 * \bar{X}_1 - b_2 * \bar{X}_2$$

où b_1 et b_2 sont les solutions de:

$$\begin{cases} b_1 * \sum x_1^2 + b_2 * \sum x_1 x_2 = \sum x_1 y \\ b_1 * \sum x_2 x_1 + b_2 * \sum x_2^2 = \sum x_2 y \end{cases}$$

les solutions de ce système d'équations peuvent être calculées en utilisant les formules:

$$b_1 = \frac{D_1}{D}$$

et

$$b_2 = \frac{D_2}{D}$$

où:

$$\begin{aligned} D &= \sum x_1^2 * \sum x_2^2 - \sum x_1 x_2 * \sum x_1 x_2 \\ D_1 &= \sum x_1 y * \sum x_2^2 - \sum x_1 x_2 * \sum x_2 y \\ D_2 &= \sum x_1^2 * \sum x_2 y - \sum x_1 y * \sum x_1 x_2 \end{aligned}$$

11 Corrélation simple

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

12 Tests non-paramétriques

12.1 Mann-Whitney

$$Z = \frac{|R_1 - n_1 * \frac{n_1+n_2+1}{2}|}{\sqrt{\frac{n_1*n_2*(n_1+n_2+1)}{12}}}$$

R_1 est la somme des rangs du groupe de taille n_1

12.2 Kruskal-Wallis

$$KW = \frac{12}{n(n+1)} \sum_{i=1}^k n_i \bar{r}_i^2 - 3(n+1)$$

12.3 Wilcoxon

$$Z = \frac{|T - \frac{N(N+1)}{4}|}{\sqrt{\frac{N(N+1)(2N+1)}{24}}}$$

T = La plus petite somme de rangs (+ ou -)

N = Le nombre de différences non-nulles

12.4 Spearman

$$r_S = 1 - \frac{6 * \sum_{i=1}^N (R_X(i) - R_Y(i))^2}{N * (N^2 - 1)}$$

Test de la corrélation de Spearman:

$$t_{N-2} = r_S * \sqrt{\frac{N-2}{1-r_S^2}}$$