TESTS D'HYPOTHESE

1 Etude sur un seul caractère.

1.1 Test de conformité à une loi

On considère une v.a. X et on désire tester l'ajustement de sa loi notée $\mathcal{L}$ à une loi connue $\mathcal{L}_0$. On va alors tester : $H_0 : \mathcal{L} = \mathcal{L}_0$ contre $H_1 : \mathcal{L} \neq \mathcal{L}_0$.

1. **Test du Khi2 :** On a n observations de la variable X partagées en k classes. Soit O_i l'effectif observé de la classe i. On calcul l'effectif théorique par $C_i = n \cdot \mathbb{P}\left(X \in classe_i / X \rightsquigarrow \mathcal{L}_0\right)$.

   On calcule la valeur de la variable aléatoire $K = \sum_{i=1}^{k} \frac{(O_i - C_i)^2}{C_i} \rightsquigarrow \chi^2\left(\nu = k - 1 - r\right)$, avec r le nombre de paramètres de la loi $\mathcal{L}_0$ qu'on a du estimer et on rejette H_0 si $\chi_c^2 > \chi_\alpha^2$.

2. **Test de Kolmogorov :** La statistique de Kolmogorov-Smirnov est $D_n = \sup_{x \in \mathbb{R}} \left|F_n(x) - F(x)\right|$. Si $D_n > \frac{c}{\sqrt{n}}$, on rejette H_0, c étant une valeur tabulée dépendant de n et de α.

1.2 Tests de conformité

1.2.1 Population gaussienne ou taille d'échantillon $n > 30$.

Test de conformité d'une	Hypothèses	Statistique	Intervalle d'acceptation
Moyenne avec variance σ^2 connue	$H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$	$Z = \frac{\overline{X} - \mu_0}{\sigma_0 / \sqrt{n}} \rightsquigarrow \mathcal{N}(0,1)$	$I_\alpha = \left[-z_{1-\frac{\alpha}{2}}, +z_{1-\frac{\alpha}{2}}\right]$
	$H_0 : \mu \leq \mu_0$ vs $H_1 : \mu > \mu_0$		$I_\alpha = \left]-\infty, z_{1-\alpha}\right]$
	$H_0 : \mu \geq \mu_0$ vs $H_1 : \mu < \mu_0$		$I_\alpha = \left[-z_{1-\alpha}, +\infty\right[$
Moyenne avec variance σ^2 inconnue	$H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$	$T = \frac{\overline{X} - \mu_0}{S / \sqrt{n-1}} \rightsquigarrow \mathcal{T}_{n-1}$	$I_\alpha = \left[-t_{1-\frac{\alpha}{2}}, +t_{1-\frac{\alpha}{2}}\right]$
	$H_0 : \mu \leq \mu_0$ vs $H_1 : \mu > \mu_0$		$I_\alpha = \left]-\infty, t_{1-\alpha}\right]$
	$H_0 : \mu \geq \mu_0$ vs $H_1 : \mu < \mu_0$		$I_\alpha = \left[-t_{1-\alpha}, +\infty\right[$
Proportion	$H_0 : p = p_0$ vs $H_1 : p \neq p_0$	$Z = \frac{\sqrt{n}\left(\overline{X} - p\right)}{\sqrt{p(1-p)}} \rightsquigarrow \mathcal{N}(0,1)$	$I_\alpha = \left[-z_{1-\frac{\alpha}{2}}, +z_{1-\frac{\alpha}{2}}\right]$
	$H_0 : p \leq p_0$ vs $H_1 : p > p_0$		$I_\alpha = \left]-\infty, z_{1-\alpha}\right]$
	$H_0 : p \geq p_0$ vs $H_1 : p < p_0$		$I_\alpha = \left[-z_{1-\alpha}, +\infty\right[$
Variance	$H_0 : \sigma = \sigma_0$ contre $H_1 : \sigma \neq \sigma_0$	$K = \frac{nS^2}{\sigma_0^2} \rightsquigarrow \chi_{n-1}^2$	$I_\alpha = \left[k_{\frac{\alpha}{2}}, k_{1-\frac{\alpha}{2}}\right]$
	$H_0 : \sigma \leq \sigma_0$ contre $H_1 : \sigma > \sigma_0$		$I_\alpha = \left[0, k_{1-\alpha}\right]$
	$H_0 : \sigma \geq \sigma_0$ contre $H_1 : \sigma < \sigma_0$		$I_\alpha = \left[k_{1-\alpha}, +\infty\right[$

1.3 Tests de comparaison

1.3.1 Population gaussienne ou tailles d'échantillons $n_1 > 30$ et $n_2 > 30$.

Test de comparaison de 2	Hypothèses	Statistique	Intervalle d'acceptation
Moyennes avec variance σ^2 connue	$H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$	$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1}{n_1} + \frac{\sigma_2}{n_2}}} \rightsquigarrow \mathcal{N}(0,1)$	$I_\alpha = \left[-z_{1-\frac{\alpha}{2}}, +z_{1-\frac{\alpha}{2}}\right]$
	$H_0 : \mu_1 \leq \mu_2$ vs $H_1 : \mu_1 > \mu_2$		$I_\alpha = \left]-\infty, z_{1-\alpha}\right]$
	$H_0 : \mu_1 \geq \mu_2$ vs $H_1 : \mu_1 < \mu_2$		$I_\alpha = \left[-z_{1-\alpha}, +\infty\right[$
Moyennes avec variance σ^2 inconnue	$H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$	$T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\left(\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}\right)}} \rightsquigarrow \mathcal{T}_{n_1+n_2-2}$	$I_\alpha = \left[-t_{1-\frac{\alpha}{2}}, +t_{1-\frac{\alpha}{2}}\right]$
	$H_0 : \mu_1 \leq \mu_2$ vs $H_1 : \mu_1 > \mu_2$		$I_\alpha = \left]-\infty, t_{1-\alpha}\right]$
	$H_0 : \mu_1 \geq \mu_2$ vs $H_1 : \mu_1 < \mu_2$		$I_\alpha = \left[-t_{1-\alpha}, +\infty\right[$
Proportions	$H_0 : p_1 = p_2$ vs $H_1 : p_1 \neq p_2$	$Z = \frac{f_1 - f_2}{\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}} \rightsquigarrow \mathcal{N}(0,1)$ avec $\widehat{p} = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$	$I_\alpha = \left[-z_{1-\frac{\alpha}{2}}, +z_{1-\frac{\alpha}{2}}\right]$
	$H_0 : p_1 \leq p_2$ vs $H_1 : p_1 > p_2$		$I_\alpha = \left]-\infty, z_{1-\alpha}\right]$
	$H_0 : p_1 \geq p_2$ vs $H_1 : p_1 < p_2$		$I_\alpha = \left[-z_{1-\alpha}, +\infty\right[$
Variances	$H_0 : \sigma_1 = \sigma_2$ vs $H_1 : \sigma_1 \neq \sigma_2$	$F = \frac{n_1 S_1^2}{n_2 S_2^2} \rightsquigarrow \mathcal{F}(n_1 - 1, n_2 - 1)$	$I_\alpha = \left[f_{\frac{\alpha}{2}}, f_{1-\frac{\alpha}{2}}\right]$
	$H_0 : \sigma_1 \leq \sigma_2$ vs $H_1 : \sigma_1 > \sigma_2$		$I_\alpha = \left[0, f_{1-\alpha}\right]$
	$H_0 : \sigma_1 \geq \sigma_2$ vs $H_1 : \sigma_1 < \sigma_2$		$I_\alpha = \left[f_{1-\alpha}, +\infty\right[$

1.3.2 Population inconnue et tailles d'échantillons $n_1 > 30$ ou $n_2 > 30$.

1. **Test de Mann et Whitney (**deux échantillons indépendants de tailles différentes)

Soit U la variable aléatoire de valeur $u = \min\{u_1, u_2\}$.

- Pour n_1 et n_2 quelconques, on lit dans les tables du test de Mann et Whitney le nombre m_α tel que, sous H_0, $\mathbb{P}(U \leq m_\alpha) = \alpha$ et $I_\alpha = \left]m_\alpha, +\infty\right[$

- Si n_1 et n_2 sont assez grands (20 en général), sous H_0, U suit approximativement la loi $\mathcal{N}(\mu, \sigma)$ avec $\mu = \frac{n_1 n_2}{2}$ et $\sigma = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$.

On lit u_α dans la table de l'écart réduit de la loi normale tel que $\mathbb{P}(|U| \geq u_\alpha) = \alpha$, on calcule $z_\alpha = \frac{u_\alpha - \mu}{\sigma}$ et $I_\alpha \in \left]-z_\alpha, z_\alpha\right[$.

2. **Test de Wilcoxon (**deux échantillons appariés i.e. de même taille)

Soit W la variable aléatoire de valeur $w = \min\{w_+, w_-\}$.

- Si $N \leq 25$, on lit dans les tables du test de Wilcoxon le nombre w_α tel que, sous H_0, $\mathbb{P}(U \geq w_\alpha) = \alpha$. On rejette H_0 au risque d'erreur α si $w \geq w_\alpha$.

- Si $N > 25$, sous H_0, W suit approximativement la $\mathcal{N}(\mu, \sigma)$ avec $\mu = \frac{N(N+1)}{2}$ et $\sigma = \sqrt{\frac{N(N+1)(2N+1)}{24}}$, on calcule $z_\alpha = \frac{u_\alpha - \mu}{\sigma}$ et $I_\alpha \in]-z_\alpha, z_\alpha[$.

1.3.3 On dispose de k échantillons indépendants.

On imerai comparer simultanément les moyennes

- Populations gaussiennes de même variance : **ANOVA** à un facteur et on vérifie l'ègalité des variances par le **test de Bartlett** pour cela on calcule

$$\begin{aligned} B &= \frac{1}{\lambda}\left[(n-k)\ln S_R^2 - \sum_{i=1}^{k}(n_i - 1)\ln S_i^2\right] \rightsquigarrow \chi^2(k-1) \\ \text{avec } \lambda &= 1 + \frac{1}{3(k-1)}\left[\left(\sum_{i=1}^{k}\frac{1}{n_i - 1}\right) - \frac{1}{n-k}\right]. \end{aligned}$$

On calcule la valeur b prise par la v.a. B. Si $b \geq \chi^2_\alpha(k-1)$ on rejette l'hypothèse H_0.

Pour l'**ANOVA** on calcule la valeur de

$$F = \frac{\frac{S_F^2}{k-1}}{\frac{S_R^2}{n-k}} \rightsquigarrow \mathcal{F}(k-1, n-k)$$

au seuil α on rejette H_0 si $F_c = \frac{s_F^2}{k-1}\frac{n-k}{s_R^2} \geq F_{k-1,n-k}(\alpha)$, avec s_F^2 est la variance factorielle et S_R^2 est la variance résiduelle.

- Populations inconnues: **Test de Kruskal et Wallis**. On calcule le nombre

$$h = \frac{12}{n(n+1)}\left(\sum_{i=1}^{k}\frac{r_i^2}{n_i}\right) - 3(n+1).$$

Soit H la v.a. qui prend la valeur h.

  - Si $\forall i, n_i > 5$, alors sous H_0, $H \rightsquigarrow \chi^2(k-1)$. Dans la table de Khi2, on lit la valeur χ^2_α et on rejette H_0 si $h \geq \chi^2_\alpha$.
  - Si les $n_i \leq 5$, on a des tables de Kruskal-Wallis qui donnent la valeur h_α telle que $\mathbb{P}(h \geq h_\alpha)$. On rejette H_0 si $h \geq h_\alpha$.

2 Etude simultanée de deux caractères

2.1 Etude de l'indépendance

1. les deux caractères sont qualitatifs : On calcule la valeur de la variable

$$\sum_{i,j} \frac{(n_{ij} - C_{ij})^2}{C_{ij}} = \chi_c^2 \rightsquigarrow \chi^2(\nu) \text{ où } \nu = (l-1) \times (k-1).$$

On cherche la valeur critique χ_α^2 dans la table de la loi du $Khi2$ à ν degrés de liberté. Décision : si $\chi_c^2 < \chi_\alpha^2$, on accepte l'hypothèse H_0.

2. Les deux caractères sont quantitatifs

   - Populations gaussiennes: On calcule la valeur de la variable $T = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}} \rightsquigarrow \mathcal{T}(\nu = n-2)$ .Puis on déterminera t_α ou $t_{\frac{\alpha}{2}}$ de la table de la loi de students à $\nu = n-2$ degrés de liberté et on adoptera la règle de décision suivante :

- Si $H_1 : \rho \neq 0$: rejet de H_0 au risque α si $t_c \notin \left]-t_{\frac{\alpha}{2}}; t_{\frac{\alpha}{2}}\right[$;
- Si $H_1 : \rho > 0$: rejet de H_0 au risque α si $t_c > t_\alpha$;
- Si $H_1 : \rho < 0$: rejet de H_0 au risque α si $t_c < -t_\alpha$.

1. - Populations de lois inconnues: **Test de Spearman** on calcule $r_S = 1 - \frac{6\sum(x_i' - y_i')^2}{n(n^2-1)}$ à partir des couples des rangs.

- Si $n \leq 13$. Pour un risque α on détermine la valeur de r_α telle $P(|R_S| > r_\alpha) = \alpha$ lue dans la table de Spearman. Si $|r_S| > r_\alpha$ on rejette H_0 avec un risque α de se tromper.
- Si $n > 13$. Dans ce cas si H_0 est vraie, la statistique $T = \frac{R_S\sqrt{n-2}}{\sqrt{1-R_s^2}}$ suit approximativement la loi de Student à $n-2$ ddl. La décision se fera à l'aide la table du coefficient de corrélation linéaire qui donne la valeur de r_α telle que $P(|R_S| > r_\alpha) = \alpha$.

3. Un caractère qualitatif et l'autre est quantitatif:

   - Populations gaussiennes: **ANOVA** à un facteur
   - Populations de lois inconnues: **Test de Kruskal et Wallis.**