Tests Statistiques sur un Echantillon (1)

On suppose en préalable que des différences peuvent intervenir entre l'échantillon et la population générale. Le test permet de déterminer si ces différences sont l'effet du hasard (différence non significative) ou si elles trouvent leur origine dans la population générale (différence significative).

Type de Variable	Tests utilisables	Formule du test	Conditions du test
Variables qualitatives catégorielles *comparaison entre fréquence observée et fréquence théorique*	Ecart réduit	$Z = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}$	Grand échantillon : n π ET n(1-π) sup. ou = à 5
	Test du χ^2 (k-1 ddl)	$\frac{(N1-\upsilon 1)^2}{\upsilon 1} + \frac{(N2-\upsilon 2)^2}{\upsilon 2} + \ldots \frac{(Nk-\upsilon k)^2}{\upsilon k}$	Grand échantillon : $\upsilon 1 =$ n π ET $\upsilon 2 =$ n(1-π) sup. ou = à 5
Variables qualitatives catégorielles *test de lien entre deux variables*	Ecart réduit	$p = \frac{p1n1 + p2n2}{n1 + n2}$ $Z = \frac{\lvert p1 - p2 \rvert}{\sqrt{\frac{p(1-p)}{n1} + \frac{p(1-p)}{n2}}}$	Grand échantillon : n1 π ET n2 Sup. ou = à 5
	Test du χ^2 de Pearson (Grands échantillons) cf. Fiche suivante «Tests Statistiques sur un Echantillon (2) »		
	Test exact de Fisher (Petits échantillons) cf. Fiche suivante «Tests Statistiques sur un Echantillon (2) »		
Variables qualitatives catégorielles *mesure de la force d'un lien entre deux variables* → Odds-ratio (Cf. Cours d'Epidémiologie)			
Variables quantitatives *comparaison entre une moyenne observée (m) et une valeur théorique (μ)*	Ecart réduit	$Z = \frac{\lvert m - \mu \rvert}{\frac{s}{\sqrt{n}}}$	Grand échantillon (n >ou = à 30)
	T de student (n-1 ddl)	$t_{n-1ddl} = \frac{\lvert m - \mu \rvert}{\frac{s}{\sqrt{n}}}$	Petits échantillons (n < à 30) Possible aussi si grands échantillons.
Variables quantitatives *comparaison entre deux moyennes observées sur des échantillons indépendants*	Ecart réduit = Test de Behrens Fisher	$Z = \frac{\lvert m1 - m2 \rvert}{\sqrt{\frac{s1^2}{n1} + \frac{s2^2}{n2}}}$	Grand échantillon (n >ou = à 30)
	T de student (n1+n2-2 ddl)	$s^2 = \frac{s1^2(n1-1) + s2^2(n2-1)}{n1 + n2 - 2}$ $t_{n1+n2-2} = \frac{\lvert m1 - m2 \rvert}{\sqrt{\frac{s^2}{n1} + \frac{s^2}{n2}}}$	Petits échantillons (n < à 30) Possible aussi si grands échantillons.

Variables quantitatives *comparaison entre les distributions de deux séries indépendantes A et B = tests non paramétriques (Pour le détail du calcul des rangs, cf. cours)*	Test de Mann et Whitney	U = somme des nombres de valeur B à gauche de chaque valeur de A. $$Z = \frac{\left\lvert U - \frac{1}{2} n_A n_B \right\rvert}{\sqrt{n_A n_B \frac{n_A + n_B + 1}{12}}}$$	U+U' = $n_A \cdot n_B$ n_A et n_B > ou = à 10
	W de Wilcoxon	S=somme des rangs observée de A. $$Z = \frac{\lvert S - n_A(n_A + n_B + 1)/2 \rvert}{\sqrt{n_A n_B (n_A + n_B + 1)/12}}$$	U+U' = $n_A \cdot n_B$ n_A et n_B > ou = à 10
Variables quantitatives *comparaison entre plus de deux moyennes observées sur des échantillons indépendants*	ANOVA = Analyse de Variance	$$F_{n-k}^{k-1} = \frac{\text{var}\,iance.entre.colonnes}{\text{var}\,iance.résiduelle}$$ $$F_{n-k}^{k-1} = \frac{(1)/k-1}{(3)-(1)/n-k}$$ (Cf. tableau ci-dessous)	n = nb total de sujets ni = nb de sujets dans la catégorie i k = nb de catégories Ti = somme des valeurs sur les ni sujets de la catégorie i Tg = total des valeurs sur l'ensemble des n sujets *Interprétation en utilisant la table du F de Fisher*

ANOVA				
Origine de la fluctuation	Somme des carrés des écarts par rapport à la moyenne	ddl	Variance = carré moyen	Rapport des variances F_{n-k}^{k-1}
Entre colonnes	$(1) = \sum_{i=1}^{k} \frac{Ti^2}{ni} - \frac{Tg^2}{n}$	k-1	(1) / k-1	$\frac{(1)/k-1}{(3)-(1)/n-k}$
Intra colonne ou résiduelle	(3) - (1)	n-k	(3) – (1) / n-k	
Totale	$(3) = \sum_{j=1}^{n} x_j{}^2 - \frac{Tg^2}{n}$	n-1		

Variables quantitatives *Liaison entre deux variables gaussiennes*	Coef. de corrélation de Pearson (n-2 ddl)	$$r = \frac{Sxy - \frac{Sx.Sy}{n}}{\sqrt{(Sx^2 - \frac{(Sx)^2}{n}).(Sy^2 - \frac{(Sy)^2}{n})}}$$	*n = nb de variables* *Sx = somme des x* *Sx² = somme des x²* *Sy = somme des y* *Sy² = somme des y²* *Sxy = somme ds produits xy*
	Passage du r de Pearson en t de student (n-2 ddl)	$$t_{n-2ddl} = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$	