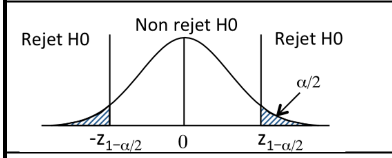
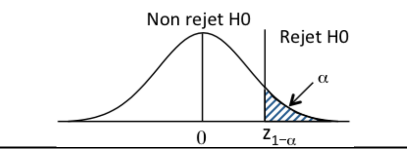
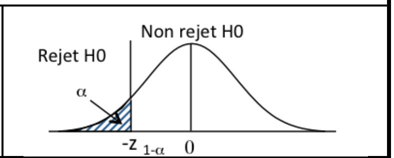
Formulaire d'inférence statistique pour 1 ou 2 variables

Une variable quantitative Normale X ~ N(μ,σ^2)

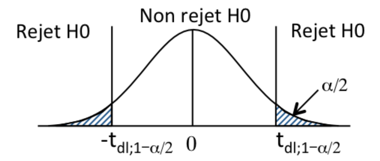
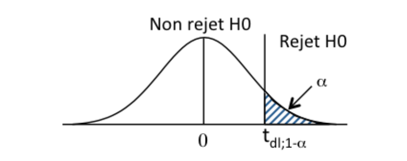
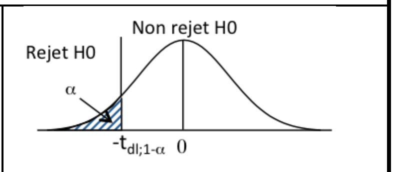
Intervalle de confiance sur une moyenne μ à <u>σ^2 connu</u> : $\left[\bar{X} - z_{1-\alpha/2}\sigma/\sqrt{n}, \bar{X} + z_{1-\alpha/2}\sigma/\sqrt{n}\right]$

Test Z sur une moyenne μ à σ^2 connu :

<table>
<tr><td>H_0 $\mu=\mu_0$ H_1 $\mu\neq\mu_0$</td><td>H_0 $\mu\leq\mu_0$ H_1 $\mu>\mu_0$</td><td>H_0 $\mu\geq\mu_0$ H_1 $\mu<\mu_0$</td></tr>
<tr><td colspan="3">Statistique de test : $z_{obs} = \dfrac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ ~ N(0,1) sous H_0</td></tr>
<tr><td>
</td><td>
</td><td>
</td></tr>
<tr><td>p-valeur = 2 x P(Z > $|z_{obs}|$)</td><td>p-valeur = P(Z > z_{obs})</td><td>p-valeur = P(Z < z_{obs})</td></tr>
</table>

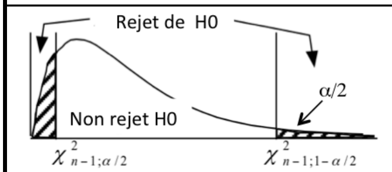
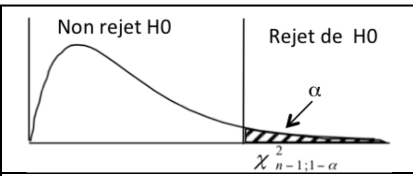
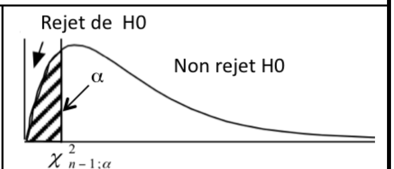
Intervalle de confiance sur une moyenne μ à <u>σ^2 inconnu</u> : $\left[\bar{X} - t_{n-1;1-\alpha/2}s/\sqrt{n}, \bar{X} + t_{n-1;1-\alpha/2}s/\sqrt{n}\right]$

Test t sur une moyenne μ à σ^2 inconnu :

<table>
<tr><td>H_0 $\mu=\mu_0$ H_1 $\mu\neq\mu_0$</td><td>H_0 $\mu\leq\mu_0$ H_1 $\mu>\mu_0$</td><td>H_0 $\mu\geq\mu_0$ H_1 $\mu<\mu_0$</td></tr>
<tr><td colspan="3">Statistique de test : $t_{obs} = \dfrac{\bar{X}-\mu_0}{s/\sqrt{n}}$ ~ t_{dl} sous H_0 avec dl = n-1</td></tr>
<tr><td>
</td><td>
</td><td>
</td></tr>
<tr><td>p-valeur = 2 x P(t_{dl} > $|t_{obs}|$)</td><td>p-valeur = P(t_{dl} > t_{obs})</td><td>p-valeur = P(t_{dl} < t_{obs})</td></tr>
</table>

Intervalle de confiance sur une variance σ^2 : $\left[\dfrac{(n-1)s^2}{\chi^2_{n-1;1-\alpha/2}}, \dfrac{(n-1)s^2}{\chi^2_{n-1;\alpha/2}}\right]$

Test sur une variance σ^2 :

<table>
<tr><td>H_0 $\sigma^2=\sigma_0^2$ H_1 $\sigma^2\neq\sigma_0^2$</td><td>H_0 $\sigma^2\leq\sigma_0^2$ H_1 $\sigma^2>\sigma_0^2$</td><td>H_0 $\sigma^2\geq\sigma_0^2$ H_1 $\sigma^2<\sigma_0^2$</td></tr>
<tr><td colspan="3">Statistique de test : $\chi^2_{obs} = \dfrac{(n-1)s^2}{\sigma_0^2} \sim \chi^2_{\text{n-1}}$ sous H_0</td></tr>
<tr><td>
</td><td>
</td><td>
</td></tr>
<tr><td>p-valeur = pas utile</td><td>p-valeur = pas utile</td><td>p-valeur = pas utile</td></tr>
</table>

Une variable qualitative (et v.a. de comptage associée binomiale (k=2) ou multinomiale (k≥2))

Intervalle de confiance sur une proportion en grand échantillon (X_{obs} et $(n-X_{obs})$ ≥5)

$$\left[\frac{X_{obs}}{n}-z_{1-\alpha/2}\sqrt{\frac{(X_{obs}/n)(1-X_{obs}/n)}{n}},\frac{X_{obs}}{n}+z_{1-\alpha/2}\sqrt{\frac{(X_{obs}/n)(1-X_{obs}/n)}{n}}\right]$$

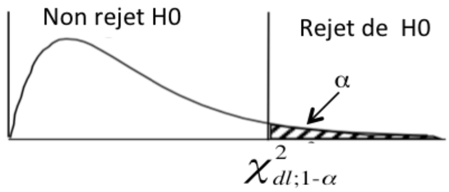
Test binomial exact relatif à une proportion (en petit échantillon)

$H_0\ \pi = \pi_0 \quad H_1\ \pi \neq \pi_0$	$H_0\ \pi \leq \pi_0 \quad H_1\ \pi > \pi_0$	$H_0\ \pi \geq \pi_0 \quad H_1\ \pi < \pi_0$
Statistique de test : X_{obs} ~ Bi(n, π_0) sous H_0		
Règle de décision basée sur la p-valeur		
P-valeur = 2xminimum (P(Bi(n,π_0)≤X_{obs}),P(Bi(n,π_0)≥X_{obs}))	P-valeur = P(Bi(n,π_0)≥X_{obs})	P-valeur = P(Bi(n,π_0)≤X_{obs})

Test Z relatif à une proportion (pour grands échantillons : $n\pi_0$ et $n(1-\pi_0)$≥5)

$H_0\ \pi = \pi_0 \quad H_1\ \pi \neq \pi_0$	$H_0\ \pi \leq \pi_0 \quad H_1\ \pi > \pi_0$	$H_0\ \pi \geq \pi_0 \quad H_1\ \pi < \pi_0$
Statistique de test : $z_{obs} = \dfrac{X_{obs}-n\pi_0}{\sqrt{n\pi_0(1-\pi_0)}}$ ≈ N(0,1) sous H_0		
Règle de décision et p-valeurs : voir test Z sur une moyenne à variance connue		

Test d'ajustement chi-carré à un critère de classification

H_0 : $\pi_1 = \pi_{10}, \pi_2 = \pi_{20}, \ldots \pi_k = \pi_{k0}$	H_1 : au moins une des égalités n'est pas vérifiée
Statistique de test : $\chi^2_{obs} = \sum_{i=1}^{k}\frac{(X_i-n\pi_{i0})^2}{n\pi_{i0}} = \sum_{i=1}^{k}\frac{(Obs_i-Att_i)^2}{Att_i} \approx \chi^2_{dl}$ sous H_0 avec dl = k - 1	
	p-valeur = $P(\chi^2_{dl} \geq \chi^2_{obs})$

Une variable quantitative VD et une variable qualitative VI (k=2 et groupes indépendants)

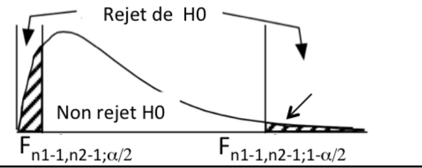
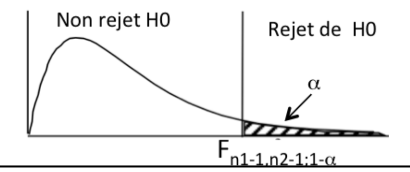
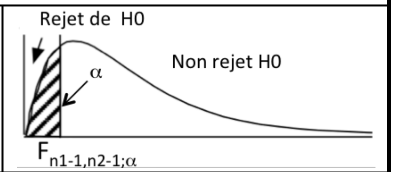
Test t de comparaison de deux moyennes, variances inconnues et supposées égales (données Normales)

$H_0\ \mu_1 = \mu_2 \quad H_1\ \mu_1 \neq \mu_2$	$H_0\ \mu_1 \leq \mu_2 \quad H_1\ \mu_1 > \mu_2$	$H_0\ \mu_1 \geq \mu_2 \quad H_1\ \mu_1 < \mu_2$
Statistique de test : $t_{obs} = \dfrac{\bar{X}_1-\bar{X}_2}{\sqrt{s^{*2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim t_{dl}$ sous H_0 avec $dl = n_1+n_2-2$, $s^{*2} = \dfrac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$		
Règle de décision et p-valeurs : voir test t sur une moyenne à variance inconnue		

Test t de comparaison de deux moyennes, variances inconnues et supposées ≠ (données Normales)

H_0 $\mu_1 = \mu_2$ H_1 $\mu_1 \neq \mu_2$	H_0 $\mu_1 \leq \mu_2$ H_1 $\mu_1 > \mu_2$	H_0 $\mu_1 \geq \mu_2$ H_1 $\mu_1 < \mu_2$
Statistique de test : $t_{obs} = \dfrac{(\overline{X}_1 - \overline{X}_2)}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$ ~ t_{dl} sous H_0 avec $dl = \dfrac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\dfrac{\left(s_1^2/n_1\right)^2}{n_1 - 1} + \dfrac{\left(s_2^2/n_2\right)^2}{n_2 - 1}}$		
Règle de décision et p-valeurs : voir test t sur une moyenne à variance inconnue		

Test F de comparaison de deux variances (données Normales)

H_0 $\sigma_1^2 = \sigma_2^2$ H_1 $\sigma_1^2 \neq \sigma_2^2$	H_0 $\sigma_1^2 \leq \sigma_2^2$ H_1 $\sigma_1^2 > \sigma_2^2$	H_0 $\sigma_1^2 \geq \sigma_2^2$ H_1 $\sigma_1^2 < \sigma_2^2$
Statistique de test : $F_{obs} = \dfrac{s_1^2}{s_2^2}$ ~ $F_{n1-1,n2-1}$ sous H_0		
		
p-valeur = pas utile	p-valeur = pas utile	p-valeur = pas utile

Test de la somme des rangs de Wilcoxon de comparaison de 2 tendances centrales (petit échantillon)

H_0 $C_1 = C_2$ H_1 $C_1 \neq C_2$	H_0 $C_1 \leq C_2$ H_1 $C_1 > C_2$	H_0 $C_1 \geq C_2$ H_1 $C_1 < C_2$
Statistique de test : - Transformer l'ensemble des données en rangs - Calculer la somme des rangs W_S du groupe 1 (qui est celui qui a la plus petite taille si les groupes sont de tailles inégales) - Calculer la somme W_S' des rangs du groupe 1 si les données avaient été classées en ordre décroissant. $W_S' = 2\overline{W} - W_S$ où $2\overline{W} = n_1(n_1 + n_2 + 1)$ - Calculer le minimum en W_S et W_S' : $\min(W_S, W_S')$		
Règle de décision : Trouver le seuil critique dans la table en fonction de n_1, n_2 et α ou $\alpha/2$ selon que le test est unilatéral ou bilatéral. Rejeter H_0 si le $\min(W_S, W_S')$ est ≤ que le seuil critique. Pour le test unilatéral, vérifier aussi que les sommes des rangs vont dans le sens de H_1.		
Statistique de test de Mann Whitney équivalente : $U = \text{minimum}(W_S, W_S') - \dfrac{n_1(n_1+1)}{2}$		

Test de la somme des rangs de Wilcoxon de comparaison de 2 tendances centrales (grand échantillon)

H_0 $C_1 = C_2$ H_1 $C_1 \neq C_2$	H_0 $C_1 \leq C_2$ H_1 $C_1 > C_2$	H_0 $C_1 \geq C_2$ H_1 $C_1 < C_2$
Statistique de test : $z_{obs} = \dfrac{W_s - \dfrac{n_1(n_1 + n_2 + 1)}{2}}{\sqrt{\dfrac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \approx N(0,1)$ sous H_0		
Règle de décision et p-valeurs : voir test Z sur une moyenne à variance connue		

Test t de comparaison de deux moyennes pour données pairées (Normales)

H_0 $\mu_1 = \mu_2$ H_1 $\mu_1 \neq \mu_2$ H_0 $\mu_D = 0$ H_1 $\mu_D \neq 0$	H_0 $\mu_1 \leq \mu_2$ H_1 $\mu_1 > \mu_2$ H_0 $\mu_D \leq 0$ H_1 $\mu_D > 0$	H_0 $\mu_1 \geq \mu_2$ H_1 $\mu_1 < \mu_2$ H_0 $\mu_D \geq 0$ H_1 $\mu_D < 0$
Déroulement du test - Calculer les différences entre les données pairées $D_i = X_{1i} - X_{2i}$ - Calculer la moyenne $\overline{D}$ des D_i et la variance s_D^2 des D_i - Appliquer un test t sur une moyenne à variance inconnue sur la moyenne des différences μ_D		

Test (non-paramétrique) des rangs de Wilcoxon de comparaison de 2 tendances centrales pour données pairées (petit échantillon)

H_0 $C_1 = C_2$ H_1 $C_1 \neq C_2$	H_0 $C_1 \leq C_2$ H_1 $C_1 > C_2$	H_0 $C_1 \geq C_2$ H_1 $C_1 < C_2$
Statistique de test : - Calculer les ≠ entre couples de données $D_i = X_{2i} - X_{1i}$, - Calculer les rangs des différences prises en valeur absolue. - Calculer la somme T+ des rangs correspondant à une différence positive - Calculer la somme T- des rangs correspondant à une différence négative		
Règle de décision : Rechercher le seuil dans la table en fonction de n et α ou $\alpha/2$ selon que le test est unilatéral ou bilatéral. Prendre le seuil donné à la 1ère ligne		
Rejeter H_0 si minimum (T+,T-) ≤ $T_{\alpha/2}$	Rejeter H_0 si T+ ≤ T_α	Rejeter H_0 si T- ≤ T_α

Test des rangs de Wilcoxon de comparaison de 2 tendances centrales pour données pairées (grand échantillon)

H_0 $C_1 = C_2$ H_1 $C_1 \neq C_2$	H_0 $C_1 \leq C_2$ H_1 $C_1 > C_2$	H_0 $C_1 \geq C_2$ H_1 $C_1 < C_2$
Statistique de test : - Calculer T- avec la méthode du test de rangs de Wilcoxon pour petits échantillons $z_{obs} = \dfrac{T_- - \dfrac{n(n+1)}{4}}{\sqrt{\dfrac{n(n+1)(2n+1)}{24}}} \sim N(0,1)$ sous H_0		
Règle de décision et p-valeurs : voir test Z sur une moyenne à variance connue		

Test de signe de comparaison de 2 tendances centrales pour données pairées

H_0 $C_1 = C_2$ H_1 $C_1 \neq C_2$	H_0 $C_1 \leq C_2$ H_1 $C_1 > C_2$	H_0 $C_1 \geq C_2$ H_1 $C_1 < C_2$
Calcul de la statistique de test - Calculer les différences entre les couples de données $D_i = X_{2i} - X_{1i}$ - Calculer D+ et D- : le nombre de différences positives et négatives (en ignorant celles qui sont nulles) - Soit n* le nombre de différences non nulles		
Règle de décision par la méthode de la p-valeur Tester avec un test binomial sur une proportion si la proportion de signes + et - est différente de 1/2		
p-valeur = 2xP(Bi(n*,1/2)≤min(D+,D-))	p-valeur = P(Bi(n*,1/2)≤D+)	p-valeur = P(Bi(n*,1/2)≤D-)

Deux variables qualitatives

Test Z de comparaison de deux proportions (groupes indépendants et grands échantillons)

$H_0\ \pi_1 = \pi_2$ $H_1\ \pi_1 \neq \pi_2$	$H_0\ \pi_1 \leq \pi_2$ $H_1\ \pi_1 > \pi_2$	$H_0\ \pi_1 \geq \pi_2$ $H_1\ \pi_1 < \pi_2$
Statistique de test : $z_{obs} = \dfrac{\dfrac{X_1}{n_1} - \dfrac{X_2}{n_2}}{\sqrt{\dfrac{X_1+X_2}{n_1+n_2}\left(1-\dfrac{X_1+X_2}{n_1+n_2}\right)\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}}$ ≈ N(0,1) sous H_0		
Règle de décision et p-valeurs : voir test Z sur une moyenne à variance connue		

Test χ^2 d'homogénéité des proportions d'une variable qualitative VD sur les niveaux d'une variable VI

Test équivalent au test χ^2 de Pearson d'indépendance

Test χ^2 de Pearson d'indépendance de deux variables qualitatives VD

H_0 : Les deux variables sont indépendantes H_1 : Les deux variables ne sont pas indépendantes
Statistique de test : $\chi^2_{obs} = \sum_{j=1}^{g}\sum_{i=1}^{k}\dfrac{(X_{ij}-n\hat{\pi}_{ij})^2}{n\hat{\pi}_{ij}} = \sum_{ij}\dfrac{(Obs_{ij}-Att_{ij})^2}{Att_{ij}} \sim \chi^2_{dl}$ sous H_0 avec $\hat{\pi}_{ij} = \dfrac{X_{i.}}{n}\dfrac{X_{.j}}{n}$ dl = (k-1)(g-1)
Règle de décision et p-valeurs : test chi-carré d'ajustement
Relation entre la statistique Chi-carré et le coefficient d'association V de Cramer $\chi^2_{obs} = V^2 n(t-1)$ et $V = \sqrt{\dfrac{\chi^2_{obs}}{n(t-1)}}$ et t=minimum(k,g)

Deux variables quantitatives VD

Intervalle de confiance sur un coefficient de corrélation ρ

Transformer r en r' à l'aide de la table

Calculer l'intervalle sur ρ' avec la formule : $\left[\rho'_m, \rho'_M\right] = \left[r' - z_{1-\alpha/2}\dfrac{1}{\sqrt{n-3}}\,;\, r' + z_{1-\alpha/2}\dfrac{1}{\sqrt{n-3}}\right]$

Puis transformer cet intervalle sur ρ' en un intervalle sur ρ à partir de la table (inverse)

Test t de nullité d'un coefficient de corrélation (sous l'hypothèse de Normalité des données)

$H_0\ \rho = 0$ $H_1\ \rho \neq 0$	$H_0\ \rho \leq 0$ $H_1\ \rho > 0$	$H_0\ \rho \geq 0$ $H_1\ \rho < 0$
Statistique de test : $t_{obs} = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{dl}$ sous H_0 et avec dl = n - 2		
Règle de décision et p-valeurs : voir test t sur une moyenne à variance inconnue		

Supplément : Calcul de la puissance d'un test sur une ou deux moyenne(s) en population normale

Puissance d'un test unilatéral sur une moyenne à variance connue

$$1-\beta = 1-\mathrm{P}\left(Z \le z_{1-\alpha} - \frac{|\mu_1 - \mu_0|}{\sigma/\sqrt{n}} \right) \text{avec } \overline{X} \sim N(\mu_1, \frac{\sigma^2}{n}) \text{ sous } \mathrm{H}_1$$

Fonction δ pour différents tests pour moyennes

Test t ou z sur une moyenne : $\delta = d\sqrt{n} = \dfrac{|\mu_1 - \mu_0|}{\sigma}\sqrt{n}$

Test t ou z sur deux moyennes avec $n_1=n_2=n$: $\delta = \dfrac{|\mu_2 - \mu_1|}{\sigma}\sqrt{\dfrac{n}{2}}$

Test t sur deux moyennes avec $n_1 \neq n_2$: $\delta = \dfrac{|\mu_2 - \mu_1|}{\sigma}\sqrt{\dfrac{n^*}{2}}$ où $n^* = \dfrac{2n_1n_2}{n_1 + n_2}$

Test t sur deux moyennes pour échantillons pairés : $\delta = \dfrac{|\mu_2 - \mu_1|}{\sigma}\sqrt{n}$