DOCUMENT À DISTRIBUER
À L'EXAMEN
DE STATISTIQUE INFÉRENTIELLE

INTERVALLE DE CONFIANCE POUR V.A. NORMALES

Données
Population : inconnue
Échantillon : $X_1, \ldots, X_n$, avec $X_k \sim \mathcal{N}\left(\mu, \sigma^2\right)$ de v.a.i.i.d.. On a espérance mathématique $E(X) = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, variance $S^2 = \frac{1}{n}\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2$ ou variance non biaisée $S^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2$.
Notons que $\overline{X}$ est une v.a. avec $\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$.
Tirage d'un échantillon : $x_1, \ldots, x_n$. On a moyenne empirique $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, variance empirique $s^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \overline{x})^2$ ou variance empirique non biaisée $s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \overline{x})^2$
Notation Dans la suite on notera par

$$\Sigma^2 = \begin{cases} \sigma^2, & \text{si l'écart-type est connu} \\ s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \overline{x})^2, & \text{si l'écart-type est inconnu} \end{cases}$$

Calcul d'un intervalle de confiance pour la moyenne

Problème 1.- Trouver un intervalle $[a, b] \subset \mathbb{R}$ telle que l'estimation $\widehat{\theta}$ de la moyenne μ de la population soit dans cet intervalle avec risque α, c-à-d. $P_\theta\left[a \le \theta \le b\right] = 1 - \alpha \;\; ; \;\; \alpha \in [0, 1]$.

Problème 2.- Calculer la taille de l'échantillon pour déterminer un intervalle $[a, b]$ dans lequel se trouve l'estimation $\widehat{\theta}$ de la moyenne μ avec un risque α et avec marge d'erreur $b - a = 2R_\Theta$.

Variance connue ou variance inconnue et taille de la population $n \geq 30$
Solution du problème 1 : Intervalle de confiance

$$-a = b = z_{\alpha/2}$$

et

$$P_\Theta\left\{\mu \in \left[\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]\right\} = 1 - \alpha$$

La valeur de $z_{\alpha/2}$ est lue dans les tables de la loi normale.
Solution du problème 2 : Taille de l'échantillon

$$n = \left[\frac{z_{\alpha/2} \cdot \Sigma}{R_\Theta}\right]^2$$

Variance inconnue et taille de la population $n < 30$
Solution du problème 1 : Intervalle de confiance

$$-a = b = t_{n-1,\alpha/2}$$

et

$$P_\mu\left\{\mu \in \left[\overline{x} - t_{n-1;\alpha/2}\frac{s}{\sqrt{n}},\ \overline{x} + t_{n-1;\alpha/2}\frac{s}{\sqrt{n}}\right]\right\} = 1 - \alpha$$

La valeur de $t_{n-1;\alpha/2}$ est lue dans les tables de la loi de Student.
Solution du problème 2 : Taille de l'échantillon

$$n = \left[\frac{t_{n-1;\alpha/2} \cdot s}{R_\Theta}\right]^2$$

Calcul d'un intervalle de confiance pour la variance

Problème 1.- Trouver un intervalle $[a,b] \subset \mathbb{R}$ telle que l'estimation $\widehat{\theta}$ de la variance σ^2 de la population soit dans cet intervalle avec risque α, c-à-d. $P_\theta\left[a \leq \sigma^2 \leq b\right] = 1-\alpha \;\; ; \; \alpha \in [0,1]$.
Problème 2.- Calculer la taille de l'échantillon pour déterminer un intervalle $[a,b]$ dans lequel se trouve l'estimation $\widehat{\theta}$ de la variance σ^2 avec un risque α etr avec marge d'erreur $b-a=R_\Theta$.

Moyenne connue ou inconnue
où χ^2_n suit la loi du χ^2 avec n ddl.
Solution du problème 1 : Intervalle de confiance

$$a=\chi^2_{\nu,1-\alpha/2}, \quad b=\chi^2_{\nu,\alpha/2}$$

et

$$P_{\sigma^2}\left\{s^2 \in \left[\frac{ns^2}{b},\; \frac{n\,s^2}{a}\right]\right\} = 1-\alpha$$

où on noté
nombre de degrés de liberté $\nu = \begin{cases} n, & \text{si moyenne connue} \\ n-1 & \text{si moyenne inconnue} \end{cases}$ et $s^2 = \begin{cases} \frac{1}{n-1}\sum\limits_{k=1}^{n}(x_k-\mu)^2, & \text{si moyenne } \mu \text{ connue} \\ \frac{1}{n-1}\sum\limits_{k=1}^{n}(x_k-\overline{x})^2 & \text{si moyenne inconnue} \end{cases}$
Solution du problème 2 : Taille de l'échantillon

$$n=\frac{R_\Theta}{s^2}\frac{1}{\frac{1}{b}-\frac{1}{a}}$$

INTERVALLE DE CONFIANCE POUR V.A. BINOMIALES

Données
Population : inconnue
Échantillon : $X_1,\ldots,X_n$ de v.a.i.i.d., avec $X_k \sim B(1,p)$. On a $\overline{X}=\frac{1}{n}\sum\limits_{i=1}^{n}X_i$.
Tirage d'un échantillon : $x_1,\ldots,x_n$. On a $\overline{x}=\frac{1}{n}\sum\limits_{i=1}^{n}x$
Problème : Trouver un intervalle $[a,b] \subset \mathbb{R}$ telle que l'estimation $\widehat{\theta}$ de la proportion q d'articles de la population ayant une caractéristique particulière, soit dans cet intervalle avec risque α, c-à-d. $P_\theta\left[a \leq q \leq b\right] = 1-\alpha \;\; ; \; \alpha \in [0,1]$.
Remarquons que $\overline{X}$ représente le nombre d'articles dans l'échantillon qui ont une caractéristique particulière et que l'on notera par $\widehat{Q}$. On a $\widehat{Q} \sim \mathcal{N}\left(q, \frac{q(1-q)}{n}\right)$ si $n\,\widehat{q} \geq 5$ et $n(1-\widehat{q}) \geq 5$.
La v.a. standarisée correspondant à $\widehat{Q}$ est $Z_Q = \frac{\widehat{q}-q}{\sqrt{\frac{q(1-q)}{n}}}$
Solution du problème : Soit dans la formule de la v.a. Z_Q on utilise l'estimation $\widehat{q}$ de q, soit on prend $q=\frac{1}{2}$ ce qui permet de maximiser la variance et, donc, l'intervalle de confiance. On a ainsi

$$P_Q\left\{\mu \in \left[\overline{x}-z_{\alpha/2}\sqrt{\frac{w(1-w)}{n}},\; \overline{x}+z_{\alpha/2}\sqrt{\frac{w(1-w)}{n}}\right]\right\} = 1-\alpha$$

avec $w = \begin{cases} \text{soit} & \widehat{q}, \\ \text{soit} & \frac{1}{2} \end{cases}$.
La valeur de $z_{\alpha/2}$ est lue dans les tables de la loi normale.

TESTS D'HYPOTHÈSES SUR LES MOYENNES

Données
Population : inconnue
Échantillon : $X_1, \ldots, X_n$, avec $X_k \sim \mathcal{N}\left(\mu, \sigma^2\right)$ de v.a.i.i.d.. On a espérance mathématique $E(X) = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, variance $S^2 = \frac{1}{n}\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2$ ou variance non biaisée $S^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2$.
Notons que $\overline{X}$ est une v.a. avec $\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$.
Tirage d'un échantillon : $x_1, \ldots, x_n$. On a moyenne empirique $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, variance empirique $s^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j - \overline{x})^2$ ou variance empirique non biaisée $s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \overline{x})^2$
Notation Dans la suite on notera par

$$\Sigma^2 = \begin{cases} \sigma^2, & \text{si variance est connue} \\ s^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2, & \text{si la variance est inconnue} \end{cases}$$

Paramètre à estimer La moyenne $\theta = \mu$ de la population totale et nous utiliserons $\overline{X}$ comme valeur de référence pour le test l'estimation. On prend comme hypothèse nulle
$H_0 : \theta = \theta_0$
et comme hypothèse alternative
$H_1 : \theta \neq \theta_0$
Parfois on cherche à détecter si $\theta = \theta_1$, avec $\theta_1 \neq \theta_0$, qui est l'hypothèse de détection :
$H_D : \theta = \theta_1$
Notons, enfin, que selon les valeurs que θ puisse prendre, nous avons un test simple ou un test double.

Test de la moyenne avec une valeur de référence. β fixé

La procédure du test est la suivante.

1°.- Il y a deux possibilités :

1.1 Test simple de la moyenne

$$H_0 \ : \ \theta = \theta_0 \ ; \ H_1 \ : \ \theta < \theta_0 \ (ou\ \theta > \theta_0) \ ; \ H_D \ : \ \theta = \theta_1$$

1.2 Test double de la moyenne

$$H_0 \ : \ \theta = \theta_0 \ ; \ H_1 \ : \ \theta \neq \theta_0 \ ; \ H_D \ : \mid \theta_1 - \theta_0 \mid = d$$

2°.- On fixe les probabilités α et β. Dans la suite on prendra

$$\alpha' = \begin{cases} \alpha, & \text{si test simple} \\ \alpha/2, & \text{si test double} \end{cases}$$

3°.- Calcul de la taille de l'échantillon :

3.1.- Calcul de la valeur $U_{\alpha'}$. Nous avons

$$P\left(\frac{\theta_0 - \overline{X}}{\Sigma/\sqrt{n}} > U_{\alpha'}\right) = \alpha'$$

avec

$$\Sigma^2 = \begin{cases} \sigma^2, & \text{si la variance } \sigma^2 \text{ de la population est connue} \\ s^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2 & \text{si la variance } \sigma^2 \text{ de la population est inconnue} \end{cases}$$

et

$$U_\alpha = \begin{cases} Z_{\alpha'} \sim \mathcal{N}\left(0,1\right), & \text{loi normale si la variance } \sigma^2 \text{ de la population est connue} \\ T_{n-1,\alpha'} \sim t_{n-1,\alpha'} & \text{loi de Student si la variance } \sigma^2 \text{ de la population est inconnue} \end{cases}$$

Par lecture des tables de la distribution normale (resp. de Student) on établit la valeur de $Z_{\alpha'}$ (resp. $T_{n-1,\alpha'}$).

3.2.- Calcul de la valeur U_β. Nous avons

$$P\left(\frac{\overline{X} - \theta_1}{\sigma/\sqrt{n}} > U_\beta\right) = \beta$$

avec

$$U_\beta = \begin{cases} Z_\beta \sim \mathcal{N}\left(0,1\right), & \text{loi normale si la variance } \sigma^2 \text{ de la population est connue} \\ T_{n-1,\beta} \sim t_{n-1,\beta} & \text{loi de Student si la variance } \sigma^2 \text{ de la population est inconnue} \end{cases}$$

Par lecture des tables de la distribution normale (resp. de Student) on établit la valeur de Z_β (resp. $T_{n-1,\beta}$).

3.3.- Calcul de la taille minimale n_C de l'échantillon :

$$n_C = \frac{(U_{\alpha'} + U_\beta)^2 \, \Sigma^2}{(\theta_0 - \theta_1)_D^2}$$

Si $n_C > n$, où n la taille de l'échantillon, il faut recommencer l'échantillonnage avec une valeur de n plus grande.

N.B. L'indice D à la différence $(\theta_0 - \theta_1)_D^2$ indique que nous devons travailler sous l'hypothèse de détection $\theta_0 \neq \theta_1$, sinon on risque de faire une division par 0.

4°.- Calcul de la valeur du critère C :

$$C = \pm\frac{1}{2}\,(U_\beta - U_{\alpha'})\sqrt{\frac{\Sigma^2}{n}} + \frac{(\theta_0 + \theta_1)_D}{2}$$

Il y a deux possibilités :

4.1 Test simple de la moyenne :
- On calcule C avec le signe "+" si $H_1 : \theta < \theta_0$.
- On calcule C avec le signe "-" si $H_1 : \theta > \theta_0$

4.2 Test double de la moyenne : On forme l'intervalle $[C_1, C_2]$, où $C_1 = C$ et $C_2 = 2C_1 - \theta_0$. (N.B. C est calculé avec le signe "+".

5°.- Décision. Il y a deux possibilités :

5.1 Test simple de la moyenne.

Si $H_1 : \theta > \theta_0$, alors l'hypothèse H_0 n'est pas rejetée si $\overline{X} \leq -C$, avec risque de se tromper β. Sinon c'est l'hypothèse H_1 qui est acceptée avec risque de se tromper α.

Si $H_1 : \theta < \theta_0$, alors l'hypothèse H_0 n'est pas rejetée si $\overline{X} \geq C$., avec risque de se tromper β. Sinon c'est l'hypothèse H_1 qui est acceptée avec risque de se tromper α.

5.2 Test double de la moyenne.

Si $\overline{X} \in [C_1, C_2]$, alors l'hypothèse H_0 n'est pas rejetée, avec risque de se tromper β. Sinon c'est l'hypothèse H_1 qui est acceptée avec risque de se tromper α.

Test de la moyenne avec une valeur de référence. β non fixé

La procédure du test est la suivante.

1°.- Test double de la moyenne

$$H_0 \; : \; \theta = \theta_0 \; ; \; H_1 \; : \; \theta \neq \theta_0 \; ; \; H_D \; : \mid \theta_1 - \theta_0 \mid = d$$

2°.- On fixe la probabilité α.

3°.- Calcul du critère C du test.

On sait que

$$\alpha = P\,(\text{accepter } H_1 \,|H_0 \text{ vraie}) = P\left(\left|\overline{X}\right| \geq C \;\mid\; H_0\right) = P\left(\left|\frac{\overline{X} - \theta_0}{\Sigma/\sqrt{n}}\right| \geq \frac{C - \theta_0}{\Sigma/\sqrt{n}}\right)$$

qui donne

$$\alpha = F\,(U_\alpha)$$

avec

$$U_\alpha = \begin{cases} Z_\alpha \sim \mathcal{N}\,(0,1)\,, & \text{loi normale si la variance } \sigma^2 \text{ de la population est connue} \\ T_{n-1,\alpha} \sim t_{n-1,\alpha} & \text{loi de Student si la variance } \sigma^2 \text{ de la population est inconnue} \end{cases}$$

Par lecture des tables de la loi normale (resp.loi de Student) on évalue U_c et par conséquent on calcule le critère

$$C = \frac{U_\alpha \Sigma}{\sqrt{n}} + \theta_0$$

4°.- Calcul de la puissance du test $1-\beta$. On a

$$1 - \beta = P\,(\text{accepter } H_1 \,|H_1 \text{ vraie}) = P\left(\left|\overline{X}\right| \geq C \;\mid\; H_1\right) = P\left(\left|\frac{\overline{X} - \theta_1}{\Sigma/\sqrt{n}}\right| \geq \frac{C - \theta_1}{\Sigma/\sqrt{n}}\right)$$

ce qui donne

$$1 - \beta = F\,(U_\beta)$$

Ici on connaît la valeur de $\dfrac{C - \theta_1}{\Sigma/\sqrt{n}} = U_\beta$ et par conséquent on peut calculer la valeur de β par lecture des tables de la loi normale, si la variance est connue, ou des tables de Student, si la variance est inconnue.

Si la valeur de β est grande et si on veut la diminuer, il faut refaire le test soit en augmentant les effectifs de l'échantillon, soit en augmentant la valeur de α.

5°.- Calcul de la valeur du critère C :

$$C = \frac{1}{2}\,(U_\beta - U_{\alpha'})\sqrt{\frac{\Sigma^2}{n}} + \frac{(\theta_0 + \theta_1)_D}{2}$$

On forme l'intervalle $[C_1, C_2]$, où $C_1 = C$ et $C_2 = 2C_1 - \theta_0$.

6°.- Décision.

Si $\overline{X} \in [C_1, C_2]$, alors l'hypothèse H_0 n'est pas rejetée, avec risque de se tromper β. Sinon c'est l'hypothèse H_1 qui est acceptée avec risque de se tromper α.

Notons que si nous avons un test unilatéral, à savoir

$$H_0 \; : \; \theta = \theta_0 \; ; \; H_1 \; : \; \theta < \theta_0 \; (ou \; \theta > \theta_0)$$

comme la valeur de θ_1 n'est pas fixée, nous ne pouvons pas calculer la puissance du test. Tout au plus on peut tracer un graphique des différentes valeurs de β en fonction des valeurs de θ_1

COMPARAISON ENTRE DEUX MOYENNES

Nous avons deux séries de mesures représentées par les deux séries des v.a. $X_1, \ldots, X_n$ et $Y_1, \ldots, Y_m$. Ces deux séries de mesures sont issues
- soit de la même population et, par conséquent, nous voulons tester l'homogénéité de la population ;
- soit de deux populations différentes et, par conséquent, nous voulons tester l'équivalence de deux populations.

Comparaison simple ou double entre deux moyennes avec β fixé

La procédure du test est la suivante.

1°.- Il y a deux possibilités :

1.1 Comparaison simple entre deux moyennes

$$H_0 : \theta_X - \theta_Y = 0; H_1 : \theta_X - \theta_Y > 0; H_D : \theta_X - \theta_Y = d$$

1.2 Comparaison double entre deux moyennes

$$H_0 : \theta_X - \theta_Y = 0; H_1 : \theta_X - \theta_Y \neq 0; H_D : \theta_X - \theta_Y = d$$

2°.- On fixe les probabilités α et β. Dans la suite on prendra

$$\alpha' = \begin{cases} \alpha, & \text{si test simple} \\ \alpha/2, & \text{si test double} \end{cases}$$

3°.- Calcul de la taille de l'échantillon :

3.1.- Calcul de la valeur $U_{\alpha'}$. Nous avons

$$P(H_1 \mid H_0) = P\left(\frac{\overline{X} - \overline{Y} - (\theta_X - \theta_Y)}{\sigma_W} > \frac{C - (\theta_X - \theta_Y)}{\sigma_W}\right) = \alpha'$$

et, comme sous H_0 on a $\theta_X = \theta_Y$, on obtient

$$P(H_1 \mid H_0) = P\left(\frac{\overline{X} - \overline{Y}}{\sigma_W} > \sigma_W\right) = P\left(U_{\alpha'} > Z_{\alpha'}\right) = \alpha'$$

avec

$$U_\alpha = \begin{cases} Z_{\alpha'} \sim \mathcal{N}(0,1) & \text{si les variances } \sigma_X^2, \sigma_Y^2 \text{ des populations connues} \\ T_{\alpha',\nu} \sim t_{\alpha',\nu} & \text{si les variances sont inconnues et égales } (\nu = n+m-2) \\ T_{\alpha',\nu} \sim t_{\alpha',\nu} & \text{si les variances sont inconnues et non égales } \left(\nu = \dfrac{\left(\frac{s_X^2}{n} + \frac{s_Y^2}{m}\right)^2}{\frac{\left(\frac{s_X^2}{n}\right)^2}{n+1} + \frac{\left(\frac{s_Y^2}{m}\right)^2}{m+1}} - 2\right) \\ & \text{N.B. Si } \nu > n+m-2, \text{ il faut recommencer avec des échantillons plus grands.} \end{cases}$$

et

$$\sigma_W^2 = \begin{cases} \dfrac{\sigma_X^2}{n} + \dfrac{\sigma_Y^2}{m} & \text{si les variances } \sigma_X^2, \sigma_Y^2 \text{ des populations connues} \\ \dfrac{(n-1)\, s_X^2 + (m-1)\, s_Y^2}{n+m-2}\left(\dfrac{1}{n} + \dfrac{1}{m}\right) & \text{si les variances sont inconnues et égales} \\ \dfrac{s_X^2}{n} + \dfrac{s_Y^2}{m} & \text{si les variances sont inconnues et non égales} \end{cases}$$

Dans les formules précédentes, on a posé

$$s_X^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2, s_Y^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(Y_k - \overline{Y}\right)^2$$

Par lecture des tables de la distribution normale (resp. de Student) on établit la valeur de $Z_{\alpha'}$ (resp. de $T_{\alpha',\nu}$).

3.2.- Calcul de la valeur U_β. Nous avons

$$P\left(\frac{(\theta_X-\theta_Y)_D - \left(\overline{X}-\overline{Y}\right)}{\sigma_W} > U_\beta\right) = \beta$$

avec

$$U_\beta = \begin{cases} Z_\beta \sim \mathcal{N}(0,1) & \text{si les variances } \sigma_X^2, \sigma_Y^2 \text{ des populations connues} \\ T_{\beta,\nu} \sim t_{\beta,\nu} & \text{si les variances sont inconnues et égales} \end{cases}$$

Par lecture des tables de la distribution normale (resp. de Student) on établit la valeur de Z_β (resp. de $T_{\beta,\nu}$).

3.3.- Calcul de la taille minimale n_C de l'échantillon :

$$n_C = \begin{cases} \frac{(Z_\alpha+Z_\beta)^2}{(\theta_X-\theta_Y)_D^2}\left(\sigma_X^2+\sigma_Y^2\right) & \text{si les variances } \sigma_X^2, \sigma_Y^2 \text{ des populations connues} \\ \frac{(T_{\alpha,\nu}+T_{\beta,\nu})^2}{(\theta_X-\theta_Y)_D^2}\left(s_X^2+s_Y^2\right) & \text{si les variances sont inconnues et égales} \end{cases}$$

Si $n_C > n$, alors il faut recommencer avec un échantillon plus grand.

4°.- Calcul de la valeur du critère C :

$$C = \begin{cases} \frac{1}{2}\left(Z_{\alpha'} - Z_\beta\right)\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}} + \frac{(\theta_X-\theta_Y)_D}{2} & \text{si } \sigma_X^2, \sigma_Y^2 \text{ connues} \\ \frac{1}{2}\left(T_{\alpha',\nu} - T_{\beta,\nu}\right)\sqrt{\frac{(n-1)s_X^2+(m-1)s_Y^2}{n+m-2}}\cdot\sqrt{\frac{1}{n}+\frac{1}{m}} + \frac{(\theta_X-\theta_Y)_D}{2} & \text{si les variances inconnues et égales} \\ T_{\alpha';\nu}\sqrt{\frac{s_X^2}{n}+\frac{s_Y^2}{m}} & \text{si les variances inconnues et différentes} \end{cases}$$

5°.- Décision. Il y a deux possibilités :

5.1 Comparaison simple des moyennes.- Si $\overline{X} - \overline{Y} \leq C$, alors l'hypothèse H_0 n'est pas rejetée, avec risque de se tromper β. Sinon c'est l'hypothèse H_1 qui est acceptée avec risque de se tromper α.

5.2 Comparaison double des moyennes.- Si $\overline{X} - \overline{Y} \in [-C, C]$, alors l'hypothèse H_0 n'est pas rejetée, avec risque de se tromper β. Sinon c'est l'hypothèse H_1 qui est acceptée avec risque de se tromper α.

Si β non fixé, on calcule U_{alpha} comme en 3.1. Le calcul du critère se fait à l'aide de la relation

$$C = \frac{\sqrt{2}U_\alpha\sigma_w}{sqrtn}$$

La décision suit les règles du 5°. Notons que dans ce cas nous ne pouvons pas préciser le risque β.

TESTS D'HYPOTHÈSES SUR LES VARIANCES

Test simple de la variance avec une valeur de référence

La procédure du test est la suivante~ :

1°.- $H_0 : \theta = \theta_0$; $H_1 : \theta > \theta_0$; $H_D : \theta = \theta_1$

2°.- On fixe les probabilités α et β.

3°.- Calculer la valeur du rapport des variances : $\rho^2 = \dfrac{\theta_1}{\theta_0}$.

4°.- Calcul de la taille minimale nécessaire n_c de l'échantillon :

4.1.- Calcul du nombre de degrés de liberté (ddl)

$$\nu = n - 1$$

4.2.- Calcul de la valeur $\chi^2_{\alpha,\nu}$. Nous avons

$$P\left(\frac{(n-1)\,s^2}{\theta_0} > \chi^2_{\alpha,\nu}\right) = \alpha \ ; \ \chi^2_{\alpha,\nu} \sim \text{ chi deux à } \nu \text{ ddl}$$

Par lecture des tables de la distribution du chi deux on établit la valeur de $\chi^2_{\alpha,\nu}$.

4.3.- Calcul de la valeur $\chi^2_{1-\beta,\nu}$. Nous avons

$$P\left(\frac{(n-1)\,s^2}{\theta_1} > \chi^2_{1-\beta,\nu}\right) = 1 - \beta \ ; \ \chi^2_{1-\beta,\nu} \sim \text{ chi deux à } \nu \text{ ddl}$$

Par lecture des tables de la distribution du chi deux on établit la valeur de $\chi^2_{1-\beta,\nu}$.

4.4.- Calcul de la taille nécessaire n_C de l'échantillon par ajustement du rapport :

$$\frac{\chi^2_{\alpha,\nu}}{\chi^2_{1-\beta,\nu}} \simeq \rho^2$$

Si $n_C > n$ recommencer en prenant un échantillon de taille supérieure à n_C.

4°.- Calcul de la valeur du critère C :

$$C = \frac{\theta_0}{n-1}\,\chi^2_{\alpha,\nu}$$

5°.- Décision :

Si $s^2 \leq C$, alors l'hypothèse H_0 est acceptée.

Test double de la variance avec une valeur de référence

La procédure du test est la suivante :

1°.- $H_0 : \theta = \theta_0$; $H_1 : \theta \neq \theta_0$; $H_D : \theta = \theta_1$ ou $\theta = \theta_2$

2°.- On fixe les probabilités α et β.

3°.- Calculer la valeur du rapport des variances : $\rho_1^2 = \dfrac{\theta_1}{\theta_0}$; $\rho_2^2 = \frac{\theta_2}{\theta_0}$.

4°.- Calcul de la taille minimale nécessaire n_c de l'échantillon :

4.1.- Calcul du nombre de degrés de liberté (ddl)

$$\nu = n - 1$$

4.2.- Calcul de la valeur $\chi^2_{\alpha/2,\nu}$. Nous avons

$$P\left(\frac{(n-1)\,s^2}{\theta_0} > \chi^2_{\alpha/2,\nu}\right) = \frac{\alpha}{2} \;;\; \chi^2_{\alpha/2,\nu} \sim \text{ chi deux à } \nu \text{ ddl}$$

Par lecture des tables de la distribution du chi deux on établit la valeur de $\chi^2_{\alpha,\nu}$.

4.3.- Calcul de la valeur $\chi^2_{1-\beta,\nu}$. Nous avons

$$P\left(\frac{(n-1)\,s^2}{\theta_1} > \chi^2_{1-\beta,\nu}\right) = 1-\beta \;;\; \chi^2_{1-\beta,\nu} \sim \text{ chi deux à } \nu \text{ ddl}$$

Par lecture des tables de la distribution du chi deux on établit la valeur de $\chi^2_{1-\beta,\nu}$.

4.4.- Calcul de la taille n_c de l'échantillon par ajustement du rapport :

$$\frac{\chi^2_{\alpha/2,\nu}}{\chi^2_{1-\beta,\nu}} \simeq \rho_1^2$$

Si $n_c > n$ recommencer en prenant un échantillon de taille supérieure à N.

4°.- Calcul des valeurs des critères C_1, C_2 :

$$C_1 = \frac{\theta_0}{n-1}\,\chi^2_{\alpha/2,\nu} \quad ; \quad C_2 = 2\theta_0 - C_1$$

5°.- Décision :

Si $s^2 \in [C_1, C_2]$, alors l'hypothèse H_0 est acceptée.

COMPARAISON ENTRE LES VARIANCES

Nous avons deux séries de mesures représentées par les deux séries des v.a. $X_1, \ldots, X_n$ et $Y_1, \ldots, Y_m$. Ces deux séries de mesures sont issues

– soit de la même population et, par conséquent, nous voulons tester l'homogénéïté de la population ;
– soit de deux populations différentes et, par conséquent, nous voulons tester l'équivalence de deux populations.

On cherche à tester si les variances de deux populations sont identiques. Pour ce test les variances sont estimées en utilisant les valeurs des échantillons à l'aide de la formule $s^2 = \dfrac{1}{n-1}\displaystyle\sum_{k=1}^{n}\left(X_k - \overline{X}\right)^2$.

Comparaison simple entre les variances

La procédure du test est la suivante :

1°.- $H_0 : \theta_X = \theta_Y$; $H_1 : \theta_X > \theta_Y$; $H_D : \dfrac{\theta_X}{\theta_Y} =.\rho^2$.

2°.- On fixe les probabilités α , β.

3°.- Calcul de la taille minimale nécessaire n_C de l'échantillon par ajustement de la relation

$$\rho^2 = F_{\alpha;n-1,m-1} \cdot F_{\beta;m-1,n-1}$$

où $F_{\alpha;n-1,m-1}$ la distribution de Fisher.

Si $n_C > \min\{n, m\}$, alors il faut recommencer en prenant des échantillons plus grands.

4°.- Décision :

Si $\dfrac{s_X^2}{s_Y^2} \leq F_{\alpha;n-1,m-1}$, alors l'hypothèse H_0 est acceptée.

Comparaison double entre les variances

La procédure du test est la suivante :

1°.- $H_0 : \theta_X = \theta_Y$; $H_1 : \theta_X \neq \theta_Y$; $H_D : \dfrac{\theta_X}{\theta_Y} = \rho^2$ ou $\dfrac{\theta_X}{\theta_Y} = \dfrac{1}{\rho^2}$.

2°.- On fixe les probabilités α , β.

3°.- Calcul de la taille minimale nécessaire n_C de l'échantillon par ajustement de la relation

$$\rho^2 = F_{\alpha/2;n-1,m-1} \cdot F_{\beta;m-1,n-1}$$

où $F_{\alpha;n-1,m-1}$ la distribution de Fisher.

Si $n_C > \min\{n, m\}$, alors il faut recommencer en prenant des échantillons plus grands.

4°.- Décision :

Si $\dfrac{s_X^2}{s_Y^2} \in \left[F_{1-\alpha/2;n-1,m-1}, F_{\alpha/2;n-1,m-1}\right]$, alors l'hypothèse H_0 est acceptée.

TEST D'ADÉQUATION DU χ^2

Supposons que nous disposons d'un échantillon $(X_1 = x_1, \cdots, X_n = x_n)$ et nous voulons tester si les valeurs de v.a. de cet échantillon suivent une loi de probabilté $\mathcal{L}(\mu, \sigma^2)$. La procédure du test est la suivante.

1°.- Établir le modèle théorique $\mathcal{L}(\mu, \sigma^2)$, c-à-d. la loi de probabilité, par exemple Bernoulli, binomiale, Poisson, uniforme, normale, etc.) qui correspond à l'expérience.

2°.- Formuler les hypothèses :

$$H_0 \ : \ X \sim \mathcal{L}(\mu, \sigma^2) \ ; \ H_1 \ : \ X \nsim \mathcal{L}(\mu, \sigma^2)$$

3°.- On fixe la probabilité α.

4°.- Calcul de la distance $d(F_n, F)$.

4.1 On forme m sous-ensembles de l'échantillon et on calcule leurs effectifs

$$n_i = [I(x < x_i) - I(x < x_{i-1})] \, ; i = 1, \ldots, m$$

avec $I(x < x_i) =$ nombre d'éléments de l'échantillon avec une valeur inférieure à x_i.

Parfois il est plus judicieux de calculer leurs fréquences

$$p(x_i) = \frac{1}{n} [I(x < x_i) - I(x < x_{i-1})] \, ; i = 1, \ldots, m$$

4.2 Évaluation, à l'aide des tables, des fréquances théoriques p_i issues de la loi $\mathcal{L}(\mu, \sigma^2)$.

4.3 Calcul de la distance qui est donnée par

$$d(F_n, F) = \sum_{i=1}^{m} \frac{(n_i - np_i)^2}{np_i}$$

si on utilise les effectifs n_i, et par

$$d(F_n, F) = n \sum_{i=1}^{m} \frac{(p(x_i) - p_i)^2}{p_i}$$

si on utilise les fréquences $p(x_i)$.

5°.- Calcul de la valeur du critère C :

$$C = \chi^2_{m-1,\alpha}$$

par lecture des tables de la distribution du χ^2.

6°.- Décision :

Si $d(F_n, F) \leq C$, alors l'hypothèse H_0 n'est pas rejetée.

Parfois le risque α n'est pas précisé. Dans ce cas, en connaissant la valeur de la distance $d(F_n, F)$ et en examinant la table de la loi du χ^2, on peut établir la valeur minimale de α pour que l'hypothèse H_0 ne soit pas rejetée.
D'autre part on peut utiliser ce test même si tous les paramètres de la loi du modèle théorique ne sont pas connues mais seulement estimés. Dans ce cas si on suppose que nous avons r paramètres estimés, alors on utilise la loi du χ^2 avec $m - k - 1$ degrés de liberté.

TEST D'ADÉQUATION DE KOLMOGOROV-SMIRNOV

Supposons que nous disposons d'un échantillon $(X_1 = x_1, \cdots, X_n = x_n)$ et nous voulons tester si les valeurs de v.a. de cet échantillon suivent une loi continue de probabilté $\mathcal{L}\left(\mu, \sigma^2\right)$. La procédure du test est la suivante.

1°.- Établir le modèle théorique $\mathcal{L}\left(\mu, \sigma^2\right)$, c-à-d. la loi continue de probabilité qui correspond à l'expérience.

2°.- Formuler les hypothèses :

$$H_0 \ : \ X \sim \mathcal{L}\left(\mu, \sigma^2\right) \ ; \ H_1 \ : \ X \nsim \mathcal{L}\left(\mu, \sigma^2\right)$$

3°.- On fixe la probabilité α.

4°.- Calcul de la valeur de $K_{\max}$.

4.1.- On range les valeurs $(X_1 = x_1, \cdots, X_n = x_n)$ selon l'ordre croissant.

4.2.- On calcule $F_n(i) = card\{X(j)/X(j \leq i\} \, ; \, i = 1, \ldots, n$

4.3.- On calcule les valeurs

$$K^-(i) = F_n(i-1) - nF\left(X_i\right), \quad K^+(i) = F_n(i) - nF\left(X_i\right); \, i = 1, \ldots, n$$

4.4.- Calcul des valeurs maximales

$$K^-_{\max} = \max\left\{K^-(1), \ldots K^-(n)\right\}; \quad K^+_{\max} = \max\left\{K^+(1), \ldots K^+(n)\right\}$$

4.5.- Calcul de la distance $d(F_n, F)$: On a $K_{\max} = \max\{K^-_{\max}, K^+_{\max}\}$ d'où $d(F_n, F) = \frac{K_{\max}}{n}$

5°.- Calcul de la valeur du critère C par lecture de la table du test de Kolmogorov–Smirnov, avec α risque de 1er espèce.

6°.- Décision :

Si $K_n \leq C$, alors l'hypothèse H_0 n'est pas rejetée.

TEST D'INDÉPENDANCEDE DEUX ÉCHANTILLONS

Supposons que nous disposons les valeurs des modalités, issues d'un échantillon, de deux caractères $(X = x_1, \cdots, X = x_n)$ et $(Y = y_1, \cdots, Y = y_n)$ et nous voulons tester si ces deux caractères sont indépendants. La procédure du test est la suivante.

1°.- On calcule les effectifs

$$n_{ij} = \text{card}\left\{(x_i, y_j) \ / \ X = x_i, Y = y_j\right\} \ ; i = 1, \ldots, p; \ j = 1, \ldots q$$

et les effectifs marginaux

$$n_{i\cdot} = \sum_{j=1}^{q} n_{ij} \ ; i = 1, \ldots, p$$

et

$$n_{ij} = \sum_{i=1}^{p} n_{ij} \ ; j = 1, \ldots, q$$

et aussi

$$n_T = \sum_{i=1}^{p} \sum_{j=1}^{q} n_{ij}$$

2°.- On fixe la probabilité α.

3°.- Calcul de la distance $d(X, Y)$

$$d(X, Y) = \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{\left(n_{ij} - \frac{n_{i\cdot} \times n_{\cdot j}}{n_{\cdot j}}\right)^2}{\frac{n_{i\cdot} \times n_{\cdot j}}{n_{\cdot j}}}$$

4°.- Calcul de la valeur du critère C :

$$C = \chi^2_{\alpha,(p-1)(q-1)}$$

par lecture des tables de la distribution du χ^2.

6°.- Décision :

Si $d(X, Y) \leq C$, alors l'hypothèse H_0 n'est pas rejetée.