Première partie

Théorie d'estimation

0.1 Echantillonnage :

On s'intéresse à l'étude d'un caractère dans une population P à laquelle on n'a pas accés. Si on extrait plusieurs échantillons reprèsentatif de taille n fixées, les différences observées entre les résultats obtenus sont dues à *des flucuations d'échantillonnages.* A partir d'un échantillon, on n'a pas de certitudes mais estimations de paramètres.

L'échantillonnage est dit **non-exhaustif** si le tirade de n individus constituant l'échantillon a lieu avec remise.

Il est **exhaustif** si le tirage est réalisé sans remise. En fait, le plus souvent la taille de l'éhantillon est faible par rapport à celle de la population, aussi on considérera dans la suite que tout échantilllonnageest assimilable au cas non-exhaustif.

Si on étudie sur la population un caractére quantitatif, et si on considère plusieurs échantillon de taille n , on peut calculer la moyenne et la variancede chaque échantillon. On définit ainsi trois variables aléatoires.

$\overline{X}$ **moyenne aléatoire de l'échantillon** qui prend pour valeurs les moyenne des échantillons de taille n.

S_e^2 **variance aléatoire de l'échantillon** qui prend pour valeurs les variance des échantillons de taille n.

Si la population est formée des individus ayant ou non un caractére A.

F **fréquence des échantillons** qui prend pour valeurs les fréquences de A des échantillons de taille n.

0.2 Estimation ponctuelle de la moyenne et la variance :

Soit X une variables aléatoire définie sur la population avec la moyenne μ et la variance σ^2.

c'est μ et σ qu'il s'agit d'estimer

Théorème 1 $E(\overline{X}) = \mu \qquad V(\overline{X}) = \frac{\sigma^2}{n}$; $E(S_e^2) = \frac{n-1}{n}\sigma^2$

0.2.1 Interprétation des résultats :

D'une maniére générale,on dit qu'une variable aléatoire R est **estimateur sans biais** d'un paramétre r si $E(R) = r$

Dans le cas contraire on parle **d'estimateur biaisé**.

$\overline{X}$ est estimateur sans biais de μ.Mais S_e^2 est un estimateur biaisé de σ^2.

Par contre $S^2 = \frac{n}{n-1} S_e^2$ est un estimateur sans biais de σ^2.

En pratique ; on dispose d'un seul échantillon de taille n , alors la meilleur estimation ponctuelle :

La meilleur estimation ponctuelle de μ est la moyenne $\overline{x}$ de l'échantillon.

la meilleur estimation ponctuelle de σ^2 est le nombre s^2 :

$s^2 = \frac{n}{n-1} \times varaince\ de\ l'\acute{e}chantillon.$

0.3 Estimation ponctuelle d'un pourcentage :

la population est formée d'individus ayant ou non un caractére donné A. Soit p la probabilité pour qu'un individu pris au hasard dans la population présente le caractére A.

Théorème 2 $E(F) = p \quad V(F) = \frac{p(1-p)}{n}$

0.3.1 Utilisation :

F est un estimateur sans biais de p.Quand on dispose d'un seul échantillon de taille n, la meilleur estimation ponctuelle de p est donc la fréquence f observée sur l'échantiollon.

0.4 Estimation d'une moyenne par un intervalle de confiance :

L'estimation ponctuelle n'est pas commaude ; c'est pour cette raison on cherche un intervalle I telle que : on ait de forte probabilité contenue μ

c-à-d $I =]a, b[\quad P(\mu \in I) = 1 - \alpha$; où α =risque

par exemple si $\alpha = 5\% \Rightarrow 1 - \alpha = 95\%$

On a 3 cas se présentent :

1^{er}cas : $P \curvearrowright N(\mu;\sigma)$ où $n \geq 30$

$P \curvearrowright N(\mu;\sigma) \Rightarrow \overline{X} \curvearrowright N(\mu;\frac{\sigma}{\sqrt{n}})$

$\overline{X} \curvearrowright N(\mu;\frac{\sigma}{\sqrt{n}})$; posons $z = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \curvearrowright N(0,1)$

$P(-z_\alpha < z < +z_\alpha) = 1-\alpha \Rightarrow p(-z_\alpha < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < +z_\alpha) = 1-\alpha$

$\Rightarrow P(-z_\alpha\frac{\sigma}{\sqrt{n}} < \overline{X}-\mu < +z_\alpha\frac{\sigma}{\sqrt{n}}) = 1-\alpha$

$\Rightarrow P(\overline{X}-z_\alpha\frac{\sigma}{\sqrt{n}} < \mu < \overline{X}+z_\alpha\frac{\sigma}{\sqrt{n}}) = 1-\alpha$

$$\mu \in I =]\overline{x}-z_\alpha\frac{\sigma}{\sqrt{n}};\overline{x}+z_\alpha\frac{\sigma}{\sqrt{n}}[$$

$2^{ème}$ cas: $P \curvearrowright N(\mu;\sigma)$ où $n \geq 30$ et σ de la population est inconnue :

Si σ est inconnue, on l'estime par $s = \sqrt{\frac{n}{n-1}}s_e$

D'où : $I =]\overline{x}-z_\alpha\frac{s_e}{\sqrt{n-1}};m+z_\alpha\frac{s_e}{\sqrt{n-1}}[$

$3^{éme}$ cas : P quelconque et $n < 30$ et σ inconnue

$\overline{X}$ suit la loi de student à n-1 degrés de liberté $\overline{X} \curvearrowright St((n-1)d.d.l)$

D'où $I =]\overline{x}-t_\alpha\frac{s_e}{\sqrt{n-1}};m+t_\alpha\frac{s_e}{\sqrt{n-1}}[$

t_α est détermina par la table de student.

Exemple :

Un dosage de sucre dans une solution effecuée sur 8 prélévement provenant d'une meme fabrication a donné les résultats suivants exprimé en g/l :

19,5 19,7 19,8 20,2 20,2 20,3 20,4 20,8

trouver I au risque 5%

On ait au $3^{éme}$ cas : $n < 30$ et $\sigma = ?$

$\overline{x} = \frac{\sum x_i}{n} = 20,11g/l;\ V = s_e^2 = \frac{\sum x_i^2}{n} - \overline{x}^2 = 0,178(g/l)^2 \Rightarrow s_e = 0,422g/l$

$n-1 = 8-1 = 7$ et $\alpha = 5\% = 0,05$ d'aprés la table de student $t_\alpha = 2,365$

$\mu \in]20,11 - \frac{(2,365)(0,422)}{7};20,11+\frac{(2,365)(0,422)}{7}[\Rightarrow \mu \in]19,75;20,46[$

0.5 Estimation d'un pourcentage par un intervalle de confiance :

Soit dans une population P ; une variable quantitative présentée dans la partition p(absente dans la partition $q = 1 - p$).Soit un échantillon de taille n sur lequel on dénombre np fois la présence du caractére A (nq fois son absence) avec : $n = np + nq$

1^{er} **cas :** si $n \geq 30; np \geq 5$ et $nq \geq 5 \implies F \curvearrowright N(p, \sqrt{\frac{pq}{n}})$

posons $z = \frac{F-p}{\sqrt{\frac{pq}{n}}} \curvearrowright N(0,1)$

$P(-z_\alpha < z < +z_\alpha) = 1 - \alpha \implies P(-z_\alpha < \frac{F-p}{\sqrt{\frac{pq}{n}}} < +z_\alpha) = 1 - \alpha \implies$ $P(-z_\alpha\sqrt{\frac{pq}{n}} < F - p < +z_\alpha\sqrt{\frac{pq}{n}}) = 1 - \alpha$

$\implies P(-z_\alpha\sqrt{\frac{pq}{n}} < F - p < +z_\alpha\sqrt{\frac{pq}{n}}) = 1 - \alpha$

$\implies P(F - z_\alpha\sqrt{\frac{pq}{n}} < p < F + z_\alpha\sqrt{\frac{pq}{n}}) = 1 - \alpha$

Alors l'intervalle de confiane de p est : $I =]f - z_\alpha\sqrt{\frac{pq}{n}}; f + z_\alpha\sqrt{\frac{pq}{n}}[$

et comme $\frac{f(1-f)}{n}$ est une estimation non biaisée de $\frac{p(1-p)}{n}$

D'où $I =]f - z_\alpha\sqrt{\frac{f(1-f)}{n}}; f + z_\alpha\sqrt{\frac{f(1-f)}{n}}[$

$2^{éme}$**cas :** Si $n < 30; np < 5$ où $nq < 5$

dans ce cas l'intervalle de confiance de la proprotion p est :

$I =]f - t_\alpha\sqrt{\frac{f(1-f)}{n}}; f + t_\alpha\sqrt{\frac{f(1-f)}{n}}[$

le t_α lu dans la table de student à $(n-1)$ degrés de libereté.

Remarque 3 *lorsque le nombre de degré de liberté tend vers l'infini, la fonction de répartition de la loi de student tends vers celle de la loi normale centrée réduite(en pratique $n \geq 30$)*

0.6 Estimation d'une variance par un intervalle de confiance :

Théorème 4 *Si X suit une loi normale, la variable aléatoire $Y^2 = \frac{n-1}{\sigma^2}S^2$ suit la loi de χ^2 à $n-1$ degrés de libertés.*

0.6.1 Utilisation si $n \leq 31$:

Le coefficient de risque α étant choisi et le nombre de degrés de liberté étant connu, la table de χ^2 permet de déterminer a et b tels que :

$P(a < Y^2 < b) = 1 - \alpha$ avec $P(Y^2 \geq b) = \frac{\alpha}{2}$ et $P(Y^2 \geq a) = 1 - \frac{\alpha}{2}$

La seule valeur connue de S^2 est s^2,on obtient comme intervalle de confiance de σ^2 au risque α :$]\frac{(n-1)}{b}s^2; \frac{(n-1)}{a}s^2[$

Remarque 5 *On a* $(n-1)s^2 = ns_e^2$ *où* s_e^2 *est la variance de l'échantillon.*

0.6.2 Utilisation si $n > 31$:

Le théorème cité est vrai quel que soit n. Mais les tables de χ^2 s'arrete au degrés de liberté 30. On ne peut pas les utiliser si $n > 31$.

Théorème 6 *Si* Y^2 *est une variable aléatoire qui suit une loi de* χ^2 *de degrés de liberté* ν *et si* $\nu > 30$, *alors la variable aléatoire* $U = \sqrt{Y^2} - \sqrt{2\nu - 1}$ *suit sensiblement la loi normale centrée réduite.*

Utilisation :

Ici on a $U = \sqrt{\frac{2(n-1)}{\sigma^2}S^2} - \sqrt{2n-3}$, Aprés avoir choisi α on détermine z_α et on déduit l'intervalle de confiance de σ^2 :

$]\frac{2(n-1)s^2}{\sqrt{2n-3+z_\alpha^2}}; \frac{2(n-1)s^2}{\sqrt{2n-3-z_\alpha^2}}[$

Deuxième partie

Introduction aux test statistiques

0.7 Test d'hypothéses :

On est souvent conduit à prendre une décision au sujet d'une population à partir des informations données sur un échantillon. Il s'agit de faire un choix entre plusieurs hypothéses.

On met en avant une hypothése dite *hypothése nulle* et notée (H_0). on souhaite vérifier si (H_0) est vraie alors que deux hypothéses sont possibles (H_0) et (H_1) ; où (H_1) est dite *hypothése alternative.*

Pour effectue notre choix entre (H_0) et (H_1) ; on effectue un test statistique lequel comporte les étapes suivantes :

a) Définir (H_0) et (H_1).

b) Choisir un certain paramétre sous l'hypothése (H_0).

c)Choisir un seuil α.

d) Lire dans la table de la distributuion du parametre choisi en (b) la valeur correspendante à α.

e) Définir une région critique.

f) Calculer la valeur du paramétre choisi en (b)

g) Adopter la régle suivante : si la valeur du paramétre calculer en (b) appartienne à la région critique (H_1) si non on accepte (H_0).

On est alors conduit à choisir parmi deux décisions :

-Le résultat obetenue conduit à rejeter (H_0). Mais on court ainsi un risque de se tromper dit **erreur de première espèce** et noté α.

α est la probabilité de se tromper quand on rejete (H_0)

-Le résultat obtenu conduit à accepter (H_0). Mais il y a là encore un risque d'erreur. Il est dit **erreur de deuxiéme espèce** et noté β.

β est la probabilité de se tromper quand on accepte (H_0).

0.7.1 Test de conformité :

Il s'agit de comparer un échantillon à une loi théorique. L'hypothése (H_0) consiste à supposer que les différences observées ne sont pas significative et sont dues aux fluctuations d'échantillonnage.

0.7.2 Test d'homogénétié :

Il s'agit de comparer deux échantillons. L'hypothése (H_0) est qu'ils proviennent d'une meme population.

0.7.3 Comparaison d'une moyenne observée à une moyenne théorique :

Soit $\overline{x}$ la moyenne observée d'une variable quantitative dans un échantillon E. Soit d'autre part P_0 une population de moyenne μ_0.

On désire savoir : **Si l'échantillon E provient de P_0 ou s'il est extrait d'une autre population P de moyenne $\mu \neq \mu_0$**

On désire savoir si la différence $\overline{x} - \mu_0$ entre les deux moyennes est significative (en l'est pas extrait de P_0) où bien la différence n'est pas significative (extrait de P_0).

Test bilatéral :

On test (H_0) l'hypothése nulle contre (H_1) l'hypothése alternative.

(**H**$_0$) **l'hypothése nulle** : $\mu = \mu_0$

(**H**$_1$) **l'hypothése alternative** : $\mu \neq \mu_0$

On calcule $z = \frac{\overline{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$

En choisissant un seuil significative α est on est amené à prendre la décision suivante :

$z \in]-z_\alpha, +z_\alpha[$ on accepte (H_0)

$z \notin]-z_\alpha, +z_\alpha[$ on rejette (H_0).

Test unilatéral :

On test (H_0) l'hypothése nulle contre l'hypothése (H_1) :

(**H**$_0$) **l'hypothése nulle** : $\mu = \mu_0$

(**H**$_1$) **l'hypothése** : $\mu > \mu_0$ où $\mu < \mu_0$

On calcule $z = \frac{\overline{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$

Pour le cas unilatéral à droite :

$z \in]-\infty, +z_\alpha]$ on accepte (H_0)

$z \in]+z_\alpha, +\infty[$ on rejette (H_0).

Pour le cas unilatéral à gauche :

$z \in [-z_\alpha, +\infty[$ on accepte (H_0)

$z \in]-\infty, -z_\alpha[$ on rejette (H_0).

Soit la population P_0 de moyenne μ_0 et d'écart type σ ; on extrait un echantiollon E de taille n. la moyenne d'échantillon est $\overline{x}$.On a 4 cas :

1^{er} cas : $n \geq 30$ et σ connue

$\overline{x} \curvearrowright N(\mu_0; \frac{\sigma}{\sqrt{n}}) \Rightarrow z_\alpha = \frac{\overline{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} \curvearrowright N(0,1)$

Supposon que $\alpha = 5\% \Rightarrow z_\alpha = 1,96$

Si $z_\alpha \in]-1,96, +1,96[$ on accepete l'hypothése (H_0) (E extrait de la population P_0)

Si $z_\alpha \in]-\infty, -1,96[$ où $z_\alpha \in]+1,96, +\infty[$ (la région critique) On accepete (H_1) (E extrait de P et non P_0)

$2^{éme}$ cas : $n \geq 30$ et σ inconnue

σ de la population est inconnue on l'estime par $s = \sqrt{\frac{n}{n-1}} s_e$; s_e l'écart type de l'échantillon E.

$z_\alpha = \frac{\overline{x}-\mu_0}{\frac{s}{\sqrt{n}}} \curvearrowright N(0,1)$

$3^{éme}$ cas : $N < 30$ et σ inconnue

$t_l = \frac{\overline{x}-\mu_0}{\frac{S}{\sqrt{N}}} \curvearrowright St((n-1)d.d.l)$

$4^{ème}$ cas : la population est normale et σ connue

c'est le 1^{er} cas.

0.7.4 Comparaison d'une fréquence observée à une fréquence théorique :

Dans une population P on étudie un caractére statistique à deux modalités A et $\overline{A}$. Soit p le pourcentage d'apparition de A dans la population, et f est la fréquence d'apparition de A dans un échantillon de taille n. notons F la variable aléatoire qui prend la valeur f sur chaque échantiollon de taille n.

Est ce que la difference entre f et p est explicable par les aléatoires dus à l'échantillonnage ?

Test bilatéral :

On test (H_0) l'hypothése nulle contre l'hypothése alternative (H_1) :

($\mathbf{H}_0$) **l'hypothése nulle** : $p = p_0$

($\mathbf{H}_1$) **l'hypothése alternative** : $p \neq p_0$

On calcule $z = \frac{f-p}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

En choisissant un seuil significative α est on est amené à prendre la décision suivante :

$z \in]-z_\alpha, +z_\alpha[$ on accepte (H_0)

$z \notin]-z_\alpha, +z_\alpha[$ on rejette (H_0).

Test unilatéral :

On test (H_0) l'hypothése nulle contre l'hypothése (H_1)

($\mathbf{H}_0$) **l'hypothése nulle** : $p = p_0$

($\mathbf{H}_1$) **l'hypothése** : $p > p_0$ où $p < p_0$

, On calcule $z = \frac{f-p_0}{\sqrt{\frac{(p_0(1-p_0)}{n}}}$

Pour le cas unilatéral à droite :

$z \in]-\infty, +z_\alpha]$ on accepte (H_0)

$z \in]+z_\alpha, +\infty[$ on rejette (H_0).

Pour le cas unilatéral à gauche :

$z \in [-z_\alpha, +\infty[$ on accepte (H_0)

$z \in]-\infty, -z_\alpha[$ on rejette (H_0).

On a deux cas à distingué :

1^{er} cas : si $n \geq 30; np \geq 5$ et $nq \geq 5 \implies F \curvearrowright N(p, \sqrt{\frac{pq}{n}})$

posons $z = \frac{F-p}{\sqrt{\frac{pq}{n}}} \curvearrowright N(0,1) \Rightarrow z = \frac{f-p_0}{\sqrt{\frac{(p_0(1-p_0)}{n}}}$

$2^{éme}$cas : Si $n < 30; np < 5$ où $nq < 5$

le t_α lu dans la table de Student à $(n-1)$ degrés de libereté.

0.7.5 Comparaison d'une variance expérimentale et d'une variance théorique :

Soit X une variable aléatoire suit la loi normale $N(\mu, \sigma)$; on extrait un échantillon de taille n, $\overline{x}$ est la moyenne de l'échantillon et $s^2 = \frac{n}{n-1}s_e^2$

s_e^2 est la variance de l'échantillon.

Alors $Y^2 = \frac{n-1}{\sigma^2}S^2$ suit la loi khi-deux à $(n-1)$ degrés de liberté.

Test bilatérale :

On test (H_0) l'hypothése nulle contre (H_1) l'hypothése alternative.

$(\mathbf{H}_0)$ **l'hypothése nulle** : $\sigma^2 = \sigma_0^2$

$(\mathbf{H}_1)$ **l'hypothése alternative** : $\sigma^2 \neq \sigma_0^2$

On calcule $y^2 = \frac{n-1}{\sigma^2}s^2$

Le risque α étant fixé et le nombre de degrés de liberté étant connu, la table de χ^2 permet de déterminer a et b tels que :

$P(a < Y^2 < b) = 1 - \alpha$ avec $p(Y^2 \geq b) = \frac{\alpha}{2}$ et $P(Y^2 \geq a) = 1 - \frac{\alpha}{2}$

Si $y^2 \in]a, b[$ alors on accepte (H_0).

Si $y^2 \notin]a, b[$ alors on rejette (H_0).

Test unilatérale :

On test (H_0) l'hypothése nulle contre l'hypothése (H_1) :

($\mathbf{H}_0$) l'hypothése nulle : $\sigma^2 = \sigma_0^2$

($\mathbf{H}_1$) l'hypothése : $\sigma^2 > \sigma_0^2$

On détermine b tel que $P(Y^2 \geq b) = \alpha$, on décide que :

Si $y^2 < b$, alors on accepte (H_0).

Si $y^2 \geq b$, alors on rejette (H_0).

Test unilatérale :

On test (H_0) l'hypothése nulle contre l'hypothése (H_1).

($\mathbf{H}_0$) l'hypothése nulle : $\sigma^2 = \sigma_0^2$

($\mathbf{H}_1$) l'hypothése : $\sigma^2 < \sigma_0^2$

On détermine a tel que $P(Y^2 \geq a) = 1 - \alpha$, on décide que :

Si $y^2 > a$, alors on accepte (H_0).

Si $y^2 \leq a$, alors on rejette (H_0) avec une probabilité α de se tromper.

0.7.6 Test d'homogénétié :

Il s'agit de comparer deux échantillons. L'hypothése (H_0) est qu'ils proviennent d'une meme population.

0.7.7 Comparaison de deux moyennes :

Dans deux populations P_1 et P_2 ; on étudie une variable aléatoire X.

On note :

μ_1 et σ_1 la moyenne et l'écart-type de X dans P_1.

μ_2 et σ_2 la moyenne et l'écart-type de X dans P_2.

De P_1 on extrait un échantillon E_1 de taille n_1; pour lequel on calcule sa moyenne $\overline{x}_1$ et s_1 estimation de σ_1.

De P_2 on extrait un échantillon E_2 de taille n_2; pour lequel on calcule sa moyenne $\overline{x}_2$ et s_2 estimation de σ_2.

les échantillons E_1 et E_2 sont supposés indépendantes. le problème est de savoir si les différence entre les moyennes expirémentales $\overline{x}_1$ et $\overline{x}_2$ est significative, ou contraire explicable par les flucuations d'échantillonnage.

1)Cas de grand échantillons indépendants :

On suppose que $n_1 > 30$ et $n_2 > 30$

Théorème 7 *Quelle que soit la loi suivie par* X *, la v.a* $Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ *suit la loi normale centrée réduite.*

On calcule $z = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

0.7.8 Test bilatéral :

(H_0) $\mu_1 = \mu_2$ c'est à dire $P_1 = P_2$ sont homogénes.

(H_1) $\mu_1 \neq \mu_2$

α étant fixé, on lit z_α tel que : $P(|Z| \geq z_\alpha) = \alpha$

Si $z \in]-z_\alpha, z_\alpha[$; on accepte ($\mathrm{H}_0$)

Si $z \notin]-z_\alpha, z_\alpha[$; on rejette ($\mathrm{H}_0$).

0.7.9 Test unilatérale :

(H_0) : $\mu_1 = \mu_2$ conte (H_1) $\mu_1 > \mu_2$

On détermine z_α tel que $P(Z < z_\alpha) = 1 - \alpha$ et on décide que :

Si $z < z_\alpha$, alors on accepte (H_0).

Si $z \geq z_\alpha$, alors on rejette (H_0).

(H_0) : $\mu_1 = \mu_2$ conte (H_1) $\mu_1 < \mu_2$

On détermine z'_α tel que $P(Z \geq z_\alpha) = 1 - \alpha$ et on décide que :

Si $z > z_\alpha$, alors on accepte (H_0).

Si $z \leq z_\alpha$, alors on rejette (H$_0$).

2) **Cas de petits échantillons extraits de populations gaussiennes :**

On suppose $n_1 < 30$ et $n_2 < 30$ et $\sigma = \sigma_1 = \sigma_2$

Théorème 8 *Si X suit une loi normale dans P_1 et P_2 et si $\sigma = \sigma_1 = \sigma_2$;Alors la variable aléatoire $T = \frac{\overline{X}_1 - \overline{X}_2}{\sigma\sqrt{\frac{1}{n_2}+\frac{1}{n_2}}}$ suit la loi de student à (n_1+n_2-2) degrés de liberté.*

Remarque 9 *: Comme on ne connait pas σ_1 et $,\sigma_2$; il faut d'abord tester l'égalités des deux variances. Si l'hypothèses $\sigma_1 = \sigma_2$ est retenue, cette valeur commune σ est alors estimée par : $\widehat{\sigma} = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$*

On calcul $t = \frac{\overline{x}_1 - \overline{x}_2}{\widehat{\sigma}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$

0.7.10 Test bélatéral :

α étant fixé, on lit t_α tel que : $P(|T| \geq t_\alpha) = \alpha$

Si $t \in]-t_\alpha, t_\alpha[$; on accepte (H$_0$)

Si $t \notin]-t_\alpha, t_\alpha[$; on rejette (H$_0$).

0.7.11 Test unilatéral :

(H$_0$) : $\mu_1 = \mu_2$ contre (H$_1$) $\mu_1 > \mu_2$

On détermine t_α tel que $P(T < t_\alpha) = 1 - \alpha$ et on décide que :

Si $t < t_\alpha$, alors on accepte (H$_0$).

Si $t \geq t_\alpha$, alors on rejette (H$_0$).

(H$_0$) : $\mu_1 = \mu_2$ contre (H$_1$) $\mu_1 < \mu_2$

On détermine t_α tel que $P(T \geq t_\alpha) = 1 - \alpha$ et on décide que :

Si $t > t_\alpha$, alors on accepte (H$_0$).

Si $t \leq t_\alpha$, alors on rejette (H$_0$).

3) Cas des échantillons appariés :

Deux échantillons E_1 et E_2 sont dits **appariés** lorsque chaque observation $x_{1,i}$ de E_1 est asociée à une valeur $x_{2,i}$ de E_2 (appariés=associés par paires). C'est par exemple le cas lorsque E_1 et E_2 proviennent d'un meme groupe de malades avant et aprés traitement. Deux échantillons appariés ont donc la meme taille $n = n_1 = n_2$.

3.1)Cas des grands échantillons appariés :

On suppose que $n = n_1 = n_2 > 30$, et que les échantillons E_1 et E_2 sont appariés.

On considère la variable aléatoire $D = X_1 - X_2$, dont un échantillon est $(D_1, D_2, ..., D_n)$ avec $D_i = X_{1,i} - X_{2,i}$ la moyenne et la variance d'échantillon sont :

$\overline{D} = \frac{\sum D_i}{n}$ et $S_d^2 = \frac{n}{n-1} S_{d,e}^2$ avec $S_{d,e}^2 = \frac{\sum D_i^2}{n} - (\overline{D})^2$

Désignons par $\mu = \mu_1 - \mu_2$ la moyenne de D.

Puisque $n > 30$, $U = \frac{\overline{D}-\mu}{\frac{S_d}{\sqrt{n}}}$ suit la loi normale $N(0,1)$.

Test bélatéral :

L'hypothése $(\mathrm{H}_0) : \mu_1 = \mu_2$ contre (H_1) $\mu_1 \neq \mu_2$

Ce test est équivalent au test bilatéral de $(\mathrm{H}_0) :$ $\mu = 0$ contre (H_1) $: \mu \neq 0$

Test unilatéral :

L'hypothése $(\mathrm{H}_0) : \mu_1 = \mu_2$ contre (H_1) $\mu_1 > \mu_2$

Ce test est équivalent au test bilatéral de $(\mathrm{H}_0) :$ $\mu = 0$ contre (H_1) $: \mu > 0$

3.2) Cas des petits échantillons appariés extraits de populations gaussiennes :

On suppose que $n = n_1 = n_2 \leq 30$, et que les échantillons E_1 et E_2 sont appariés. et que X_1 et X_2 suivent les lois normales $N(\mu_1, \sigma_1)$ et $N(\mu_2, \sigma_2)$

Dans ce cas $T = \frac{\overline{D}-\mu}{\frac{S_d}{\sqrt{n}}}$ suit la loi de Student à $(n-1)$ degrés de liberté.

0.7.12 Comparaison de deux fréquences :

Dans deux populations P_1 et P_2 on étudié un caractére statistique à deux modalités A et $\overline{A}$.Les pourcentages d'apparition de A dans les populations P_1 et P_2 sont p_1 et p_2.De P_1 et P_2 on extrait deux échantillons E_1 et E_2 de taille n_1 et n_2 dans lesquels les fréquences d'appartition de A sont respectivement f_1 et f_2.

Notons F_1 et F_2 les variables aléatoires qui prennent les valeurs f_1 et f_2 sur chaque échantillon.

Test bilatérale :

(H_0) $p_1 = p_2 = p$

(H_1) $p_1 \neq p_2$

Théorème 10 *Supposons que :* $n_1 \geq 30$ *et* $n_2 \geq 30$ *;* $n_1 f_1 \geq 5$ *et* $n_1(1-f_1) \geq 5; n_2 f_2 \geq 5$ *et* $n_2(1-f_2) \geq 5;$*Alors la v.a* $Z = \frac{F_1 - F_2}{\sqrt{\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}}}$ *suit la loi normale centrée réduite.*

Estimation de p : on l'estime p par $\widehat{p} = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$

On calcule $z = \frac{f_1 - f_2}{\sqrt{\widehat{p}(1-\widehat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$

Puis on choisit α le risque, on lit dans la table z_α tel que : $P(|Z| \geq z_\alpha) = \alpha$

Si $z \in]-z_\alpha, z_\alpha[$; on accepte ($\mathrm{H}_0$)

Si $z \notin]-z_\alpha, z_\alpha[$; on rejette ($\mathrm{H}_0$).

Test unilatérale :

On a (H_0) $p_1 = p_2$ contre (H_1) $p_1 > p_2$

Dans ce cas on a toujour $z > 0$

On détermine alors z_α tels que : $P(Z \geq z_\alpha) = \alpha$

Si $z < z_\alpha$; on accepte (H_0)

Si $z \geq z_\alpha$; on rejette (H_0)

0.7.13 Comparaison de deux variances :

Dans deux populations P_1 et P_2 ; on étudie une variable aléatoire X.

On note :

μ_1 et σ_1 la moyenne et l'écart-type de X dans P_1.

μ_2 et σ_2 la moyenne et l'écart-type de X dans P_2.

De P_1 on extrait un échantillon E_1 de taille n_1; pour lequel on calcule sa moyenne $\overline{x}_1$ et s_1 estimation de σ_1.

De P_2 on extrait un échantillon E_2 de taille n_2; pour lequel on calcule sa moyenne $\overline{x}_2$ et s_2 estimation de σ_2.

les échantillons E_1 et E_2 sont supposés indépendantes. le problème est de savoir si les différence entre les moyennes expirémentales s_1^2 et s_2^2 est significative, ou au contraire explicable par les flucuations d'échantillonnage.

On teste $(\mathbf{H}_0)$: $\sigma_1^2 = \sigma_2^2$ contre $(\mathbf{H}_1)$: $\sigma_1^2 \neq \sigma_2^2$

Théorème 11 *Si les deux population sont gaussiennes, la variable aléatoire* $F = \frac{S_1^2}{S_2^2}$ *suit la loi de Snédécor à* $(n_1 - 1, n_2 - 1)$ *degrés de liberté.*

On calcule $f = \frac{s_1^2}{s_2^2}$; Si nécesaire, on permute les échantillons de sorte que $f \geq 1$.le risque α étant fixé ; on détermine f_α tel que :

$P(F \geq f_\alpha) = \frac{\alpha}{2}$ (on utilise la table de Snédécor) et on décide que :

Si $f < f_\alpha$ alors on accepte (H_0).

Si $f \geq f_\alpha$ on rejette (H_0).

Lois de Snédécor :

Une loi de Snédécor est une loi de probabilité continue dont la densité est nulle pour $x < 0$ et dépend de deux paramétres appelés degrés de liberté.

Le risque α étant fixé, la table de Snédécor permettent de déterminer f_α et f'_α tel que : $P(f'_\alpha < F < f_\alpha) = 1 - \alpha$

avec $P(F \leq f'_\alpha) = \frac{\alpha}{2}$ et $P(F \geq f_\alpha) = \frac{\alpha}{2}$.