

Tests d'hypothèse

L'objectif du chapitre est de forger des outils de statistique, dont les applications sont maintenant largement diffusées dans les processus industriels de production les plus performants, en particulier en contrôle de qualité. Les applications des outils que nous allons mettre en place couvrent aussi des domaines plus larges, englobant la théorie des sondages, et la mise en place d'outils d'aide à la prise de décision, élaborés à partir d'outils statistiques sur divers indicateurs.

La plupart des résultats de ce chapitre ont constitué un des cœurs de l'activité de recherche mathématiques du XX^e siècle, dans les domaines des Probabilités et des Statistiques : la majorité des outils exposés ont moins de 150 ans, ce qui est tout-à-fait remarquable, et montre, outre l'intérêt d'entreprendre des recherches scientifiques, la richesse des résultats auxquels elles parviennent.

1 Tests d'hypothèses : principes généraux et objets d'étude

Le problème principal est de parvenir à ériger un outil de contrôle d'un caractère traduisant la qualité d'une démarche. Il peut par exemple s'agir de contrôler une production industrielle, de mener une politique populaire. Dans cette perspective, on souhaite parvenir à établir des règles de décisions pré-établies, fixées une fois pour toutes, basées sur des observations régulières, les moins coûteuses possibles, qui indiqueront, en temps réel ou presque, la conduite à tenir pour conserver la qualité mesurée d'un caractère. On équipe ainsi par exemple une chaîne de production d'appareils de mesures, qui donnent à tout instant la valeur moyenne d'un ensemble de paramètres que l'on souhaite voir rester dans des limites pré-définies (normes). À partir des résultats ainsi collectés, on pourra, une fois ces règles de décision établies, opter très rapidement pour des remédiations, et corriger les problèmes survenus, avant qu'ils ne soient trop longtemps répétés.

Il n'est pas possible de déterminer des critères de décision parfaitement fiables, car la mesure que l'on réalise n'est pas exempte d'erreur (puisque l'on se base sur des sondages), aussi commence-t-on par se fixer un niveau de risque maximal acceptable. Puis, on s'attache à déterminer les valeurs critiques d'un certain nombre de paramètres (les "signaux d'alerte"), au niveau de risque requis, qui doivent entraîner des modifications dans la poursuite du processus de production, d'une politique économique, etc.

Nous nous intéressons à deux grands types de tests, qui sont parmi les plus simples : les tests de comparaison par rapport à un nombre de référence (norme), et les tests de comparaisons entre deux populations, relativement à une grandeur caractéristique donnée. Pour chacun d'eux enfin, nous détaillons les tests relatifs à une moyenne, et les tests relatifs à une proportion (ou une fréquence, ceci revenant au même).

2 Tests bilatéraux de comparaison à un nombre de référence

2.1 Tests bilatéraux de comparaison à une moyenne de référence

Commençons par l'étude d'une situation concrète.

Exemple 1 Situation : *Le conserveur Beurckduelle indique sur ces boîtes de conserve de haricots que celle-ci contiennent 250 g de haricots (la norme de masse visée est donc $m_0 = 250$). Il indique également à ses clients que la machine qui remplit les boîtes est réglée de telle manière que l'écart-type des masses de haricots mis dans une boîte est de 20 g.*

Le magazine "Quoi Choisir ?" a acheté 50 boîtes de haricots de la marque Beurckduelle, et a pesé les haricots contenus dans celles-ci : il a obtenu une moyenne de 243 g de haricots par boîte. Ce magazine peut-il affirmer que la société Beurckduelle trompe ses clients ?

Éléments de réponse : *Le magazine "Quoi Choisir ?", en achetant 50 boîtes, a constitué un échantillon de la population de l'ensemble de la production Beurckduelle. Appelons m la masse moyenne réelle des boîtes fabriquées. Nous savons que tous les échantillons de taille 50 ne sont pas constitués rigoureusement de la même*

manière, et que donc l'observation d'une masse de 243 g dans un échantillon particulier n'est pas forcément significative, car d'un échantillon à l'autre la masse moyenne est variable (ce sont les résultats sur la fluctuation d'échantillonnage qui imposent cette prudence).

Formulons l'hypothèse suivante :

(H_0) : "le conserveur remplit ses boîtes d'en moyenne 250 g de haricots" (ceci revient à $m = m_0$),

et l'hypothèse alternative (H_1) :

(H_1) : "le conserveur ne remplit pas ses boîtes d'en moyenne 250 g de haricots" (ceci revient à $m \neq m_0$),

Alors, toujours d'après les résultats sur la fluctuation d'échantillonnage, nous savons que si M est la masse moyenne de haricots dans une boîte, calculée sur un échantillon de taille 50, et si (H_0) est satisfaite, alors M suit approximativement une loi normale $\mathcal{N}\left(250; \frac{20}{\sqrt{50}}\right)$, soit une loi $\mathcal{N}(250; 2,83)$.

Nous pouvons par suite déterminer la fourchette de masses moyennes de haricots qui recouvre (par exemple) 95% des échantillons : soit pour cela $t \in \mathbb{R}$ tel que $2\Pi(t) - 1 = 0,95$. Alors $t \simeq 1,96$, et on a :

$$0,95 = 2\Pi(t) - 1 = \mathbb{P}\left(t \leq \frac{M - 250}{2,83} \leq t\right) = \mathbb{P}(250 - 2,83t \leq M \leq 250 + 2,83t)$$

car $\frac{M - 250}{2,83}$ suit une loi $\mathcal{N}(0; 1)$ (centrage-réduction).

Autrement dit, 95% des échantillons vont conduire à observer une masse moyenne empirique de haricots M située dans l'intervalle $[244,45; 255,55]$ grammes. Notons que m_e est une valeur de M , et que les calculs du magazine montrent que m_e n'est pas dans la fourchette attendue !

Le magazine "Quoi Choisir ?" est donc face à l'alternative suivante :

- soit il est en présence d'un échantillon non représentatif de la population globale, et il est possible que m_e ne soit pas significatif pour conclure sur la validité de (H_0) : on ne peut pas écarter clairement le fait que (H_0) soit fausse, et on fait confiance au conserveur (on accepte donc (H_0)) ;
- soit il est en présence d'un échantillon représentatif de la population globale, et donc le fait que $m_e \notin [244,45; 255,55]$ (qui est significatif à 95%) doit conduire à rejeter l'hypothèse (H_0) , et donc accepter l'hypothèse alternative (H_1) : la société Beurckduelle ment, et ne remplit pas ses boîtes de 250 g de haricots (en moyenne).

Notons sur cet exemple que le risque de rejeter l'hypothèse (H_0) alors qu'elle est vraie est de 0,05 : c'est l'erreur de première espèce du test considéré.

Nous pouvons maintenant donner le théorème qui régit ce type de test, et la règle de décision qui en découle :

Théorème 1 Dans une population, un paramètre devrait avoir la valeur moyenne normale de m_0 , tandis que sa véritable moyenne (inconnue) est m . On suppose que l'écart-type de cette grandeur est connu, et vaut σ pour l'ensemble de la population.

Soient (H_0) l'hypothèse : "La valeur m est conforme à la norme m_0 " (en résumé : $m = m_0$), et (H_1) l'hypothèse alternative "La valeur m n'est pas conforme à la norme m_0 " (en résumé : $m \neq m_0$).

On prélève un échantillon de taille n , et on calcule la moyenne empirique m_e des valeurs du paramètre observé dans cet échantillon.

Soit $\alpha \in \mathbb{R}$ le seuil de signification du test. Soit $t \in \mathbb{R}$ tel que $2\Pi(t) - 1 = 1 - \alpha$.

Alors l'intervalle $I_\alpha = \left[m_0 - t\frac{\sigma}{\sqrt{n}}; m_0 + t\frac{\sigma}{\sqrt{n}}\right]$ est l'intervalle d'acceptation de l'hypothèse (H_0) au seuil de risque α .

(Ceci signifie que la règle de décision est la suivante :

- si $m_e \in I_\alpha$, alors on accepte l'hypothèse (H_0) ;
- si $m_e \notin I_\alpha$, alors on rejette l'hypothèse (H_0) et on accepte l'hypothèse (H_1) .

De plus, la probabilité de rejeter (H_0) alors qu'elle est vraie est de α .)

La preuve de ce résultat reprend les grandes lignes de l'exemple précédent :

Preuve

◀ Le test est basé sur un échantillon de taille n , prélevé dans une population pour laquelle le paramètre étudié suit une loi $\mathcal{N}(m; \sigma)$. Dans ces conditions, et si (H_0) est remplie, les résultats sur l'échantillonnage montrent que la moyenne empirique M calculée dans l'échantillon suit une loi $\mathcal{N}\left(m_0; \frac{\sigma}{\sqrt{n}}\right)$. La probabilité pour que $M \in I_\alpha$ est donc :

$$\mathbb{P}(M \in I_\alpha) = \mathbb{P}\left(m_0 - t\frac{\sigma}{\sqrt{n}} \leq M \leq m_0 + t\frac{\sigma}{\sqrt{n}}\right) = \mathbb{P}\left(-t \leq \frac{\sqrt{n}}{\sigma}(M - m_0) \leq t\right) = 2\Pi(t) - 1 = 1 - \alpha,$$

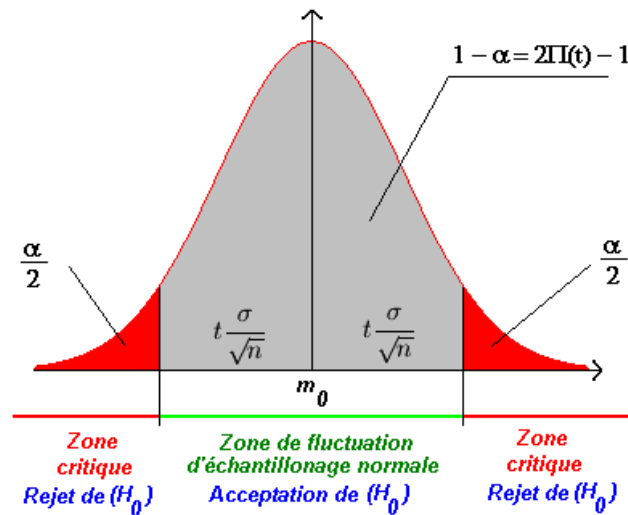


FIG. 1 – Construction du test bilatéral pour une moyenne testée m_0 au seuil α

car, par centrage-réduction, $\frac{\sqrt{n}}{\sigma}(M - m_0) = \frac{M - m_0}{\frac{\sigma}{\sqrt{n}}}$ suit une loi $\mathcal{N}(0; 1)$.

Dès lors, pour l'échantillon considéré, m_e est une des valeurs possibles de M , et on a l'alternative suivante (en gardant à l'esprit que α est proche de 0) :

- si $m_e \in I_\alpha$, la probabilité que l'échantillon soit non représentatif de la population globale est de α : on ne peut pas écarter clairement le fait que (H_0) soit fausse, donc on accepte (H_0) ;
- si $m_e \notin I_\alpha$, la probabilité que l'échantillon soit représentatif de la population globale est de $1 - \alpha$: dans ce cas, on doit rejeter l'hypothèse (H_0) , et donc accepter l'hypothèse alternative (H_1) . ►

Remarque 1 Dans le cas (fréquent !) où l'écart-type σ n'est pas connu, on utilise alors une estimation $\hat{\sigma}$ de σ obtenue à partir de l'écart-type σ_e relevé dans l'échantillon. L'intervalle d'acceptation de l'hypothèse (H_0) au seuil de risque α est alors

$$I_\alpha = \left[m_0 - t \frac{\sigma_e}{\sqrt{n-1}}; m_0 + t \frac{\sigma_e}{\sqrt{n-1}} \right],$$

puisque $\hat{\sigma} = \sqrt{\frac{n}{n-1}} \sigma_e$.

Exercice 1 Un fabricant d'allumettes souhaite placer 100 allumettes dans les boîtes qu'il commercialise. Soit m le nombre moyen d'allumettes contenues dans sa production du jeudi 18 mars. Il sait par ailleurs que la machine qui remplit les boîtes le fait avec un écart-type constant de 2,5.

Au cours de cette journée du jeudi 18 mars, il a constitué un échantillon de 30 boîtes, qui contiennent en moyenne 99 allumettes. Au seuil de risque 0,05, le fabricant peut-il considérer que la machine a mal travaillé, et qu'il doit en conséquence jeter la production de la journée ? et au seuil de 0,01 ?

Réponse : Soit (H_0) l'hypothèse : "la production est bonne" (c'est-à-dire $m = m_0$), et (H_1) l'hypothèse alternative "la production n'est pas bonne" (c'est-à-dire $m \neq m_0$).

Au seuil de 0,05, on a $t = 1,96$, et l'intervalle d'acceptation de (H_0) est donc

$$I_{0,05} = \left[100 - 1,96 \frac{2,5}{\sqrt{30}}; 100 + 1,96 \frac{2,5}{\sqrt{30}} \right] = [99,1; 100,9].$$

Comme $m_e = 99 \notin [99,1; 100,9]$, au seuil de risque de 0,05, on rejette (H_0) et on accepte (H_1) : la production n'est pas conforme à la norme ; il faut donc la détruire, et on est certain à 95% de le faire à bon escient (la probabilité de détruire la production alors qu'elle est bonne est de $\alpha = 0,05$).

Au seuil de signification de 0,01, on a $t = 2,58$, et l'intervalle d'acceptation de (H_0) devient alors

$$I_{0,01} = \left[100 - 2,58 \frac{2,5}{\sqrt{30}}; 100 + 2,58 \frac{2,5}{\sqrt{30}} \right] = [98,8; 101,2].$$

Comme $m_e = 99 \in [98,8; 101,2]$, au seuil de risque de 0,01, on ne peut pas rejeter clairement (H_0) , donc on accepte (H_0) : la production est pas significativement non-conforme à la norme, on peut la conserver.

2.2 Tests bilatéraux de comparaison à une fréquence de référence

Exemple 2 Situation : Le pourcentage de personnes atteintes du virus de la grippe parmi les employés d'une entreprise est traditionnellement de 20% à la fin du mois d'octobre. Certaines années cependant, le virus touche plus ou moins d'employés, suivant la virulence du virus. Afin de savoir à l'avance si l'épidémie de grippe de l'hiver à venir sera comparable ou pas aux années précédentes, le chef d'entreprise interroge, à la fin octobre, 25 de ces salariés et voudrait savoir, au vu des résultats de ce sondage, si l'épidémie sera de virulence "normale" ou pas.

Éléments de réponse : Le chef d'entreprise a constitué un échantillon de taille 25 en interrogeant 25 de ses salariés. Il calcule à partir de cet échantillon la fréquence empirique f_e des employés touchés par la grippe parmi les 25 employés, et trouve par exemple $f_e = 0,38$. Il sait que la "norme" pour cette fréquence est $f_0 = 0,2$ en cette période.

Nous savons que tous les échantillons de taille 25 ne sont pas constitués rigoureusement de la même manière, et que donc l'observation d'une proportion de grippés de 0,38 dans un échantillon particulier n'est pas forcément significative, car d'un échantillon à l'autre la proportion de grippés est variable (ce sont les résultats sur la fluctuation d'échantillonnage qui imposent cette prudence).

Formulons l'hypothèse suivante :

$$(H_0) : \text{"la fréquence des grippés est } f \text{ " (soit } f = f_0),$$

et l'hypothèse alternative (H_1) :

$$(H_1) : \text{"la fréquence des grippés n'est pas } f \text{ " (soit } f \neq f_0).$$

Alors, toujours d'après les résultats sur la fluctuation d'échantillonnage, nous savons que si F est la fréquence empirique des grippés dans un échantillon de taille 25, et sous l'hypothèse (H_0) , alors F suit approximativement

une loi normale $\mathcal{N}\left(0,2; \sqrt{\frac{0,2 \cdot 0,8}{25}}\right)$, soit une loi $\mathcal{N}(0,2; 0,08)$.

Nous pouvons par suite déterminer la fourchette de fréquences empiriques qui recouvre (par exemple) 95% des échantillons : soit pour cela $t \in \mathbb{R}$ tel que $2\Pi(t) - 1 = 0,95$. Alors $t \simeq 1,96$, et on a :

$$0,95 = 2\Pi(t) - 1 = \mathbb{P}\left(t \leq \frac{F - 0,2}{0,08} \leq t\right) = \mathbb{P}(0,2 - 0,08t \leq F \leq 0,2 + 0,08t)$$

car $\frac{F - f_0}{0,08}$ suit une loi $\mathcal{N}(0;1)$ (par centrage-réduction).

Autrement dit, 95% des échantillons vont conduire à observer une fréquence empirique de grippés F située dans l'intervalle $[0,0432; 0,3568]$. Notons que f_e est une valeur de F , et que les calculs du chef d'entreprise montrent que m_e n'est pas dans la fourchette attendue !

Le chef d'entreprise est donc face à l'alternative suivante :

- soit il est en présence d'un échantillon non représentatif de la population globale, et il est possible que f_e ne soit pas significatif pour conclure sur la validité de (H_0) : on ne peut pas écarter clairement le fait que (H_0) soit fausse, et on fait comme si la grippe était semblable aux autres années (on accepte donc (H_0)) ;
- soit il est en présence d'un échantillon représentatif de la population globale, et donc le fait que $f_e \notin [0,0432; 0,3568]$ (qui est significatif à 95%) doit conduire à rejeter l'hypothèse (H_0) , et donc accepter l'hypothèse alternative (H_1) : l'épidémie de grippe sera d'une sévérité différente.

Notons sur cet exemple que le risque de rejeter l'hypothèse (H_0) alors qu'elle est vraie est de 0,05 : c'est l'erreur de première espèce du test considéré.

Nous pouvons maintenant établir une règle de décision pour la comparaison à une fréquence de référence :

Théorème 2 Dans une population, la fréquence d'apparition d'un phénomène devrait avoir la valeur normale de f_0 , tandis que sa véritable fréquence (inconnue) est f .

Soit (H_0) l'hypothèse : "La valeur f est conforme à la norme f_0 " (en résumé : $f = f_0$), et soit (H_1) l'hypothèse alternative "La valeur f n'est pas conforme à la norme f_0 " (en résumé : $f \neq f_0$).

On prélève un échantillon de taille n , et on calcule la fréquence empirique f_e des valeurs du paramètre observé dans cet échantillon.

Soit $\alpha \in \mathbb{R}$ le seuil de signification du test. Soit $t \in \mathbb{R}$ tel que $2\Pi(t) - 1 = 1 - \alpha$.

Alors l'intervalle $I_\alpha = \left[f_0 - t \sqrt{\frac{f_0(1-f_0)}{n}}; f_0 + t \sqrt{\frac{f_0(1-f_0)}{n}} \right]$ est l'intervalle d'acceptation de l'hypothèse (H_0) au seuil de risque α .

(Ceci signifie que la règle de décision est la suivante :

- si $f_e \in I_a$, alors on accepte l'hypothèse (H_0) ;
- si $f_e \notin I_a$, alors on rejette l'hypothèse (H_0) et on accepte l'hypothèse (H_1).

De plus, la probabilité de rejeter (H_0) alors qu'elle est vraie est de α .)

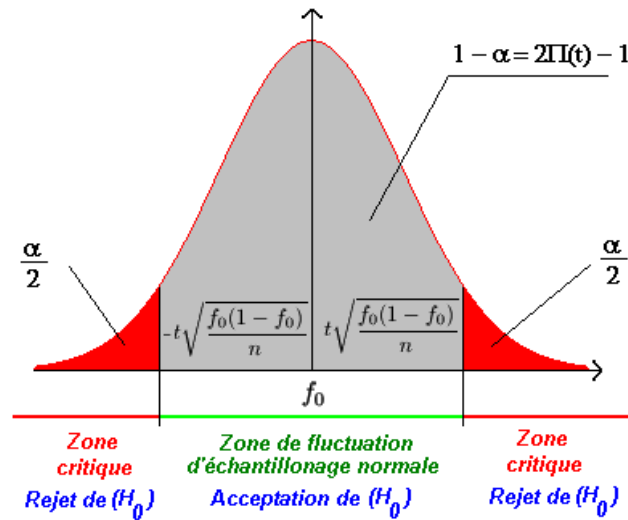


FIG. 2 – Construction du test bilatéral pour une fréquence testée f_0 au seuil α

La preuve de ce résultat reprend les grandes lignes de l'exemple précédent :

Preuve

◀ Le test est basé sur un échantillon de taille n , prélevé dans une population pour laquelle le paramètre étudié se manifeste avec fréquence f . Dans ces conditions, et si (H_0) est remplie, les résultats sur l'échantillonnage montrent que la fréquence empirique F calculée dans l'échantillon suit une loi $\mathcal{N}\left(f_0; \sqrt{\frac{f_0(1-f_0)}{n}}\right)$. La probabilité pour que $M \in I_\alpha$ est donc :

$$\mathbb{P}(F \in I_\alpha) = \mathbb{P}\left(f_0 - t\sqrt{\frac{f_0(1-f_0)}{n}} \leq F \leq f_0 + t\sqrt{\frac{f_0(1-f_0)}{n}}\right) = \mathbb{P}\left(-t \leq \frac{F - f_0}{\sqrt{\frac{f_0(1-f_0)}{n}}} \leq t\right) = 2\Pi(t) - 1 = 1 - \alpha,$$

car, par centrage-réduction, $\frac{F - m_0}{\sqrt{\frac{f_0(1-f_0)}{n}}}$ suit une loi $\mathcal{N}(0; 1)$.

Dès lors, pour l'échantillon considéré, m_e est une des valeurs possibles de F , et on a l'alternative suivante (en gardant à l'esprit que α est proche de 0) :

- si $f_e \in I_a$, la probabilité que l'échantillon soit non représentatif de la population globale est de α : on ne peut pas écarter clairement le fait que (H_0) soit fausse, donc on accepte (H_0) ;
- si $f_e \notin I_a$, la probabilité que l'échantillon soit représentatif de la population globale est de $1 - \alpha$: dans ce cas, on doit rejeter l'hypothèse (H_0), et donc accepter l'hypothèse alternative (H_1). ▶

Exercice 2 Au printemps dans le Gers, il pleut 1 jour sur 3 en année normale. Cette année, Marcel Péré¹, le plus vieil habitant de Vic, a observé le ciel durant les 15 premiers jours du printemps, et affirme qu'il a plu 3 jours (seulement). Après cette observation, il affirme partout dans le village que "le (prin-)temps est détraqué", que "tout fout le camp", qu'"il n'y a plus de saison, ma bonne dame", et que "c'était mieux avant".

Au seuil de signification de 5%, peut-on considérer que Marcel a raison ? et au seuil de 1% ?

Solution : Les 15 jours d'observation attentive de Marcel constituent un échantillon de taille $n = 15$. Soit f_0 la fréquence des jours de pluie en année ordinaire, et f_e celle des jours de pluie calculée d'après les observations de Marcel : on a donc $f_0 = \frac{1}{3} \simeq 0,333$, et $f_e = \frac{3}{15} = 0,2$.

Formulons l'hypothèse suivante :

¹Cet exercice est dédié à mon professeur d'Histoire-Géographie.

(H_0) : "ce printemps, la pluviosité est conforme à l'ordinaire" (soit $f = f_0$).

et l'hypothèse alternative :

(H_1) : "ce printemps, la pluviosité n'est pas conforme à l'ordinaire" (soit $f \neq f_0$).

Au seuil de 0,05, on a $t = 1,96$, et l'intervalle d'acceptation de (H_0) est donc

$$I_{0,05} = \left[0,333 - 1,96\sqrt{\frac{0,20,8}{15}}; 0,333 + 1,96\sqrt{\frac{0,20,8}{15}} \right] = [0,131; 0,535].$$

Comme $f_e = 0,2 \in [0,131; 0,535]$, au seuil de 5%, on ne peut pas affirmer que Marcel a tort. On considère donc qu'il a raison.

Au seuil de 0,01, on a $t = 2,58$, et l'intervalle d'acceptation de (H_0) est donc

$$I_{0,01} = \left[0,333 - 2,58\sqrt{\frac{0,20,8}{15}}; 0,333 + 2,58\sqrt{\frac{0,20,8}{15}} \right] = [0,066; 0,6].$$

Comme $f_e = 0,2 \in [0,066; 0,6]$, au seuil de 1%, on ne peut pas affirmer que Marcel a tort. On considère donc qu'il a raison.

3 Tests unilatéraux de comparaison à un nombre de référence

3.1 Tests unilatéraux de comparaison à une moyenne de référence

Exemple 3 Nous reprenons la situation de l'exemple 1. Une amélioration de notre étude s'impose en effet : le consommateur n'ayant aucune raison de protester si l'industriel lui vend plus de haricots qu'il ne l'indique sur la boîte, les conclusions du remarquable magazine "Quoi choisir ?" doivent être nuancées et adaptées. Dans l'étude menée plus haut, nous avons choisi l'hypothèse

(H_0) : "le conserveur remplit ses boîtes d'en moyenne 250 g de haricots" (ceci revient à $m = m_0$).

et nous nous sommes intéressés à l'hypothèse alternative (H_1) :

(H_1) : "le conserveur ne remplit pas ses boîtes comme annoncé" (ceci revient à $m \neq m_0$).

En réalité, ce qui intéresse le consommateur, c'est de savoir s'il se fait voler ou pas, autrement dit la considération de l'hypothèse alternative (H_1) :

(H_1) : "le conserveur lèse le consommateur" (ceci revient à $m < m_0$).

Choisissons comme d'habitude un seuil de signification, par exemple $\alpha = 0,05$. Nous savons, comme dans l'exemple 1, que les masses moyennes empiriques de haricots M dans des échantillons de taille 50 suivent une loi $\mathcal{N}(250; 2,83)$.

Nous allons chercher la fourchette des masses moyennes de haricots qui contient théoriquement 95% des échantillons satisfaisant (H_0) (et ne satisfaisant pas (H_1)), c'est-à-dire la fourchette contenant 95% des échantillons sans léser le consommateur. Pour cela, on cherche $z \in \mathbb{R}$ se sorte que $\mathbb{P}(M > 250 - z) = 0,95$. Or

$$\mathbb{P}(M > 250 - z) = 0,95 \Leftrightarrow \mathbb{P}(M \leq 250 - z) = 0,05.$$

On cherche donc z de sorte que $\mathbb{P}(M \leq 250 - z) = 0,05$, soit tel que $\mathbb{P}\left(\frac{M - 250}{2,83} \leq \frac{-z}{2,83}\right) = 0,05$. Comme par centrage-réduction $\frac{M - 250}{2,83}$ suit une loi $\mathcal{N}(0; 1)$, on a donc $\frac{-z}{2,83} \simeq -1,645$, soit donc $z \simeq 4,66$.

Ainsi, 95% des échantillons amènent à constater une masse moyenne M de haricots supérieure à $250 - 4,66 \simeq 245,34$. Or m_e est une valeur de M , et l'observation du magazine "Quoi Choisir ?" indique que m_e est inférieur à 245,34 g !

Le magazine "Quoi Choisir ?" est donc face à l'alternative suivante :

- soit il est en présence d'un échantillon non représentatif de la population globale, et il est possible que m_e ne soit pas significatif pour conclure sur la validité de (H_0) : on ne peut pas écarter clairement le fait que (H_0) soit fausse, et on fait confiance au conserveur (on accepte donc (H_0)) ;
- soit il est en présence d'un échantillon représentatif de la population globale, et donc le fait que $m_e \notin$

[245, 34; +∞[(qui est significatif à 95%) doit conduire à rejeter l'hypothèse (H_0), et donc accepter l'hypothèse alternative (H_1) : la société Beurckduelle ment, et remplit ses boîtes de moins de 250 g de haricots (en moyenne).

Notons sur cet exemple que le risque de rejeter l'hypothèse (H_0) alors qu'elle est vraie est de 0,05 : c'est l'erreur de première espèce du test considéré.

Nous avons donc deux cas à distinguer suivant l'hypothèse alternative (H_1) que nous souhaitons soumettre au test décisionnel ($m < m_0$, ou $m > m_0$). Ceci conduit aux théorèmes suivants.

3.1.1 Tester $m = m_0$ contre $m < m_0$

Théorème 3 Dans une population, un paramètre devrait avoir la valeur moyenne normale de m_0 , tandis que sa véritable moyenne (inconnue) est m . On suppose que l'écart-type de cette grandeur est connu, et vaut σ pour l'ensemble de la population.

Soient (H_0) l'hypothèse : "La valeur m est conforme à la norme m_0 " (en résumé : $m = m_0$), et (H_1) l'hypothèse alternative "La valeur m est inférieure à la norme m_0 " (en résumé : $m < m_0$).

On prélève un échantillon de taille n , et on calcule la moyenne empirique m_e des valeurs du paramètre observé dans cet échantillon.

Soit $\alpha \in \mathbb{R}$ le seuil de signification du test. Soit $t \in \mathbb{R}$ tel que $\Pi(-t) = \alpha$ ($t > 0$).

Alors l'intervalle $I_\alpha = [m_0 - t \frac{\sigma}{\sqrt{n}}; +\infty[$ est l'intervalle d'acceptation de l'hypothèse (H_0) au seuil de risque α .

(Ceci signifie que la règle de décision est la suivante :

- si $m_e \in I_\alpha$, alors on accepte l'hypothèse (H_0) ;
- si $m_e \notin I_\alpha$, alors on rejette l'hypothèse (H_0) et on accepte l'hypothèse (H_1).

De plus, la probabilité de rejeter (H_0) alors qu'elle est vraie est de α .)

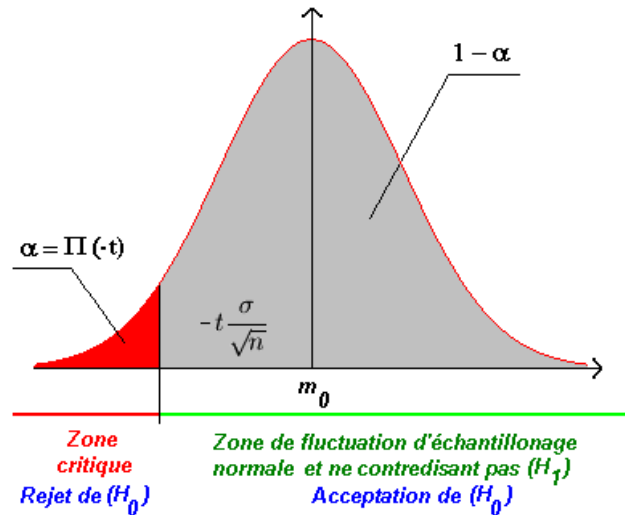


FIG. 3 – Construction du test unilatéral pour une moyenne testée m_0 au seuil α contre $m < m_0$

La preuve de ce résultat reprend les grandes lignes de l'exemple précédent :

Preuve

◀ Le test est basé sur un échantillon de taille n , prélevé dans une population pour laquelle le paramètre étudié suit une loi $\mathcal{N}(m; \sigma)$. Dans ces conditions, et si (H_0) est remplie, les résultats sur l'échantillonnage montrent que la moyenne empirique M calculée dans l'échantillon suit une loi $\mathcal{N}\left(m_0; \frac{\sigma}{\sqrt{n}}\right)$. La probabilité pour que $M \in I_\alpha$ est donc :

$$\mathbb{P}(M \in I_\alpha) = \mathbb{P}\left(m_0 - t \frac{\sigma}{\sqrt{n}} \leq M\right) = \mathbb{P}\left(-t \leq \frac{\sqrt{n}}{\sigma}(M - m_0)\right) = 1 - \mathbb{P}\left(\frac{\sqrt{n}}{\sigma}(M - m_0) \leq -t\right) = \Pi(-t) = \alpha,$$

car, par centrage-réduction, $\frac{\sqrt{n}}{\sigma}(M - m_0) = \frac{M - m_0}{\frac{\sigma}{\sqrt{n}}}$ suit une loi $\mathcal{N}(0; 1)$.

Dès lors, pour l'échantillon considéré, m_e est une des valeurs possibles de M , et on a l'alternative suivante (en

gardant à l'esprit que α est proche de 0) :

- si $m_e \in I_a$, la probabilité que l'échantillon soit non représentatif de la population globale est de α : on ne peut pas écarter clairement le fait que (H_0) soit fausse, donc on accepte (H_0) ;
- si $m_e \notin I_a$, la probabilité que l'échantillon soit représentatif de la population globale est de $1 - \alpha$: dans ce cas, on doit rejeter l'hypothèse (H_0) , et donc accepter l'hypothèse alternative (H_1) . ►

Remarque 2 Dans le cas (fréquent!) où l'écart-type σ n'est pas connu, on utilise alors une estimation $\hat{\sigma}$ de σ obtenue à partir de l'écart-type σ_e relevé dans l'échantillon. L'intervalle d'acceptation de l'hypothèse (H_0) au seuil de risque α est alors

$$I_\alpha = [m_0 - t \frac{\sigma_e}{\sqrt{n-1}}; +\infty[,$$

puisque $\hat{\sigma} = \sqrt{\frac{n}{n-1}} \sigma_e$.

Exercice 3

3.1.2 Tester $m = m_0$ contre $m > m_0$

Théorème 4 Dans une population, un paramètre devrait avoir la valeur moyenne normale de m_0 , tandis que sa véritable moyenne (inconnue) est m . On suppose que l'écart-type de cette grandeur est connu, et vaut σ pour l'ensemble de la population.

Soient (H_0) l'hypothèse : "La valeur m est conforme à la norme m_0 " (en résumé : $m = m_0$), et (H_1) l'hypothèse alternative "La valeur m est supérieure à la norme m_0 " (en résumé : $m > m_0$).

On prélève un échantillon de taille n , et on calcule la moyenne empirique m_e des valeurs du paramètre observé dans cet échantillon.

Soit $\alpha \in \mathbb{R}$ le seuil de signification du test. Soit $t \in \mathbb{R}$ tel que $\Pi(t) = 1 - \alpha$ ($t > 0$).

Alors l'intervalle $I_\alpha =]-\infty; m_0 + t \frac{\sigma}{\sqrt{n}}]$ est l'intervalle d'acceptation de l'hypothèse (H_0) au seuil de risque α .

(Ceci signifie que la règle de décision est la suivante :

- si $m_e \in I_a$, alors on accepte l'hypothèse (H_0) ;
- si $m_e \notin I_a$, alors on rejette l'hypothèse (H_0) et on accepte l'hypothèse (H_1) .

De plus, la probabilité de rejeter (H_0) alors qu'elle est vraie est de α .)

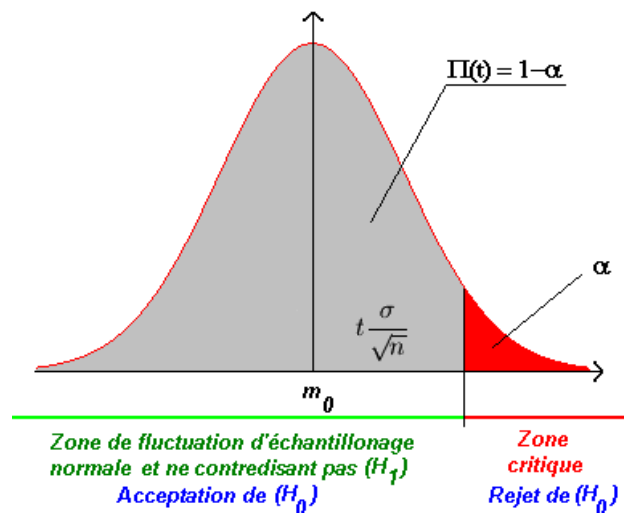


FIG. 4 – Construction du test unilatéral pour une moyenne testée m_0 au seuil α contre $m > m_0$

La preuve de ce résultat reprend les grandes lignes de l'exemple précédent :

Preuve

► Le test est basé sur un échantillon de taille n , prélevé dans une population pour laquelle le paramètre étudié suit une loi $\mathcal{N}(m; \sigma)$. Dans ces conditions, et si (H_0) est remplie, les résultats sur l'échantillonnage montrent que

la moyenne empirique M calculée dans l'échantillon suit une loi $\mathcal{N}\left(m_0; \frac{\sigma}{\sqrt{n}}\right)$. La probabilité pour que $M \in I_\alpha$ est donc :

$$\mathbb{P}(M \in I_\alpha) = \mathbb{P}\left(M \leq m_0 - t \frac{\sigma}{\sqrt{n}}\right) = \mathbb{P}\left(\frac{\sqrt{n}}{\sigma}(M - m_0) \leq t\right) = \Pi(t) = 1 - \alpha,$$

car, par centrage-réduction, $\frac{\sqrt{n}}{\sigma}(M - m_0) = \frac{M - m_0}{\frac{\sigma}{\sqrt{n}}}$ suit une loi $\mathcal{N}(0; 1)$.

Dès lors, pour l'échantillon considéré, m_e est une des valeurs possibles de M , et on a l'alternative suivante (en gardant à l'esprit que α est proche de 0) :

- si $m_e \in I_\alpha$, la probabilité que l'échantillon soit non représentatif de la population globale est de α : on ne peut pas écarter clairement le fait que (H_0) soit fausse, donc on accepte (H_0) ;
- si $m_e \notin I_\alpha$, la probabilité que l'échantillon soit représentatif de la population globale est de $1 - \alpha$: dans ce cas, on doit rejeter l'hypothèse (H_0) , et donc accepter l'hypothèse alternative (H_1) . ►

Remarque 3 Dans le cas (fréquent!) où l'écart-type σ n'est pas connu, on utilise alors une estimation $\hat{\sigma}$ de σ obtenue à partir de l'écart-type σ_e relevé dans l'échantillon. L'intervalle d'acceptation de l'hypothèse (H_0) au seuil de risque α est alors

$$I_\alpha =]-\infty; m_0 + t \frac{\sigma_e}{\sqrt{n-1}}],$$

puisque $\hat{\sigma} = \sqrt{\frac{n}{n-1}} \sigma_e$.

Exercice 4

3.2 Tests unilatéraux de comparaison à une fréquence de référence

Exemple 4 Nous reprenons la situation de l'exemple 2. Une amélioration de notre étude s'impose en effet : le chef d'entreprise n'ayant aucune raison de s'inquiéter si la grippe devait être moins virulente que d'habitude : il pourrait dans ce cas, il n'a pas besoin de ralentir sa production, ou de se préoccuper d'embaucher des remplaçants. Dans l'étude menée plus haut, nous avons choisi l'hypothèse

$$(H_0) : \text{la fréquence des grippés est } f.$$

et l'hypothèse alternative (H_1) :

$$(H_1) : \text{"la fréquence des grippés est différente d'ordinaire"} \text{ (ceci revient à } f \neq f_0).$$

En réalité, ce qui intéresse le chef d'entreprise, c'est de savoir s'il doit prévoir des remplaçants ou pas, autrement dit la réalisation, ou non, de l'hypothèse (H_1) :

$$(H_1) : \text{"la fréquence des grippés sera plus forte que d'ordinaire"} \text{ (ceci revient à } f > f_0).$$

Choisissons comme d'habitude un seuil de signification, par exemple $\alpha = 0,05$.

Nous avons donc deux cas à distinguer suivant l'hypothèse alternative (H_1) que nous souhaitons soumettre au test décisionnel ($f < f_0$, ou $f > f_0$). Ceci conduit aux théorèmes suivants.

3.2.1 Tester $f = f_0$ contre $f < f_0$

Théorème 5 Dans une population, la fréquence d'apparition d'un phénomène devrait avoir la valeur normale de f_0 , tandis que sa véritable fréquence (inconnue) est f .

Soit (H_0) l'hypothèse : "La valeur f est conforme à la norme f_0 " (en résumé : $f = f_0$), et soit (H_1) l'hypothèse alternative "La valeur f est inférieure à la norme f_0 " (en résumé : $f < f_0$).

On prélève un échantillon de taille n , et on calcule la fréquence empirique f_e des valeurs du paramètre observé dans cet échantillon.

Soit $\alpha \in \mathbb{R}$ le seuil de signification du test. Soit $t \in \mathbb{R}$ tel que $\Pi(-t) = \alpha$ ($t > 0$).

Alors l'intervalle $I_\alpha = [f_0 - t \sqrt{\frac{f_0(1-f_0)}{n}}; +\infty[$ est l'intervalle d'acceptation de l'hypothèse (H_0) au seuil de risque α .

(Ceci signifie que la règle de décision est la suivante :

- si $f_e \in I_\alpha$, alors on accepte l'hypothèse (H_0) ;
- si $f_e \notin I_\alpha$, alors on rejette l'hypothèse (H_0) et on accepte l'hypothèse (H_1) .

De plus, la probabilité de rejeter (H_0) alors qu'elle est vraie est de α .)

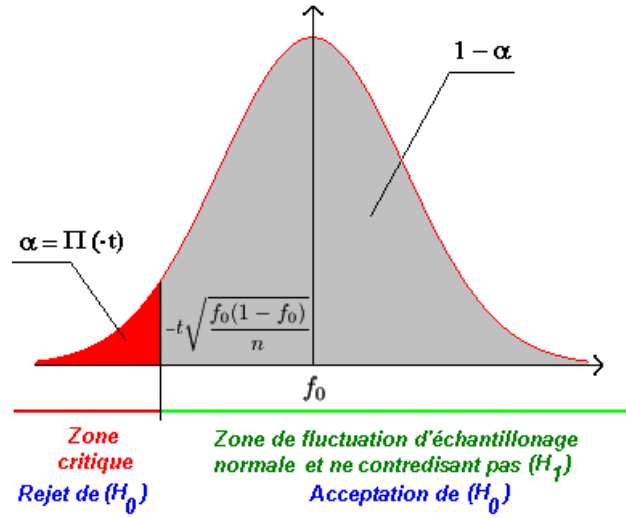


FIG. 5 – Construction du test unilatéral pour une fréquence testée f_0 au seuil α contre $f < f_0$

La preuve de ce résultat reprend les grandes lignes de l'exemple précédent :

Preuve

► Le test est basé sur un échantillon de taille n , prélevé dans une population pour laquelle le paramètre étudié se manifeste avec fréquence f . Dans ces conditions, et si (H_0) est remplie, les résultats sur l'échantillonnage montrent que la fréquence empirique F calculée dans l'échantillon suit une loi $\mathcal{N}\left(f_0; \sqrt{\frac{f_0(1-f_0)}{n}}\right)$. La probabilité pour que $M \in I_\alpha$ est donc :

$$\mathbb{P}(F \in I_\alpha) = \mathbb{P}\left(f_0 - t\sqrt{\frac{f_0(1-f_0)}{n}} \leq F\right) = \mathbb{P}\left(-t \leq \frac{F - f_0}{\sqrt{\frac{f_0(1-f_0)}{n}}}\right) = 1 - \mathbb{P}\left(\frac{F - f_0}{\sqrt{\frac{f_0(1-f_0)}{n}}} \leq -t\right) = \Pi(-t) = \alpha,$$

car, par centrage-réduction, $\frac{F - m_0}{\sqrt{\frac{f_0(1-f_0)}{n}}}$ suit une loi $\mathcal{N}(0; 1)$.

Dès lors, pour l'échantillon considéré, m_e est une des valeurs possibles de F , et on a l'alternative suivante (en gardant à l'esprit que α est proche de 0) :

- si $f_e \in I_\alpha$, la probabilité que l'échantillon soit non représentatif de la population globale est de α : on ne peut pas écarter clairement le fait que (H_0) soit fausse, donc on accepte (H_0) ;
- si $f_e \notin I_\alpha$, la probabilité que l'échantillon soit représentatif de la population globale est de $1 - \alpha$: dans ce cas, on doit rejeter l'hypothèse (H_0) , et donc accepter l'hypothèse alternative (H_1) . ►

Exercice 5

3.2.2 Tester $f = f_0$ contre $f > f_0$

Théorème 6 Dans une population, la fréquence d'apparition d'un phénomène devrait avoir la valeur normale de f_0 , tandis que sa véritable fréquence (inconnue) est f .

Soit (H_0) l'hypothèse : "La valeur f est conforme à la norme f_0 " (en résumé : $f = f_0$), et soit (H_1) l'hypothèse alternative "La valeur f est supérieure à la norme f_0 " (en résumé : $f > f_0$).

On prélève un échantillon de taille n , et on calcule la fréquence empirique f_e des valeurs du paramètre observé dans cet échantillon.

Soit $\alpha \in \mathbb{R}$ le seuil de signification du test. Soit $t \in \mathbb{R}$ tel que $\Pi(t) = 1 - \alpha$ ($t > 0$).

Alors l'intervalle $I_\alpha =]-\infty; f_0 + t\sqrt{\frac{f_0(1-f_0)}{n}}]$ est l'intervalle d'acceptation de l'hypothèse (H_0) au seuil de risque α .

(Ceci signifie que la règle de décision est la suivante :

- si $f_e \in I_\alpha$, alors on accepte l'hypothèse (H_0) ;

- si $f_e \notin I_a$, alors on rejette l'hypothèse (H_0) et on accepte l'hypothèse (H_1) .
De plus, la probabilité de rejeter (H_0) alors qu'elle est vraie est de α .)

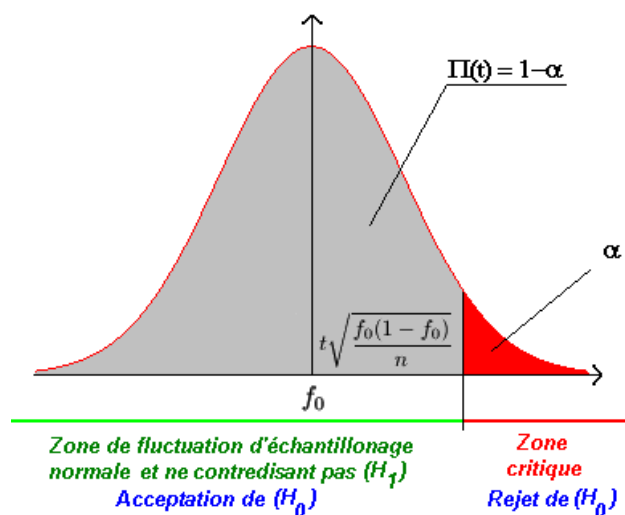


FIG. 6 – Construction du test unilatéral pour une fréquence testée f_0 au seuil α contre $f > f_0$

La preuve de ce résultat reprend les grandes lignes de l'exemple précédent :

Preuve

◀ Le test est basé sur un échantillon de taille n , prélevé dans une population pour laquelle le paramètre étudié se manifeste avec fréquence f . Dans ces conditions, et si (H_0) est remplie, les résultats sur l'échantillonnage montrent que la fréquence empirique F calculée dans l'échantillon suit une loi $\mathcal{N}\left(f_0; \sqrt{\frac{f_0(1-f_0)}{n}}\right)$. La probabilité pour que $M \in I_\alpha$ est donc :

$$\mathbb{P}(F \in I_\alpha) = \mathbb{P}\left(F \leq f_0 + t\sqrt{\frac{f_0(1-f_0)}{n}}\right) = \mathbb{P}\left(\frac{F - f_0}{\sqrt{\frac{f_0(1-f_0)}{n}}} \leq t\right) = \Pi(t) = 1 - \alpha,$$

car, par centrage-réduction, $\frac{F - m_0}{\sqrt{\frac{f_0(1-f_0)}{n}}}$ suit une loi $\mathcal{N}(0; 1)$.

Dès lors, pour l'échantillon considéré, m_e est une des valeurs possibles de F , et on a l'alternative suivante (en gardant à l'esprit que α est proche de 0) :

- si $f_e \in I_a$, la probabilité que l'échantillon soit non représentatif de la population globale est de α : on ne peut pas écarter clairement le fait que (H_0) soit fausse, donc on accepte (H_0) ;
- si $f_e \notin I_a$, la probabilité que l'échantillon soit représentatif de la population globale est de $1 - \alpha$: dans ce cas, on doit rejeter l'hypothèse (H_0) , et donc accepter l'hypothèse alternative (H_1) . ►

Exercice 6

4 Tests de comparaison entre deux populations

Ces tests sont particulièrement utiles pour mesurer une variation, une amélioration (ou une baisse!) de la qualité d'une production à partir de mesures opérées sur des échantillons, donc relativement rapides à exécuter.

4.1 Tests bilatéraux comparant des moyennes calculées dans deux populations

Exemple 5

Théorème 7

4.2 Tests bilatéraux comparant des fréquences calculées dans deux populations

Exemple 6

Théorème 8

4.3 Tests unilatéraux comparant des moyennes calculées dans deux populations

Exemple 7

Théorème 9

4.4 Tests unilatéraux comparant des fréquences calculées dans deux populations

Exemple 8

Théorème 10

Table des matières

1	Tests d'hypothèses : principes généraux et objets d'étude	1
2	Tests bilatéraux de comparaison à un nombre de référence	1
2.1	Tests bilatéraux de comparaison à une moyenne de référence	1
2.2	Tests bilatéraux de comparaison à une fréquence de référence	4
3	Tests unilatéraux de comparaison à un nombre de référence	6
3.1	Tests unilatéraux de comparaison à une moyenne de référence	6
3.1.1	Tester $m = m_0$ contre $m < m_0$	7
3.1.2	Tester $m = m_0$ contre $m > m_0$	8
3.2	Tests unilatéraux de comparaison à une fréquence de référence	9
3.2.1	Tester $f = f_0$ contre $f < f_0$	9
3.2.2	Tester $f = f_0$ contre $f > f_0$	10
4	Tests de comparaison entre deux populations	11
4.1	Tests bilatéraux comparant des moyennes calculées dans deux populations	11
4.2	Tests bilatéraux comparant des fréquences calculées dans deux populations	12
4.3	Tests unilatéraux comparant des moyennes calculées dans deux populations	12
4.4	Tests unilatéraux comparant des fréquences calculées dans deux populations	12