Practice of statistical tests of conformity

Keita, Moussa

April 2016

Pratique des tests statistiques de conformité

Par
Moussa Keita, PhD*
Avril 2016
(Version 1)

*Ecole d'Economie, Université d'Auvergne Clermont Ferrand 1 ;
Ecole Nationale Supérieure de Statistique et d'Economie Appliquée ENSEA-C.I

Contact info: Email : keitam09@ymail.com

Codes JEL: C1

Mots clés: statistique, test, probabilités.

AVANT-PROPOS

Ce fascicule s'inscrit dans la continuité du premier document intitulé «*Introduction à la méthode statistique et probabiliste* » et vise à apporter quelques précisions et des détails sur la conduite des tests paramétriques. Il se focalise, à cet effet sur le cas particulier des tests de conformité. Un accent particulier est mis sur le calcul des statistiques des tests, leurs lois de distributions ainsi que les règles de décision quant à la conclusion du test. Nous menons également une ample discussion sur les notions de risque d'erreur de première espèce mais aussi sur la notion de p-valeur ou "p-value" qui sont des outils nécessaires pour la construction et la conclusion d'un test.

Table des matières

1. Introduction

1.1. Généralités sur les tests

La plupart des problèmes statistiques sont des problèmes de décision ou de choix. Dans le cas des tests d'hypothèses le choix (ou la décision) consiste à déterminer, parmi deux éventualités, celle qui peut être considérée comme « vraie ».

On dispose d'un modèle statistique dans lequel les variables suivent des lois de probabilité.

Comme dans tout problème de décision, dans un test statistique, il faut avoir des règles de décision. La méthode de construction de ces règles décision constitue l'objet de cet exposé. La construction d'un règle de décision dans un test statistique fait entrer en jeux plusieurs notions statistiques que sont notamment, l'hypothèse nulle, l'hypothèse alternative, les risque d'erreurs, la statistique de tests ou encore la valeur p.

En statistique, pour examiner une hypothèse nulle contre une alternative, on dispose généralement de deux types de tests : paramétriques ou non paramétriques. En effet, lorsque l'on connaît les lois de probabilités des variables analysées, on utilise des tests paramétriques. En revanche, lorsque l'on ne connaît pas les lois de probabilités des variables de test, on utilise plutôt des tests non-paramétriques.

Dans ce document, nous nous intéressons uniquement aux tests paramétriques en se focalisant sur les tests de conformité qui sont des tests dans lesquels on cherche à comparer la valeur d'un paramètre à une valeur théorique donnée.

1.2. Formulation de l'hypothèse d'un test statistique

Pour pouvoir décider entre plusieurs éventualités statistiques, on met en avant une hypothèse particulière appelée l'hypothèse nulle (notée H0) tout en formulant une hypothèse alternative qui notée H1.

Très souvent l'hypothèse H1 traduit la situation contraire de H0. Mais cela n'est pas nécessairement le cas ; il arrive que l'hypothèse H1 soit plus restrictive.

Par exemple, si l'hypothèse H0 est

$$H_0 : \theta = \theta_0$$

l'hypothèse H1 pourrait être

$$H_1 : \theta \neq \theta_0$$

$$H_1 : \theta < \theta_0$$

$$H_1 : \theta > \theta_0$$

Dans le premier cas, on parle de test bilatéral et dans les deux autres cas de test unilatéral (à gauche ou à droite).

1.3. Risques d'erreur

Dans la prise de décision lors d'un test, on court un double risque d'erreur. D'une part, on peut décider que H0 est fausse alors qu'elle est vraie. C'est le risque de première espèce, noté α. D'autre part, on peut décider que H0 est vraie alors qu'elle est fausse. C'est le risque de deuxième espèce, noté β.

Le risque α intéresse l'utilisateur du test : pour lui, H0 est rejetée ou pas au risque α. C'est le risque de rejeter à tort H0. Il est courant de fixer α = 0,05 (5%) ou α = 0,01 (1%). Dans certains domaines comme l'économie, on peut fixer le seuil d'erreur jusqu'à 10%.

Le risque β intéresse le concepteur du test : c'est le risque d'accepter à tort H0. Par conséquent, pour lui, 1-β représente la puissance du test puisque c'est le risque d'accepter à juste titre.

Le tableau suivant résume la situation :

Décision	**H0 Vraie (H1 Fausse)**	**H0 Fausse (H1 Vraie)**
Accepter H0	$1-\alpha$ Décision correcte	β Erreur de 2ème espèce
Rejeter H0	α Erreur de 1ère espèce	1-β Décision correcte

Remarque :

1. Les procédures de tests fixent une limite supérieure au risque de première espèce. On prend souvent la valeur 5% (significatif) ou 1% (très significatif). Cette valeur (aussi appelée seuil d'erreurs) représente le niveau de signification du test.

2. On souhaite minimiser à la fois les risques α et β mais, pour un échantillon donné, une diminution du risque α conduit à une augmentation du risque β.

3. L'erreur de seconde espèce diminue si la taille de l'échantillon augmente.

4. Un test unilatéral est plus puissant qu'un test bilatéral.

1.4. Région de rejet et région d'acceptation

Les tests procèdent tous schématiquement de la même manière : on dispose d'une variable de décision X suivant une loi théorique donnée lorsque l'hypothèse H0 est vraie.

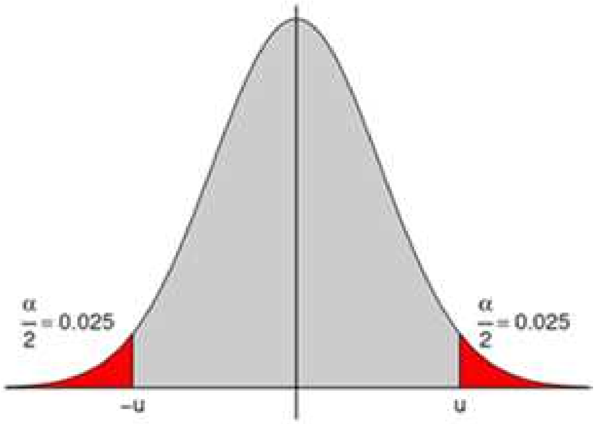
La région de rejet est l'ensemble des valeurs de la variable de décision qui sont "improbables" lorsque H0 est vraie. Ce sont des valeurs qui conduisent à rejeter cette hypothèse. Mais, si H0 était vraie, ce serait un rejet à tort. Or justement c'est la probabilité que mesure le risque α.

Par conséquent, le seuil α définit la taille de la région de rejet. C'est une région située sous la courbe de la distribution d'échantillonnage. Cette région peut prendre deux formes différentes :

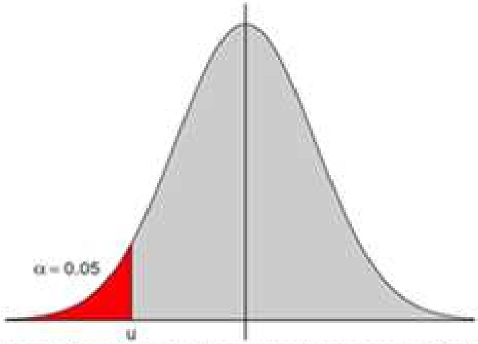
Si on fait un test unilatéral, elle est entièrement à une extrémité de la distribution de probabilité ;

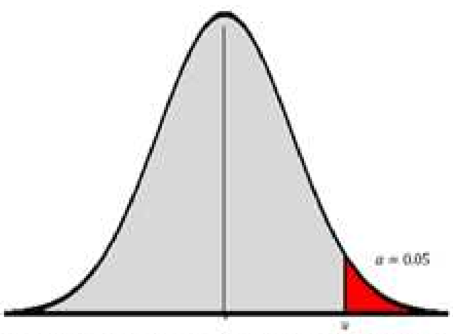
Si on fait un test bilatéral, elle est en deux morceaux de surface $\frac{\alpha}{2}$ à chaque extrémité de la distribution.

Les graphique ci-dessous sont les illustrations dans le cas où la variable de décision X suit une loi normale centrée réduite N(0; 1).



Région de rejet dans le cas du test bilatéral





Région de rejet dans le cas du test unilatéral à gauche   Région de rejet dans le cas du test unilatéral à droite

Remarque :

Dans le cas d'un test unilatéral à gauche, pour un risque d'erreur fixé à 0,05, la limite de la région de rejet est le quantile u tel que :

$$P(X < u) = 0{,}05$$

Par exemple, pour une distribution normale, on trouve u = -1,64.

Dans le cas d'un test unilatéral à droite, pour un risque d'erreur fixé à 0.05, la limite de la région de rejet est le quantile u tel que :

$$P(X < u) = 0{,}95$$

Pour une distribution normale, on trouve u = +1, 64.

Dans le cas d'un test bilatéral, les limites de la région de rejet sont les quantiles -u et u tels que :

$$P(-u < X < u) = 0{,}95$$

et

$$P(X < u) = 0{,}975$$

On trouve, pour une distribution normale, u = 1,96.

La région d'acceptation est le complémentaire de la région de rejet. C'est l'ensemble des valeurs observées de la variable de décision pour lesquelles l'hypothèse H0 est acceptable.

La règle de décision des tests consiste à regarder si la valeur de la variable de décision se trouve dans la région d'acceptation ou dans la région de rejet.

1.5. Démarche générale d'un test

Les étapes de la mise en œuvre d'un test d'hypothèse sont les suivantes :

1. Choix du risque α
2. Choix des hypothèses H0 et H1
3. Détermination de la variable de décision i.e la variable sur laquelle porte le test (moyenne, proportion, variance, etc..)
4. Détermination de la région critique i.e la région de rejet de H0 : Choisir RC telle que $P(RC\ |H0\ )\ = \alpha$
5. Calcul éventuel de la puissance du test : $P(RC|H0\ ) = 1 - \alpha$
6. Calcul, sur l'échantillon, de la valeur expérimentale de la variable de décision i.e. la statistique du test.
7. Conclusion du test : rejet ou acceptation de H0 par comparaison de la valeur expérimentale à la valeur théorique de la statistique de test.

2. Tests de conformité sur la moyenne

2.1. Présentation

Exemple introductif

Le montant moyen des achats par client dans toutes les succursales d'une enseigne commerciale est de 50 € avec un écart-type de 5 €. Afin de vérifier les performances d'un point de vente, le directeur prélève un échantillon des achats effectués par 20 consommateurs. Les résultats observés sont les suivants :

44.49	43.66	48.32	48.95	49.30	53.05	44.52	51.50	46.94	50.28
47.73	53.38	43.72	49.95	45.78	48.56	38.14	46.82	50.78	49.38

Les résultats fournis par cet échantillon sont-ils conformes aux valeurs globales ? En d'autres termes, la moyenne dans ce point de vente est-elle égale, supérieure ou inférieure à la moyenne générale ? La réponse à chacune de ces questions nécessite la mise en place d'un test de conformité qui peut se décliner en un test bilatéral, unilatéral (à gauche) ou unilatéral (à droite).

On suppose que les achats suivent une loi normale N(50,5).

La réalisation d'un test de conformité est toujours basée sur le calcul d'une statistique dite « *statistique du test* » en utilisant d'une part l'espérance et d'autre part la variance du paramètre étudié et en appliquant la loi des grandes (ou le théorème central limite).

Cependant, il faut avoir à l'esprit que le fait que la variance de la variable de test soit connue ou pas a une implication importante dans la conduite du test. En effet lorsque la variance de la variable de test n'est pas connue, il faut alors l'estimer à partir de l'échantillon et utiliser cette valeur dans le test. Cela peut modifier la loi de distribution de la statistique de test. Pour le montrer, considérons par exemple une variable aléatoire X suivant une loi normale de moyenne m et de variance σ^2. On montre d'abord que si : $X \rightsquigarrow N(m, \sigma^2)$ alors : $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \rightsquigarrow N\left(m, \frac{\sigma^2}{n}\right)$. Ainsi lorsque la variance σ^2 est connue, en appliquant le théorème central limite, on a

$$Z = \frac{\bar{X}-m}{\frac{\sigma}{\sqrt{n}}} = \frac{\sqrt{n}(\bar{X}-m)}{\sigma} \rightsquigarrow N(0,1).$$

Lorsque σ^2 est connue, la statistique de test définie par $\frac{\sqrt{n}(\bar{X}-m)}{\sigma}$ est distribuée selon une loi normale (0, 1).

Cependant lorsque la variance σ^2 n'est pas connue, on utilise l'expression de la variance estimée telle que : $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Cherchons en à intégrer cette valeur dans la construction de la statistique de test.

On sait que:

$$X_i \leadsto N(m, \sigma^2) \Rightarrow (X_i - \bar{X}) \leadsto N(0, \sigma^2)$$

$$\Rightarrow \quad \frac{(X_i - \bar{X})}{\sigma} \leadsto N(0,1)$$

En élevant au carré et en prenant la somme on a :

$$\Rightarrow \quad \sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sigma}\right)^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} = \frac{(n-1)\hat{\sigma}^2}{\sigma^2} \leadsto \chi^2(n-1)$$

Ainsi, on peut écrire cette expression telle que :

$$\frac{\hat{\sigma}^2}{\sigma^2} \leadsto \frac{\chi^{2(n-1)}}{n-1}$$

Cette expression montre que le rapport entre la variance estimée et la vraie variance est distribuée selon une loi de khi-deux à n-1 degré de liberté (normée par son degré de liberté).

Maintenant faisons le rapport entre Z et la racine carrée de cette loi de khi-deux. On a :

$$\frac{\frac{\sqrt{n}(\bar{X} - m)}{\sigma}}{\frac{\hat{\sigma}}{\sigma}} = \frac{N(0,1)}{\sqrt{\frac{\chi^{2(n-1)}}{n-1}}}$$

$$\frac{\sqrt{n}(\bar{X} - m)}{\hat{\sigma}} = \frac{N(0,1)}{\sqrt{\frac{\chi^{2(n-1)}}{n-1}}}$$

Mais on sait que le rapport entre une loi normale N(0, 1) et la racine carrée d'une loi de khi-deux divisée par son degré de liberté est distribuée selon une loi de Student. On peut alors écrire :

$$\frac{\sqrt{n}(\bar{X} - m)}{\hat{\sigma}} \leadsto t(n-1)$$

Cette propriété montre que la connaissance de la variance a donc une implication importante dans la conduite des tests d'hypothèses.

En reprenant l'exemple ci-dessus, le montant moyen des achats par client dans toutes les succursales d'une enseigne commerciale est de 50 € avec un écart-type de 5 €. Mais la moyenne calculée sur un échantillon de 20 consommateur est $\bar{X} = 47{,}7625$.

Nous allons utiliser ces informations pour mettre en œuvre trois types de tests de conformité : le test bilatéral, le test unilatéral à gauche et le test unilatéral à droite.

2.2. Test bilatéral sur la moyenne

Soit X une variable aléatoire suivant une loi normale de moyenne m et de variance σ^2. On souhaite tester l'hypothèse suivante :

$$\begin{cases} H_0 \;\; m = m_0 \\ H_1 \;\; m \neq m_0 \end{cases}$$

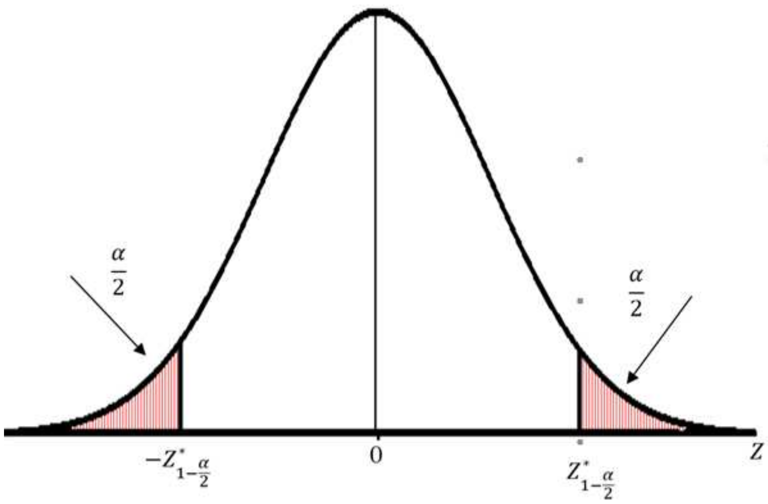
Ici m est la moyenne théorique dont $\bar{X}$(moyenne empirique constitue une estimation ponctuelle calculée à partir de l'échantillon) et m_0 est la valeur théorique avec laquelle on teste la conformité.

2.2.1. Cas où la variance σ^2 est connue

Lorsque σ^2 est connue, avec le théorème central limite, on peut poser :

$$Z = \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} \rightsquigarrow N(0,1)$$

En fixant le seuil de première espèce α, on lit d'abord dans la table de la loi normale centrée réduite le fractile correspondant à $1 - \frac{\alpha}{2}$. (noté $Z^*_{1-\frac{\alpha}{2}}$). Ensuite, on positionne cette valeur sur l'échelle d'une distribution normale telle que illustrée ci-dessous



Puisqu'il s'agit d'un test bilatéral et que la distribution de la loi normale est symétrique, on repartit le risque d'erreur entre $Z^*_{1-\frac{\alpha}{2}}$ et son opposée $-Z^*_{1-\frac{\alpha}{2}}$. Cette répartition du risque d'erreur est représentée par les deux zones hachurées en rouges où chaque zone correspond à la moitié du risque α. Ainsi, étant donné cette répartition, on peut écrire que :

$$P\left(\frac{\bar{X}-m}{\frac{\sigma}{\sqrt{n}}} < -Z^*_{1-\frac{\alpha}{2}}\right) + P\left(\frac{\bar{X}-m}{\frac{\sigma}{\sqrt{n}}} > Z^*_{1-\frac{\alpha}{2}}\right) = \alpha$$

$$P\left(Z < -Z^*_{1-\frac{\alpha}{2}}\right) + P\left(\mathrm{Z} > Z^*_{1-\frac{\alpha}{2}}\right) = \alpha$$

Mais sachant que $P\left(Z < -Z^*_{1-\frac{\alpha}{2}}\right) = P\left(Z > Z^*_{1-\frac{\alpha}{2}}\right)$ (propriété d'une loi symétrique), on a :

$$2\,P\left(\mathrm{Z} > Z^*_{1-\frac{\alpha}{2}}\right) = \alpha$$

Où Z est la statistique du test calculée $\frac{\bar{X}-m}{\frac{\sigma}{\sqrt{n}}}$ et $Z^*_{1-\frac{\alpha}{2}}$ le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi normale centrée réduite (qui doit être lue dans la table de loi normale centrée réduite).

Aussi, comme $P\left(Z < -Z^*_{1-\frac{\alpha}{2}}\right) + P\left(\mathrm{Z} > Z^*_{1-\frac{\alpha}{2}}\right) = \alpha$, on peut aussi écrire que :

$$P\left(-Z^*_{1-\frac{\alpha}{2}} < Z < Z^*_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

Cette expression équivaut mathématique à celle-ci :

$$P\left(|Z| < Z^*_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

Dès lors on peut utiliser l'une des deux expressions pour prendre la décision du test : soit $2\,P\left(\mathrm{Z} > Z^*_{1-\frac{\alpha}{2}}\right) = \alpha$ qui exprime le seuil d'erreur ou $P\left(|Z| < Z^*_{1-\frac{\alpha}{2}}\right) = 1-\alpha$ qui exprime le seuil de confiance.

Dans l'un ou l'autre cas, on compare la valeur Z calculée à la valeur de $Z^*_{1-\frac{\alpha}{2}}$ lue dans la table de la loi normale.

Règle de décision :

Lorsque $|Z| > Z^*_{1-\frac{\alpha}{2}}$, on rejette l'hypothèse H0. En revanche lorsque $|Z| < Z^*_{1-\alpha}$, on ne peut pas rejeter H0.

La région critique de ce test (encore appelée région de rejet de H0) se définit telle que :

$$RC = \left\{ \left| \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} \right| > Z^*_{1-\frac{\alpha}{2}} \right\} \ soit \ \ \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} \notin \left[-Z^*_{1-\frac{\alpha}{2}} \ ; \ Z^*_{1-\frac{\alpha}{2}} \right]$$

Ou

$$RC = \left\{ |\bar{X} - m| > Z^*_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right\} \ soit \ \ \bar{X} \notin \left] m - Z^*_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \ ; \ \mathrm{m} + Z^*_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$$

Connaissant donc la région critique, on peut définir la région d'acceptation de H0 sachant que $\left(\left| \frac{\bar{X}-m}{\frac{\sigma}{\sqrt{n}}} \right| < Z^*_{1-\frac{\alpha}{2}} \right) = 1 - \alpha$. Dès lors, la région d'acceptation se définit comme suit.

$$RA = \left\{ -Z^*_{1-\alpha} < \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} < Z^*_{1-\frac{\alpha}{2}} \right\}$$

Ou

$$RA = \left\{ \mathrm{m} - Z^*_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \ \bar{X} < \mathrm{m} + Z^*_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right\}$$

Ainsi connaissant la région de rejet ou la région d'acceptation, on peut donner une autre règle de décision par rapport au test. En effet après avoir calculée la moyenne $\bar{X}$ sur l'échantillon, on regarde si sa valeur appartient ou pas à la région d'acception. Ainsi si $\bar{X}$ appartient à l'intervalle RA, on ne peut pas rejeter H0. Par contre si $\bar{X}$ appartient de RC, on rejette H0.

Application :

La variance est connu ($\sigma = 5$), $\alpha = 0{,}05$ et m $= 50$, $\bar{X} = 47{,}7625$ et n=20.

Ainsi

$$Z = \frac{47{,}7625 - 50}{\frac{5}{\sqrt{20}}} = -2{,}0013$$

On sait que dans la table de la loi normale $Z^*_{1-\frac{\alpha}{2}} = 1{,}96$ pour $\alpha = 0{,}05$.

On peut obtenir cette valeur avec Excel en utilisant la fonction suivante :

=LOI.NORMALE.STANDARD.INVERSE.N(0,975)

Ainsi, comme $|Z| > Z^*_{1-\frac{\alpha}{2}}$, on rejette H0 au seuil de 5%.

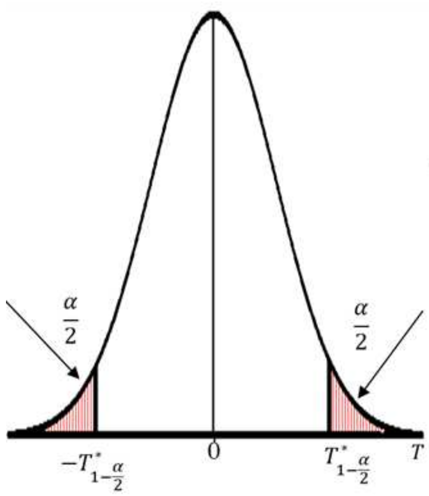
2.2.2. Cas où la variance σ^2 n'est pas connue

Lorsque la variance σ^2 n'est pas connue, on utilise la variance estimée telle que: $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Dès lors, en suivant les étapes décrites à la section 2.1, on obtient la statistique de test qui se présente comme suit :

$$\frac{\sqrt{n}(\bar{X} - m)}{\hat{\sigma}} \leadsto T(n-1)$$

Cette statistique suit une loi Student

En fixant le seuil de première espèce α, on lit d'abord dans la table de la loi de Student le fractile correspondant à $1 - \frac{\alpha}{2}$. (noté $T^*_{1-\frac{\alpha}{2}}$). Ensuite, on positionne cette valeur sur l'échelle d'une distribution normale telle que illustrée ci-dessous (notons que la courbe de la loi de Student est très proche de celle de la loi normale. Mais elle est légèrement plus étirée que la loi normale)



Puisqu'il s'agit d'un test bilatéral et que la distribution de la loi de Student est symétrique, on repartit le risque d'erreur entre $T^*_{1-\frac{\alpha}{2}}$ et son opposé $-T^*_{1-\frac{\alpha}{2}}$. Cette répartition du risque d'erreur est représentée par les deux zones hachurées en rouges. Ainsi, on peut écrire que :

$$P\left(\frac{\sqrt{n}(\bar{X} - m)}{\hat{\sigma}} < -T^*_{1-\frac{\alpha}{2}}\right) + P\left(\frac{\sqrt{n}(\bar{X} - m)}{\hat{\sigma}} > T^*_{1-\frac{\alpha}{2}}\right) = \alpha$$

$$P\left(T < -T^*_{1-\frac{\alpha}{2}}\right) + P\left(T > T^*_{1-\frac{\alpha}{2}}\right) = \alpha$$

Mais sachant que $P\left(T < -T^*_{1-\frac{\alpha}{2}}\right) = P\left(T > T^*_{1-\frac{\alpha}{2}}\right)$ (propriété d'une loi symétrique), on a :

$$2\,P\left(T > T^*_{1-\frac{\alpha}{2}}\right) = \alpha$$

Où T est la statistique du test calculée et $T^*_{1-\frac{\alpha}{2}}$ le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi de Student.

Par ailleurs sachant que $P\left(T < -T^*_{1-\frac{\alpha}{2}}\right) + P\left(T > T^*_{1-\frac{\alpha}{2}}\right) = \alpha$, cela signifie que :

$$P\left(-T^*_{1-\frac{\alpha}{2}} < T < T^*_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

$$P\left(|T| < T^*_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

Dès lors on peut utiliser l'une des deux expressions pour prendre la décision du test : soit $2\,P\left(T > T^*_{1-\frac{\alpha}{2}}\right) = \alpha$ qui exprime le seuil d'erreur ou $P\left(|T| < T^*_{1-\frac{\alpha}{2}}\right) = 1-\alpha$ qui exprime le seuil de confiance. Dans l'un ou l'autre des cas, on compare la valeur T calculée à la valeur de $T^*_{1-\frac{\alpha}{2}}$ lue dans la table de la loi de Student.

Règle de décision :

Lorsque $|T| > T^*_{1-\frac{\alpha}{2}}$, on rejette l'hypothèse H0. En revanche lorsque $|T| < T^*_{1-\alpha}$, on ne peut pas rejeter H0.

La région critique de ce test (encore appelée région de rejet de H0) se définit telle que :

$$RC = \left\{\left|\frac{\sqrt{n}(\bar{X}-m)}{\hat{\sigma}}\right| > T^*_{1-\frac{\alpha}{2}}\right\} \; soit \;\; \frac{\sqrt{n}(\bar{X}-m)}{\hat{\sigma}} \notin \left[-T^*_{1-\frac{\alpha}{2}}\,;\,T^*_{1-\frac{\alpha}{2}}\right]$$

Ou

$$RC = \left\{|\bar{X}-m| > T^*_{1-\frac{\alpha}{2}}\frac{\hat{\sigma}}{\sqrt{n}}\right\} \; soit \;\; \bar{X} \notin \left]\mathrm{m} - T^*_{1-\frac{\alpha}{2}}\frac{\hat{\sigma}}{\sqrt{n}}\,;\,\mathrm{m} + T^*_{1-\frac{\alpha}{2}}\frac{\hat{\sigma}}{\sqrt{n}}\right]$$

Connaissant donc la région critique, on peut définir la région d'acceptation de H0 sachant que $\left(\left|\frac{\sqrt{n}(\bar{X}-m)}{\hat{\sigma}}\right| < T^*_{1-\frac{\alpha}{2}}\right) = 1-\alpha$. Dès lors, la région d'acceptation se définit comme suit.

$$RA = \left\{-T^*_{1-\alpha} < \frac{\sqrt{n}(\bar{X} - m)}{\hat{\sigma}} < T^*_{1-\frac{\alpha}{2}}\right\}$$

Ou

$$RA = \left\{\mathrm{m} - T^*_{1-\frac{\alpha}{2}}\frac{\hat{\sigma}}{\sqrt{n}} < \ \bar{X} < \mathrm{m} + T^*_{1-\frac{\alpha}{2}}\frac{\hat{\sigma}}{\sqrt{n}}\right\}$$

2.3. Test unilatéral (à droite)

D'une manière générale, le test unilatéral à droite se présente comme suit :

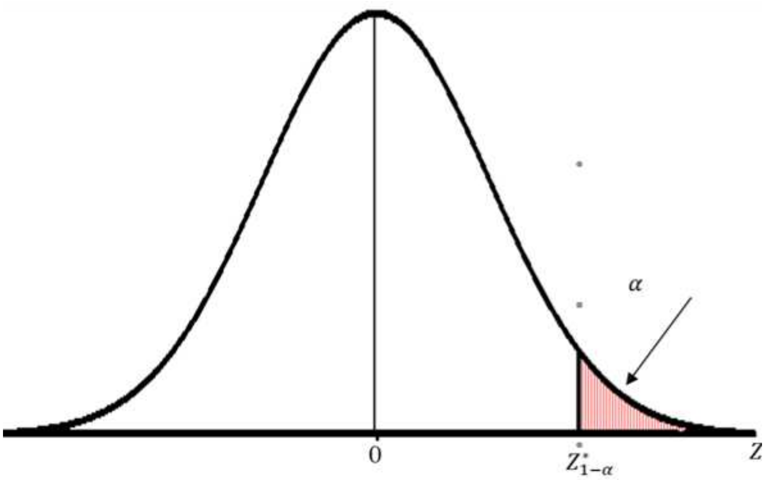
$$\begin{cases} H_0 \ \ m = m_0 \\ H_1 \ \ m > m_0 \end{cases}$$

Pour conduire ce test, deux cas se présentent : cas où la variance est connue et cas où la variance n'est pas connue.

2.3.1. Cas où la variance σ^2 est connue

Lorsque la variance est connue la statistique du test sous H_0 suit une loi normale$N(0,1)$.

Ainsi en fixant un seuil d'erreur α, la région critique et la région d'acceptation peuvent être illustrées comme suit :



Ainsi connaissant le seuil d'erreur α on définit la région critique telle que :

$$P\left(\frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} > Z^*_{1-\alpha}\right) = \alpha$$

$$P(Z > Z^*_{1-\alpha}) = \alpha$$

Où Z est la statistique du test calculée et $Z_{1-\alpha}^{*}$ le quantile d'ordre $1-\alpha$ de la loi normale centrée réduite.

Règle de décision :

Lorsque $Z > Z_{1-\alpha}^{*}$, on rejette l'hypothèse H0. En revanche lorsque $Z < Z_{1-\alpha}^{*}$, on ne peut pas rejeter H0.

NB : Il ne s'agit pas ici de comparer $|Z|$ à $Z_{1-\alpha}^{*}$ mais simplement de comparer Z à $Z_{1-\alpha}^{*}$ (qu'il soit positif ou négatif).

La région critique de ce test (région de rejet de H0) se définit alors comme suit :

$$RC = \left\{ \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} > Z_{1-\alpha}^{*} \right\}$$

Ou

$$RC = \left\{ \bar{X} > \mathrm{m} + Z_{1-\alpha}^{*} \frac{\sigma}{\sqrt{n}} \right\}$$

Connaissant donc la région critique, on peut définir la région d'acceptation de H0 sachant que $\left(\frac{\bar{X}-m}{\frac{\sigma}{\sqrt{n}}} < Z_{1-\alpha}^{*} \right) = 1-\alpha$. Dès lors, la région d'acceptation se définit comme suit.

$$RA = \left\{ \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} < Z_{1-\alpha}^{*} \right\} \; soit \; \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} \in \left]-\infty\,;\, Z_{1-\alpha}^{*}\right]$$

Ou

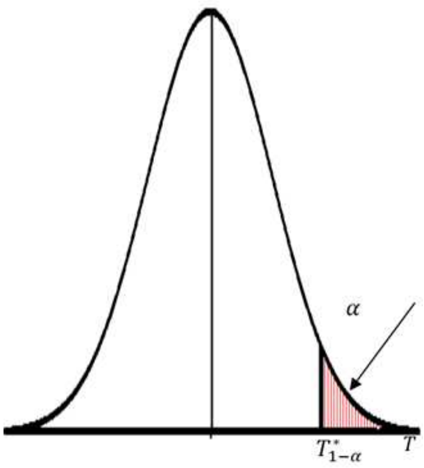
$$RA = \left\{ \bar{X} < \mathrm{m} + Z_{1-\alpha}^{*} \frac{\sigma}{\sqrt{n}} \right\} \; soit \; \bar{X} \in \left]-\infty\,;\, \mathrm{m} + Z_{1-\alpha}^{*} \frac{\sigma}{\sqrt{n}}\right]$$

2.3.2. Cas où la variance σ^2 n'est pas connue :

Lorsque σ^2 n'est pas connue, appliquant le théorème central-limite, on obtient une loi de Student :

$$\frac{\sqrt{n}(\bar{X} - m)}{\hat{\sigma}} \rightsquigarrow T(n-1)$$

Ainsi en fixant un seuil d'erreur α, la région critique et la région d'acceptation peuvent être illustrées comme suit :



Ainsi connaissant le seuil d'erreur α on définit la région critique du test telle que :

$$P\left(\frac{\sqrt{n}(\bar{X}-m)}{\hat{\sigma}} > T^*_{1-\alpha}\right) = \alpha$$

$$P(T > T^*_{1-\alpha}) = \alpha$$

Où T est la statistique du test calculée et $T^*_{1-\alpha}$ le quantile d'ordre $1-\alpha$ de la loi de la loi de Student.

Règle de décision :

Lorsque $T > T^*_{1-\alpha}$, on rejette l'hypothèse H0. Et lorsque $T < T^*_{1-\alpha}$, on ne peut pas rejeter H0.

NB : Là aussi, il ne s'agit pas de comparer $|T|$ à $T^*_{1-\alpha}$ mais simplement de comparer T à $T^*_{1-\alpha}$ (qu'il soit positif ou négatif).

La région critique du test se définit comme suit :

$$RC = \left\{\frac{\sqrt{n}(\bar{X}-m)}{\hat{\sigma}} > T^*_{1-\alpha}\right\}$$

Ou

$$RC = \left\{\bar{X} > \mathrm{m} + T^*_{1-\alpha}\frac{\hat{\sigma}}{\sqrt{n}}\right\}$$

Connaissant donc la région critique, on peut définir la région d'acceptation de H0 sachant que $\left(\frac{\sqrt{n}(\bar{X}-m)}{\hat{\sigma}} < T^*_{1-\alpha}\right) = 1-\alpha$.

$$RA = \left\{\frac{\sqrt{n}(\bar{X} - m)}{\hat{\sigma}} < T^*_{1-\alpha}\right\} \ soit \quad \frac{\sqrt{n}(\bar{X} - m)}{\hat{\sigma}} \in \left]-\infty\, ;\, T^*_{1-\alpha}\right]$$

Ou

$$RA = \left\{\bar{X} < \mathrm{m} + T^*_{1-\alpha}\frac{\hat{\sigma}}{\sqrt{n}}\right\} \ soit \ \ \bar{X} \ \in \left]-\infty\, ;\, \mathrm{m} + T^*_{1-\alpha}\frac{\hat{\sigma}}{\sqrt{n}}\right]$$

2.4. Test unilatéral (à gauche)

La forme générale d'un test de conformité unilatérale à gauche sur la moyenne se présente comme suit :

$$\begin{cases} H_0 \ \ m = m_0 \\ H_1 \ \ m < m_0 \end{cases}$$

Tout comme pour le test bilatéral ou le test unilatéral à droite, on peut distinguer deux cas selon que la variance soit connue ou pas.

2.4.1. Cas où la variance σ^2 est connue

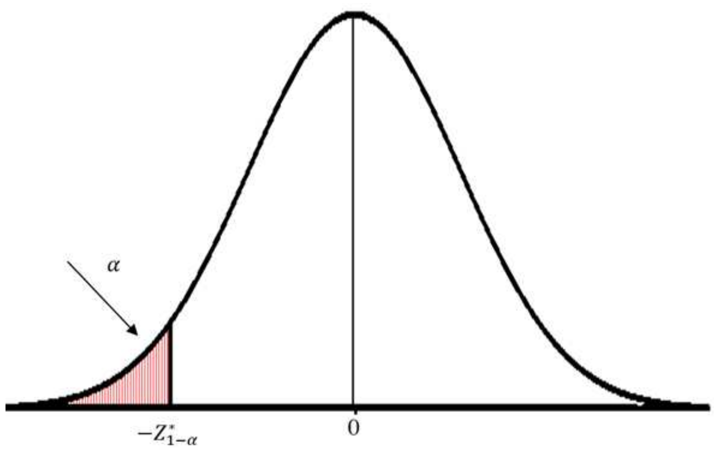
Lorsque la variance est connue la statistique du test sous H_0 suit une loi normale $N(0,1)$. Ainsi connaissant le seuil d'erreur α on définit la région critique telle que :

$$P\left(\frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} < -Z^*_{1-\alpha}\right) = \alpha$$

$$P(Z < -Z^*_{1-\alpha}) = \alpha$$

Où Z est la statistique du test calculée et $Z^*_{1-\alpha}$ le quantile d'ordre $1 - \alpha$ de la loi normale centrée réduite.

La région d'acceptation et la région critique sont illustrées sur le graphique ci-dessous.



Règle de décision :

Lorsque $Z < -Z^*_{1-\alpha}$, on rejette l'hypothèse H0. Et lorsque $Z > -Z^*_{1-\alpha}$, on ne peut pas rejeter H0.

NB : On peut remarquer ici que contrairement au test unilatéral à droite dans lequel on compare Z à $Z^*_{1-\alpha}$, dans le test unilatéral à gauche, on compare Z à $-Z^*_{1-\alpha}$ (c'est-à-dire l'opposée de $Z^*_{1-\alpha}$ obtenue par lecture dans la table de loi normale).

La région critique du test se définit alors comme suit :

$$RC = \left\{ \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} < -Z^*_{1-\alpha} \right\}$$

Ou

$$RC = \left\{ \bar{X} < \mathrm{m} - Z^*_{1-\alpha} \frac{\sigma}{\sqrt{n}} \right\}$$

Sachant que $\left(\frac{\bar{X}-m}{\frac{\sigma}{\sqrt{n}}} > -Z^*_{1-\alpha} \right) = 1 - \alpha$, on peut définir la région d'acceptation de H0 comme suit.

$$RA = \left\{ \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} > -Z^*_{1-\alpha} \right\} \ soit \ \ \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} \in \left]-Z^*_{1-\alpha} \, ; +\infty\right]$$

Ou

$$RA = \left\{ \bar{X} > \mathrm{m} - Z^*_{1-\alpha} \frac{\sigma}{\sqrt{n}} \right\} \ soit \ \ \bar{X} \in \left] \mathrm{m} - Z^*_{1-\alpha} \frac{\sigma}{\sqrt{n}} \, ; +\infty \right]$$

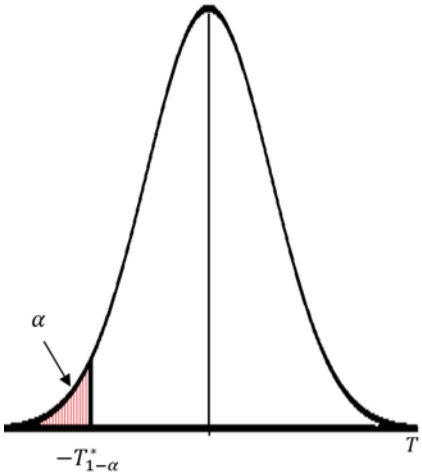
2.4.2. Cas où la variance σ^2 n'est pas connue

Lorsque σ^2 n'est pas connue, sachant que $\frac{\sqrt{n}(\bar{X}-m)}{\hat{\sigma}} \backsim T(n-1)$ et connaissant le seuil d'erreur α on définit la région critique du test telle que :

$$P\left(\frac{\sqrt{n}(\bar{X} - m)}{\hat{\sigma}} < -T^*_{1-\alpha} \right) = \alpha$$

$$P(T < -T^*_{1-\alpha}) = \alpha$$

Où T est la statistique du test calculée et $T^*_{1-\alpha}$ le quantile d'ordre $1-\alpha$ de la loi de la loi de Student. La région critique de ce test est illustrée sur le graphique ci-dessous.



Règle de décision :

Lorsque $T < -T^*_{1-\alpha}$, on rejette l'hypothèse H0. Et lorsque $T > -T^*_{1-\alpha}$, on ne peut pas rejeter H0.

La région critique du test se définit comme suit :

$$RC = \left\{\frac{\sqrt{n}(\bar{X}-m)}{\hat{\sigma}} < -T^*_{1-\alpha}\right\}$$

Ou

$$RC = \left\{\bar{X} < \mathrm{m} - T^*_{1-\alpha}\frac{\hat{\sigma}}{\sqrt{n}}\right\}$$

Connaissant donc la région critique, on peut définir la région d'acceptation de H0 sachant que $\left(\frac{\sqrt{n}(\bar{X}-m)}{\hat{\sigma}} > -T^*_{1-\alpha}\right) = 1-\alpha$. Dès lors, la région d'acceptation se définit comme suit.

$$RA = \left\{\frac{\sqrt{n}(\bar{X}-m)}{\hat{\sigma}} > -T^*_{1-\alpha}\right\}\ soit\ \ \frac{\sqrt{n}(\bar{X}-m)}{\hat{\sigma}} \in \left]-T^*_{1-\alpha};\ +\infty\ \right]$$

Ou

$$RA = \left\{\bar{X} > \mathrm{m} - T^*_{1-\alpha}\frac{\hat{\sigma}}{\sqrt{n}}\right\}\ soit\ \ \bar{X}\ \in \left]\mathrm{m} - T^*_{1-\alpha}\frac{\hat{\sigma}}{\sqrt{n}};+\infty\right]$$

2.5. Notion de « valeur p »

La valeur p (ou p-value) d'un test est la probabilité associée à la statistique calculée lors de la construction du test. En prenant le cas d'une statistique Z qui suit une loi normale, la p-value de Z est la probabilité lue dans la table de la loi normale correspondant à cette valeur Z.

La p-value fournit une règle décision dans le test. En effet, lorsque la p-value est inférieure au seuil α, on rejette H0. Mais lorsque la p-value est supérieure au seuil α on ne peut pas rejeter H0.

2.6. Relation entre le risque d'erreur α, le seuil critique Z^*, la statistique du test Z et la p-value p

Le risque d'erreur α, le seuil critique Z^*, la statistique du test Z et la p-value p sont les principaux indicateurs permettant de conclure un test statistique. Les deux premiers indicateurs sont des valeurs théoriques (fixé par l'analyste) alors que les deux derniers sont des valeurs calculées lors de la construction du test. Toutefois, il existe une relation de translation entre les deux couples de valeurs.

En effet, « *la p-value est à la statistique du test ce que le risque d'erreur est au seuil critique* ». En d'autres termes la correspondance entre le risque d'erreur et le seuil critique est de même nature que la correspondance entre la p-value et la statistique calculée. Ces correspondances ont une interprétation directe en vue de la conclusion du test. Par exemple, dans le cas d'un test bilatéral, on rejette systématiquement l'hypothèse H0 lorsque la valeur absolue de la **statistique calculée** est **supérieure** au **seuil critique** lue dans la table statistique. Par contre on ne rejette pas l'hypothèse nulle H0 lorsque la valeur absolue de la **statistique calculée** est **inférieure** au **seuil critique** lue dans la table. Cette prise de décision peut aussi être basée sur le risque d'erreur et la p-value. En effet, on rejette l'hypothèse nulle lorsque la p-value est inférieure au seuil d'erreur α. Et on accepte l'hypothèse l'hypothèse nulle lorsque la p-value est supérieure au seuil d'erreur. La correspondance entre les valeurs et les probabilités est résumée dans le tableau suivant :

Soit α le risque d'erreur et Z^* le seuil théorique correspondant à ce risque d'erreur (seuil bilatéral ou unilatéral). Et soit Z la statistique de test calculée et p la p-value correspondant à cette statistique. On peut construire les correspondances suivantes :

$$\alpha \rightarrow Z^*$$
$$p \rightarrow Z$$

$$\alpha > p \;\Leftrightarrow\; Z^* < |Z| \rightarrow \text{Rejet de H0}$$

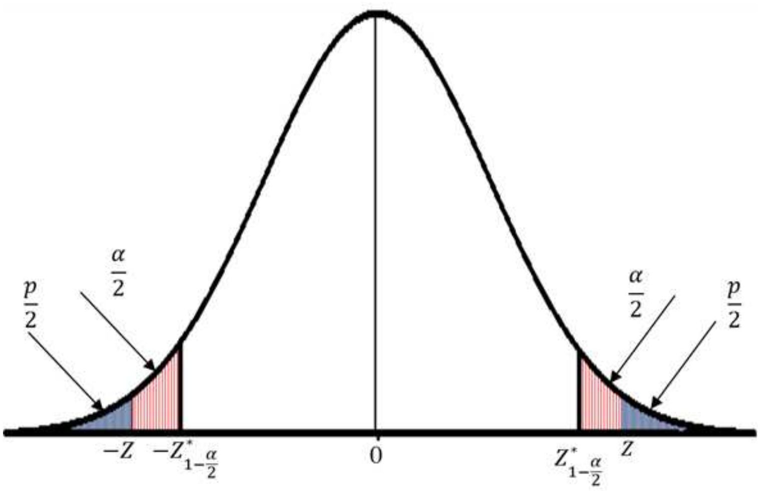
$$\alpha < p \Leftrightarrow\; Z^* > |Z| \;\rightarrow \text{Non rejet de H0}$$

Cette conclusion peut être généralisée à partir d'un graphique de décision selon qu'il s'agisse d'un test bilatéral ou un test unilatéral.

2.6.1. Graphique de décision dans le cas d'un test bilatéral

Considérons un test bilatéral où la statistique de test suit une loi normale centrée réduite Z (on pourra utiliser la démarche pour les autres lois comme la loi de Student par analogie). On peut illustrer le seuil d'erreur, le seuil critique, la statistique de test et la p-value p sur le même graphique.

Cas où l'hypothèse nulle est rejetée :



Dans un test bilatéral, lorsque l'hypothèse nulle est rejetée, cela veut dire que la statistique $|Z| > Z^*_{1-\frac{\alpha}{2}}$ c'est-à-dire que $Z \notin \left]-Z^*_{1-\frac{\alpha}{2}}, Z^*_{1-\frac{\alpha}{2}}\right[$. Cette situation implique alors que la p-value p soit inférieure au risque d'erreur α. La p-value correspond à la surface totale hachurée en bleue alors que le risque d'erreur correspond à la surface totale hachurée en rouge qui balaie également toute la zone en bleue.

Pour rappel

$$P\left(z < -Z^*_{1-\frac{\alpha}{2}}\right) + P\left(z > Z^*_{1-\frac{\alpha}{2}}\right) = \alpha$$

$$P(z < -Z) + P(z > Z) = p$$

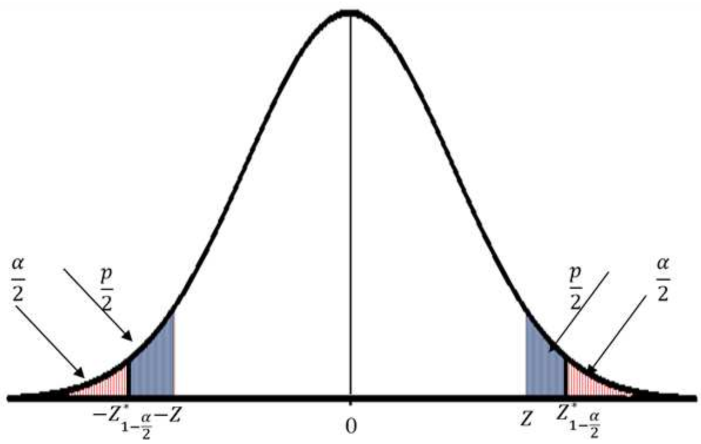
Etant donné que la loi normale est symétrique, on a :

$$P\left(Z < -Z^*_{1-\frac{\alpha}{2}}\right) = P\left(Z > Z^*_{1-\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$$

$$P(z < -Z) = P(\mathrm{z} > Z) = \frac{p}{2}$$

Cas où l'hypothèse nulle n'est pas rejetée

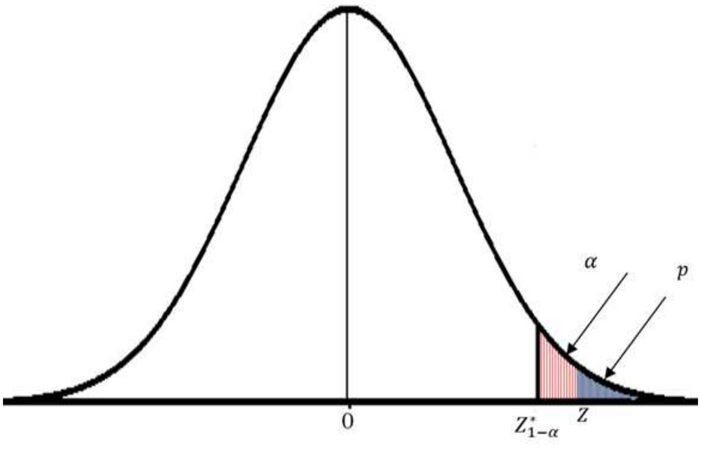
Dans le cas où l'hypothèse nulle n'est pas rejetée, le graphique de décision se présente comme suit :



Lorsque l'hypothèse nulle n'est pas rejetée, cela veut dire que la statistique $|Z| < Z^*_{1-\frac{\alpha}{2}}$ c'est-à-dire que $Z \in]-Z^*_{1-\frac{\alpha}{2}}, Z^*_{1-\frac{\alpha}{2}}[$. Cette situation implique alors que la p-value p soit supérieure au risque d'erreur α. Ainsi, contrairement au cas précédent, c'est la p-value (zone en bleue) qui balaie toute la zone en rouge (zone correspondant au risque d'erreur)..

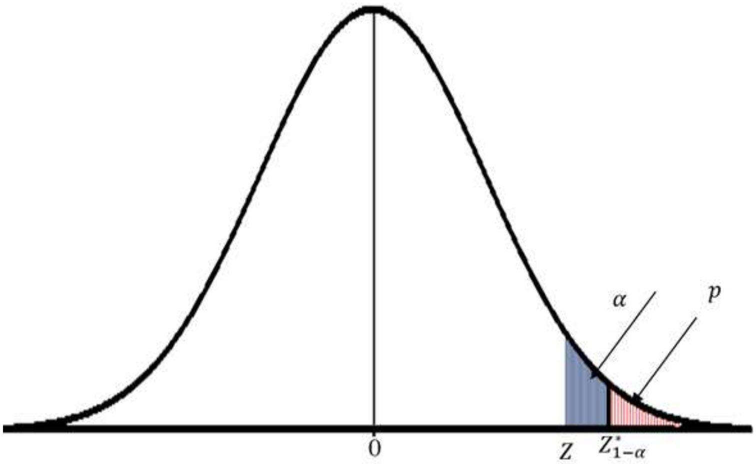
2.6.2. Graphique de décision dans le cas d'un test unilatéral à droite

Dans le cas d'un test unilatéral à droite, lorsque l'hypothèse nulle est rejetée, le graphique de décision se résume comme suit :



Le rejet de l'hypothèse nulle signifie que la statistique de test $Z > Z^*_{1-\alpha}$ c'est-à-dire que $Z \notin]-\infty\ , Z^*_{1-\alpha}[$. Ce implique alors que la p-value p (zone en bleu) soit inférieure au risque d'erreur α (zone en rouge qui balaie aussi la zone bleue).

Mais dans le cas où l'hypothèse nulle n'est pas rejetée, le graphique se présente telle que la zone hachurée en bleue (p-value) balaie toute la zone hachurée en rouge (risque d'erreur).



On peut représenter les graphiques de décision dans le cas du test unilatéral à gauche en faisant une analogie avec le cas du test unilatéral à droite en comparant Z et $-Z^*_{1-\alpha}$. On rappelle que l'hypothèse nulle est rejetée lorsque $Z < -Z^*_{1-\alpha}$.

2.7. Détermination du seuil critique Z^* connaissant le risque d'erreur α

La conduite de tout test statistique nécessite de fixer un seuil d'erreur permettant de déterminer la valeur critique qui sert de repère dans la conclusion du test. Il est donc important d'effectuer une brève présentation sur la méthode détermination de cette valeur critique.

D'une manière générale, il n'y a pas formule mathématique pour calculer la valeur critique d'un test. Celle-ci doit être lue directement à partir des tables statistiques lorsque la loi de distribution de la variable de test est connue. Par exemple dans un test de conformité sur la moyenne, lorsque la statistique de test suit une loi normale (0,1), ainsi en fixant un seuil d'erreur α, on obtient la valeur critique du test en recherchant dans la table de loi normale (0, 1) le fractile correspondant à ce risque d'erreur. Toutefois, la valeur à choisir dans cette correspondance diffère selon qu'il s'agisse d'un test bilatéral ou d'un test unilatéral.

Dans cette section, nous présentons les différentes méthodes de sélection des valeurs critiques en fonction de la nature du test mais aussi en fonction de loi statistique considérée.

2.7.1. Détermination de la valeur critique dans le cas d'une loi normale N(0,1)

Dans un test de conformité d'une moyenne la statistique du test suit une loi normale lorsque la variance de la variable de test est connue (σ^2 est connue). On peut alors faire une distinction selon qu'il s'agisse d'un test bilatéral ou unilatéral.

Valeur critique dans le cas d'un test bilatéral (loi normale)

Dans le cas d'un test bilatéral, le seuil critique correspondant à un seuil d'erreur α est le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi normale centrée réduite noté $Z^*_{1-\frac{\alpha}{2}}$. $1-\frac{\alpha}{2}$ est appelée la probabilité du seuil de critique. On utilise $1-\frac{\alpha}{2}$ car dans un test bilatéral, le risque d'erreur se divise en deux parts égales aux deux extrémités de la loi normale. $1-\frac{\alpha}{2}$ correspond à la probabilité de la borne supérieure définie par le seuil de confiance. En fixant par exemple $\alpha = 0{,}05$, on a : $1-\frac{\alpha}{2} = 0{,}975$. Il faut alors déterminer le fractile (quantile) correspondant à cette probabilité dans la table de la loi normale centrée réduite (voir Annexe pour une aide à la lecture des tables statistiques).

Microsoft EXCEL offre des fonctions capables de lire directement ces valeurs. Par exemple, pour obtenir le fractile correspondant à la probabilité 0,975, on utilise la fonction suivante :

=LOI.NORMALE.STANDARD.INVERSE.N(0.975)

On trouve 1.96

Pour généraliser cette fonction quel que soit seuil α on utilise la forme suivante:

=LOI.NORMALE.STANDARD.INVERSE.N($1-\frac{\alpha}{2}$)

Valeur critique dans le cas d'un test unilatéral à droite (loi normale)

Pour le cas d'un test unilatéral à droite, le seuil critique correspondant à un seuil d'erreur α est le quantile d'ordre $1-\alpha$ de la loi normale centrée réduite noté $Z^*_{1-\alpha}$. Ici, contrairement au test bilatéral, on utilise $1-\alpha$ comme probabilité du seuil critique car le risque d'erreur est située d'un seul côté de la distribution en l'occurrence à droite.

Pour $\alpha = 0{,}05$, on a : $1 - \alpha = 0{,}95$. Il faut alors lire le fractile correspondant à cette probabilité dans la table de la loi normale centrée réduite (Cf. Annexe pour une aide à la lecture des tables statistiques).

On peut aussi utiliser EXCEL pour calculer la valeur critique comme suit :

=LOI.NORMALE.STANDARD.INVERSE.N(0.95)

On trouve 1.64.

Pour généraliser cette fonction pour tout seuil α on utilise :

=LOI.NORMALE.STANDARD.INVERSE.N($1 - \alpha$)

Valeur critique dans le cas d'un test unilatéral à gauche (loi normale)

Le seuil critique d'un test unilatéral à gauche correspondant à un seuil d'erreur α est le quantile d'ordre α de la loi normale centrée réduite noté Z_α^* et non le quantile $1 - \alpha$ comme pour le test à droite. Ici, aussi, on utilise α comme probabilité du seuil critique au lieu de $1 - \alpha$ pour le test unilatéral à droite ou $1 - \frac{\alpha}{2}$ pour le test bilatéral. Ici le risque d'erreur est située d'un seul côté de la distribution en l'occurrence à gauche (et au début de la distribution).

Par exemple, pour $\alpha = 0{,}05$, on lit directement le fractile correspondant à cette probabilité dans la table de la loi normale centrée réduite (Cf. Annexe pour une aide à la lecture des tables statistiques).

Avec EXCEL ce fractile se calcule comme suit :

=LOI.NORMALE.STANDARD.INVERSE.N(0.05)

On trouve -1.64.

Pour généraliser cette fonction pour tout seuil α on utilise :

=LOI.NORMALE.STANDARD.INVERSE.N(α)

On peut alors noter que dans le cas du test unilatéral à gauche, le risque d'erreur est égale à la probabilité du seuil critique.

2.7.2. Détermination de la valeur critique dans le cas d'une loi de Student

Dans un test de conformité d'une moyenne, la statistique du test suit une loi de Student lorsque la variance de la variable de test n'est pas connue (σ^2 inconnue) et qu'il faut l'estimer en utilisant la variance empirique modifiée $\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

Ainsi connaissant la statistique du test et sa distribution, on peut déterminer la valeur critique du test en faisant une distinction selon qu'il s'agisse d'un test bilatéral ou unilatéral.

Valeur critique dans le cas d'un test bilatéral (loi de Student)

Dans le cas d'un test bilatéral, le seuil critique correspondant à un seuil d'erreur α est le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi de Student noté $T^*_{1-\frac{\alpha}{2}}$. On utilise $1-\frac{\alpha}{2}$ car dans un test bilatéral, le risque d'erreur se divise en deux parts égales aux deux extrémités de la loi de Student. Toutefois, le quantile $T^*_{1-\frac{\alpha}{2}}$ se lit non pas au niveau de la colonne de probabilité $1-\frac{\alpha}{2}$ mais plutôt au niveau de la colonne de probabilité α.

En effet, pour le cas de loi de Student, pour déterminer le quantile correspondant à un risque d'erreur α, on suit quatre étapes :

Dans un premier temps, on calcule la probabilité du seuil critique en se servant du risque d'erreur. Ici cette probabilité du seuil critique est $1-\frac{\alpha}{2}$; Dans second temps, on soustrait cette valeur de 1 c'est-à-dire $1-\left(1-\frac{\alpha}{2}\right)$; Dans un troisième temps, on multiplie cette probabilité résiduelle par 2. Ainsi on a : $2\times\left[1-\left(1-\frac{\alpha}{2}\right)\right]=\alpha$; Et enfin dans la quatrième étape, on recherche cette valeur obtenue dans la colonne des probabilités. Pour cela, on cherche le nombre de degré de liberté fournit dans l'énoncé sur la ligne des degrés de liberté. Le croisement entre cette ligne et la colonne de la probabilité α permet d'identifier le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi de Student $T^*_{1-\frac{\alpha}{2}}$.

Par exemple en fixant $\alpha = 0{,}05$ et en choisissant le nombre de degré de liberté ddl=30, on souhaite déterminer le quantile d'ordre $1-\frac{\alpha}{2}$ de la loi de Student $T^*_{1-\frac{\alpha}{2}}$ correspondant à un test bilatéral. En suivant les étapes décrites ci-dessous, on a :

$$\alpha = 0{,}05$$

$$\left(1-\frac{\alpha}{2}\right) = \left(1-\frac{0{,}05}{2}\right) = 0{,}975$$

$$\left[1-\left(1-\frac{\alpha}{2}\right)\right] = \left[1-\left(1-\frac{0{,}05}{2}\right)\right] = 1-0{,}975 = 0{,}025$$

$$2 \times \left[1 - \left(1 - \frac{\alpha}{2}\right)\right] = 2 \times \left[1 - \left(1 - \frac{0,05}{2}\right)\right] = 2 \times 0,025 = 0,05$$

En recherchant le croisement entre la colonne p=0,05 et la ligne n=30 dans la loi de Student, on trouve 2,042. Ainsi $T^*_{1-\frac{\alpha}{2}} = 2,042$. Ce qui représente la valeur critique du test pour $\alpha = 0,05$ et ddl=30.

D'une manière générale, dans le test bilatéral avec la loi de Student, on utilise directement le seuil comme la probabilité servant à la lecture de la table. Cela est illustrée avec les démonstrations faites dans les quatre étapes décrites ci-dessous. (voir en Annexe pour plus de détail sur la lecture d'une table de Student).

A la place d'une lecture manuelle de la table, on peut aussi utiliser la fonction excel présenté comme suit :

=LOI.STUDENT.INVERSE.BILATERALE(0.05;30)

Où 0,05 correspond au risque d'erreur alors que 30 correspond au nombre de degrés de liberté. Cette formule donne 2,042.

Pour généraliser cette fonction quelle que soit la valeur de α et du ddl, on a :

=LOI.STUDENT.INVERSE.BILATERALE(α;ddl)

Il faut aussi noter que les quatre étapes dans le calcul de la probabilité du seuil critique précédemment présentées pour le test bilatéral sont aussi valables pour le cas des tests unilatéraux avec la seule différence que le calcul de probabilité ne se base pas sur $2 \times \left[1 - \left(1 - \frac{\alpha}{2}\right)\right] = \alpha$ mais plutôt sur $2 \times [1 - (1 - \alpha)] = 2\alpha$ pour le test unilatéral à droite et $2 \times [1 - (\alpha)] = 2(1 - \alpha)$ pour le test unilatéral à gauche (voir ci-dessous).

Valeur critique dans le cas d'un test unilatéral à droite (loi de Student)

S'agissant d'un test unilatéral à droite, le seuil critique correspondant à un seuil d'erreur α est le quantile d'ordre $1 - \alpha$ de la loi de Student $Z^*_{1-\alpha}$. Ici, contrairement au test bilatéral, on utilise $1 - \alpha$ car le risque d'erreur est positionné sur un seul côté de la distribution qui est ici la droite.

Comme signalé précédemment, pour lire le quantile d'ordre $1 - \alpha$, on calcule d'abord $2 \times [1 - (1 - \alpha)]$. Ce qui donne 2α. C'est donc cette valeur que nous utilisons pour lire dans la table de Student au croisement de la ligne du nombre de degré de liberté.

Par exemple, pour $\alpha = 0{,}05$ et un ddl=30, on a $2 \times [1 - (1 - \alpha)] = 2\alpha = 0{,}10$. Il faut alors lire le fractile correspondant à cette probabilité dans la table de la loi de Student en utilisant ddl=30 et P=0,10. On trouve 1.697.

On pouvait simplement utiliser EXCEL comme suit :

=LOI.STUDENT.INVERSE.BILATERALE(0.10;30)

Où 0,10 correspond au risque d'erreur alors que 30 correspond au nombre de degrés de liberté. Cette formule donne 1,697.

Pour généraliser cette fonction quelle que soit la valeur de α et du ddl, on a :

=LOI.STUDENT.INVERSE.BILATERALE(2α;ddl)

Valeur critique dans le cas d'un test unilatéral à gauche (loi de Student)

Le seuil critique d'un test unilatéral à gauche correspondant à un seuil d'erreur α est le quantile d'ordre α de la loi de Student T_{α}^{*} et non le quantile $1 - \alpha$ comme pour le test à droite. Ici, aussi, on utilise α au lieu de $1 - \alpha$ pour le test unilatéral à droite ou $1 - \frac{\alpha}{2}$ pour le test bilatéral. Ici le risque d'erreur est située d'un seul côté de la distribution en l'occurrence à gauche (et au début de la distribution). Néanmoins, on suit la même démarche de détermination comme dans les deux autres tests. En effet, connaissant α, on calcul d'abord $2 \times [1 - (\alpha)] = 2(1 - \alpha)$. Ensuite connaissant le ddl, on recherche la valeur se trouvant au croisement entre $2(1 - \alpha)$ et ddl. Toutefois, lorsque α est faible alors $2(1 - \alpha)$ sera supérieure à 1 et devient impossible de lire à partir de la table. Il faut alors déterminer l'opposé du quantile recherché en utilisant les probabilités tabulées. En effet puisque la loi de Student est symétrique, on a $T_{\alpha}^{*} = -T_{1-\alpha}^{*}$. Il faut alors déterminer $T_{1-\alpha}^{*}$ en suivant les étapes du test du unilatéral à droite.

Avec le test unilatéral à droite, nous avons vu que $T_{1-\alpha}^{*}$ se détermine avec 2α et ddl puisque $2 \times [1 - (1 - \alpha)] = 2\alpha$. Dès lors, pour $\alpha = 0{,}05$ et ddl=30, on lit directement le fractile correspondant la probabilité 0.10 avec ddl=30. On trouve 1.697. D'où $T_{\alpha}^{*} = -T_{1-\alpha}^{*} = -1{,}697$.

Finalement, le quantile d'ordre T_{α}^{*} de la loi de Student dans un test unilatéral à gauche au seuil de 0,05 est de -1,697.

Notons qu'on pouvait directement obtenir cette valeur avec Excel, en utilisant :

=LOI.STUDENT.INVERSE.BILATERALE(1.9;30)

Où 1,9 correspond à $2(1 - \alpha)$ et 30 correspond au degré de liberté.

Cette fonction peut être utilisable quelle que soit la valeur de α sans avoir besoin de passer par le calcul de l'opposé du quantile. On a :

=LOI.STUDENT.INVERSE.BILATERALE($2(1-\alpha)$;ddl)

En conclusion, on peut dire que :

- Pour le test bilatéral, la détermination du quantile $T^*_{1-\frac{\alpha}{2}}$ se base sur α
- Pour le test unilatéral à droite, la détermination du quantile $T^*_{1-\alpha}$ se base sur 2α
- Et pour le test unilatéral à gauche, la détermination du quantile T^*_{α} se base sur $2(1-\alpha)$. Mais dans ce dernier cas, lorsque α est faible, c'est-à-dire lorsque $2(1-\alpha) > 1$, on peut être amener à déterminer l'opposé du quantile $T^*_{1-\alpha}$ sachant que $T^*_{\alpha} = -T^*_{1-\alpha}$. Mais on peut aussi opter pour la fonction Excel qui fournit directement T^*_{α} sans être obligé de passer par le calcul de $T^*_{1-\alpha}$.

2.8. Détermination de la p-value p connaissant la statistique de test Z

La méthode de détermination de la p-value d'une statistique de test dépend de la nature du test (test bilatéral ou unilatéral).

2.8.1. Détermination de la p-value dans un test bilatéral

Dans le cas d'un test bilatéral, la p-value se définit comme la probabilité d'obtenir une statistique z qui soit supérieure à la valeur absolue de statistique calculée $|Z|$ multipliée par 2.

$$Pvalue = 2 * P(z \geq |Z|)$$

$$Pvalue = 2 * [1 - P(z < |Z|)]$$

Ainsi pour obtenir la p-value, on lit d'abord dans la table la probabilité $P(z < |Z|)$. Ensuite, on calcule la p-value.

On peut facilement obtenir la p-value par une simple lecture de la table statistique ou en utilisant Microsoft EXCEL.

Par exemple, pour **la loi normale** on peut calculer $P(z < |Z|)$ en utilisant la fonction EXCEL suivante :

=LOI.NORMALE.STANDARD.N(ABS(Z);VRAI) où Z est la statistique calculée. Ensuite on calcule la p-value telle que $Pvalue = 2 * [1 - P(z < |Z|)]$

Pour **la loi de Student**, on peut utiliser l'une des trois fonctions suivantes :

=LOI.STUDENT.BILATERALE(ABS(T);ddl) où T est la statistique de Student calculée et ddl le nombre de degré de libertés de la loi de Student. Cette fonction donne directement la p-value telle que $Pvalue = 2 * [1 - P(t < |T|)]$

=LOI.STUDENT.N(ABS(T);ddl;VRAI) où T est la statistique de Student calculée, ddl le nombre de degrés de liberté. Cette fonction calcule d'abord $P(t < |T|)$. Ensuite, on calcule la p-value telle que : $Pvalue = 2 * [1 - P(t < |T|)]$

=LOI.STUDENT.DROITE(ABS(T);ddl) . Cette fonction donne la moitié de la p-value qu'il faut simplement multiplier par 2 pour obtenir la p-value.

Pour **la loi de Khi-deux**, on peut directement obtenir la p-value d'une statistique de test quelconque en utilisant l'une des deux fonctions suivantes :

=LOI.KHIDEUX.DROITE(K;ddl) où K est la statistique de Khi-deux calculée, ddl le nombre de degrés de liberté. Cette fonction calcule $P(k > K)$ qui correspond directement à la p-value.

=1-LOI.KHIDEUX.N(K;ddl;VRAI) Cette fonction calcule d'abord $P(k < K)$. Ensuite, on calcule $P(k > K)$=1- $P(k < K)$ pour obtenir la p-value

Il faut noter que la loi de Khi-deux n'est pas une loi symétrique. On dispose de deux bornes différentes. Chacune de ces deux bornes peut être utilisée pour calculer la p-value car elles correspondent à chacune à la moitié de la p-value.

2.8.2. Détermination de la p-value dans un test unilatéral à droite

Dans le cas d'un test unilatéral à droite, la p-value se définit comme la probabilité d'obtenir une statistique z qui soit supérieure à la statistique calculée Z.

$$Pvalue = P(z \geq Z)$$

$$Pvalue = [1 - P(z < Z)]$$

Ainsi pour obtenir la p-value, on lit d'abord dans la table la probabilité $P(z < Z)$. Ensuite, on calcule la p-value.

En utilisant Microsoft EXCEL, on peut faire les distinctions suivantes :

Pour **la loi normale** on calcule $P(z < Z)$ en utilisant:

=LOI.NORMALE.STANDARD.N(Z;VRAI) où Z est la statistique calculée. Ensuite, on calcule la p-value telle que $Pvalue = [1 - P(z < Z)]$

Attention : pour le test unilatéral, on n'utilise pas de valeur absolue.

Pour **la loi de Student**, on peut utiliser l'une des deux fonctions suivantes :

=LOI.STUDENT.N(T;ddl;VRAI) où T est la statistique de Student calculée, ddl le nombre de degrés de liberté. Cette fonction calcule d'abord $P(t < T)$. Ensuite, on calcule la p-value telle que : $Pvalue = [1 - P(t < T)]$

=LOI.STUDENT.DROITE(T;ddl) . Cette fonction donne directement la p-value.

2.8.3. Détermination de la p-value dans un test unilatéral à gauche

Dans le cas d'un test unilatéral à gauche, la p-value se définit comme la probabilité d'obtenir une statistique z qui soit inférieure à la statistique Z.

$$Pvalue = P(z < Z)$$

Cette valeur peut être obtenue en lisant dans la table de la probabilité $P(z < Z)$.

En utilisant Microsoft EXCEL, on a les cas suivants :

Pour **la loi normale** on peut calculer $P(z < Z)$ en utilisant:

=LOI.NORMALE.STANDARD.N(Z;VRAI) où Z est la statistique calculée. La valeur fournie par cette fonction correspond directement à la p-value.

Pour **la loi de Student**, on peut utiliser l'une des deux fonctions suivantes :

=LOI.STUDENT.N(T;ddl;VRAI) où T est la statistique de Student calculée, ddl le nombre de degrés de liberté. Cette fonction calcule directement la p-value.

=LOI.STUDENT.DROITE(T;ddl). Cette fonction permet de calculer d'abord $P(z > Z)$. Pour obtenir la p-value, on calcule $P(z < Z) = 1 - P(z > Z)$.

3. Tests de conformité sur la proportion

Le test de conformité de la proportion se met en œuvre de la même façon que le test de conformité de la moyenne en supposant que l'échantillon est distribué selon une loi normale.

Par exemple, en utilisant un échantillon, on souhaite savoir si la proportion est égale à une valeur donnée. Ce test se présente sous la forme suivante

$$H_0 \ p = p_0$$

La proportion empirique est :

$$F_n = \frac{1}{n}\sum_{i=1}^{n_1} 1_{X_i=1} = \frac{n_1}{n}$$

Avec $E(F_n) = p$ et $Var(F_n) = \frac{p(1-p)}{n}$ où n_1 est le nombre de succès, n la taille de l'échantillon. En utilisant la loi des grands nombres, on peut écrire :

$$\left(\frac{F_n - p_0}{\sqrt{\frac{p(1-p_0)}{n}}}\right) = \sqrt{n}\left(\frac{F_n - p_0}{\sqrt{p_0(1-p_0)}}\right) \rightsquigarrow N(0,1)$$

Ainsi, à un risque d'erreur, on peut alors calculer la statistique de test comme suit :

$$Z = \sqrt{n}\left(\frac{F_n - p_0}{\sqrt{p_0(1-p_0)}}\right)$$

- Dans le cas d'un test bilatéral, on rejette l'hypothèse nulle lorsque $|Z| > Z^*_{1-\frac{\alpha}{2}}$, c'est-à-dire lorsque $Z \notin]-Z^*_{1-\frac{\alpha}{2}}, Z^*_{1-\frac{\alpha}{2}}[$.
- Dans le cas d'un test unilatéral à gauche, on rejette l'hypothèse nulle lorsque $Z < -Z^*_{1-\alpha}$.
- Et dans le cas d'un test unilatéral à droite, on rejette l'hypothèse nulle lorsque $Z > Z^*_{1-\alpha}$.

Conditions de validité du test sur la proportion

Le test de proportion n'est valable que sous les conditions suivantes :

$$\begin{cases} n \geq 30 \\ np_0 \geq 5 \\ n(1-p_0) \geq 5 \end{cases}$$

Lorsque ces conditions ne sont pas vérifiées, il faut alors utiliser le Test exact ou des tests non paramétriques. Ces types de tests ne sont pas abordés dans ce document.

4. Tests de conformité sur la variance

Le test de conformité sur la variance permet de tester la valeur de la variance $Var(X)$ d'un caractère X dans la population au vu de la variance empirique d'un échantillon. On suppose que la variable est distribuée selon une loi normale.

L'hypothèse H0 est que la variance au niveau de la population a une certaine valeur σ^2. On pose ainsi :

$$H0 : \mathrm{Var(X)} = \sigma^2$$

En notant S^2 la variance empirique (modifiée) de l'échantillon, on montre que sous l'hypothèse H0, la statistique du test est la suivante :

$$Y = \frac{n-1}{\sigma^2} S^2$$

Cette statistique suit une loi de khi-deux à n-1 degrés de liberté.

Ainsi, pour déterminer la région d'acceptation (respectivement la région de rejet) de H0, il faut utiliser les quantiles de la loi du χ^2. Par exemple, dans le cas d'un test bilatéral au seuil 5% pour avoir la région de rejet, il faut trouver les bornes a et b telles que :

$$P(Y \leq a) = \frac{\alpha}{2}$$

$$P(Y \geq b) = \frac{\alpha}{2}$$

Exemple : Une société fabrique un câble en acier trempé galvanisé dont la charge de rupture est de 210 kg avec une marge de 5 kg (écart-type). Un contrôle de qualité effectué sur 10 bobines a conduit aux résultats suivants :

203.70	211.80	201.60	226.00	213.30
201.80	214.90	217.40	215.80	206.90

Cet échantillon confirme-t-il la marge annoncée ?

On calcule la moyenne et la variance modifiée de l'échantillon :

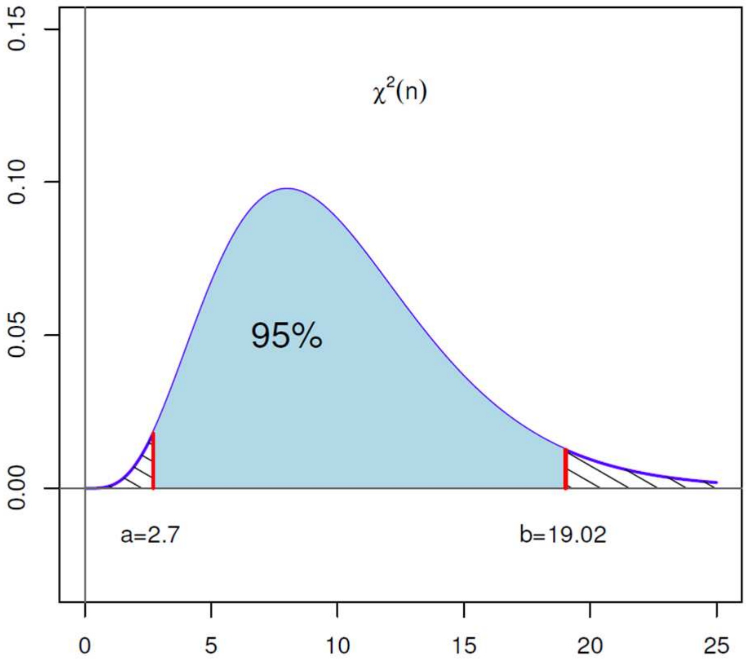
$\bar{X} = 211{:}32$ et Var(X) = S^2 =61,357

La statistique du test de variance vaut :

$$Y = \frac{n-1}{\sigma^2} S^2 = \frac{10-1}{5^2} 61{,}357^2 = 22{,}088$$

Avec un échantillon de taille n = 10, on a n -1 = 9 degrés de liberté et les tables de la loi du χ^2 donnent les valeurs suivantes pour les quantiles : a = 2,70 et b = 19,02 Car $P(Y \leq 2{,}70) = \frac{0{,}05}{2}$ et $P(Y \geq 19{,}02\,) = \frac{0{,}05}{2}$ (voir Annexe pour la lecture de la table de khi-deux).

La figure ci-dessous illustre la région de rejet et d'acceptation dans le cadre d'une loi de khi-deux.



Cette valeur de Y se trouve dans la région de rejet, c'est-à-dire à l'extérieur de l'intervalle [2,70 ; 19,02]. On doit donc rejeter l'hypothèse et considérer, au risque 5% de se tromper, que l'échantillon étudié présente une variance qui ne correspond pas à la variance annoncée.

Bibliographie

Dagnelie P., (2011), Statistique théorique et appliquée. Tome 2. Inférence statistique à une et à deux dimensions, De Boeck, 736 p.

Dagnelie P. (2013), Statistique théorique et appliquée. Tome 1. Statistique descriptive et bases de l'inférence statistique, De Boeck, 517 p.

Lecoutre J.-P., (2002), Statistiques et probabilités, manuel et exercices corrigés, Dunod, 3e édition, 296p

Saporta G. (1990), Probabilités, analyse de données et statistique, Technip, 3e édition, 656p.

Tillé Y., (2001), Théorie des sondages – Échantillonnage et estimation en populations finies Cours et exercices avec solutions, Collection: Sciences Sup, Dunod, 296p.

Annexe

Aide à la lecture des tables statistiques usuelles

Utilisation de la table de la loi normale centrée réduite

La table de la loi normale centrée réduite présente sur la première ligne et la première colonne les valeurs des fractiles Z (encore appelés quantiles). Dans la première colonne, on lit la valeur du quantile Z à un décimal près. Et sur la première ligne, on lit le nombre de décimaux restants à 10-2 près. Ainsi c'est en faisant la somme d'un élément en colonne et d'un élément en ligne qu'on obtient la valeur de Z. Par exemple $Z = 1.96$ s'obtient en faisant 1,9+0,06 où 1,9 provient de la première colonne alors que 0,06 provient de la première ligne. Ainsi pour trouver la valeur de n'importe quel fractile, on procède à cette décomposition. Par exemple $Z = 3.37$ se lit en décomposant 3.3 (en colonne) et 0,07 (en ligne).

En plus de la première colonne extérieure et la première ligne extérieure (qui permettent de connaitre la valeur de Z), on se réfère aux cellules intérieures pour lire les probabilités associées aux fractiles. La probabilité associée à un fractile correspond à la valeur contenue dans la cellule qui se trouve au croisement des deux membres qui forment la valeur de la fractile. Par exemple, sachant que 2,47 est formée par 2,4(en ligne) et 0,07 (en colonne), alors, la probabilité correspondant à 2,47 se trouve au croisement entre 2,4 et 0,07. Cette valeur est égale à 0,993.

Cette compréhension de la structure de la table de loi normale est extrêmement importante car elle servira à déterminer les fractiles (lorsque l'on connait les probabilités) ou à l'inverse déterminer les probabilités (lorsque l'on connait les fractiles).

Lecture des fractiles connaissant les probabilités α

Dans une optique de détermination de la statistique d'un test bilatérale suivant une loi normale et dont le seuil d'erreur est α, on lit le fractile correspondant à α. Pour cela, on calcule d'abord$(1-\alpha)+\frac{\alpha}{2}$ c'est-à-dire $1-\frac{\alpha}{2}$. Ensuite, on recherche cette valeur dans les cellules intérieures de la table. Une fois cette valeur identifiée, on fait la somme des deux cellules extérieures (en ligne et en colonne) dont le croisement correspond à cette valeur $1-\frac{\alpha}{2}$ lue dans la table. Par exemple, pour le trouver la statistique (le fractile) correspondant à α =5 %, on calcule d'abord $1-\frac{\alpha}{2}$ (soit 0,975). Ensuite, en recherchant 0,975 dans les cellules intérieures de la table, on constate que cette valeur se trouve au croisement entre 1,9 et 0,06. Par conséquent le fractile correspondant à 5% est 1,96.

Notons aussi que cette valeur peut être obtenue avec les fonctions de Microsoft™ Excel®. En effet, pour obtenir le fractile correspondant au seuil α, on utilise la formule suivante :

=LOI.NORMALE.STANDARD.INVERSE$(1-\frac{\alpha}{2})$

Où α représente la probabilité (et correspond généralement au seuil d'erreur).

Remarque :

Puisque la loi normale est une loi symétrique, si l'on veut déterminer la valeur opposée du fractile (en vue par exemple de la détermination d'un intervalle de confiance (ou autre), on considère juste l'opposé de ce fractile pour trouver la borne inférieure de l'encadrement.

Lecture des probabilités α connaissant les fractiles

Lorsque l'on connait le fractile, pour déterminer la probabilité correspondante par lecture d'une table de la loi normale centrée et réduite, on décompose d'abord ce fractile en deux éléments (selon la méthode précédemment discutée). Ensuite, on recherche la cellule intérieure de la table se trouvant au croisement (de la ligne et de la colonne extérieure) des deux valeurs. Ainsi, après avoir déterminé cette valeur notée P, on calcule α telle que $1-\frac{\alpha}{2}=P$ soit $\alpha=2(1-P)$. Cette valeur α correspond donc à la probabilité recherchée. Par exemple, pour trouver la probabilité correspondant à 1,53, on décompose d'abord cette valeur entre 1,5 et 0,03. En recherchant dans les cellules intérieures de la table la probabilité se trouvant au croisement de ces deux valeurs, on trouve 0,937. Ainsi puisque cette valeur équivaut à $1-\frac{\alpha}{2}$, on peut alors en déduire α comme $\alpha = 2(1-0{,}937)$. Soit $\alpha = 0{,}126 = 12{,}6\%$

Notons aussi que valeur de la probabilité peut s'obtenir en utilisant également les fonctions de Microsoft Excel. Pour cela, on peut utiliser la formule suivante :

=LOI.NORMALE.STANDARD.N $(q; VRAI)$

Où q représente la valeur du fractile dont on cherche à déterminer la probabilité.

Remarque :

Dans une démarche de test (bilatéral), la valeur de la probabilité ainsi calculée correspond généralement à la p.value lorsque la détermination de la probabilité porte sur une statistique de test. En effet dans un test, on connaît à priori le seuil théorique α. Ce seuil est utilisé pour lire la statistique théorique (ou seuil critique) du test. On a alors le couple (α ; S^*). Ensuite en construisant le test et en calculant la statistique du test sous l'hypothèse nulle, on obtient S. Ce qui manque alors c'est la probabilité associée à cette statistique calculée. Elle est dénommée la p.value p. Dès lors, pour déterminer la p.value afin de former le couple (p_0 ; S), on lit la probabilité correspondant à S dans la table. C'est donc après avoir définie cette probabilité qu'on forme la règle de décision du test :

- Si $|S| > S^* \Rightarrow p < \alpha$ alors on rejette H0.
- Si $|S| < S^* \Rightarrow p > \alpha$ alors on ne peut pas rejeter H0.

Lecture de la table de la loi normale dans le cas d'un encadrement de la fractile

- **Cas d'un encadrement de type** : $P\,(-b\ <\ Z\ <\ b)$

Lorsque le fractile Z est une valeur encadrée par une borne supérieure b et une borne inférieure -b, pour lire la probabilité pour que Z soit compris entre -b et b, on procède d'abord à un développement comme suit :

$$P\,(-b\ <\ Z\ <\ b)\ = P(Z < b) - P(Z < -b)$$

Or, sachant les propriété d'une loi symétrique, on a: $P(Z < -b) = P(Z > b)$. Mais on sait aussi que $P(Z > b) = 1 - P(Z < b)$. Dès lors, on a : $P\,(-b\ <\ Z\ <\ b) = P(Z < b) - [\,1 - P(Z < b)]$. Au final, après développement, on trouve :

$$P\,(-b\ < Z <\ b)\ = 2P(Z < b) - 1$$

Cela montre donc que dans une loi symétrique, pour trouver la probabilité d'un encadrement symétrique (qui correspond en général au seuil de confiance), il faut simplement multiplier par 2 la probabilité obtenue en considérant uniquement la borne supérieure (en suivant les méthodes de lectures précédemment présentées). Ensuite retrancher 1 pour trouver la probabilité de l'encadrement (au seuil de confiance). Par exemple, quand on demande de calculer la probabilité pour que Z soit comprise entre -2,72 et 2,72. On lit d'abord la probabilité associée à 2,72 (soit 0,9967). Ensuite, on multiplie cette valeur par 2 et on retranche 1. On trouve alors 0,9934. Ainsi le seuil d'erreur α s'obtient simplement comme est 1-0,9934 soit 0,66%. Il faut noter que dans un encadrement de type intervalle de confiance α n'est pas calculée telle que $1 - \frac{\alpha}{2} = P$ mais comme $1 - \alpha = P$.

- **Cas d'un encadrement de type** $P\ (Z\ <\ b)$ ou de type $P\ (Z >\ b)$

Lorsqu'il s'agit d'un encadrement de type $P\ (Z\ <\ b)$, on garde sans aucune transformation la valeur lue dans la table (ou obtenue par la fonction : $= loi.normale.standard.n(q; VRAI)$. Ainsi, le seuil d'erreur α s'obtient en utilisant la relation $1 - \alpha = P$.
Mais quand il s'agit d'un encadrement de type $P(U > b)$, on lit d'abord $P(U < b)$, ensuite on calcule $P(U > b) = 1 - P(U < b)$. Ainsi, le seuil d'erreur α s'obtient en utilisant la relation $1 - \alpha = P$.

Utilisation de la table de Student

Lecture des fractiles connaissant les probabilités α

En général la table de Student se présente de telle sorte que les lignes correspondent aux degrés de liberté et les colonnes correspondent aux valeurs des probabilités. Pour lire les fractiles dans une table se présentant sous ce format, on retrouve d'abord le degré de liberté, ensuite, on identifie la colonne de probabilité correspondant à la probabilité considérée. Le croisement entre la ligne du degré de liberté et la colonne de probabilité permet d'identifier le fractile recherché.

Néanmoins, il faut noter que l'identification de la colonne de probabilité dépend de la nature du test effectuée en amont. Par exemple dans le cas d'un test bilatéral, le quantile d'ordre $1 - \frac{\alpha}{2}$ se lit à partir de la colonne $P = \alpha$. Dans le cas d'un test unilatéral (à droite) le quantile d'ordre $1 - \alpha$ se lit à partir de la colonne $P = 2\alpha$. Et enfin dans le cadre d'un test unilatéral (à gauche), le quantile d'ordre α se lit à partir de la colonne $P = 2(1 - \alpha)$.

En prenant par exemple le cas du test bilatéral où le seuil d'erreur α et le nombre de degré de liberté sont donnés, les étapes de détermination de la colonne de probabilité sont les suivantes. Dans un premier temps, on calcule la probabilité du seuil critique. Ici $1-\frac{\alpha}{2}$; Dans second temps, on soustrait cette valeur de 1 c'est-à-dire $1-\left(1-\frac{\alpha}{2}\right)$; Dans un troisième temps, on multiplie cette probabilité résiduelle par 2. Ainsi on a : $2\times\left[1-\left(1-\frac{\alpha}{2}\right)\right]=\alpha$. Cette expression montre bien que le cas d'un test bilatéral, le quantile d'ordre $1-\frac{\alpha}{2}$ se lit à partir de la colonne $P=\alpha$. En suivant le même raisonnement pour le cas d'un test unilatéral à droite, on peut aussi montrer que le quantile d'ordre $1-\alpha$ se lit à partir de la colonne $P=2\alpha$ alors que pour le cas du test unilatéral à gauche, le quantile d'ordre α se lit à partir de la colonne $P=2(1-\alpha)$. Toutefois, dans ce dernier cas, il peut arriver que $P=2(1-\alpha)$ soit supérieure à 1. Dans ce cas, il est n'est pas possible d'identifier une colonne de probabilité. Il faut alors se servir de la propriété de symétrie de la loi de Student. En effet, on sait que $T_{\alpha}^{*}=-T_{1-\alpha}^{*}$. Il faut alors calculer $T_{1-\alpha}^{*}$ en utilisant les étapes d'un test unilatéral à droite. Ensuite calculée l'opposée de la valeur obtenue.

Notons aussi que ces fractiles peuvent être obtenus des fonctions Excel suivantes :
=LOI.STUDENT.INVERSE.BILATERALE(α;ddl) pour le cas d'un test bilatéral
=LOI.STUDENT.INVERSE.BILATERALE(2α;ddl) pour le cas d'un test unilatéral à droite
=LOI.STUDENT.INVERSE.BILATERALE(2(1-α);ddl) pour le test unilatéral à gauche.

Aussi lorsque l'on veut déterminer la valeur symétrique (opposée) d'un fractile en vue, par exemple, de la détermination d'un intervalle de confiance, etc, on prend juste l'opposé du fractile calculée puisque la loi de Student est une loi symétrique.

Par ailleurs, il faut aussi noter que lorsque n est grand (n>30), on peut approximer la loi de Student par la loi normale. Dès lors, on peut utiliser la table de la loi normale comme décrite précédemment.

Lecture des probabilités connaissant les fractiles

Pour déterminer la probabilité α dans une table de la loi de Student, on se sert uniquement du fractile et du nombre de degré de liberté en prenant le chemin inverse qui conduit à la détermination du fractile. En effet, on se place sur la ligne correspondant au nombre de degrés de liberté et on se déplace de gauche vers la droite en essayant d'identifier la valeur la plus proche possible du fractile recherché. Une fois la valeur du fractile identifiée, on retrouve la valeur de la

probabilité P en lisant dans le libellé de la colonne correspondant en haut de la table. Ensuite, on en déduit la probabilité recherchée en tenant compte de la nature du test. Dans le cas du test bilatéral, la probabilité recherchée est directement égale à P ($\alpha = P$). Dans le cas du test unilatéral à droite, la probabilité α est égale à la moitié P. ($\alpha = \frac{P}{2}$). Et enfin dans le cas d'un test unilatéral à gauche, la probabilité, on a : $\alpha = 1 - \frac{P}{2}$.
La valeur de la probabilité peut être directement obtenue à partir des fonctions Excel comme suit :
=LOI.STUDENT.BILATERALE(ABS(T);ddl) pour le cas d'un test bilatéral
=1-LOI.STUDENT.N(T;ddl;VRAI) pour le cas d'un test unilatéral à droite
=LOI.STUDENT.N(T;ddl;VRAI)) pour le test unilatéral à gauche.
Où T représente le fractile dont on cherche la probabilité ; ddl représente le nombre de degrés de liberté.

Utilisation de la table de khi-deux

Lecture des fractiles connaissant les probabilités $\boldsymbol{\alpha}$

La lecture d'une table de khi-deux est très proche de la lecture de la table de Student discutée précédemment notamment. La différence réside dans le fait que les probabilités utilisées sont multipliées par 2 dans le cas de la loi de Student. En effet, pour la loi de Student, pour rechercher un quantile d'ordre $1 - \frac{\alpha}{2}$ dans le cadre d'un test bilatéral, la lecture se fait non pas au niveau de la colonne de probabilité $1 - \frac{\alpha}{2}$ mais plutôt au niveau de la colonne de probabilité $P = \alpha$ qui est obtenue à travers l'expression $2 \times \left[1 - \left(1 - \frac{\alpha}{2}\right)\right] = \alpha$. Mais pour le cas de la loi de Khi-deux le facteur multiplicateur 2 est ignorée et la lecture se fait à partir à partir de la probabilité$\left[1 - \left(1 - \frac{\alpha}{2}\right)\right] = \frac{\alpha}{2}$. Par exemple, pour $\alpha = 0.05$, la lecture du quantile d'ordre 0.975 se fait à partir de la colonne de probabilité 0.025 et de la ligne correspondant au degré de liberté.

Cette lecture peut aussi directement se faire à partir d'Excel en utilisant la fonction :

=LOI.KHIDEUX.INVERSE.DROITE(0.025;ddl).

D'une manière générale, la lecture avec Excel du fractile d'une loi de Khi-deux dans un test bilatéral au risque d'erreur α et au degré de liberté ddl peut se généraliser comme suit :

=LOI.KHIDEUX.INVERSE.DROITE($\frac{\alpha}{2}$;ddl).

Pour ce qui concerne le cas d'un test unilatéral à droite au risque d'erreur et au ddl donné, la lecture du quantile d'ordre $1-\alpha$ se fait à partir de la colonne de probabilité définie par l'expression $[1-(1-\alpha)]=\alpha$ au lieu $2\times[1-(1-\alpha)]=2\alpha$ pour la loi de Student. Le quantile se détermine simplement en recherchant le point de croisement entre la ligne du degré de liberté et la colonne de probabilité correspondant à $P=\alpha$.

Avec Excel, on peut utiliser la fonction suivante :

=LOI.KHIDEUX.INVERSE.DROITE(α;ddl).

S'agissant maintenant de la lecture dans le cas d'un test unilatéral à gauche au risque d'erreur α et au degré de liberté donné, le quantile d'ordre α de la loi de Khi-deux se lit partir de la colonne de probabilité définie par $[1-(\alpha)]=1-\alpha$. Il suffit alors simplement de recherchée dans la table la valeur se trouvant au croisement entre entre la ligne du ddl et la colonne correspondant à $P=1-\alpha$.

Avec Excel, on peut utiliser la fonction suivante :

=LOI.KHIDEUX.INVERSE.DROITE($1-\alpha$;ddl).

Lecture des fractiles dans la table de Khi-deux en vue des encadrements

Même s'il existe une certaine proximité entre la méthode de lecture dans la table de Student et dans la table de Khi-deux, les deux procédures diffèrent très nettement lorsqu'il s'agit des encadrements. En effet, à la différence de la loi de Student ou de la loi normale, la loi de khi-deux n'est pas une loi symétriquement distribuée autour de 0. Par conséquent lorsqu'on veut procéder, à un encadrement, on doit calculer deux bornes distinctes en utilisant les probabilités associées à chaque fractile constituant les bornes. Par exemple pour encadrer une valeur U dans la perspective de la détermination d'un intervalle de confiance etc.., on calcule d'abord deux probabilités :

$$p1=\frac{\alpha}{2}\ et\ \ p2=(1-\alpha)+\frac{\alpha}{2}=1-\frac{\alpha}{2}$$

Ensuite, on lit le fractile correspondant directement à chaque probabilité directement en utilisant les degrés de liberté. On obtient alors $q1$ et $q2$. où $q1$ et $q2$ représentent respectivement les fractiles correspondants à $p1$ et $p2$.

Ensuite, on encadre U telle que $q1<U<q2$. Cet encadrement se fait donc de telle sorte que $P(q1<U<q2)=1-\alpha$. Où $1-\alpha$ est le seuil de confiance.

On pour lire les fractiles correspondant à $p1$ et $p2$, on peut utiliser les fonctions suivantes :

=LOI.KHIDEUX.INVERSE($p1;ddl$) et

$= \text{LOI.KHIDEUX.INVERSE}(p2; ddl)$.

Les valeurs obtenues servent donc à construire l'intervalle de confiance.

Remarquons que la méthode de lecture de la loi de khi-deux diffère selon qu'il s'agisse de déterminer la valeur critique d'un test ou de déterminer un encadrement.

Lecture des probabilités connaissant les fractiles

Pour déterminer la probabilité α dans une table de la loi de Khi-deux, on se sert du fractile et du nombre de degré de liberté en prenant le chemin inverse qui conduit à la détermination du fractile. En effet, on se place sur la ligne correspondant au nombre de degrés de liberté et on se déplace de gauche vers la droite en essayant d'identifier la valeur la plus proche possible du fractile recherché. Une fois la valeur du fractile identifiée, on retrouve la valeur de la probabilité P en lisant dans le libellé de la colonne correspondant en haut de la table. Ensuite, on en déduit la probabilité recherchée en tenant compte de la nature du test. Dans le cas du test bilatéral, la probabilité recherchée est égale au double de P ($\alpha = 2P$). Dans le cas du test unilatéral à droite, la probabilité α est égale à P ($\alpha = P$). Et enfin dans le cas d'un test unilatéral à gauche, la probabilité, on a : $\alpha = 1 - P$.
La valeur de la probabilité peut être directement obtenue à partir des fonctions Excel comme suit :
=2*LOI.KHIDEUX.DROITE(q;ddl) pour le cas d'un test bilatéral
=LOI.KHIDEUX.DROITE(q;ddl) pour le cas d'un test unilatéral à droite
=1-LOI.KHIDEUX.DROITE(q;ddl) pour le test unilatéral à gauche.
Où q représente le fractile dont on cherche la probabilité ; ddl représente le nombre de degrés de liberté.

Lecture de probabilité dans le cas d'un encadrement

Pour déterminer les probabilités lorsqu'on dispose d'un encadrement, la procédure est un peu particulière mais très simple. En effet, puisque, nous avons deux bornes, pour obtenir la probabilité correspondant à la fractile inférieure (c'est-à-dire pour obtenir $\frac{\alpha}{2}$), on recherche juste cette valeur dans la colonne où l'on identifie le fractile en se servant de la ligne de degré de liberté. Ensuite, on multiplie par 2 pour obtenir α. De la même manière, on peut se servir de la fractile supérieure pour déterminer la probabilité correspondant à $\frac{\alpha}{2} + (1 - \alpha)$ qui permet ensuite d'obtenir α. Toutes ces procédures peuvent être mises en œuvre sous Excel, en utilisant l'une des formules suivantes :
= LOI.KHIDEUX.N(q1;ddl)
= LOI.KHIDEUX.N(q2;ddl)
L'une ou l'autre de ces deux valeurs obtenues permettra alors de calculer α.

Test de conformité et calcul des seuils critiques et des p-values avec les fonctions de Microsoft Excel

	Variance	Seuil d'erreur	Statistique du test	Seuil critique	P-value
TEST BILATERAL	σ^2 connue	α	$Z = \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} \rightsquigarrow N(0,1)$	=LOI.NORMALE.STANDARD.INVERSE.N($1 - \frac{\alpha}{2}$)	=2*(1-LOI.NORMALE.STANDARD.N(ABS(Z);VRAI))
	σ^2 inconnue		$\frac{\bar{X} - m}{\frac{\hat{\sigma}}{\sqrt{n-1}}} \rightsquigarrow T(n-1)$	=LOI.STUDENT.INVERSE.BILATERALE(α;ddl)	=LOI.STUDENT.BILATERALE(ABS(T);ddl)
TEST UNIILATERAL A DROITE	σ^2 connue	α	$Z = \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} \rightsquigarrow N(0,1)$	=LOI.NORMALE.STANDARD.INVERSE.N($1 - \alpha$)	=1-LOI.NORMALE.STANDARD.N(Z;VRAI)
	σ^2 inconnue		$\frac{\bar{X} - m}{\frac{\hat{\sigma}}{\sqrt{n-1}}} \rightsquigarrow T(n-1)$	=LOI.STUDENT.INVERSE.BILATERALE(2α;ddl)	=1-LOI.STUDENT.N(T;ddl;VRAI)
TEST UNIILATERAL A GAUCHE	σ^2 connue	α	$Z = \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} \rightsquigarrow N(0,1)$	=LOI.NORMALE.STANDARD.INVERSE.N(α)	=LOI.NORMALE.STANDARD.N(Z;VRAI)
	σ^2 inconnue		$\frac{\bar{X} - m}{\frac{\hat{\sigma}}{\sqrt{n-1}}} \rightsquigarrow T(n-1)$	=LOI.STUDENT.INVERSE.BILATERALE(2(1-α);ddl)	=LOI.STUDENT.N(T;ddl;VRAI)