

Tests d'hypothèses

6 .TESTS D'HYPOTHESES	2
6.1. Exemple introductif	2
6.2. Généralisation	6
6.3. Tests non paramétriques	7
6.4. Tests d'une moyenne	8
6.5. Test de comparaison de moyennes.....	11
6.6. Test d'une proportion	13
6.7. Test de comparaison de proportions.....	13
6.8. Test d'une variance	14
6.9. Test de comparaison de variances.....	14

6 .Tests d'hypothèses

6.1. Exemple introductif

Emprunté au domaine du contrôle de fabrication (renouvelé par E. Pearson dans les années 1930) cet exemple permet de comprendre la problématique des tests paramétriques.

Soit un lot de produits dans lequel il y a une certaine proportion p inconnue de produits défectueux (pièces non conformes à une norme, fruits avariés, lampes grillées) Le vendeur prétend que cette proportion est de 10 %, c'est l'hypothèse « nulle » :

$$H_0 : p = p_0 = 0,10$$

Dans l'approche inférentielle de Fisher cette hypothèse est la seule prise en compte.

Dans l'approche décisionnelle, Neyman et Pearson introduisent un second acteur – l'acheteur – et une seconde hypothèse. L'acheteur craint que cette proportion soit plus élevée : c'est l'hypothèse alternative :

$$H_1 : p = p_1 > p_0$$

Dans cet exemple, l'hypothèse H_0 est simple (elle spécifie une valeur ponctuelle du paramètre p caractérisant la population) et l'hypothèse H_1 est multiple (ensemble de valeurs possibles) et unilatérale. Elle est incompatible avec H_0 mais n'est pas forcément sa complémentaire. Ces deux hypothèses traduisent les états possibles de la nature dans l'approche décisionnelle.

Le vendeur ne veut pas rater une vente sous le prétexte que cette proportion serait importante. L'acheteur ne veut pas conclure la vente si le lot est de mauvaise qualité. Ils doivent prendre une décision rationnelle sur la seule base de l'observation d'un échantillon que nous supposons de taille $n = 50$ et tiré au hasard. L'échantillon fortuit (x_1, x_2, \dots, x_n) est la réalisation d'un échantillon aléatoire de n variables de Bernoulli (indépendantes et de même loi si les tirages sont avec remise, ce que nous supposons).

Dans l'approche classique de Laplace et Fisher, on cherche à valider ou invalider la seule hypothèse H_0 sur la base de cet échantillon : le test, dit de significativité, doit conclure qu'il est vraisemblable ou pas, au vu des observations, que cet échantillon reste compatible avec H_0 . Si la proportion de pièces défectueuses est par exemple de 50% dans l'échantillon, on jugera que cela est trop peu probable sous cette hypothèse H_0 , que l'écart avec 10% est significatif, et qu'il faut rejeter cette hypothèse. Dans le cas inverse on dira seulement que l'hypothèse H_0 n'est pas remise en cause. Mais supposons que cet échantillon comprenne 10 pièces défectueuses, soit 20% des observations. Que faut-il conclure ? Que ce taux obtenu fortuitement reste compatible avec l'hypothèse H_0 (10%) ou que ce taux est significativement différent de 10% ce qui nous amène à rejeter H_0 ?

Dans l'approche décisionnelle de Neyman et Pearson, les seules décisions possibles sont :

$$D_0 = \text{décider que } H_0 \text{ est vraie}$$

$$D_1 = \text{décider que } H_1 \text{ est vraie}$$

Prendre une décision revient donc à réaliser une application T de l'ensemble Ω des échantillons possibles dans l'ensemble des deux décisions possibles $\{D_0, D_1\}$, soit encore à scinder cet ensemble Ω en deux parties (on dit 2 régions) : Ω_0 la région d'acceptation (de H_0) et Ω_1 la région critique (de rejet de H_0). Un test = une **règle de décision** = une bipartition de Ω .

Il n'y a pas *a priori* une seule règle de décision possible. Toute bipartition de Ω en est une, et donc dans le cas fini il y a $2^{\text{card}(\Omega)}$ tests possibles. Je peux par exemple décider que le vendeur ment si la fréquence des produits défectueux dans l'échantillon est supérieur à 12 %, si elle est supérieure à 25 %, si les 3 premiers tirés sont défectueux, si le premier et le dernier sont défectueux, si plus de deux tirages consécutifs ont donné des produits défectueux, etc... . Quelle est la meilleure règle de décision ? Celle qui bien sûr me conduit le moins souvent à prendre une décision erronée.

Mais il y a deux sortes d'erreurs dans un test d'hypothèse à la Neyman et Pearson.

Décision	Etats de la nature	
	H_0 vraie	H_1 vraie
D_0	OK	β, C_1
D_1	α, C_0	OK

L'erreur de première espèce consiste à décider D_1 alors que H_0 est vraie. Cette erreur pénalise le vendeur qui ne vendra pas un lot de bonne qualité. On appelle risque de première espèce sa probabilité :

$$\alpha = \text{prob}\{D_1 / H_0\} = P_0(\Omega_1)$$

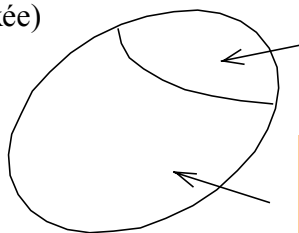
L'erreur de seconde espèce consiste à décider D_0 alors que H_1 est vraie. Cette erreur pénalise l'acheteur qui achètera un lot de mauvaise qualité. On appelle risque de seconde espèce sa probabilité :

$$\beta = \text{prob}\{D_0 / H_1\} = P_1(\Omega_0)$$

En théorie de la décision, on associe également deux coûts C_0 et C_1 à ces deux erreurs. Ici par exemple, C_0 pourrait être le coût pour le vendeur d'une recherche d'un nouvel acheteur, et C_1 le coût pour l'acheteur du remplacement des produits défectueux du lot. Ces deux erreurs sont donc de nature et de conséquences très différentes. Si les produits sont des médicaments, la première n'a que des conséquences commerciales, la seconde peut avoir des conséquences graves sur la santé des consommateurs. De plus si, comme dans cet exemple, H_1 est une hypothèse multiple, β n'est pas calculable comme une probabilité simple mais comme une fonction des valeurs de p supérieures à $p_0 = 0,1$. On appelle courbe de puissance du test la courbe de $(1 - \beta)$ en fonction des valeurs du paramètre (ici, p) compatibles avec H_1 .

La bipartition de Ω qui constitue la règle de décision peut se faire plus facilement à l'aide d'une statistique S , c'est à dire d'une fonction S de l'échantillon (X_1, X_2, \dots, X_n) , que directement sur l'ensemble des échantillons possibles. Si cette fonction est à valeur dans R (ensemble des réels) alors on pourra faire correspondre aux régions Ω_0 et Ω_1 des intervalles I_0 et I_1 de R que l'on appellera intervalles d'acceptation et de rejet pour la statistique du test. Là où les valeurs bornant ces intervalles s'appelleront **valeur(s) critique(s)**, car elles sont critiques au sens où notre décision s'inverse quand la valeur de la statistique S traverse ces valeurs.

Ensemble des échantillons
(taille n fixée)



région critique Ω_1

ensemble des échantillons pour lesquels $S \in I_1$
de probabilité α si H_0 est vraie

région d'acceptation Ω_0

ensemble des échantillons pour lesquels $S \in I_2$
de probabilité $1 - \alpha$ si H_0 est vraie

Dans notre exemple, il est naturel de prendre pour statistique la fréquence observée dans l'échantillon : soit la fréquence absolue des produits défectueux, $S = \sum_{i=1}^n X_i$ (on rappelle que les X_i sont ici des variables de Bernoulli), soit la fréquence relative $F = S/n$. Mais l'on pourrait aussi utiliser une statistique d'ordre. Nous montrerons plus loin que le test le plus puissant utilise bien la fréquence. Toujours intuitivement, plus cette fréquence sera faible, plus nous accorderons de confiance à l'hypothèse H_0 mise en avant par le vendeur. Mais cette confiance n'est jamais totale puisqu'il est possible que « par hasard » cette fréquence soit élevée dans le cas où il a raison (c'est le risque β). Inversement, plus cette fréquence sera élevée, donc supérieure à une certaine valeur critique, plus nous aurons tendance à croire en l'hypothèse H_1 , mais cette défiance vis à vis du vendeur peut être due au hasard (c'est le risque α). Notre règle de décision prend donc la forme suivante :

Si $S > c$ nous décidons D_1 (c'est à dire H_1 vraie) avec un risque
 $\alpha = \text{prob}\{D_1 / H_0\}$

Si $S \leq c$ nous décidons D_0 (c'est à dire H_0 vraie) avec un risque
 $\beta = \text{prob}\{D_0 / H_1\}$

Comment dès lors fixer la valeur critique c qui déterminera entièrement notre règle de décision ? Fisher avait construit des tables de certaines statistiques (t , F , χ^2) pour les valeurs de 5 % et 1 %. C'est de cette pratique que les statisticiens (y compris Neyman et Pearson) ont gardé 5 % comme valeur conventionnelle du risque α . De cette valeur, que l'on peut bien sûr modifier pour un contexte particulier, il est possible de déduire la valeur de c pourvu qu'on connaisse la loi de la statistique S . Ici par exemple, $S = nF$

suit une loi $\text{Binomiale}(n, p)$, que l'on peut approcher par la loi normale $N(np, \sqrt{np(1-p)})$ compte tenu de la taille de l'échantillon. Sous l'hypothèse H_0 ($p = p_0$) cette loi est parfaitement connue. On peut donc écrire :

$$\alpha = 0,05 = \text{prob}(S > c / H_0) = P_0 \left(\frac{S - np_0}{\sqrt{np_0(1-p_0)}} > \frac{c - np_0}{\sqrt{np_0(1-p_0)}} \right)$$

D'où l'on déduit que la quantité qui est à droite de l'inégalité est le fractile d'ordre $(1 - \alpha)$ de la loi normale centrée réduite :

$$\frac{c - np_0}{\sqrt{np_0(1-p_0)}} = f_{0,95} = 1,645$$

$$\text{d'où } c = 50 \times 0,10 + 1,645 \sqrt{50 \times 0,10 \times 0,90} = 8,49.$$

En fait le calcul fait avec la loi binomiale donnerait : $P_0(S > 7) = 0,128$; $P_0(S > 8) = 0,058$; $P_0(S > 9) = 0,009$. Donc avec cette règle (rejeter H_0 si le nombre de produits défectueux est supérieur à 8), l'erreur de première espèce sera en fait de 5,8 %.

Nous pourrions avoir envie de réduire cette erreur de première espèce, mais il faut voir que nous augmenterions alors l'erreur de seconde espèce :

$$\beta = \text{prob}\{D_0 / H_1\} = \text{prob}(S \leq 5 / p = p_1)$$

Cette erreur ne dépend que de la valeur de p_1 choisie pour l'hypothèse alternative :

$$\beta = \text{prob}(S \leq c / p = p_1) = P_1 \left(\frac{S - np_1}{\sqrt{np_1(1-p_1)}} \leq \frac{c - np_1}{\sqrt{np_1(1-p_1)}} \right)$$

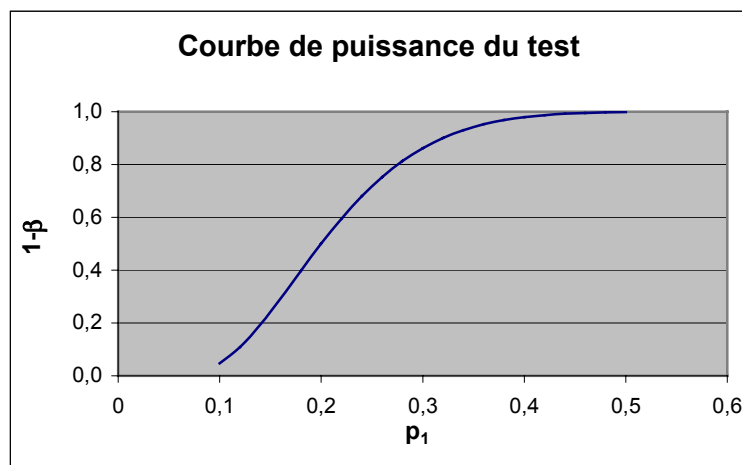
Mais il est clair, pour n'importe quelle valeur p_1 , que si on réduit α , alors on augmente la valeur critique c , et on augmente donc la probabilité β de lui être inférieur.

Comment varie ce risque de seconde espèce avec p_1 ?

- pour $p_1 = 0,16$, nous trouvons $\beta = \text{prob}(N(0,1) \leq 0) = 0,5$
- pour $p_1 = 0,3$, nous trouvons $\beta = \text{prob}(N(0,1) \leq -2,16) = \text{prob}(N(0,1) \geq 2,16) = 0,015$

Cette erreur décroît lorsque p augmente. Le test est de plus en plus puissant (discriminant) quand les valeurs relatives aux deux hypothèses entre lesquelles il faut trancher s'éloignent l'une de l'autre. Plus généralement nous pouvons construire pas à pas la fonction puissance du test en donnant des valeurs successives à p_1 . Nous obtenons pour $(1 - \beta)$ une fonction croissante de p_1 .

p_1	t	β	$1-\beta$
0,1	1,414	0,921	0,079
0,12	0,870	0,808	0,192
0,14	0,408	0,658	0,342
0,16	0,000	0,500	0,500
0,18	-0,368	0,356	0,644
0,2	-0,707	0,240	0,760
0,22	-1,024	0,153	0,847
0,24	-1,325	0,093	0,907
0,26	-1,612	0,053	0,947
0,28	-1,890	0,029	0,971
0,3	-2,160	0,015	0,985
0,32	-2,425	0,008	0,992
0,34	-2,687	0,004	0,996
0,36	-2,946	0,002	0,998
0,38	-3,205	0,001	0,999
0,4	-3,464	0,000	1,000
0,42	-3,725	0,000	1,000
0,44	-3,989	0,000	1,000
0,46	-4,256	0,000	1,000
0,48	-4,529	0,000	1,000
0,5	-4,808	0,000	1,000



Remarquons que nous aurions tout aussi bien pu prendre comme statistique du test $F = S/n$ et comme région critique $\{F > c'\}$. Avec, pour le même α , $c' = c/n$. Ce test, utilisant la même bipartition de Ω , aurait produit la même décision avec les mêmes risques.

Une fois la règle de décision parfaitement connue, ainsi que les erreurs qui lui sont attachées pour une taille donnée d'échantillon, il reste le plus simple mais l'essentiel, à savoir appliquer cette règle de décision à l'échantillon (unique en général) que l'on a obtenu. Dans notre exemple, si l'échantillon a donné 4 produits défectueux, nous décidons que le vendeur a raison (H_0 est vraie) c'est à dire que le lot contient 10 % de défectueux; et non pas qu'il en a $4/25 = 15\%$ ce qui serait la juste estimation de p avec l'échantillon observé. Si l'échantillon contient 10 pièces défectueuses (20%) nous décidons au contraire que le vendeur a tort (rejet de H_0), et dans l'optique décisionnelle, que cette proportion étant vraisemblablement supérieure à 10%, nous n'achetons pas ce lot.

6.2. Généralisation

Etapes d'une procédure de test statistique

1. On spécifie les hypothèses H_0 et H_1 .
2. On se donne les conditions du test à savoir :
 - les conditions de tirage de l'échantillon, méthode, taille ;
 - loi L connue ou inconnue ;
 - les valeurs obtenues dans l'échantillon fortuit ;
 - la valeur du risque de première espèce α ;
 - éventuellement, les valeurs des coûts associés à chacune des deux erreurs.
3. On détermine la statistique S du test, c'est à dire celle qui va servir à définir la règle de décision, soit en partant de l'estimateur usuel, soit en utilisant la règle de Neyman et Pearson sur les tests UMP (hors programme de ce cours).

4. On détermine la forme de la région critique, c'est à dire celle de l'intervalle I_1 des valeurs de la statistique S qui définissent la région critique. Elle se déduit de la forme des hypothèses et du bon sens (plus la fréquence est élevée, plus je crois que la proportion dans la population est élevée) ou plus formellement du théorème de Neyman et Pearson.
5. Calcul de la valeur de la statistique pour l'échantillon fortuit observé : S prend la valeur s .
6. Calcul de la valeur critique (ou des valeurs critiques) qui définissent la région critique à l'aide de la loi de probabilité de la statistique sous l'hypothèse H_0 :
$$\alpha = \text{prob}\{D_1 / H_0\} = \text{prob}\{S \in I_1 / H_0\}$$
7. Conclusion du test, prise de la décision. Si l'échantillon $\omega \in \Omega_1$, ou équivalent, si la valeur de la statistique du test $s \in I_1$ alors nous rejetons l'hypothèse H_0 . Sinon nous l'acceptons.
8. Nous calculons si possible le risque de second espèce ou la fonction puissance du test.

Valeur critique et probabilité critique

La présentation classique des tests selon l'approche de Neyman et Pearson suppose que l'on calcule d'abord une **valeur critique** pour un risque α donné, puis que l'on calcule la valeur prise par la statistique du test dans l'échantillon, et qu'enfin on place cette valeur calculée par rapport à la valeur critique pour prendre la décision. Cette méthode correspond à l'idée que la règle précède l'observation, et au fait que les tables ne sont établies que pour quelques valeurs de α : 10 %, 5 %, 2,5 %, 1 %.

Une autre approche tout à fait équivalente quant aux résultats en terme de décision, consiste à observer d'abord l'échantillon et calculer la statistique du test sur cet échantillon. Puis à calculer une **probabilité critique** qui est la probabilité que la statistique ait une valeur supérieure (pour une région critique de la forme $\{S > c\}$) à la valeur observée de la statistique. On compare alors cette probabilité critique à la valeur que l'on veut du niveau de confiance. Par cette méthode, grandement facilitée par les programmes informatiques qui ont remplacé les tables, on imprime la probabilité critique et on laisse à chacun prendre une décision suivant le seuil de confiance choisi. Si celui-ci est de 5%, alors une probabilité critique de 0,5% conduit à rejeter H_0 .

Attention, plus la probabilité critique est faible, plus la valeur critique serait facilement dépassée, et plus le rejet de H_0 a des chances d'être décidé.

6.3. Tests non paramétriques

Un test est dit paramétrique quand la loi des variables observées dans l'échantillon est une loi connue dépendant d'un (ou plusieurs) paramètre inconnu. Dans certains cas l'hypothèse H_0 ne porte pas sur le paramètre d'une loi connue mais sur la nature de la loi elle-même. C'est le cas des tests d'adéquation. Ou bien elle porte sur une propriété de cette loi qui n'est pas paramétrique. C'est le cas du test d'indépendance. Nous verrons ces deux tests dans la leçon L7. Mais c'est aussi le cas de tests très simples sur la loi des observations. Prenons le seul exemple du test des signes.

Sur deux sous populations nous avons des échantillons appariés X_1, \dots, X_n et Y_1, \dots, Y_n . Par exemple X_i et Y_i sont les ventes mensuelles du magasin i avant et après une campagne de publicité. On ne connaît pas la loi de probabilité de X ou de Y . Mais on comprend bien que tester l'efficacité de la publicité peut se faire en testant seulement l'hypothèse que $Y_i - X_i$ est positif avec une probabilité supérieure à $\frac{1}{2}$ (H_1). L'hypothèse H_0 sera au contraire celle d'un effet nul donc d'un signe qui s'établirait au hasard ($p = \frac{1}{2}$). Sous cette hypothèse H_0 , le nombre de signes positifs suit une loi Binomiale $(n, \frac{1}{2})$ que l'on peut approximer par une loi normale de moyenne $n/2$ et de variance $n/4$.

Dans l'exemple suivant nous avons les 15 paires de valeurs X, Y observées suivantes :

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	23	34	12	65	87	43	8	55	13	19	92	25	34	47	33
Y	26	32	14	60	77	48	11	51	17	22	85	28	37	44	36
Signe	+	-	+	-	-	+	+	-	+	+	-	+	+	-	+

Il y a 9 cas sur 15 pour lesquels y est supérieur à x . On pourrait dire que cette majorité suffit à croire en l'effet de la publicité. Mais dans l'hypothèse H_0 la probabilité de dépasser cette valeur est 0,15. Si l'on choisit un seuil de 5%, cette probabilité critique est trop importante pour rejeter H_0 . Par contre si on avait trouvé 11 signes +, elle serait de 5% et on pourrait commencer à rejeter H_0 et accepter H_1 , c'est-à-dire accepter un effet significatif de la publicité.

Evidemment ce test ne tient pas compte de la grandeur des écarts et de ce fait il n'est pas très puissant, c'est-à-dire pas très discriminant. Un autre test nécessiterait l'hypothèse d'une certaine loi pour X et Y dont un paramètre serait en question : on identifierait par exemple un succès de la campagne publicitaire au fait que la moyenne de Y est supérieure à celle de X (H_1). Cf infra le test de comparaison de moyennes qui serait ici unilatéral. Le test non paramétrique est en général moins puissant mais plus robuste (indépendant d'hypothèses supplémentaires).

Nous allons maintenant nous concentrer sur des tests paramétriques usuels.

6.4. Tests d'une moyenne

Soit (X_1, X_2, \dots, X_n) un échantillon de variables indépendantes et identiquement distribuées (*i.i.d.*) suivant une loi quelconque L , avec $E(X) = m$ et $V(X) = \sigma^2$.

Hypothèses à tester $H_0 : m = m_0$ contre $H_1 : m = m_1 > m_0$

Recherchons la forme de la région critique. Puisque \bar{X} est un estimateur sans biais et convergent de m nous utilisons ce résumé statistique pour le test.

Plus \bar{X} est grand plus nous "croyons" à H_1 et plus il est petit, plus nous croyons à H_0 . Donc la forme de la région critique (rejet de H_0) est : $\sum_{i=1}^n x_i \geq K \Leftrightarrow \bar{x} \geq c$.

Ce résultat intuitif serait confirmé par le théorème de Neyman-Pearson (que nous admettrons) qui dit que le test le plus puissant est celui dont la région critique est définie par l'inégalité $\frac{L(x, m_0)}{L(x, m_1)} \geq K$ dans laquelle $L(x, m_0)$ et $L(x, m_1)$ représentent respectivement la vraisemblance de l'échantillon x sous les hypothèses H_0 et H_1 .

On sait que la variable \bar{X} suit une loi $N(m, \sigma/\sqrt{n})$ si la loi L est déjà une loi normale ou si n est grand.

- si σ^2 est connu, on utilisera la loi normale pour déterminer c :

$$\alpha = \text{prob}(\bar{X} \geq c / m = m_0) = P_{m_0} \left(\frac{\bar{X} - m_0}{\sigma/\sqrt{n}} \geq \frac{c - m_0}{\sigma/\sqrt{n}} \right) = 5\%$$

$$\Leftrightarrow \frac{c - m_0}{\sigma/\sqrt{n}} = F_{N(0,1)}^{-1}(1 - \alpha) = 1,645$$

$$\Leftrightarrow c = m_0 + 1,645 \sigma / \sqrt{n}$$

- si σ^2 est inconnu, on utilisera son estimation s'^2 et le fractile d'ordre $1 - \alpha$ de la loi de *Student*($n - 1$) :

$$c = m_0 + F_{St(n-1)}^{-1}(1 - \alpha) s' / \sqrt{n}$$

- Dans le cas où aucune de ces conditions ne serait satisfaite, on suivra le même type de raisonnement avec la loi exacte suivie par la moyenne. Par exemple si X suit une loi de Poisson de paramètre $m = E(X)$, la règle de N.P. conduit encore à une région

critique de la forme $\sum_{i=1}^n X_i \geq k$. Or la somme d'une suite de variables *i.i.d.* de Poisson est encore une variable de Poisson de paramètre nm , même si n est petit. La loi de $\sum_{i=1}^n X_i = n\bar{X}$ est donc connue et l'on peut faire le test.

Exemple numérique

1. Sur un échantillon de taille 10, on veut tester $H_0 : m = 6$ contre $H_1 : m > 6$.
2. On choisit $\alpha = 5\%$. L'échantillon fortuit est : 7,9,3,11,10,7,9,12,8,4. La loi L est connue (loi normale), σ^2 est inconnu.
3. et 4. Théorème de N. et P. \Rightarrow La région critique est de la forme $\bar{X} \geq c$.
5. La statistique du test prend sur l'échantillon la valeur $\bar{x} = 8$ et l'estimation de la variance est $s'^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} = 74/9 = 8,22$.
6. La valeur critique est $c = m_0 + F_{St(9)}^{-1}(0,95) s' / \sqrt{n} = 6 + 1,833 \sqrt{8,22/10} = 7,66$.
7. La valeur observée 8 est dans l'intervalle critique. On décide donc de rejeter H_0 .

8. Le risque de second espèce se calculerait pour chaque valeur de m_1 par la formule :

$$\beta = \text{prob}\{D_0 / H_1\} = \text{prob}(\bar{X} < c / m = m_1) = P_{m_1}\left(\frac{\bar{X} - m_1}{s'/\sqrt{n}} < \frac{c - m_1}{s'/\sqrt{n}}\right) = F_{St(n-1)}\left(\frac{c - m_1}{s'/\sqrt{n}}\right)$$

Inversons l'hypothèse $H_1 : m = m_1 < m_0$

En reprenant les calculs ci-dessus avec $m_0 > m_1$, on voit que l'on obtiendra une région critique de la forme : $\sum_{i=1}^n x_i \leq K \Leftrightarrow \bar{x} \leq c$

$$\alpha = \text{prob}(\bar{X} \leq c / m = m_0) = 5\% \Leftrightarrow \frac{c - m_0}{\sigma/\sqrt{n}} = F_{N(0,1)}^{-1}(\alpha) = -1,645 \Leftrightarrow c = m_0 - 1,645 \sigma/\sqrt{n}$$

Si σ^2 est inconnu, on utilisera son estimation s'^2 et le fractile d'ordre $1 - \alpha$ de la loi de *Student*($n - 1$) : $c = m_0 - F_{St(n-1)}^{-1}(1 - \alpha) s'/\sqrt{n}$

Troisième cas usuel : $H_1 : m = m_1 \neq m_0$

Dans ce cas (test bilatéral), il est assez intuitif (et démontrable) qu'il faut choisir une région critique qui soit la réunion des deux précédentes, et donc répartir le risque α entre ces deux sous régions. En clair la région critique est définie par :

$$\{(\bar{X} \leq c_1) \cup (\bar{X} \geq c_2)\}$$

et il y a 2 valeurs critiques c_1 et c_2 :

$$\frac{\alpha}{2} = \text{prob}(\bar{X} \leq c_1 / m = m_0) \Rightarrow c_1 = m_0 - 1,96 \sigma/\sqrt{n} \text{ avec } -1,96 = F_{N(0,1)}^{-1}(\alpha/2)$$

$$\frac{\alpha}{2} = \text{prob}(\bar{X} \geq c_2 / m = m_0) \Rightarrow c_2 = m_0 + 1,96 \sigma/\sqrt{n}$$

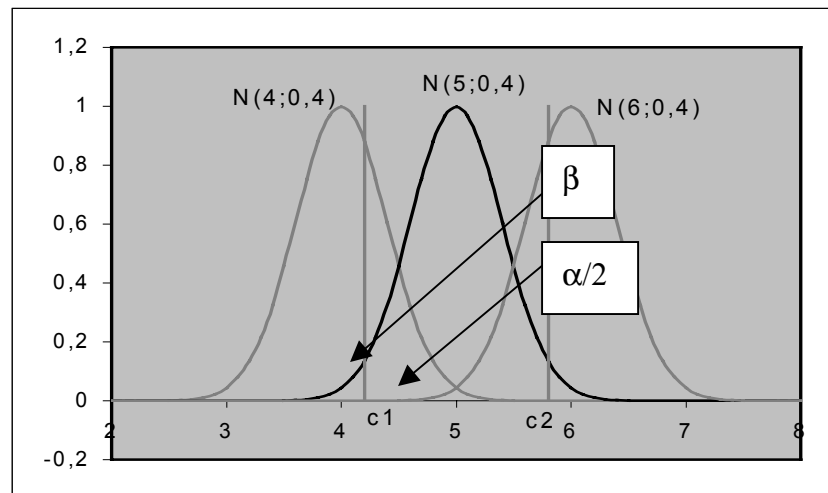
Exemple numérique :

Soit $H_0 : m_0 = 5$ à tester contre $H_1 : m \neq 5$ sachant que $n = 25$ et $\sigma = 2$.

Sous H_0 : $\bar{X} \mapsto N(m, \sigma/\sqrt{n}) = N(5; 0,4)$.

Les valeurs critiques sont : $c = 5 \pm 1,96 \times 0,4 = 5 \pm 0,8$.

La figure ci-dessus représente aussi les lois de \bar{X} dans l'hypothèse H_1 pour



$m_1 = 4$ et $m_1 = 6$ ainsi que le risque α associé à ce test et les deux risques β associés à ces deux hypothèses alternatives.

6.5. Test de comparaison de moyennes

Dans un test de comparaison paramétrique, on a deux populations P_1 et P_2 dans lesquelles une variable X est supposée suivre une loi L de même famille, mais dont on se demande si elle est de même paramètre ou non. On tire de façons indépendantes un échantillon dans chacune de ces populations :

$(X_1^{(1)}, X_2^{(1)}, \dots, X_{n_1}^{(1)})$ est l'échantillon *i.i.d.* tiré de la population P_1 et $(X_1^{(2)}, X_2^{(2)}, \dots, X_{n_2}^{(2)})$ l'échantillon *i.i.d.* tiré (indépendamment du premier) dans P_2 .

Les moyennes d'échantillon seront \bar{X}_1 et \bar{X}_2 , et les variances (non corrigées) S_1^2 et S_2^2 . Si \bar{X} et S^2 sont la moyenne et variance de l'échantillon de taille $n = n_1 + n_2$ formé à partir des deux échantillons prélevés dans P_1 et P_2 , on a les relations :

$$n\bar{X} = n_1\bar{X}_1 + n_2\bar{X}_2$$

$$nS^2 = n_1S_1^2 + n_2S_2^2 + \frac{n_1n_2}{n}(\bar{X}_1 - \bar{X}_2)^2$$

Peut-on sur la base d'une comparaison de deux statistiques similaires prises sur ces deux échantillons – ici les moyennes empiriques – tester l'égalité des paramètres de L dans P_1 et P_2 (ici les moyennes théoriques) ? On suppose que X suit une loi normale, $X^{(1)} \mapsto N(m_1, \sigma_1)$ et $X^{(2)} \mapsto N(m_2, \sigma_2)$.

A : On veut tester $H_0 : m_1 = m_2$ contre $H_1 : m_1 \neq m_2$ en supposant les variances connues

Alors puisque $\bar{X}_1 \mapsto N(m_1, \sigma_1/\sqrt{n})$ et $\bar{X}_2 \mapsto N(m_2, \sigma_2/\sqrt{n})$ nous avons :

$$\bar{X}_1 - \bar{X}_2 \mapsto N\left(m_1 - m_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

Tester H_0 , c'est donc tester la nullité de la moyenne théorique $(m_1 - m_2)$ à l'aide de la statistique $\bar{X}_1 - \bar{X}_2$. Le test est analogue au test de moyenne développé précédemment. Nous aurons tendance à rejeter H_0 lorsque l'écart entre les moyennes empiriques est trop grand. La région critique est donc :

$$\left\{ |\bar{X}_1 - \bar{X}_2| \geq c \right\} \text{ avec } c = f_{1-\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$f_{1-\alpha/2}$ désigne le fractile d'ordre $1-\alpha/2$ de la loi $N(0,1)$:
 $f_{1-\alpha/2} = F_{N(0,1)}^{-1}(1-\alpha/2)$.

B : Mêmes hypothèses avec variances inconnues mais supposées égales ($\sigma_1^2 = \sigma_2^2$)

La région critique ne change pas mais cette variance commune inconnue est estimée par la statistique :

$$S'^2 = \frac{\sum_{i=1}^{n_1} (X_i^{(1)} - \bar{X}_1)^2 + \sum_{i=1}^{n_2} (X_i^{(2)} - \bar{X}_2)^2}{n_1 + n_2 - 2} = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2} = \frac{(n_1 - 1)S_1'^2 + (n_2 - 1)S_2'^2}{n_1 + n_2 - 2}$$

$$\frac{(\bar{X}_1 - \bar{X}_2) - (m_1 - m_2)}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} S''} \text{ suit alors une loi de } Student(n_1 + n_2 - 2) \text{ et la région}$$

critique est :

$$\left\{ |\bar{X}_1 - \bar{X}_2| \geq c \right\} \text{ avec } c = t_{n_1+n_2-2; 1-\alpha/2} \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}} \text{ (Test de Student)}$$

$t_{n_1+n_2-2; 1-\alpha/2}$ désigne le fractile d'ordre $1-\alpha/2$ de la loi de $Student(n_1 + n_2 - 2)$.

C : Mêmes hypothèses avec des variances inconnues et inégales

Celles-ci seront estimées séparément sur chaque échantillon par :

$$S_1'^2 = \frac{\sum_{i=1}^{n_1} (X_i^{(1)} - \bar{X}_1)^2}{n_1 - 1} = \frac{n_1 S_1^2}{n_1 - 1} \text{ et } S_2'^2 = \frac{\sum_{i=1}^{n_2} (X_i^{(2)} - \bar{X}_2)^2}{n_2 - 1} = \frac{n_2 S_2^2}{n_2 - 1}$$

et l'on poursuivra le test sous la forme A (loi normale) dans le cas d'échantillons de grande taille, et sinon sous la forme B (loi de Student avec un degré de liberté ν qui n'est plus n_1+n_2-2 mais une fonction complexe de s_1 et de s_2) :

$$\nu = \left[\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1} \right]^2 \left/ \left[\frac{s_1^4}{(n_1 - 1)^3} + \frac{s_2^4}{(n_2 - 1)^3} \right] \right.$$

$$\left\{ |\bar{X}_1 - \bar{X}_2| \geq c \right\} \text{ avec } c = f_{1-\alpha/2} \sqrt{\frac{s_1'^2}{n_1} + \frac{s_2'^2}{n_2}} \equiv f_{1-\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

Si la loi des observations n'est pas normale, les tests de comparaison peuvent encore être faits de la même façon qu'en C pour des échantillons de grande taille grâce au théorème central limite (TCL).

6.6. Test d'une proportion

On veut tester $H_0 : p = p_0$ contre $H_1 : p > p_0$ où p représente une proportion dans une population d'individus ayant la propriété P .

L'échantillon aléatoire (X_1, X_2, \dots, X_n) est une suite de variables de Bernoulli de paramètre p (qui est aussi égal à $E(X_i)$). Nous sommes dans le cas particulier du test précédent d'une moyenne théorique de variables non normales. L'intuition et la règle de Neyman et Pearson conduisent à utiliser la statistique $\bar{X} = F = \text{fréquence}$ dans l'échantillon.

Compte tenu du fait que $p_1 > p_0$ et donc que les deux facteurs de $\sum x_i$ sont négatifs. L'exemple développé en introduction montre comment se fait ensuite ce test, sur le modèle du test de moyennes dont c'est un cas particulier.

6.7. Test de comparaison de proportions

Les variables $X_i^{(1)}$ et $X_i^{(2)}$ sont maintenant des variables de Bernoulli de paramètres respectifs p_1 et p_2 . Nous voulons tester $H_0 : p_1 = p_2$ contre $H_1 : p_1 \neq p_2$.

Puisque $p_1 = E(X_i^{(1)})$ et $p_2 = E(X_i^{(2)})$ nous sommes ramenés au test de comparaison de deux moyennes (de lois de Bernoulli) avec $\bar{X}_1 - \bar{X}_2 = F_1 - F_2$ (différence des fréquences relatives dans les deux échantillons) et avec des variances inconnues, puisque égales à $p_1(1-p_1)$ et $p_2(1-p_2)$, mais égales entre elles dans l'hypothèse H_0 . On estime cette variance commune par :

$$F(1-F) \text{ avec } F = \frac{n_1 F_1 + n_2 F_2}{n_1 + n_2}$$

Exemple :

Les fumeurs sont-ils en même proportion chez les hommes et chez les femmes sachant que j'ai observé sur deux échantillons tirés au hasard 24 fumeurs sur 60 hommes et 18 fumeuses sur 40 femmes ?

$$\text{Sous } H_0, F \underset{\text{appr.}}{\sim} N\left(0, \sqrt{\left(\frac{1}{60} + \frac{1}{40}\right) f(1-f)}\right) = N(0, 1/10) \text{ avec } f = \frac{24+18}{100} = 0,42 \text{ et}$$

$$\alpha/2 = 0,025 = \text{prob}(F_1 - F_2 \leq c / H_0) = P_0\left(\frac{F_1 - F_2}{0,1} \leq \frac{c}{0,1}\right) \Rightarrow c = -0,196$$

La région critique est donnée par $\{|F_1 - F_2| > 0,196\}$.

Or $f_1 = 24/60 = 0,40$ et $f_2 = 18/40 = 0,45$ donc $f_1 - f_2 = -0,05$.

La différence n'est pas significative. Nous acceptons $H_0 : p_1 = p_2$.

6.8. Test d'une variance

Nous voulons tester $H_0 : \sigma = \sigma_0$ contre $H_1 : \sigma > \sigma_0$ où σ^2 est la variance des variables X_i de lois normales observées dans un échantillon aléatoire de taille n .

L'intuition, confirmée par la règle de Neyman et Pearson conduit à utiliser la variance empirique pour tester la variance théorique et à choisir une région critique de la forme $\left\{ \sum_{i=1}^n (X_i - m)^2 \geq c \right\}$. Il faut faire une l'hypothèse que la loi L des X_i est normale pour continuer et utiliser la loi du χ^2 :

$$\text{- si } m \text{ est connue, } \frac{\sum_{i=1}^n (X_i - m)^2}{\sigma^2} \mapsto \chi^2(n) \text{ et } \alpha = \text{prob} \left(\frac{\sum_{i=1}^n (X_i - m)^2}{\sigma_0^2} \geq \frac{c}{\sigma_0^2} \right)$$

On peut donc en déduire $c = \chi_{n;\alpha}^2 \sigma_0^2$.

$$\text{- si } m \text{ est inconnue, on utilisera } \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} = \frac{nS^2}{\sigma^2} = \frac{(n-1)S'^2}{\sigma^2} \mapsto \chi^2(n-1).$$

Si la loi des X_i n'est pas normale, il faut avoir un grand échantillon et utiliser une loi asymptotique de S'^2 qui est $N(\sigma^2, \sigma^2 \sqrt{2/n})$.

6.9. Test de comparaison de variances

Les variables $X_i^{(1)}$ et $X_i^{(2)}$ sont supposées normales dans les deux populations. Nous voulons tester $H_0 : \sigma_1^2 = \sigma_2^2$ contre $H_1 : \sigma_1^2 \neq \sigma_2^2$.

Le rapport des estimations s_1^2 et s_2^2 des variances dans les deux échantillons fortuits devrait être proche de 1 sous H_0 .

Or $(n_1 - 1)S_1'^2 / \sigma_1^2$ suit une loi du $\chi^2(n_1 - 1)$ et $(n_2 - 1)S_2'^2 / \sigma_2^2$ suit indépendamment $\chi^2(n_2 - 1)$. Donc $\frac{S_1'^2 / \sigma_1^2}{S_2'^2 / \sigma_2^2} ; F(n_1 - 1, n_2 - 1)$ (loi de Fisher-Snedecor).

Sous l'hypothèse H_0 cette quantité est simplement égale à $\frac{S_1'^2}{S_2'^2} = \frac{S_1^2}{S_2^2}$.

La région critique est donc de la forme $\left\{ S_1^2 / S_2^2 < c_1 \text{ ou } S_1^2 / S_2^2 > c_2 \right\}$ avec c_1 et c_2 qui sont les fractiles d'ordre $\alpha/2$ et $1-\alpha/2$ de cette loi de Fisher.

Exemple :

Soient les deux échantillons d'une grandeur X supposée normale :

2, 7, 5, 9, 12 ($n_1 = 5$) et 3, 5, 11, 6, 3, 9, 4, 5, 4, 10 ($n_2 = 10$).

Moyennes 7 et 6. Variances empiriques corrigées S^2 : 14,5 et 8,66.

1°) Test de comparaison de variances. Peut-on dire (H_0) que $\sigma_1 = \sigma_2$?

S_1^2/S_2^2 suit une loi de Fisher dont les degrés de liberté sont 4 et 9.

Le rapport calculé vaut 1,674 et la probabilité qu'une telle valeur soit dépassée est 0,24 (fonction Loi.F) du tableur EXCEL. Comme elle est nettement supérieure à 2,5% (pour un test bilatéral à 5%) on rejette l'hypothèse d'égalité des variances. La fonction TEST.F du même tableur donne directement la probabilité critique du test à partir des deux séries (On obtient 0,48).

2°) Test de comparaison de moyennes. Peut on dire (H_0) que $m_1 = m_2$?

Compte tenu des résultats du test précédent, on est dans le cas de variances inconnues et différentes. La Statistique du test est la différence des moyennes

empiriques $\bar{X}_1 - \bar{X}_2 \mapsto N\left(m_1 - m_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$. Sous H_0 , $m_1 - m_2 = 0$ et si l'on remplace

les variances inconnues par leurs estimations 14,5 et 8,66 la variable centrée réduite prend la valeur 0,533. Or elle suit une loi $N(0,1)$ pour de grands échantillons, et une loi proche de Student(13) pour nos deux petits échantillons dont les valeurs critiques à 5% sont de l'ordre de 2. On est dans la région d'acceptation de l'hypothèse d'égalité des moyennes théorique.