Eric Depiereux, Benoît DeHertogh, Grégoire Vincke

www.fundp.ac.be/biostats **Module 110**

Eric Depiereux, Benoît DeHertogh, Grégoire Vincke

110 Principes du test d'hypothèse

110.1 Le test d'hypothèse

110.1.1 Modèles

Le but d'un test d'hypothèse est de déterminer s'il est vraisemblable ou non qu'un paramètre d'un modèle prenne des valeurs différentes dans des populations distinctes, afin de mettre en évidence, ou non, l'effet d'un facteur expérimental sur une population.

La différence fondamentale entre l'intervalle de confiance et le test d'hypothèse est que le premier est centré sur une statistique connue (mesurée dans un échantillon) tandis que le second est centré sur le paramètre d'une population.

Un écologiste suspecte que les Tilapia grandissent plus vite dans une eau plus chaude ; il étudie la croissance (X : taille à un an) d'échantillons de Tilapia élevés à 15°C (population 1) et à 25°C (population 2).

Le modèle Ho : $x_{(i)j} = \mu + E_{(i)j}$ décrit la situation dans l'hypothèse où la t° n'influence pas la croissance. Dans ce cas les populations 1 et 2 sont caractérisées par la même moyenne μ.

Le modèle H_1 : $x_{(i)j} = \mu + a_i + E_{(i)j}$ décrit la situation dans l'hypothèse où la t° influence la croissance. Dans ce cas la population 1 est caractérisée par la moyenne $\mu_1=\mu+ a_1$ et la population 2 est caractérisée par la moyenne $\mu 2=\mu+ a_2$ (avec ici a1 = -a2).

Selon H_1, deux constantes a_1 et a_2 sont nécessaires pour exprimer que la croissance est inférieure à μ à 15°C et supérieure à μ à 25°C.

Selon Ho, ces constantes a_i ne sont pas nécessaires puisque aux deux températures, les deux populations de poissons ont une même croissance moyenne μ.

Le test repose sur l'hypothèse nulle (Ho) qui établit que la valeur du paramètre est la même dans différentes populations. L'expérience aura généralement pour but de montrer que cette affirmation Ho est invraisemblable.

Le caractère invraisemblable est défini en référence à la probabilité α : une situation est invraisemblable si elle a une probabilité de se produire < α..

L'hypothèse nulle suppose que la température n'explique pas la variabilité du poids. Le modèle suivant Ho décrit que la variabilité de la taille d'un individu j à la t° i dépend uniquement la variabilité résiduelle des individus autour de la moyenne μ, $E_{(i)j}$ v.a. $N(0,\sigma^2)$.

L'hypothèse alternative H_1 suppose que la variabilité du poids comprend une composante factorielle dépendant de la t°. Le modèle H_1 comprend donc un facteur spécifique de la température [ai] dont l'expérimentateur souhaite démontrer l'existence.

Le test est une technique permettant de quantifier le risque d'erreur lié au choix de l'une ou l'autre hypothèse.

110.1.2 Confiance

Comme l'intervalle de confiance, le test d'hypothèse est caractérisé par une confiance 1 -α. Ho est valide[1] aussi longtemps que les valeurs expérimentales ont une probabilité 1 -α d'être observées fortuitement[2] dans le modèle proposé par Ho.

La confiance est fixée par l'expérimentateur, en fonction des circonstances.

95% est une valeur de référence arbitraire unanimement acceptée et souvent utilisée. Toutefois, de nombreuses situations, telle la mise d'un médicament sur le marché, exigent une plus grande confiance.

Ho est invalidée lorsque les valeurs expérimentales ont une probabilité < α d'être observées fortuitement dans le modèle proposé par Ho.

Il y a donc trois élements en jeu dans le test d'hypothèse :
- un modèle hypothétique, Ho, centré sur un paramètre connu ou supposé tel ;
- un modèle hypothétique, H_1, centré sur un paramètre inconnu supposé différent du paramètre proposé par Ho;
- une mesure réelle, statistique d'un échantillon, qui est confrontée au modèle Ho.

[1] *Valide : recevable comme vrai.*

[2] *Fortuitement : qui arrive par hasard.*

H_1 étant de paramètre inconnu, il est impossible de le confronter à la mesure. Si l'on veut préciser le paramètre de H_1, il faut recourir à l'intervalle de confiance.

Un médecin teste sur 10 patients pris au hasard une nouvelle substance hypo -cholesterolémiante[3], et il observe une moyenne de cholesterolémie[4] qui aurait moins de 5% de chances de se rencontrer sur 10 patients pris au hasard dans la population non traitée.

le modèle suivant Ho suppose que le traitement n'a aucun effet : Ho est invalidée par l'expérience car la moyenne observée a moins de 5% de chances d'être observée dans le modèle suivant Ho.

Cependant, Ho n'est pas <u>impossible</u>, elle est seulement improbable : la moyenne de cholesterolémie observée sur les individus traités a plus de chances de se rencontrer sur 10 patients pris au hasard dans la population non traitée (moins de 5% [α], mais tout de même).

α représente la probabilité de commettre l'erreur d'invalider Ho, (Rho) alors que le modèle suivant Ho est le bon modèle.

Le test d'hypothèse est aussi caractérisé par sa puissance ($1 - \beta$) qui est la probabilité de pouvoir invalider Ho, si elle est fausse.

La puissance est liée aux circonstances de l'expérience. Si le nombre d'observations est très petit, et/ou si la différence recherchée (a_i) est très petite et/ou si la variabilité résiduelle $E_{(i)j}$ est très grande, la puissance est faible.

Tout expérimentateur souhaite réaliser un test puissant avec le minimum d'observations possible.

Un médecin teste sur quelques patients un médicament générique[5], et il observe des valeurs qui ont 30% de chances de se rencontrer dans la population traitée avec le médicament original. Ho pose que le générique ne présente aucune différence avec la spécialité brevetée : Ho reste donc valide jusqu'à preuve du contraire.

---

[3] hypo -cholesterolémiant : qui fait diminuer la cholesterolémie .

[4] Cholesterolémie : taux sanguin du cholestérol (compris normalement entre 1,50 et 2,80 g par litre).

[5] Médicament générique : médicament dont la formule est tombée dans le domaine public et qui est vendu sous sa dénomination commune à un prix inférieur à celui de la spécialité correspondante.

110.1.3 Puissance

Ho n'est cependant pas certaine, elle est seulement valide jusqu'à preuve du contraire. Il est possible qu'il y ait une différence entre les médicaments qui n'ait pas été mise en évidence par l'expérience. La probabilité que cela se produise est appelée β.

> β représente la probabilité de commettre l'erreur de ne pas invalider Ho, (Aho) alors que le modèle suivant H_1 est le bon modèle.

Le test d'hypothèse se présente comme le jugement d'un suspect par un tribunal. Tant qu'une « preuve suffisante » de sa culpabilité n'aura pas été avancée, le suspect est présumé innocent (Ho).La « preuve suffisante » de sa culpabilité ne laisse -t -elle pas le moindre doute ? Evidemment, si... Il y a toujours une probabilité de condamner un innocent (α). L'absence de preuve donne -t -elle la certitude de son innocence ? Evidemment non. Il y a toujours une probabilité d'innocenter un coupable (β).

	Ho vraie	Ho fausse
Ho validée	$1 - \alpha$	β
Ho invalidée	α	$1 - \beta$

Tableau 110 -110-1 Probabilités de validation et d'invalidation de Ho en fonction de sa nature vraie ou fausse.

	Ho vraie	Ho fausse
AHo	confiance	Erreur type II
RHo	Erreur type I	puissance

Tableau 110 -110-2 Qualités et type d'erreur du test d'hypothèse en fonction de la nature vraie ou fausse de Ho.

> En statistique, alpha (α) représente une probabilité petite et connue tandis que bêta (β) représente une probabilité peut être grande et/ou inconnue.

En cas de rejet de Ho, on parlera d'un
Test significatif (S) si α est fixé à 5% ; on notera parfois (*)
Test hautement significatif (SS) si α est fixé à 1% ; on notera parfois (**)
Test très hautement significatif (SSS) si α est fixé à 1%° ; on notera parfois (***)

Il est cependant de plus en plus courant d'indiquer dans les tableaux la probabilité α, correspondant à la valeur observée dans l'échantillon, calculée par intégration numérique. Dans le texte, les résultats seront suivis de la mention $P < 0,05$, $P < 0,01$, suivant le seuil le plus proche de celui indiqué dans le tableau.

110.2 Exemple

110.2.1 Formulation

Lorsqu'un échantillon est prélevé dans une population, c'est généralement dans le but de <u>tester une hypothèse</u> au sujet de cette population.

A l'occasion d'un jeu de pile ou face, pour lequel le succès est "pile", il me semble qu'en utilisant la pièce de mon adversaire, la probabilité d'obtenir "pile" soit plus grande que celle d'obtenir face, et je mets en doute que sa pièce soit équilibrée. Je prétends donc tester sur 100 lancers si oui ou non cette pièce peut être considérée comme étant équilibrée.

Définissons la variable :

X : le nombre de fois que l'on obtient "pile" sur 100 lancers Bi(n;π) ≈ v.a.N(n,n.π.(1 - π)) si $\pi \approx 0,5$

Formulons les hypothèses nulle (Ho) et alternative (H_1) :

Ho : la pièce est équilibrée $\pi = 0,5$, $\mu = n\,\pi = 50$ et $\sigma^2 = n\,\pi\,(1 - \pi) = 25$. Selon Ho X suit approximativement une v.a. N(50;25)

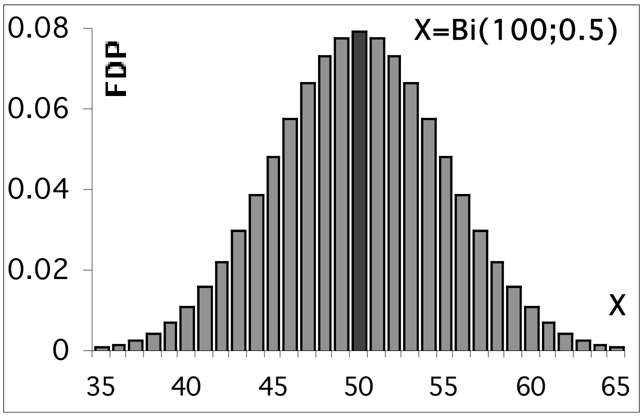

Figure 110 -1 Distribution de probabilité du nombre de 'pile' sur 100 lancers d'une pièce de monnaie selon le modèle de Ho.

Cette distribution représente toutes les valeurs <u>possibles</u> pour X selon Ho.

H_1 : la pièce n'est PAS équilibrée : $\pi > 0,5$: X est Bi(100;π)

Selon H_1, X suit une distribution de paramètre π indéterminé

π est supposé > 0,5 car si mon adversaire triche, je suppose qu'il le fait à son avantage.

110.2.2 Interprétation

Supposons que j'obtienne 100 fois pile.

Suivant la loi des probabilités composées pour des événements indépendants, la probabilité a priori d'obtenir 100 fois pile avec une pièce de monnaie équilibrée est de $0{,}5^{100}$ soit 10^{-30} environ. Elle est tellement petite que je la considère comme nulle. Néanmoins, même dans ce cas extrême, il me reste une probabilité non nulle de me tromper en affirmant que la pièce n'est pas équilibrée.

Il semble donc évident que la pièce n'est pas équilibrée, mais ce n'est pas certain : en matière de test d'hypothèse, nous n'obtiendrons jamais que des quasi -certitudes.

Supposons maintenant une situation plus réaliste : j'obtiens 65 fois "pile". Ici la décision est plus difficile à prendre : ce score a -t -il été obtenu par hasard avec une pièce équilibrée, ou est -il assez fortuit que pour établir que la pièce n'est pas équilibrée?

Il est impossible pour l'expérimentateur d'en décider raisonnablement en se fiant à sa seule intuition. Du reste, il est aussi impossible d'obtenir une ***certitude*** *par calcul. Il faut donc choisir en faveur de la situation* ***la plus probable,*** *sur base de la distribution de probabilités de la variable aléatoire.*

L'inférence statistique permet à l'expérimentateur de prendre une décision réfléchie, sur base de la connaissance de son risque d'erreur.

Remarquons que des deux hypothèses mises en opposition, Ho permet de calculer des probabilités (μ = 50, σ^2 = 25), l'autre non (π indéterminé).

Nous considérons comme valide l'hypothèse qui nous permet de calculer des probabilités. C'est l'<u>hypothèse nulle</u> (H_0). Tant que nous n'aurons pas acquis la quasi -certitude qu'elle est fausse, elle sera acceptée.

L'autre hypothèse est l'<u>hypothèse alternative</u> (H_1). Si je peux montrer que le résultat est improbable sous Ho, nous la rejetterons (RHo), sinon nous l'accepterons (AHo).

Ho : $\mu = 50$
H_1 : $\mu > 50$

Définissons "une petite probabilité" α, qui est le risque que nous accepterons de prendre en rejetant Ho . Soit α = 5%.

Nous divisons dès lors la distribution en deux zones : la zone dans laquelle il est <u>probable</u> d'observer le résultat si Ho est vraie (1 -α) et la zone dans laquelle il est <u>improbable</u> d'observer le résultat si Ho est vraie (α). La limite *arbitraire* entre ces deux zones est la valeur de X à partir de laquelle nous allons rejeter Ho en faveur de H_1. Cette valeur x_{seuil} est appelée <u>seuil de rejet</u>. Nous considérons Ho comme valide tant que le résultat ne dépasse pas le seuil de rejet x_{seuil}.

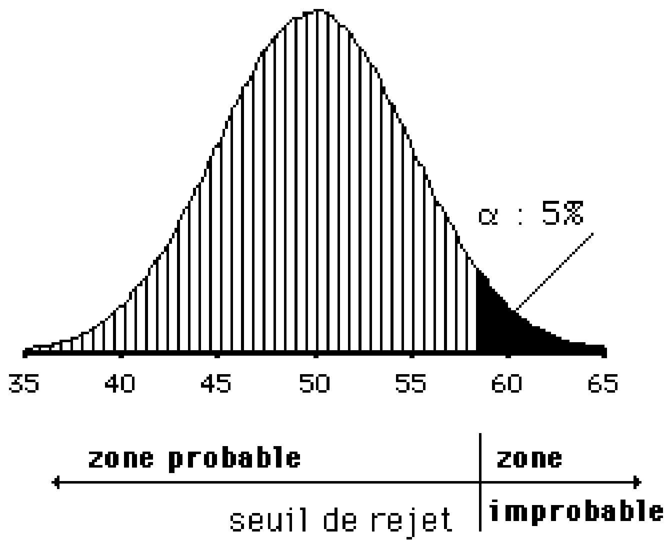

Figure 110 -2 Une petite probabilité α détermine le seuil à partir duquel Ho sera considérée comme invraisemblable.

La valeur du seuil de rejet x_{seuil} peut être calculée en référence à la distribution normale réduite. Dans notre exemple :

$P(Z \leq z) = 0,95 \Leftrightarrow z = 1,65$
$x_{seuil} = z\sigma + \mu = 1,65 * 5 + 50 = 58,25$

La valeur observée dans l'échantillon x_{obs} est comparée au seuil de rejet x_{seuil}.

Si x_{obs} appartient à la zone probable, le résultat est compatible avec la vraisemblance de H_0 : nous n'avons rien observé d'extraordinaire, rien qui permette d'affirmer avec une probabilité suffisamment élevée que H_0 soit fausse.

Si X_{obs} appartient à la zone improbable, je vais considérer que Ho est fort peu vraisemblable, et je vais la rejeter en faveur de H_1.

Je ne peux pas être certain qu'elle soit fausse, puisque le résultat reste possible si elle est vraie. Rappelons -nous que même si on observe 100 lancers pile, on n'obtient pas de certitude absolue. On doit donc tolérer un certain risque d'erreur, fixé à la valeur α. Nous verrons par la suite les critères qui guident le choix de la valeur de α.

110.3 L'échantillonnage

La première étape de la démarche statistique consiste à prélever un échantillon dans la population. Il est évident que la validité de toutes les statistiques subséquentes sera liée à la qualité de cet échantillon.

Il faut suivre des règles d'échantillonnage si l'on veut pouvoir chiffrer, sur base de la théorie des probabilités, les risques d'erreur liés à l'inférence.

110.3.1 Echantillon aléatoire simple.

L'échantillon est <u>aléatoire</u> si tous les individus ont la même probabilité de faire partie de l'échantillon

Il s'est déjà trouvé dans la presse une enquête sur la motivation des lecteurs à répondre à un sondage d'opinion. L'enquête était réalisée à partir d'un questionnaire que le lecteur devait remplir et renvoyer à la rédaction. Il en ressortait que l'immense majorité des lecteurs étaient motivés pour répondre à un sondage : effectivement, on imagine difficilement qu'un lecteur non motivé renvoie son questionnaire pour dire qu'il ne répond jamais à des enquêtes. Un tel échantillon ne représente pas du tout la population des lecteurs, car il n'est pas aléatoire.

L'échantillon est <u>simple</u> si toutes les observations sont indépendantes les unes des autres, c'est -à -dire si le résultat (Ai) obtenu pour une observation n'est en rien influencé par le résultat (Aj) obtenu pour une autre observation.

$$P(Ai/Aj) = P(Ai)$$

Je demande à un étudiant jobiste de réactualiser le fichier qui reprend les anciens étudiants de Biologie, dans le but de déterminer le pourcentage d'entre eux qui travaillent dans l'enseignement, dans la recherche, dans l'industrie… Comme de nombreux numéros de téléphone sont manquants, à l'occasion d'un appel réussi, le jobiste demande si son correspondant connaît les références de tel ou tel de sa promotion.

Supposons qu'au départ de 2 numéros connus, le jobiste récolte 10 numéros sur les 30 étudiants de la promotion. Parmi les 12 réponses obtenues, 6 sont enseignants, dont les deux anciens appelés au départ.

Cette proportion de 50% d'enseignants (6/12) ne pourra pas être considérée comme une statistique représentative de l'ensemble de la promotion, car il y a plus de chances que les deux premiers anciens contactés aient conservés plus spécifiquement les références d'autres enseignants : les 10 réponses obtenues ne sont donc pas indépendantes des deux premières.

Deux échantillons A et B sont <u>indépendants</u> si toutes les observations de l'un sont indépendantes des observations de l'autre, c'est -à -dire si le résultat (Ai) obtenu pour une observation quelconque de l'échantillon A n'est en rien influencé par le résultat (Bj) obtenu pour une observation quelconque de l'échantillon B

$$P(Ai/Bj) = P(Ai)$$

Un médecin étudie le risque de récidive[6] après un infarctus du myocarde[7], pour des patients ayant souffert d'un ou deux infarctus.
Si les patients du groupe 1 sont différents des patients du groupe 2, on peut supposer que les échantillons sont indépendants.
Par contre si tous, ou seulement certains patients se trouvent dans les groupes 1 et 2, les échantillons ne sont pas indépendants.

Les dépendances inévitables ou délibérément introduites dans le plan expérimental doivent être contrôlées, systématiques, et prises en compte dans le plan de l'analyse statistique (cfr. Critères de classification en analyse de la variance).

Dans notre exemple, si le médecin a systématiquement suivi les patients ayant eu deux infarctus, avant et après le second accident cardiaque, donc si tous les patients du groupe 1 se retrouvent dans le groupe 2, le plan de l'analyse peut (et doit) en tenir compte.

Si les dépendances sont désordonnées, l'analyse peut être rendue impossible.

Dans le cas où le groupe 2 comprendrait des patients provenant du groupe 1 et de nouveaux patients, l'analyse serait impossible.

Dans tous les tests statistiques, l'indépendance des observations est une condition essentielle de la validité des tests.

---

[6] *Récidive : réapparition d'une maladie, d'un mal dont un sujet déjà atteint avait complètement guéri.*

[7] *Infarctus du myocarde, lésion du cœur de gravité variable, consécutive à l'oblitération d'une artère coronaire.*

110.3.2 Taille de l'échantillon

- **La taille optimale**

d'un échantillon est celle qui permet de donner la puissance nécessaire avec la confiance voulue. Nous verrons par ailleurs une technique qui permet de la déterminer.

> *Le plan expérimental doit limiter les coûts, le temps, l'utilisation d'animaux de laboratoires… Si l'expérimentateur dispose de 50 observations, il est souvent préférable de répondre à 5 questions avec n = 10 et une certitude de 95% plutôt qu'à une seule question avec n = 50 et une certitude de 99,8%, sauf cas particuliers où une très grande confiance est exigée.*

- **Petits échantillons**

Dans de nombreux cas, comme l'étude de certaines maladies rares, le nombre d'observations disponibles est forcément limité. Un plan expérimental bien contrôlé et une technique d'analyse puissante peuvent permettre d'en tirer des conclusions fiables.

- **Sur -échantillonnage**

Une attention particulière doit être apportée aux échantillons de très grande taille, qui ont tendance à se produire fréquemment depuis la banalisation de l'accès à d'immenses banques de données, via internet par exemple.

> *Imaginons l'étude de l'apparition de AAAAA dans une banque génomique dans laquelle les nucléotides ACGT seraient répartis au hasard. Cet événement a une probabilité $0{,}25^5 \approx 10^{-3}$ de se produire, ce qui au regard de α = 5% est une très petite probabilité. Cependant, une recherche dans une banque de 10^8 nucléotides produira environ 100.000 résultats « fortuits » !*

La valeur α représente un faible risque <u>relatif</u>, non <u>absolu</u>.

- **Egalité des effectifs**

Pour simplifier les calculs et pouvoir appliquer toutes les techniques, l'expérimentateur doit veiller à récolter des échantillons de même taille.

Plus les plans expérimentaux se complexifient, plus il est difficile, et parfois impossible, de prendre en compte des échantillons de taille différente.

Dans tous les cas, les formules à utiliser pour effectuer les tests sur des échantillons de tailles différentes sont plus complexes, car l'égalité des effectifs permet de nombreuses simplifications.

Pour faciliter la compréhension des méthodes en évitant les approches trop techniques nous n'examinerons dans le cadre de ce cours, que des échantillons de même taille.