

Chapitre 4

Les tests d'hypothèses

1. Vocabulaire lié aux tests d'hypothèses

En statistiques, les réponses ne sont jamais tranchées. Il y a toujours une part d'aléatoire, d'imprévu, d'écart. C'est normal car sans cette part, les statistiques ne seraient pas nécessaires. Si un événement est certain, il n'y a pas besoin de loi de probabilités ou de test pour le prédire.

En statistiques, les tests, ou plus précisément **les tests d'hypothèses**, sont un outil, une règle de décision, qui permet de rejeter ou non une hypothèse, une réponse avec un risque connu. Le risque zéro dans l'absolu n'existe que dans les domaines où les statistiques ne sont pas nécessaires.

Le terme "hypothèse" est important. Tous les tests réalisés partent d'une affirmation à confirmer ou à infirmer. Par exemple, avec le test de Shapiro-Wilk, l'affirmation est : les données suivent une loi normale. Cette affirmation est appelée **hypothèse nulle** et écrite H_0 . Rejeter H_0 revient à accepter l'**hypothèse alternative** H_1 : les données ne suivent pas une loi normale.

Le test d'hypothèses permet de dire "À 95 % (ou 90 % ou 99 %), la réponse est..." Cela signifie qu'il y a une possibilité de se tromper. Deux types d'erreurs existent : l'**erreur de première espèce** et l'**erreur de deuxième espèce** (cf. figure 04-01). L'erreur de première espèce survient lorsque H_0 est rejetée alors qu'elle est vraie. La probabilité associée à cette erreur est le **seuil de significativité**, α , généralement fixé à 5 %, soit $\alpha = 0,05$. L'erreur de deuxième espèce survient lorsque H_0 n'est pas rejetée ou est gardée alors qu'elle est fausse. La probabilité associée à cette erreur est liée à la **puissance du test**, $1 - \beta$ qui calcule la probabilité de rejeter H_0 alors qu'elle est bien fausse avec β l'erreur de deuxième espèce.

■ Remarque

Le choix entre l'hypothèse nulle et alternative ne dépend que de la question posée et du contexte, même s'il est souvent conseillé de minimiser en priorité l'erreur de première espèce. Le but est souvent de rejeter H_0 de manière certaine.

Test d'hypothèses		Réalité	
		H_0 est vraie	H_0 est fausse
Décision	H_0 est conservée	Vrais positifs Décision juste <i>H_0 est acceptée alors qu'elle est vraie</i> Seuil de confiance : Probabilité = $1 - \alpha$	Faux positifs Erreur de 2^e espèce <i>H_0 est acceptée alors qu'elle est fausse</i> Probabilité = β
	H_0 est rejetée	Faux négatifs Erreur de 1^{re} espèce <i>H_0 est rejetée alors qu'elle est vraie</i> Seuil de significativité : Probabilité = α	Vrais négatifs Décision juste <i>H_0 est rejetée alors qu'elle est fausse</i> Puissance du test : Probabilité $1 - \beta$

Le seuil de significativité α – ou le seuil de confiance $1 - \alpha$ – est fixé en amont par la personne réalisant le test, mais il peut aussi être fixé pour répondre à des normes industrielles, pharmaceutiques...

La puissance du test $1 - \beta$ est calculée après la réalisation du test à partir de données (moyenne, dispersion et taille des échantillons), du seuil de significativité α et du type de test d'hypothèses réalisé :

- La puissance du test est corrélée à la population étudiée, c'est-à-dire que si les groupes testés sont très différents, la puissance du test est meilleure.

Par exemple, le test d'hypothèses cherche à voir si la taille moyenne des femmes est identique à la taille moyenne des hommes. La puissance du test est dépendante de la moyenne et de la variabilité des tailles mesurées au sein de chaque genre. Si les deux genres ont des moyennes très différentes comme c'est le cas en Suède (femmes = 167 cm et hommes = 182 cm), la puissance du test est meilleure qu'à Singapour où la différence entre les genres est plus faible (femmes = 160 cm et hommes = 171 cm).

- La puissance du test est positivement corrélée à la taille de l'échantillon, c'est-à-dire que plus la taille de l'échantillon analysé est grande, plus la puissance du test augmente pour une même population étudiée.
- La puissance du test est aussi positivement corrélée à la valeur du seuil de significativité, c'est-à-dire que plus il y a d'erreurs de première espèce, moins il y a d'erreurs de deuxième espèce pour une même population.

Pour augmenter à la fois le seuil de confiance $1 - \alpha$ (donc diminuer le seuil de significativité α) et la puissance du test $1 - \beta$, il faut augmenter la taille de l'échantillon, mais la population en elle-même aura toujours une erreur liée à ses données, et le test d'hypothèses une erreur liée à sa nature. De plus, l'augmentation de la taille de l'échantillon est parfois limitée par des caractéristiques physiques (nombre d'individus disponibles), techniques (temps de la reproduction) ou financières (coût de séquençage d'ADN).

Souvent, les études médicales et biologiques donnent la **sensibilité** et la **spécificité** d'un test. La sensibilité est la capacité à mesurer la proportion de vrais positifs si H_0 est vraie, alors que la spécificité mesure la proportion de vrais négatifs si H_0 est rejetée :

$$- \text{ Sensibilité } = \frac{\text{Vrais positifs}}{\text{Vrais positifs} + \text{Faux négatifs}} = 1 - \alpha$$

$$- \text{ Spécificité } = \frac{\text{Vrais négatifs}}{\text{Vrais négatifs} + \text{Faux positifs}} = 1 - \beta$$

Cas d'utilisation de la puissance statistique sur les autotests COVID-19

Pour qu'un autotest COVID-19 soit autorisé, la HAS (Haute Autorité de Santé) impose qu'il soit en mesure de détecter la présence du virus chez une personne effectivement malade dans 80 % des cas a minima, et qu'ils soient en mesure de constater l'absence de virus chez une personne effectivement non malade dans plus de 99 % des cas.

Un nouvel autotest sort sur le marché et communique les résultats suivants :

	COVID détecté par PCR	COVID non détecté par PCR
COVID détecté par autotest	1236	35
COVID non détecté par autotest	6	1649

L'autotest détecte la présence de virus dans $\frac{1236}{1236+6} = \frac{1236}{1242} = 0,995 = 99,5\%$ des cas, donc dans plus de 99 % des cas confirmés par PCR.

L'autotest détecte l'absence de virus dans $\frac{1649}{1649+35} = \frac{1649}{1684} = 0,979 = 97,9\%$ des cas, donc dans plus de 97 % des cas infirmés par PCR.

Bien que ce test semble très efficace, il ne respecte pas les conditions définies par la HAS et ne pourra pas être mis sur le marché. En effet, le seuil de détection d'absence de virus est inférieur à l'attendu de 99 %.

Les tests d'hypothèses peuvent être utilisés pour valider une technique ou un outil (ici, un autotest COVID), le respect de norme de production, mais aussi la comparaison entre groupes de données. Plus d'exemples sont fournis dans les chapitres La comparaison à une valeur théorique et La comparaison de deux groupes.

Le seuil de significativité et la puissance du test peuvent être fixés en amont grâce à des données fictives pour connaître le nombre minimal d'individus statistiques nécessaires à la réalisation de l'étude. Dans ce cas, la puissance statistique minimale est généralement supérieure à 80 %, voire à 90 %.

Calcul de la puissance de deux tests

Deux tests A et B sont envisagés pour analyser les prélèvements. Le risque de première espèce est fixé à $\alpha = 0,05$ et le risque de deuxième espèce est calculé : 0,21 pour A et 0,14 pour B. Quel est le meilleur test ?

Il faut commencer par calculer la puissance de chacun des tests : $P_A = 1 - \beta_A = 1 - 0,20 = 0,80$ et $P_B = 1 - \beta_B = 1 - 0,14 = 0,86$. Pour une même erreur de première espèce α , la puissance est plus grande (ou l'erreur de deuxième espèce β est plus petite) pour le test B, donc il est meilleur.

■ Remarque

Le seuil de significativité revient à rejeter une partie des données vraies mais peu courantes (cf. figure 04-01).

2. Démarche du test d'hypothèses

La démarche du test d'hypothèses est toujours la même, quel que soit le test d'hypothèses réalisé :

– 1^{re} étape : Définir les hypothèses.

Cas de l'autotest COVID : l'autotest permet de déterminer la présence de COVID ou l'autotest permet de déterminer l'absence de COVID. Afin de minimiser l'erreur de première espèce en suivant les attendus de la Haute Autorité de santé, l'autotest permet de déterminer l'absence de COVID.

- 2^e étape : Formaliser les hypothèses.

Cas de l'autotest COVID :

H_0 = La COVID n'est pas présente dans l'échantillon.

H_1 = La COVID est présente dans l'échantillon.

- 3^e étape : Définir le seuil de significativité (α) et la puissance minimale acceptée ($1 - \beta$). Il faut aussi définir le type de test : bilatéral, unilatéral à droite ou à gauche. Le type de test est aussi dépendant de la statistique de test choisie.

Cas de l'autotest COVID :

$1 - \alpha = 0,99$, donc $\alpha = 0,001$.

$1 - \beta = 0,080$, la puissance minimale acceptée.

- 4^e étape : Calculer la **statistique de test** et la probabilité que H_0 soit vraie en fonction des données et des hypothèses définies pour savoir si l'hypothèse nulle est rejetée et/ou calculer la sensibilité et la spécificité pour valider le test réalisé.

Cas de l'autotest COVID : la sensibilité du test clinique est égale à 0,979 alors que la spécificité est égale à 0,995.

- 5^e étape : Conclure en comparant la statistique de test et la valeur critique ou la probabilité calculée et le seuil de significativité (α). Si $p < \alpha$, alors H_0 est rejetée, sinon elle est conservée.

Cas de l'autotest COVID : la sensibilité du test (98 %) est inférieure au seuil de significativité imposé (99 %). L'autotest ne permet pas de détecter de façon fiable l'absence de la COVID.

La statistique de test est calculée en fonction du type et du nombre de données et des hypothèses. Il existe deux grandes catégories de tests : **les tests d'hypothèses paramétriques et non paramétriques**. Les tests d'hypothèses paramétriques s'appuient sur une loi de probabilités, comme la loi normale, la loi de Poisson... ce qui permet une meilleure sensibilité et une plus grande puissance, mais oblige à respecter certaines conditions. Les tests d'hypothèses non paramétriques ne nécessitent pas d'hypothèse sur les lois de probabilités, ils sont donc moins restrictifs, mais aussi moins sensibles et moins puissants.

Les données vraies, mais peu courantes, rejetées par le seuil de significativité peuvent être distribuées dans les valeurs extrêmes (test bilatéral, A sur la figure 04-01) ou uniquement dans les valeurs minimales (test unilatéral à gauche, B sur la figure 04-01) ou uniquement dans les valeurs maximales (test unilatéral à droite, C sur la figure 04-01). Les droites verticales en traitillés représentent les valeurs critiques, c'est-à-dire les valeurs déterminées par le seuil de significativité et le type de test.

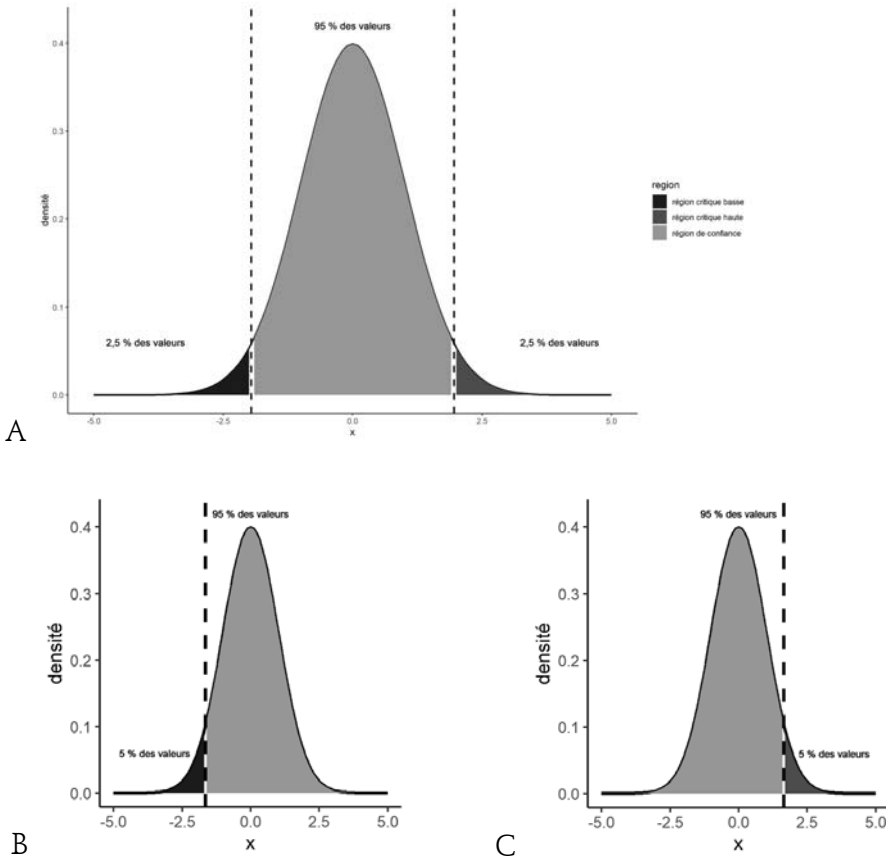


Figure 04-01 : Test d'hypothèse (ici basé sur une loi normale centrée réduite pour l'exemple) bilatéral, (A), unilatéral à gauche (B) et unilatéral à droite (C). Le seuil de significativité (α) est fixé à 5%.