♦ **Remarque**
La valeur de α doit être choisie *a priori*, jamais en fonction des données observées.

5·6·2 Le risque de deuxième espèce

Le rejet de l'hypothèse H_0 se fait au bénéfice d'une autre hypothèse, dite *hypothèse alternative*, que nous noterons H_1.

Comme sous H_0, la statistique étudiée suit sous H_1 une distribution théorique, centrée sur une valeur δ et dont les caractéristiques peuvent être étudiées.

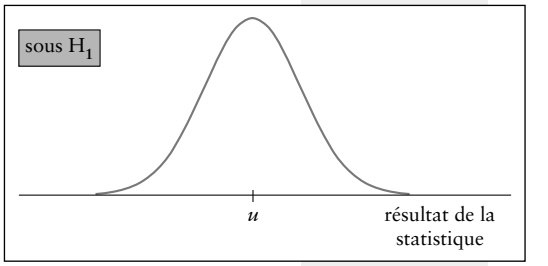


Figure 5.12 *Distribution de la statistique étudiée* U *sous* H_1.

Une fois les données observées, une autre erreur peut être commise si le résultat de la statistique du test de comparaison appartient à la zone d'acceptation de H_0 alors que H_0 est fausse (H_1 est vraie). Ce risque, noté β, est représenté par la surface en gris foncé sur la figure 5.13.

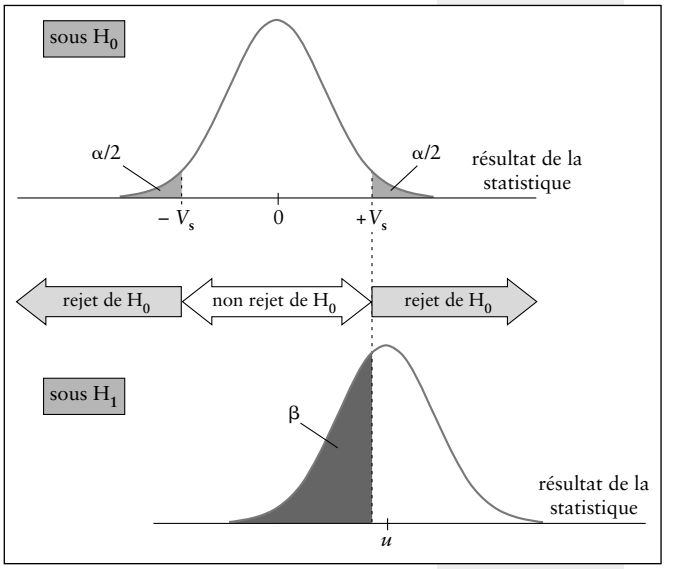


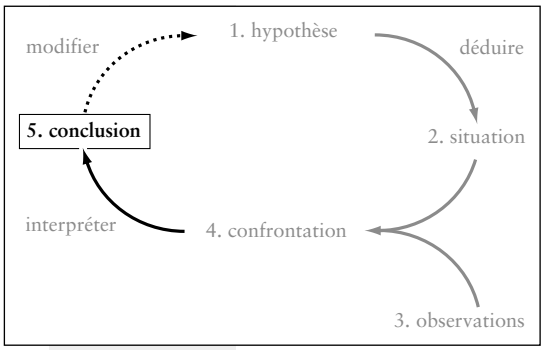
Figure 5.13 *Types d'erreurs possibles dans un test.*

Le *risque de deuxième espèce* β (ou *erreur de deuxième espèce* ou *risque d'erreur* β) est la probabilité de ne pas rejeter H_0 quand H_0 est fausse :

$$\beta = P\,(U < |V_s| \;/\; H_1 \text{ vraie})$$

Au total, une décision d'acceptation ou de rejet d'une hypothèse est toujours prise avec incertitude (la réalité n'est pas connue). Ces différentes situations décisionnelles sont représentées dans le tableau ci-dessous.

		réalité (inconnue)	
		H_0 vraie	H_0 fausse
décision retenue au vu du résultat de la statistique	H_0 crédible → non rejet de H_0	pas d'erreur	**risque** β
	H_0 non crédible → rejet de H_0	**risque** α	pas d'erreur



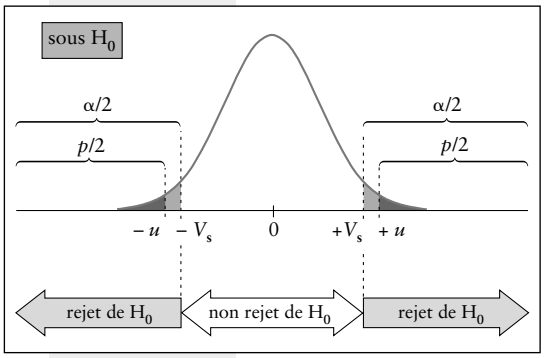


Figure 5.14 *Distribution normale centrée réduite sous H_0. Les zones de rejet de H_0 correspondant au risque de première espèce α sont repérées en gris clair. Le degré de signification* p *est repéré en gris foncé (situation où* $p \leqslant \alpha$*).*

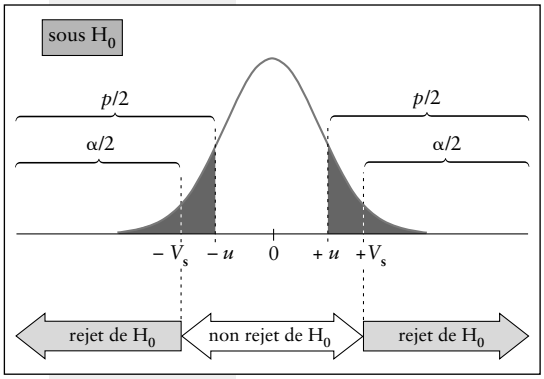


Figure 5.15 *Distribution normale centrée réduite sous H_0. Le degré de signification* p *est repéré en gris foncé (situation où* $p > \alpha$*).*

5·7 Interprétation et conclusion

Le résultat d'un test statistique s'interprète donc en fonction de la valeur seuil $|V_s|$ choisie.

Premier cas : $|u| \geqslant |V_s|$

On a :
- $p \leqslant \alpha$;
- on rejette H_0 au risque α de le faire à tort ;
- il existe une différence statistiquement significative entre l'échantillon et la population de référence ; il est souvent écrit : « le test est significatif ».

Le résultat de la procédure de rejet s'exprime avec un risque d'erreur α fixé *a priori*. Le degré de signification *p* permet de quantifier l'incompatibilité de l'observation avec l'hypothèse nulle.

Second cas : $|u| < |V_s|$

On a :
- $p > \alpha$;
- on ne rejette pas H_0 au risque β de le faire à tort ;
- on ne met pas en évidence de différence statistiquement significative entre la population d'où est issu l'échantillon et la population de référence ; il est souvent écrit : « le test n'est pas significatif ».

Dans l'exemple du paragraphe 5.5, nous avons calculé, à l'aide de l'échantillon de 30 sujets, $z = 9{,}86$. La valeur seuil pour $\alpha = 5\ \%$ est $V_s = 1{,}96$. Alors :
- $\overline{x} \approx \mu$;
- on rejette H_0 au risque $\alpha = 5\ \%$ de le faire à tort ;
- il existe une différence statistiquement significative entre l'échantillon et la population de référence (c'est-à-dire le tour de taille moyen de la population française en 2006 est différent de la valeur de référence 84,6 cm) ;
- cette différence est dans le sens d'un tour de taille moyen plus grand en 2006 qu'il ne l'était en 1997.

♦ **Remarques**

– Attention, un test statistique rejette (ou ne rejette pas) *uniquement* l'hypothèse nulle spécifiée. Il faudra donc toujours expliciter H_0 avant de faire le test et s'y reporter pour interpréter le résultat du test. De plus, nous avons dit qu'à une hypothèse nulle donnée étaient associées un certain nombre de suppositions en fonction du test statistique choisi, ou *conditions d'application* (qui seront spécifiées lors de la présentation de chaque test statistique). En effet, dans notre exemple, nous avons supposé que la distribution du tour de taille moyen était normale, ce qui nous a permis d'utiliser le test Z (ou test de l'écart réduit). Une autre condition nécessaire pour ce test est l'indépendance des sujets. Par la suite, et pour chaque test statistique, il faudra vérifier les conditions d'application.

► Paragraphe 5.11

– Il existe souvent des erreurs d'interprétation du degré de signification p. L'erreur la plus courante est d'oublier que p est calculée si H_0 est vraie. Il s'agit donc d'une probabilité conditionnelle. Une autre erreur fréquente consiste à confondre signification *statistique* et signification *biologique*. En effet, un résultat statistique significatif doit être cohérent avec les connaissances biologiques sur le sujet. De plus, l'importance de la signification statistique ne doit pas être confondue avec l'importance de l'effet étudié : p représente la force de conviction que l'observation soit incompatible avec l'hypothèse nulle ; il ne préjuge pas de l'importance de l'effet étudié. Par exemple, un tour de taille moyen observé de 84,65 cm peut être (fortement) significativement différent ($p = 0{,}0001$) de la valeur de référence 84,6 cm, bien que cet écart en cm soit très faible (0,05 cm). Inversement, un tour de taille moyen observé de 90 cm, par rapport à la même valeur de référence, ne préjuge pas du degré de signification. Il se peut aussi que p soit petit par hasard et même par erreur de première espèce. Enfin, il est également fréquent de conclure à tort que l'hypothèse nulle est vraie lorsque le test n'est pas significatif. Cela dépend, entre autre, de la puissance du test. Les assertions suivantes sont donc fausses :

► Paragraphe 5.8

  - p est la probabilité que H_0 soit vraie ;
  - $1 - p$ est la probabilité que H_1 soit vraie ;
  - p est la probabilité que les résultats sont dus au seul hasard ;
  - si $p > 5\,\%$, alors H_0 est vraie ;
  - plus p est petit, plus l'effet du facteur étudié est important.

– Il ne faut pas oublier que, pour des mécanismes complexes liés au vivant, de nombreux facteurs connus ou non entrent en jeu. Une grande part de ces mécanismes restant inconnue, il est souvent nécessaire de faire des suppositions, que l'on ne peut pas toujours vérifier. Il ne suffit donc pas d'appliquer des formules mathématiques, il faut aussi que les conclusions que l'on en tire soient réellement en cohérence avec l'ensemble des connaissances scientifiques.

5·8 Notion de puissance

Lorsque la conclusion d'un test d'hypothèse revient à ne pas rejeter H_0, il est coutume de dire que, par exemple dans le cadre de l'étude de l'efficacité d'un traitement, « l'activité du traitement n'a pas été démontrée », ou que « l'effet favorable du traitement n'a pas été mis en évidence ». D'une manière assez implicite, cette conclusion s'interprète compte tenu de l'expérience qui a été effectuée. Elle n'a de sens que si nous pouvons envisager qu'une autre expérience aurait pu mettre en évidence une différence si elle existe, c'est-à-dire que le test aurait rejeté H_0 alors qu'elle est fausse.

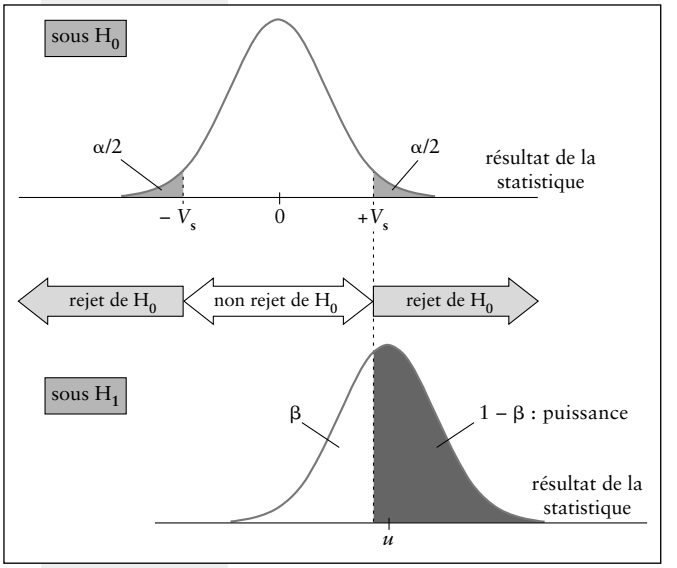


Figure 5.16 *Puissance d'un test.*

La probabilité de rejeter H_0 quand H_0 est fausse (H_1 vraie) s'appelle la *puissance d'un test* :

$$1 - \beta = P\,(U \geqslant |V_s| \ / H_1 \text{ vraie})$$

La puissance d'un test se définit sous H_1. C'est la valeur complémentaire à 1 du risque de deuxième espèce : puissance $= 1 - \beta$ (c'est la surface en gris foncé sur la figure 5.16).

5·9 Les variations de la puissance d'un test statistique

Exprimée simplement, la puissance d'un test d'hypothèses est la « probabilité de discerner une différence lorsqu'elle existe ». Nous pouvons alors facilement concevoir qu'un chercheur souhaite entreprendre son étude dans les meilleures conditions, c'est-à-dire avec une puissance *a priori* maximale.

Cependant, le calcul de la puissance d'un test d'hypothèse est relativement complexe car il fait intervenir différents paramètres statistiques. Nous allons présenter les variations de la puissance en fonction de certains de ces paramètres statistiques.

5·9·1 La variation de la puissance en fonction de α

Supposons que le même test d'hypothèse soit réalisé, d'une part, au risque α_1 et, d'autre part, au risque α_2, avec $\alpha_2 < \alpha_1$. Alors V_{s1} est la valeur seuil déterminée avec le risque α_1 et V_{s2} est la valeur seuil déterminée avec le risque α_2 et $V_{s2} > V_{s1}$.

Par exemple, si la loi de la statistique étudiée est la loi normale réduite, $\alpha_1 = 5\ \%$ donne $V_{s1} = 1{,}96$ et $\alpha_2 = 1\ \%$ donne $V_{s2} = 2{,}58$.

Toutes choses égales par ailleurs, la diminution du risque α augmente la valeur seuil, entraînant de ce fait une diminution de la puissance du test $(1 - \beta)$.

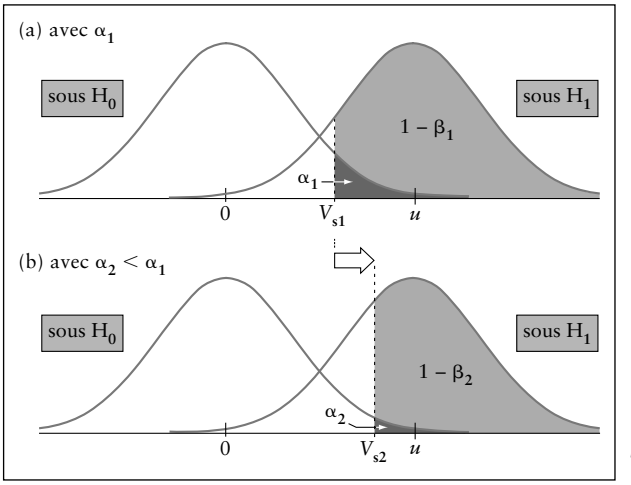


Figure 5.17 *Variation de la puissance en fonction de α.*

Toutes choses égales par ailleurs, la puissance et α varient dans le même sens.

5·9·2 La variation de la puissance en fonction de la taille de l'échantillon

► Chapitre 4

La précision d'une estimation augmente avec la taille de l'échantillon (la variabilité diminue lorsque la taille de l'échantillon augmente). D'une manière graphique, cela se traduit par un resserrement de la courbe de distribution autour de la valeur estimée.

Ainsi, sous H_0, la courbe de la distribution théorique d'un échantillon de taille $n_2 > n_1$ se resserre autour de son paramètre central et, pour conserver un risque $\alpha = 5\ \%$, la valeur seuil du test diminue. La courbe de la distribution sous H_1 se resserrant également du fait de l'augmentation de n, il s'ensuit une augmentation de la puissance (diminution de β).

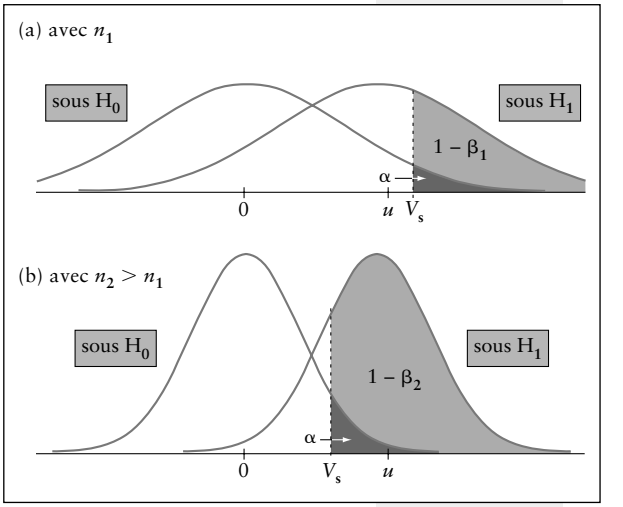


Figure 5.18 *Variation de la puissance en fonction de la taille de l'échantillon.*

Toutes choses égales par ailleurs, la puissance et n varient dans le même sens.

5·9·3 La variation de la puissance en fonction de l'écart des paramètres sous H_0 et sous H_1

Plus un test est puissant, et plus petit peut être l'écart entre le paramètre sous H_0 et celui sous H_1.

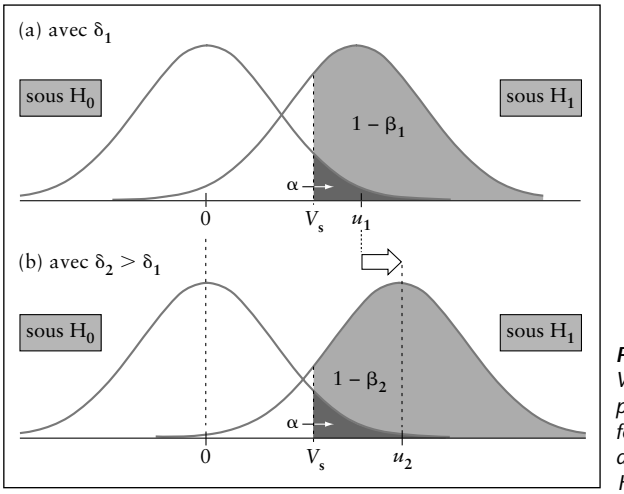


Figure 5.19
Variation de la puissance en fonction de l'écart des paramètres sous H_0 et sous H_1.

Toutes choses égales par ailleurs, la puissance et l'écart entre H_0 et H_1 varient dans le même sens.

5·9·4 La variation de la puissance en fonction de la variance

Enfin, la puissance d'un test varie également en fonction de la variance de la population (ce qui est plus difficile à représenter sous la forme de schéma). En effet, si la variance est petite, il est plus facile d'attribuer un écart entre H_0 et H_1 à un effet réel plutôt qu'à une erreur d'échantillonnage. Plus la variabilité (et donc la variance) sera grande, et plus la puissance sera faible.

♦ **Remarque**

Bien que cela puisse être difficile ou parfois peu pertinent en pratique, la variabilité peut être réduite en limitant l'étude à une sous-population (par exemple, étude limitée à un échantillon d'hommes afin de supprimer la variabilité due au sexe).

Toutes choses égales par ailleurs, la puissance et la variabilité (et donc la variance) varient en sens inverse.

5•10 Test unilatéral et test bilatéral

En début de ce chapitre, nous avons donné différents exemples concernant la formulation de l'hypothèse nulle H_0. Cependant, pour une même hypothèse H_0, il est possible de considérer deux situations permettant de définir un type d'hypothèse alternative H_1.

Exemple

On s'intéresse, d'une part à la durée moyenne vraie de séjour en réanimation après une chirurgie abdominale μ_1 et, d'autre part à la durée moyenne vraie de séjour en réanimation après une chirurgie vasculaire μ_2. On a prélevé au hasard, parmi les dossiers d'un service de réanimation, $n_1 = 300$ dossiers de malades réanimés après une chirurgie abdominale et $n_2 = 300$ dossiers de malades réanimés après une chirurgie vasculaire.

La question que l'on se pose est de savoir si la différence $m_1 - m_2$ observée sur les deux échantillons pris au hasard est significative. Dépend-elle des fluctuations d'échantillonnage ou d'une autre cause ?

L'hypothèse H_0 est $\mu_1 = \mu_2$ (ou $D = \mu_1 - \mu_2 = 0$), aux fluctuations aléatoires près.

Si nous pensons que la durée moyenne vraie de séjour en réanimation après une chirurgie abdominale est supérieure à celle après chirurgie vasculaire, alors H_1 est $\mu_1 > \mu_2$ (ou $D = \mu_1 - \mu_2 > 0$), aux fluctuations aléatoires près. Dans cette situation, l'hypothèse H_1 est formulée de manière unilatérale. Il faut donc avoir des raisons de penser que la durée moyenne vraie de séjour en réanimation après une chirurgie abdominale est supérieure à celle faisant suite à une chirurgie vasculaire. Nous avons affaire à un test d'hypothèses unilatéral.

En revanche, si nous pensons que les durées moyennes vraie de séjour sont différentes, mais sans présumer du sens de cette différence ($\mu_1 > \mu_2$ ou $\mu_2 > \mu_1$), alors H_1 est $\mu_1 > \mu_2$ ou $\mu_1 < \mu_2$ ($D = \mu_1 - \mu_2 > 0$ ou $D = \mu_1 - \mu_2 < 0$), aux fluctuations aléatoires près. Dans cette situation, l'hypothèse H_1 est formulée de manière bilatérale. Nous avons affaire à un test d'hypothèses bilatéral.

5•10•1 Le test bilatéral

Cette situation se reconnaît par des affirmations symétriques comme « différent de », « change en mieux », « change en pire », etc. On veut savoir si les deux paramètres d'intérêt sont statistiquement différents ou non, sans s'occuper de savoir lequel est supérieur à l'autre ($H_1 : \mu_1 > \mu_2$ ou $\mu_1 < \mu_2$). Dans cette situation, sous H_0, le risque erreur α doit être réparti de part et d'autre de 0 afin de tenir compte du signe de la différence ($H_1 : D = \mu_1 - \mu_2 > 0$ ou $D = \mu_1 - \mu_2 < 0$).

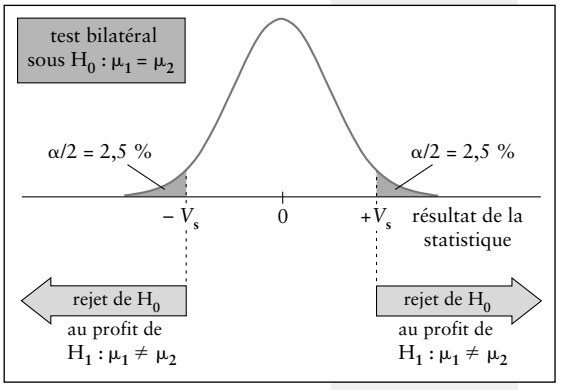


Figure 5.20 *Test bilatéral.*