Bio 2041

L'inférence statistique: les tests d'hypothèse[*]

Pierre Legendre
Département de sciences biologiques
Université de Montréal

Octobre 1991
Mises à jour: 1992, 1993, 1996, 1997, 1998, 1999, 2000, 2001, 2003, 2005

1 – À quoi servent les tests d'hypothèses statistiques?

Dans les sciences de la nature, le chercheur est appelé à prendre des décisions sur la base de résultats expérimentaux, en étant conscient qu'il y a un risque d'erreur lié à l'incertitude des observations ou des résultats expérimentaux. Avant de prendre une telle décision, il testera une hypothèse statistique *correspondant à son problème biologique*. L'issue de ce test statistique indiquera quelle décision biologique il convient de prendre.

2 – De l'hypothèse biologique à l'hypothèse statistique

Qui dit hypothèse biologique dit qu'il y a au moins deux solutions possibles, mutuellement exclusives, entre lesquelles il faut décider. Le chercheur qui désire prendre sa décision à la lumière des données provenant de l'observation ou de l'expérimentation peut avoir recours à un test statistique pour l'aider dans sa démarche, puisque dans bien des cas les résultats des expériences ou des observations ne seront pas totalement univoques, attendu les variations expérimentales importantes que connaît le matériel biologique.

• On peut tester par exemple l'hypothèse suivante (1): "La survie des alevins de truites est la même dans les eaux acides et dans les eaux bien tamponnées". On cherche à savoir si cette hypothèse est supportée ou contredite par les observations.

• Exemple (2): On peut tester l'hypothèse selon laquelle "Le taux d'amylase dans le sérum est le même chez les hépatiques et chez les sujets normaux" et on devra conclure si les données supportent cette hypothèse, ou non.

---

[*] Ce texte est en partie une adaptation des sections 3 et 4 de Legendre (1988).

L'hypothèse nulle dans un test statistique

Situation: La *pantoufle de vair* fait à Cendrillon.

Hypothèses: H_0: la personne testée est Cendrillon (test de conformité).

H_1: la personne testée n'est pas Cendrillon.

Test: On fait essayer la pantoufle à une personne. Lui fait-elle?

Issue du test

Si la pantoufle ne fait pas, on peut rejeter H_0 au profit de H_1. Cette personne n'est pas Cendrillon. Notons que le résultat du test ne donne aucune indication quant à l'identité de la personne.

Si la pantoufle fait au pied, on ne peut pas rejeter H_0. On ne peut pas non plus accepter H_0 pour les raisons suivantes:

• Manque de spécificité de H_0: d'autres personnes sont susceptibles de chausser la même pointure que Cendrillon; en particulier, la soeur jumelle de Cendrillon qui demeure à côté…

• Manque de puissance du test: la personne qui fait subir le test peut affirmer que la pantoufle fait même si elle est un peu trop petite ou trop grande pour le pied testé. En d'autres termes, le test peut ne pas être suffisamment sensible à de petites différences.

Donc, si le test est positif, on ne peut pas affirmer que la personne est Cendrillon.

• Exemple (3): On désire tester l'hypothèse selon laquelle "Le taux de cholestérol n'est pas relié à l'absorption de graisses non saturées par l'alimentation". Ici encore, on désire savoir si cette hypothèse est supportée ou contredite par les données.

Dans tous ces cas, l'hypothèse à tester formule la *négation* d'une relation: pas de différence de taille, pas de corrélation, pas d'écart entre distribution théorique et distribution observée (ce dernier exemple est un tests de conformité, "*test of goodness-of-fit*"). Comme nous le verrons plus loin, la valeur obtenue de la statistique-test sera toujours comparée à la distribution des valeurs que l'on pourrait obtenir si cette hypothèse "nulle" était vraie, prenant en compte les fluctuations normales de l'échantillonnage. Cette *hypothèse nulle* est aussi appelée l'*hypothèse principale*, ou encore "H_0". C'est elle que l'on rejettera ou non.

Si on ne peut pas rejeter l'hypothèse principale, cela ne signifie pas que celle-ci se trouve confirmée. Il y a deux raisons à cela.

• La première est qu'il peut y avoir d'autres raisons que celle explicitée dans l'hypothèse principale pour que les données soient conformes aux prédictions de H_0. Le test peut ne pas être *spécifique* à H_0. Voir l'exemple dans l'encadré.

• La seconde est que le test peut manquer de *puissance*. Il peut arriver en effet que l'hypothèse contraire (paragraphe suivant) soit vraie (i.e., un effet existe) mais que le dispositif expérimental ou le plan d'échantillonnage ne nous ait pas permis de le détecter.

3 – L'hypothèse principale et l'hypothèse contraire *Logique*

Si les résultats expérimentaux sont tels qu'ils nous forcent à rejeter l'hypothèse nulle, nous pourrons alors accepter que son contraire est au moins plausible. C'est l'*hypothèse contraire* qui correspond à la relation biologique causale que le biologiste avait en tête lorsqu'il a planifié son expérimentation ou ses observations. Celle-ci est également appelée "l'hypothèse alternative" (anglicisme), la contre-hypothèse, ou encore "H_1".

Attention: en statistique, on ne peut pas prouver que l'hypothèse H_1, qui représente le contraire de l'hypothèse principale, est vraie. En effet, on ne peut pas accepter l'hypothèse contraire comme conséquence logique du rejet de l'hypothèse principale — sauf lorsque l'hypothèses contraires est très générale, comme dans l'encadré de la page 2. Dans la plupart des cas, l'hypothèse contraire se réfère à un mécanisme particulier. De rejeter l'hypothèse nulle ne permet pas d'affirmer que le mécanisme invoqué dans l'hypothèse contraire a été démontré. Les observations peuvent en effet résulter d'un mécanisme d'action différent de celui qui est invoqué dans l'hypothèse contraire.

• Exemple (1) ci-dessus: si on devait rejeter H_0, cela signifie-t-il que c'est le degré d'acidité de l'eau qui détermine la survie des alevins de truites? Peut-être est-ce dû aux différences d'abondance de nourriture dans les deux types de lacs, à la présence de parasites, etc.

Enfin, plusieurs hypothèses "contraires" (souvent une infinité) peuvent correspondre à une même hypothèse principale (voir section 4). En voici un exemple:

• Exemple (2) ci-dessus: si on devait rejeter H_0, cela signifie-t-il que le taux d'amylase sérique des hépatiques est plus grand que celui des sujets normaux? Qu'il est inférieur? De combien plus grand ou plus petit?

Dans tous ces cas *on ne teste que l'hypothèse principale* (que l'on rejettera ou non). Pour résumer, un rejet de H_0 ne nous autorise pas à accepter l'une des solutions "contraires" plutôt qu'une autre, ni à accepter que le mécanisme invoqué dans la formulation de l'hypothèse contraire soit responsable de la relation observée.

4 – Test unilatéral et test bilatéral *Biologie*

C'est la formulation de l'hypothèse contraire, H_1, qui détermine si un test est *unilatéral* ou *bilatéral.*

Dans le cas de la comparaison de deux groupes par exemple, on réalise un *test bilatéral* (“two-tailed test”) lorsqu’on cherche une différence entre ces groupes — ou encore entre une estimation et une valeur donnée par hypothèse — sans se préoccuper du signe de la différence. Dans le cas où l’on étudierait s’il existe une relation entre deux variables, le test est bilatéral si on cherche simplement à démontrer que ladite relation existe, sans égard au signe de la corrélation. Le terme de *bilatéral* vient du fait que lorsque nous comparerons la valeur observée de la statistique-test à la distribution des valeurs qu’elle aurait pu connaître si H_0 était vraie, nous considérerons les deux extrémités de cette distribution comme des zones de rejet de H_0 (section 8).

On réalise plutôt un *test unilatéral* (“one-tailed test”) lorsque notre hypothèse biologique ne s’intéresse qu’aux différences ou aux relations ayant un signe donné. Les tests de ce type sont dits *unilatéraux* parce que la zone de rejet de H_0 se situe à une seule extrémité de la distribution de probabilités qui nous sert de référence (section 8).

• Par exemple, si on teste un nouveau médicament, on n’enclenchera le processus de mise en marché que si celui-ci s’avère significativement *meilleur* que son concurrent immédiat; s’il est moins bon (différence significative, mais dans une direction non souhaitée), ce résultat ne doit pas être considéré comme probant ou intéressant. Il en serait de même pour un autre médicament qui augmenterait la survie des patients moins que son concurrent, lorsque comparé à un placebo.

• Dans une étude de l’influence de l’acidité des pluies sur le rendement des érablières, ce n’est que si on peut prouver que la forêt *dépérit* avec l’acidification que nous pourrons demander au législateur de traduire en des lois antipollution ses politiques générales du maintien de la qualité de l’environnement. Si l’acidité des pluies n’a aucun effet sur la forêt, ou encore si l’acidification améliore leur état de santé, tout va bien.

Il est donc évident que dans tous ces cas, c’est le biologiste qui décide s’il convient de tester de façon unilatérale ou bilatérale. Certains tests particuliers (test F en analyse de variance, test χ^2) sont cependant toujours unilatéraux, de par leur construction particulière.

5 – Tester l’hypothèse principale *Statistique*

Étant conscients de la variabilité naturelle, inhérente en particulier au matériel biologique, on soumet l’hypothèse principale (H_0) à une épreuve de vérité afin de déterminer si les résultats obtenus expérimentalement sont conformes, ou non, à ce que laisserait prévoir l’hypothèse principale, en laissant place aux erreurs dues à l’échantillonnage ou à l’expérimentation. Il est important de se rappeler constamment que c’est uniquement l’hypothèse principale que l’on teste; si on n’y prend garde, on pourrait inverser les conclusions tirées du test d’hypothèse statistique, ce qui pourrait être désastreux…

Comment déterminer s’il faut rejeter H_0 ou pas? Prenons un exemple pour illustrer la démarche: supposons que nous sommes intéressés à déterminer s’il existe une différence significative entre deux groupes de mesures. Dans le groupe **A** se trouvent des mesures prises sur des individus ayant subi un premier type de traitement, alors que les individus dans le groupe **B** ont subi un deuxième type de traitement. Nous pouvons employer la statistique t de Student pour mesurer l’écart entre ces deux groupes; cette statistique est essentiellement la différence des moyennes, divisée par une valeur qui tient compte de la variance des deux groupes et qui rend la statistique *pivotale* (Scherrer, 1984, section 13.2).

Supposons que nous trouvions que la valeur de cette statistique t est 14. Nous savons par ailleurs, de par la logique de notre question biologique, que si les deux groupes de mesures sont très différents, alors la statistique t sera grande, avec un signe positif ou négatif selon le sens dans lequel a été calculée la différence; au contraire, si les deux traitements ont le même effet, la statistique t sera petite et ne reflétera que des variations aléatoires dues à l’échantillonnage de la population des réponses possibles au traitement. Question: la valeur $t = 14$ est-elle “grande” ou “petite”? On ne peut pas répondre à cette question dans l’absolu. On ne peut y répondre que par comparaison avec les valeurs que t pourrait prendre dans différentes autres expériences où H_0 est vraie. Nous allons donc nous demander: quelles valeurs pourrait prendre la statistique t si H_0 est vraie?

Il existe deux façons de répondre à cette question. La première consiste à générer une loi de distribution empirique de la variable auxiliaire t, en tenant le raisonnement décrit dans l’encadré de la page suivante; cette méthode est logiquement la plus simple, quoique lourde au niveau des calculs. La seconde est la méthode classique qui consiste à utiliser les lois de distribution théoriques (cours 8 à 12).

Le raisonnement de l’encadré constitue un *syllogisme hypothétique* en logique formelle. Il nous fait pénétrer au coeur du raisonnement qui mène à une décision statistique. Pour obtenir la distribution complète de la variable auxiliaire, il faudrait en principe calculer *toutes* les permutations possibles des objets entre les deux groupes. Pour deux groupes égaux de 10 objets chacun, le nombre de permutations atteint déjà $(n_1 + n_2)!/n_1!\, n_2! = 184\,756$. Pour deux groupes égaux de 12 données, il faudrait considérer à tour de rôle 2 704 156 permutations. Même avec les ordinateurs modernes, de tels calculs deviennent rapidement impossibles à réaliser. On peut cependant obtenir des résultats approximatifs tout à fait valables en réalisant seulement quelques milliers de permutations, choisies de façon aléatoire parmi toutes les permutations possibles (habituellement de 1000 à 10 000). Les calculs deviennent alors réalisables même par des micro-ordinateurs. Edgington (1995) et Manly (1997) montrent comment réaliser les tests statistiques courants par cette méthode, dite *méthode des permutations.* (Un autre exemple de test par permutation, utilisant la corrélation linéaire de Pearson, est développé à la section 1.2.3 de Legendre & Legendre, 1998.)

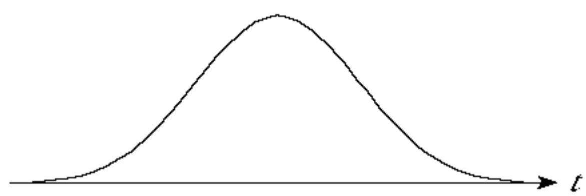
Hypothèses

H_0: les deux groupes de mesures ne sont que des variations aléatoires autour d'une même moyenne commune.

H_1: les deux traitements produisent des différences dans la variable qui a été mesurée (test bilatéral dans cet exemple).

Distribution de la variable auxiliaire (statistique du test); autrement dit, quelle serait la distribution de la statistique-test t si H_0 était vraie?

- Selon H_0, les différences de traitement n'ont aucun effet sur la variable mesurée; par conséquent, les valeurs observées ne sont que des variations aléatoires autour d'une moyenne commune, si bien que toute valeur observée dans un groupe aurait tout aussi bien pu être observée dans l'autre groupe.
- On obtient donc une réalisation de H_0 en permutant les valeurs entre les deux groupes.
- Pour les valeurs ainsi permutées, on calcule une valeur qu'aurait pu prendre la statistique-test t, en supposant que H_0 soit vraie.
- En répétant cette opération, les différentes permutations produisent un ensemble de valeurs de la statistique t obtenues sous H_0, qui représentent, toutes ensemble, une estimation de la distribution d'échantillonnage de t sous H_0.



Décision statistique

Dans tout test statistique, la décision statistique est prise en comparant la valeur obtenue de la variable auxiliaire (ici, la variable-test t) à la distribution d'échantillonnage obtenue sous H_0. La règle de décision suivante est employée:

- Si la valeur observée de la variable-test t est si élevée qu'elle est plus grande que la plupart des valeurs obtenues en simulant l'hypothèse H_0, ou encore si elle est si fortement négative que peu des valeurs obtenues en simulant H_0 sont aussi fortement négatives, alors on ne peut pas croire que les résultats expérimentaux sont compatibles avec l'hypothèse nulle et on rejette H_0.
- Si au contraire la valeur observée de la variable-test t se trouve vers le centre de la distribution des valeurs obtenues sous H_0, cela montre qu'une telle valeur aurait très bien pu être obtenue au hasard de l'échantillonnage d'une population ayant subi un traitement unique et on accepte donc que les données ne sont pas incompatibles avec l'hypothèse principale.

Tableau 1. Les 50 mesures (en mm) de la longueur de la rectrice centrale chez les gélinottes huppées mâles, juvéniles, ont été divisées en deux groupes les plus semblables possible.

```
Groupe 1:
149 150 152 153 153 155 156 157 158 158 158 158 159 159 160 162 162 162 163 163 164 164
165 171 174

Groupe 2:
140 150 151 152 153 154 155 156 157 158 158 158 158 159 160 160 162 162 162 163 164 164
165 166 171
```

Exemple 1 — Dans son tableau 3.1, Scherrer (1984) présente des mesures de la longueur de la rectrice centrale chez 50 gélinottes huppées mâles, juvéniles. Ces données ont été mises en ordre de taille croissante, puis attribuées en alternance à deux groupes égaux, de façon à s’assurer qu’il existe le moins de différences possible entre ces deux groupes (tableau 1).

Réalisons un test de différence entre les moyennes de ces deux groupes, pour nous assurer par exemple que le processus de construction des deux groupes a vraiment produit des groupes de moyenne à peu près égale. Les deux hypothèses sont les suivantes:

H_0: les deux groupes de mesures ne sont que des variations aléatoires autour d’une même moyenne commune.

H_1: les deux groupes correspondent à des populations ayant des moyennes différentes, peut-être à cause du processus d’attribution des mesures aux groupes (test bilatéral dans cet exemple).

La différence de moyenne des deux groupes est $(\bar{x}_1 - \bar{x}_2)$ = 1,07999 et la statistique t a comme valeur 0,62225. Pour tester l’hypothèse nulle, permutons au hasard les objets entre les deux groupes et calculons à nouveau les deux statistiques, $(\bar{x}_1 - \bar{x}_2)$ et t. La valeur originale de la statistique (pour les deux groupes réels), ainsi que quelques valeurs obtenues après permutations, sont présentées au tableau 2.

Sous l’hypothèse nulle, la valeur observée d’une statistique doit être comptée parmi la distribution d’échantillonnage obtenues sous H_0 (Hope, 1968; Edgington, 1995). Les résultats après 999 permutations aléatoires, plus les valeurs observées des statistiques, sont présentés à la figure 1. Pour pouvoir prendre la décision, la seule information importante, pour chacune des deux statistiques, est de déterminer combien des valeurs faisant partie de la distribution d’échantillonnage sont inférieures, égales, et supérieures à la valeur observée de la statistique dans chacune des deux queues de la distribution. Ces informations sont rassemblées dans le tableau suivant où *obs* désigne la valeur observée et *per* désigne une valeur obtenue après permutation:

	per < –*obs*	*per* = –*obs*	–*obs* < *per* < *obs*	*per* = *obs*	*per* > *obs*
$\bar{x}_1 - \bar{x}_2$	*235*	*16*	*459*	*15**	*275*
Statistique t	*235*	*16*	*459*	*15**	*275*

* L’une de ces valeurs est la valeur observée elle-même (*obs*); la valeur observée est positive dans cet exemple.

À l’examen de ce tableau, on voit que la valeur observée de chacune de ces statistiques est bien située près du centre de la distribution d’échantillonnage. Si on désire calculer une probabilité plus précise d’observer de telles données sous l’hypothèse nulle, on procède comme suit.

Tableau 2. Valeurs des deux statistiques, $(\bar{x}_1 - \bar{x}_2)$ et t, pour les deux groupes réels ainsi que pour neuf permutations aléatoires des objets en deux groupes égaux.

	$(\bar{x}_1 - \bar{x}_2)$	Statistique t
Valeurs pour les vrais groupes:	1,07999	0,62225
Valeurs pour groupes permutés:	–1,23999	–0,71535
	0,36000	0,20668
	–0,51999	–0,29867
	2,20001	1,28398
	1,32001	0,76205
	–0,36000	–0,20668
	–2,20001	–1,28398
	0,04001	0,02296
	–1,95999	–1,13986
	[etc.]	[etc.]

• On désire savoir si les données sont conformes à l'hypothèse nulle d'égalité des moyennes (H_0), en tenant compte des erreurs d'échantillonnage. On calcule donc la probabilité que la variable-test prenne une valeur aussi extrême ou plus extrême que la valeur absolue (test bilatéral) de la valeur observée, pour chacune des deux statistiques-test: $P(\,|\bar{x}_1 - \bar{x}_2| \geq 1{,}07999\,)$ ou $P(\,|\,t\,| \geq 0{,}62225\,)$. La probabilité recherchée est calculée comme suit: $P = (235 + 16 + 15 + 275)/1000 = 0{,}541$. Dans cet exemple, la probabilité est la même pour les deux statistiques, mais tel n'est pas toujours le cas; en général, il est préférable d'employer la statistique t (qui est une statistique pivotale) plutôt que la simple différence entre les moyennes. Pour un seuil de signification $\alpha = 0{,}05$ (test bilatéral), on constate que la probabilité $P = 0{,}541$ est de beaucoup supérieure à la valeur-seuil de probabilité $\alpha = 0{,}05$ d'où l'on conclut que les données ne sont pas en désaccord avec l'hypothèse nulle.

$\Rightarrow$ P est la probabilité que les données soient conformes à l'hypothèse nulle, notée $\Pr(x|H_0)$.

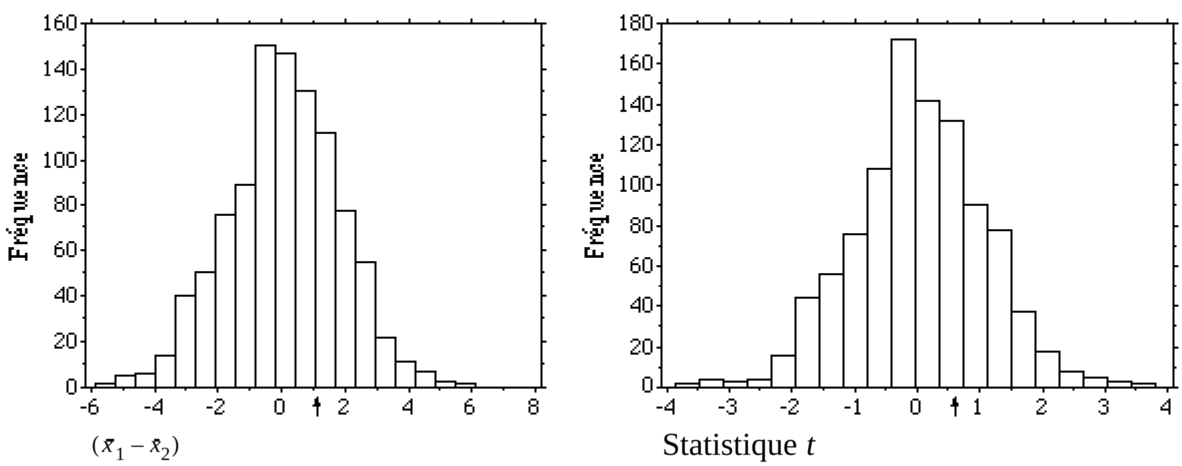


Fig. 1. Distribution d'échantillonnage des statistiques $(\bar{x}_1 - \bar{x}_2)$ et t sous H_0, estimées à l'aide de (999 + 1) permutations aléatoires. Au bas de chaque histogramme, la flèche indique la vraie valeur de la statistique.

• Pour fins de comparaison, un test *t* paramétrique classique (cours no 8) donne 0,5367 comme probabilité d’observer de telles données sous l’hypothèse nulle, pour un test bilatéral. Les deux formes de test conduisent donc à la même réponse pour cet exemple. Ceci n’est pas toujours le cas, car le test *t* paramétrique suppose que les données obéissent à des conditions de distribution (cours no 8) qui ne sont pas toujours réalisées; le test par permutation n’oblige pas de vérifier ces suppositions.

Exemple 2 — Les mêmes données sont maintenant divisées en deux groupes de la façon suivante: les 25 premières valeurs du tableau 3.1 de Scherrer (1984) forment le premier groupe alors que les 25 dernières données forment le second groupe, comme le montre le tableau 3. Les hypothèses sont les suivantes:

H_0: les deux groupes de mesures ne sont que des variations aléatoires autour d’une même moyenne commune.

H_1: les deux groupes correspondent à des populations ayant des moyennes différentes, peut-être parce que l’observateur a d’abord sélectionné les plus grosses (ou les plus petites) gélinottes (test bilatéral).

Les statistiques prennent les valeurs suivantes pour les deux groupes du tableau 3: $(\bar{x}_1 - \bar{x}_2) = -4{,}44000$ et $t = -2{,}73994$. Le tableau qui suit décrit les résultats obtenus après 999 permutations aléatoires des données entre les deux groupes:

	per < –|*obs*|	*per* = –|*obs*|	–|*obs*| < *per* < |*obs*|	*per* = |*obs*|	*per* > |*obs*|
$\bar{x}_1 - \bar{x}_2$	*6*	*1**	*991*	*0*	*2*
Statistique *t*	*6*	*1**	*991*	*0*	*2*

* Il s’agit de la valeur observée elle-même, qui est négative dans cet exemple.

Ce tableau montre que la valeur observée de chacune des statistiques est située tout près de l’extrémité gauche de la distribution d’échantillonnage. Pour calculer de façon plus précise la probabilité d’observer de telles données sous l’hypothèse nulle, on procède comme suit.

• On désire calculer la probabilité que, sous H_0, la variable-test prenne une valeur aussi extrême ou plus extrême que la valeur absolue (test bilatéral) de la valeur observée, pour chacune des deux statistiques-test: $P(|x_1 - x_2| \geq 4{,}44000)$ ou $P(|t| \geq 2{,}73994)$, pour des valeurs de $(\bar{x}_1 - \bar{x}_2) = -4{,}44000$ et $t = -2{,}73994$. La probabilité recherchée est donnée par: $P = (6 + 1 + 0 + 2)/1000 = 0{,}009$. Celle-ci est la même pour les deux statistiques, dans le cas présent, mais tel n’est pas toujours le cas. Pour un seuil de signification $\alpha = 0{,}05$ (test bilatéral), on constate que la probabilité $P = 0{,}009$ est inférieure à la valeur-seuil de probabilité $\alpha = 0{,}05$ d’où l’on conclut que les données ne supportent pas l’hypothèse nulle. Celle-ci est donc rejetée au profit de H_1, l’ordre de sélection des animaux par l’observateur n’ayant probablement pas été aléatoire.

Tableau 3. Les 50 mesures (en mm) de la longueur de la rectrice centrale chez les gélinottes huppées mâles, juvéniles. Les 25 premières données du tableau 3.1 de Scherrer (1984) forment le premier groupe, les 25 données suivantes forment le second groupe.

```
Groupe 1:
153 165 160 150 159 151 163 160 158 149 154 153 163 140 158 150 158 155 163 159 157 162
160 152 164

Groupe 2:
158 153 162 166 162 165 157 174 158 171 162 155 156 159 162 152 158 164 164 162 158 156
171 164 158
```

• Pour fins de comparaison, un test *t* paramétrique classique (cours no 8) donne 0,0086 comme probabilité d’observer de telles données sous l’hypothèse nulle, pour un test bilatéral. Les deux formes de test conduisent à la même réponse dans ce cas.

La méthode des tests par permutation avait été proposée par l’agronome R. A. Fisher dès 1935, mais les statisticiens de l’époque ne disposaient pas d’ordinateurs pour réaliser les milliers de calculs nécessaires. C’est pourquoi les statisticiens du début du siècle dernier développèrent les différentes *distributions de probabilités*, ou *lois de distribution*, qui servent plus couramment à réaliser les tests statistiques. Ceux-ci ont réussi à démontrer, avec beaucoup d’ingéniosité, que *sous certaines conditions*, certaines statistiques-test obéissent à certaines lois de distribution particulières lorsque H_0 est vraie. Par exemple,

• la variable-test *t* utilisée dans le test de comparaison des moyennes de deux échantillons obéit à une loi de Student (munie d’un certain nombre de degrés de liberté), si H_0 est vraie: Scherrer (1984), sections 13.1 et 13.2;

• la variable-test *F* utilisée dans le test de comparaison des moyennes de plusieurs groupes (analyse de variance) obéit à une loi de *F* (munie de certains degrés de liberté), si H_0 est vraie: Scherrer (1984), section 13.4;

• la variable-test *t* utilisée dans le test des coefficients de corrélation obéit à une loi de Student (munie d’un certain nombre de degrés de liberté), si H_0 est vraie: Scherrer (1984), section 17.1. Cette statistique-test *t* n’est bien sûr pas la même que celle que l’on emploie pour comparer des moyennes.

Grâce à ce subterfuge, il devient possible d’employer une loi de distribution théorique *au lieu de générer soi-même une loi de distribution empirique de la variable auxiliaire sous* H_0, ce qui épargne beaucoup de travail. La décision statistique se prend exactement de la même façon que décrit ci-dessus, c’est-à-dire en comparant la valeur observée de la statistique-test à la distribution des valeurs que l’on pourrait obtenir sous H_0; ces valeurs sont fournies par les tables des différentes lois de distribution, ou encore par des procédures informatiques capables de générer les valeurs qui se trouvent dans ces tables.

Il est fait mention ci-dessus que *sous certaines conditions*, les statistiques-test obéissent à certaines lois de distribution, lorsque H_0 est vraie. Quelles sont ces conditions? Tout dépend des tests. Dans de nombreux tests, par exemple, la loi de distribution ne s’applique au cas où H_0 est vraie que si les données sont extraites d’une population dont la distribution est *normale*; faute de pouvoir vérifier la distribution de la variable dans la population, on doit au moins tester l’hypothèse que l’échantillon que l’on étudie a vraisemblablement été tiré d’une population normale et transformer la variable si cette condition n’est pas remplie. Dans d’autres cas, comme dans l’analyse de variance ou le test *t* de la différence entre groupes, la distribution de référence ne peut strictement être employée

Tableau 4. Les deux risques d'erreur lorsqu'on réalise un test statistique.

	H_0 est vraie	H_1 est vraie
On ne rejette pas H_0	*Bonne décision*	*Erreur β (type II)*
On rejette H_0	*Erreur α (type I)*	*Bonne décision*

que si les variances des différents groupes sont égales (condition d'*égalité des variances*, ou *homoscédasticité*). Cette condition concerne également les tests par permutation.

Enfin, une condition se retrouve dans tous les tests où l'on désire avoir recours à une loi de distribution sans effectuer de correction sur les degrés de liberté: c'est la condition d'*indépendance des observations*. Cette condition signifie que chaque valeur observée doit être non influencée par les autres valeurs observées. Cette condition est rarement remplie lorsqu'on étudie des séries temporelles de données (phénomène d'autocorrélation temporelle) ou des données tirées d'un espace géographique (phénomène d'autocorrélation spatiale), puisque les observations situées à proximité les unes des autres sur une surface ou le long d'une série temporelle ont de bonnes chances d'être le résultat d'un même processus générateur. Ce problème concerne donc une partie importante des données que traitent les biologistes.

Le message à retenir est que certaines des conditions d'application particulières à chaque test résultent de notre référence aux lois de distribution; ces conditions ne sont donc pas inhérentes aux tests statistiques eux-mêmes. On peut passer outre aux conditions qui concernent la distribution des données, e.g. la normalité, en utilisant la méthode des permutations exposée plus haut. C'est la raison pour laquelle les tests par permutation connaissent présentement une grande popularité, en particulier auprès des biologistes.

6 – Les deux risques d'erreur

Si à l'issue d'un test statistique on rejette une hypothèse H_0 qui n'aurait pas dû être rejetée, on dit que l'on commet une erreur α, appelée aussi erreur de type I. Par contre, si on ne rejette pas une hypothèse H_0 qui aurait dû être rejetée, on commet une erreur β, appelée aussi erreur de type II (tableau 4).

L'erreur α est le risque de déclarer que la relation biologique existe alors qu'elle n'existerait pas (par exemple: qu'il existe des différences entre deux groupes alors qu'il n'y en a pas). On dira par exemple: “On rejette H_0 avec une probabilité d'erreur $\alpha = 5\%$”. Le chercheur connaît toujours la valeur de α puisque c'est lui qui l'a déterminée; dans bien des cas cependant, on ne connaît pas précisément β (notion de *puissance* d'un test: Scherrer, 1984, section 11.7) qui doit être déterminé par une analyse de puissance (Cohen, 1988). Un test statistique est *valide* si son erreur de type I n'est pas supérieure au seuil de signification α, et ce pour toute valeur de α (Edgington, 1995).

7 – Le seuil de signification (*“significance level”*)

En testant une hypothèse, le chercheur doit déterminer quelle est la probabilité d’erreur α qu’il est prêt à tolérer. Ce choix est arbitraire, mais on emploie la plupart du temps les seuils de signification α = 0,05 (résultat significatif), 0,01 (hautement significatif) ou 0,001 (très hautement significatif); les seuils de 5% et 1% furent proposés par Fisher (1925). Si on décide de réaliser un test au seuil α = 5% par exemple, cela signifie que l’on se donne 5 chances sur 100 de rejeter H_0 même si H_0 est vraie et ne devrait donc pas être rejetée. Ceci montre bien que le seuil α doit être choisi en fonction de la gravité des conséquences que l’on encourra si on est amené, par le test, à prendre la mauvaise décision.

• Ainsi, si des sommes d’argent importantes sont en jeu, par exemple pour le lancement d’un nouveau médicament qui serait plus efficace dans le traitement d’une maladie que les médicaments déjà sur le marché, il convient d’employer un seuil de signification α extrêmement petit lors des tests statistiques qui établissent les qualités supérieures du nouveau produit, de façon à réduire le risque de se tromper.

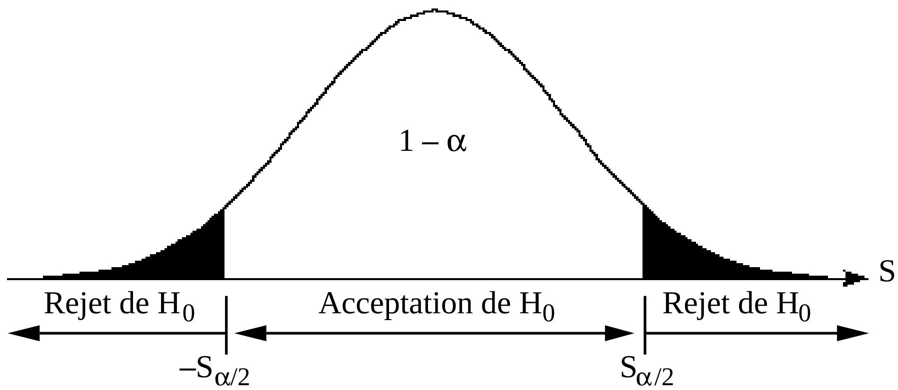
• Par contre, lors des tests d’une procédure médicale nouvelle permettant peut-être de traiter une maladie jusque là incurable, on peut se permettre une probabilité d’erreur α relativement grande, tout effet du traitement, même partiel, représentant un progrès. Une augmentation du risque d’erreur α entraîne une réduction du risque d’erreur β (Cohen, 1988).

8 – “Acceptation” et rejet de H_0

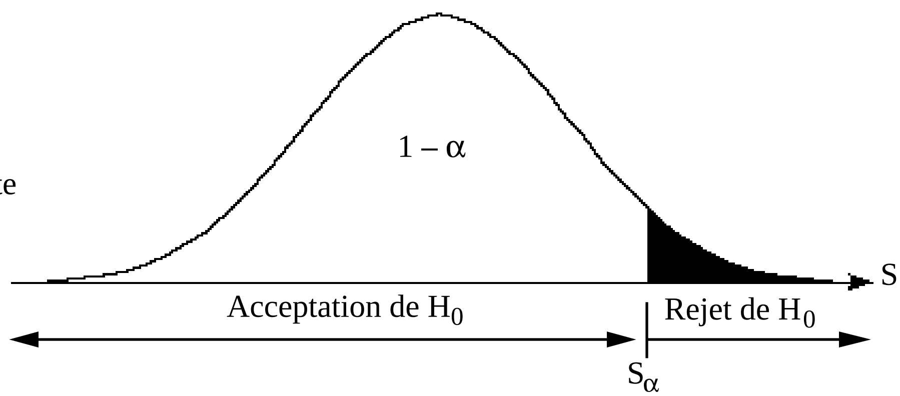
La zone dite “d’acceptation de H_0” est l’intervalle des valeurs de la statistique-test dans lequel les différences observées peuvent être attribuées aux variations dues à l’échantillonnage. La zone de rejet est au contraire la zone dans laquelle la statistique-test prend une valeur trop extrême pour qu’on puisse l’attribuer à une variation aléatoire prévisible sous l’hypothèse nulle (H_0); dans ce cas, on rejettera H_0 avec, bien sûr, un risque d’erreur égal à la valeur α choisie par le chercheur. Les figures suivantes illustrent ces concepts pour une statistique-test S quelconque; on voit clairement que les zones “d’acceptation” et de rejet ne sont pas les mêmes, selon que le test est unilatéral ou bilatéral. S_α et $S_{\alpha/2}$ sont les valeurs critiques, tirées d’une table de la loi de distribution de S (ou obtenue par permutations), pour les tests unilatéraux et bilatéraux respectivement. Dans tous les cas, $(1 - \alpha)$ est la probabilité de prendre la bonne décision si H_0 est vraie.

Nous avons vu à la section 2 qu’un test statistique ne permet jamais de conclure avec certitude que H_0 est vraie. Ce sont plutôt les conséquences, biologiques ou autres, du test statistique qu’il faudra accepter. C’est dans ce sens seulement que nous pouvons parler, par abus de langage, de la “zone d’acceptation de H_0”.

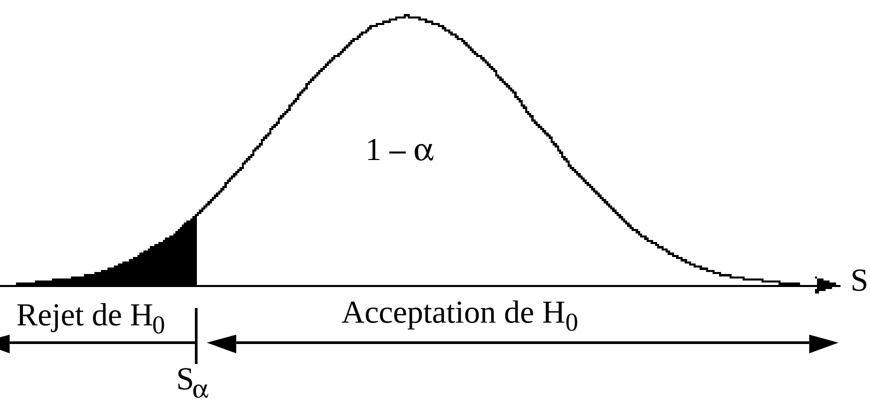
Test bilatéral
H_1: $S \geq |S_{\alpha/2}|$



Test unilatéral à droite
H_1: $S \geq S_\alpha$



Test unilatéral à gauche
H_1: $S \leq S_\alpha$



9 – Échantillons (groupes) indépendants ou appariés

La distinction entre les échantillons *indépendants* et *appariés* devient importante lors des comparaisons d'échantillons (test *t*, test *U*, analyse de variance, etc.) puisque la façon de réaliser le test varie selon l'appariement ou le non-appariement des échantillons. C'est le plan d'échantillonnage ou d'expérience qui détermine si les observations sont indépendantes ou appariées. Attention — le mot indépendant peut être employé dans au moins trois autres sens en statistique: variables indépendantes = non corrélées; variables indépendantes (= prédictives) d'un modèle de régression; observations indépendantes = non autocorrélées.

Lors de la formation des différents groupes entre lesquels on cherchera des différences, si on a sélectionné les objets devant faire partie de chacun des groupes indépendamment des objets de l'autre ou des autres groupes, on a alors constitué des échantillons indépendants ("independent samples" en anglais). Par contre, si on a prélevé (aléatoirement) des *paires*, des *triplets* ou un plus grand nombre d'objets, et non des objets séparés, on a constitué des échantillons (groupes) appariés ("related samples" ou "paired samples" en anglais, ou encore "matched pairs" ou "before-after comparisons"). Les échantillons appariés permettent de réaliser des tests statistiques plus puissants que les échantillons indépendants.

• Par exemple, si on a sélectionné systématiquement les deux conjoints de chaque demeure visitée pour former les groupes d'hommes et de femmes qui seront comparés au moyen d'un test statistique, on a constitué deux groupes appariés.

• De même, si on a visité plusieurs sites d'échantillonnage en mer et que l'on a fait des prélèvements aux mêmes profondeurs à tous ces sites, on obtient des groupes de prélèvements appariés.

• Un autre exemple courant est l'étude des mêmes sujets, avant et après une action ou un traitement; là encore, on a constitué des groupes d'observations appariées.

10 – Choix d'une méthode statistique

Le tableau 5 a pour but de vous aider à décider, parmi les tests les plus courants, quelle méthode statistique s'applique à chaque situation particulière.

Références

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences. 2nd edition.* Lawrence Erlbaum Assoc., Publ., Hillsdale, New Jersey.

Edgington, E.S. 1995. *Randomization tests. 3rd edition.* Marcel Dekker Inc., New York.

Fisher, R.A. 1925. *Statistical methods for research workers.* Oliver and Boyd, Edinburgh.

Fisher, R.A. 1935. *Design of experiments.* Oliver and Boyd, Edinburgh.

Hope, A.C.A. 1968. A simplified Monte Carlo significance test procedure. *J. Roy. Stat. Soc. Ser. B* 30: 582-598.

Legendre, P. 1988. *Statistique sur micro-ordinateur — Formation d'appoint pour scientifiques.* Centre de formation continue, Faculté de l'éducation permanente, U. de M./Association des biologistes du Québec.

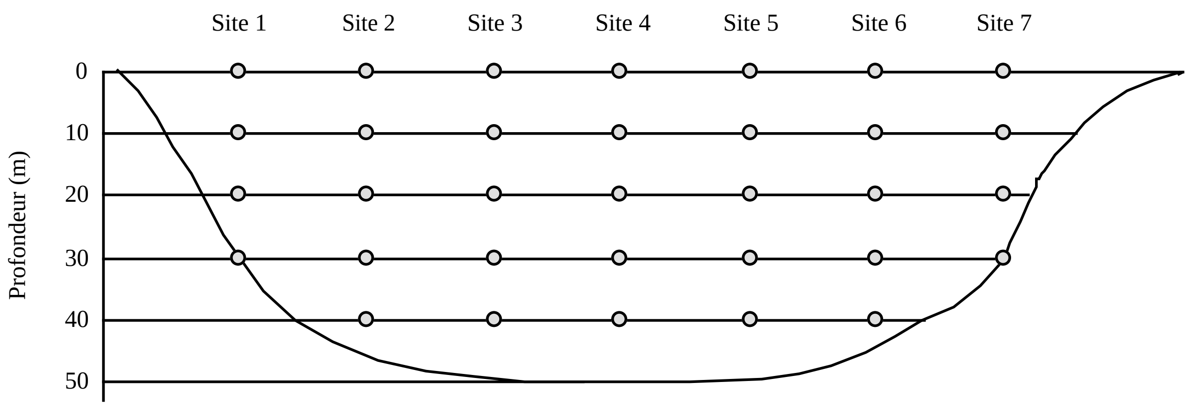
Legendre, P. & L. Legendre. 1998. *Numerical ecology. 2nd English edition.* Elsevier Science BV, Amsterdam.

Manly, B.F.J. 1997. *Randomization, bootstrap and Monte Carlo methods in biology. 2nd edition.* Chapman & Hall, London.

Scherrer, B. 1984. *Biostatistique.* Gaëtan Morin Éditeur, Chicoutimi.

Groupes (à comparer): les sites

Critère d'appariement: la profondeur



Comparaison avant-après

Groupes (à comparer): avant, après

Critère d'appariement: les individus

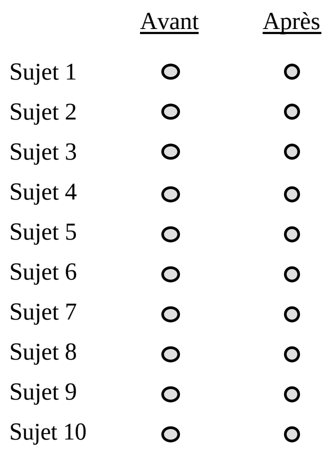


Tableau 5. Critères de choix d’une méthode statistique.

Une seule variable: recherche de différences entre des groupes

Variable quantitative

- 2 groupes: test *z* (grands groupes), test *t* (petits groupes)
- Plusieurs groupes: analyse de variance

Variable semi-quantitative, ou variable quantitative à distribution non normale

- 2 groupes: test *U* de Wilcoxon-Mann-Whitney
- Plusieurs groupes: test de Kruskal-Wallis

Variable qualitative

- 2 ou plusieurs groupes: test χ^2, test *G*

Deux variables: recherche d’une liaison entre ces deux variables

Variables quantitatives

- Recherche d’une relation linéaire: corrélation *r* de Pearson
- Équation (modèle) décrivant une relation linéaire: régression linéaire simple
- Équation (modèle) décrivant une relation non linéaire: régression polynomiale, régression non linéaire

Variables semi-quantitatives, ou variables quantitatives en relation monotone

- Recherche d’une relation monotone: corrélations non-paramétriques *r* de Spearman et τ de Kendall

Variables qualitatives, ou autres variables divisées en classes

- Recherche d’une liaison: test χ^2, test *G*

Plusieurs variables: méthodes d’analyse de données (statistique multidimensionnelle)
