Chapitre 6

Tests d'hypothèses Données quantitatives

6.1 Introduction

Une application importante de la théorie des probabilités consiste à tester des hypothèses scientifiques à partir des résultats d'expériences ou d'enquêtes. Les données issues de ces études comportent un important élément aléatoire, de sorte que certaines variations observées peuvent être purement accidentelles, sans signification aucune ; ou, au contraire, peuvent révéler des faits scientifiques ou confirmer des conjectures. Lorsqu'on conclut, par exemple, que « le médicament *Tamoxifen* est la cause de certains cancers », c'est qu'on a noté un taux de cancer anormalement élevé parmi ceux qui en prennent. Est-ce à cause du *Tamoxifen*, ou est-ce un accident? La réponse exige parfois un raisonnement complexe et sophistiqué, mais le fonds de la démarche est simple et naturel. Le prochain exemple en illustre l'essentiel.

Exemple 6.1.1 *Test sur une probabilité*

On obtient le résultat « FACE » 90 fois sur 100 lancers d'une pièce de monnaie. La pièce est-elle équilibrée?

Solution La réponse intuitive est que non, l'hypothèse que la pièce est équilibrée est invraisemblable. Pourquoi? Parce que le nombre observé de FACE présente un écart de 40 par rapport au nombre de FACE attendu d'une pièce équilibrée, 50. L'écart est trop grand, trop peu probable. Donc l'hypothèse que la pièce est équilibrée n'est pas plausible. *Nous la rejetons*. ■

L'argument est essentiellement probabiliste, bien que nous n'ayons calculé aucune probabilité explicitement. Aucun calcul n'a été nécessaire, car il suffisait de savoir que le résultat observé est très peu probable sous l'hypothèse que le dé est équilibré. On le sait intuitivement parce que le résultat est extrême. Dans l'exemple suivant, l'argument est le même, sauf que la probabilité pertinente doit être calculée explicitement.

Exemple 6.1.2 *Test sur le paramètre d'une loi exponentielle*

La durée X (en milliers d'heures) d'une pièce électronique est une variable aléatoire de loi exponentielle de moyenne β. Selon le fabriquant, $\beta = 4$. Une pièce est testée. Elle dure 3000 heures, soit $X = 3$. L'affirmation du fabriquant est-elle crédible?

Solution La pièce testée a duré moins qu'attendu. Une valeur de X aussi petite est-elle compatible avec l'hypothèse que $\beta = 4$? Quelle est la probabilité d'une valeur égale ou plus extrême que 3?
Supposons que $\beta = 4$ et calculons la probabilité $P(X \leq 3 \mid \beta = 4)$.
La fonction de répartition de X est $F(x) = 1 - e^{-x/\beta}$. Donc $P(X \leq 3 \mid \beta = 4) = 1 - e^{-3/4} = 0{,}528$. Cette probabilité n'étant pas particulièrement faible, nous concluons que la valeur observée de X est compatible avec l'hypothèse: on ne peut pas rejeter l'hypothèse que $\beta = 4$. ■

Cette analyse constitue un *test d'hypothèse*. L'hypothèse testée est que $\beta = 4$, et l'observation ne rend pas cette hypothèse invraisemblable: bien que la valeur attendue est 4 [$E(X) = 4$ sous H_o], une valeur aussi petite que $X = 3$ n'a rien d'anormal: elle pourrait bien se produire par hasard.

L'approche décrite dans ces exemples est essentiellement correcte. Mais elle est informelle et intuitive; elle ne convainc pas toujours et peut donner lieu à des controverses. C'est pour clarifier ce flou que nous développerons une approche et un langage qui exposent plus clairement la logique d'un test et définissent ses propriétés.

Un test statistique est une règle à suivre:

- on tire un échantillon;
- on calcule une certaine quantité, appelée *statistique de test*, qui mesure l'écart entre ce qui est observé et ce qu'on *s'attend* à observer sous l'hypothèse nulle;

- La statistique de test suit une loi connue sous l'hypothèse. On l'utilise afin de déterminer la probabilité d'un écart aussi grand que l'écart observé;
- Si cette probabilité est trop faible, on rejette l'hypothèse.

6.2 Développement formel

Dans cette section nous présentons les idées essentielles et la terminologie des tests d'hypothèse. Nous le ferons dans un contexte particulier — celui qui concerne la moyenne μ d'une population — afin d'éviter un langage trop abstrait. Mais il faut savoir que les idées sont générales et réapparaîtront dans les sections suivantes, avec des différences contextuelles relativement superficielles.

L'exemple suivant servira à concrétiser les idées.

Exemple 6.2.1 *Test sur une moyenne, σ connu*

Une chaîne de supermarchés commande un grand lot de boîtes de petits poids portant la marque distributeur. Le poids net moyen des boîtes du lot, μ, est censé être de 450 g. Afin de s'assurer que μ n'est pas inférieur à 450 g, on tire un échantillon de 20 boîtes et on en pèse le contenu. On suppose que la population est de loi $N(\mu\,;\,\sigma^2)$ et que σ est connu, σ = 3,2. Voici les poids obtenus:

442,70	444,90	446,86	448,80	449,15	449,96	450,23	451,18	452,39	453,61
444,19	445,02	448,16	449,11	449,57	450,02	450,65	451,67	453,04	455,35

Ces résultats permettent-ils de conclure que le poids net des boîtes du lot est inférieur à 450 ?

Solution La moyenne observée de l'échantillon est $\bar{x}$ = 449,328, inférieure donc à 450. On devine que l'écart est assez petit pour être attribué au hasard. On le verra bien en calculant la probabilité d'une moyenne $\bar{x}$ aussi petite que 449,328 (sous l'hypothèse que μ = 450).

Alors $\bar{X} \sim N\left(\mu\,;\,\frac{\sigma^2}{n}\right)$, $Z = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{20}} \sim N(0\,;\,1)$.

On calcule la probabilité: $P(\bar{X} \le 449{,}328 \mid \mu = 450) = P\left(\dfrac{\bar{X}-450}{\sigma/\sqrt{20}} \le \dfrac{449{,}328-450}{\sigma/\sqrt{20}} \,\middle|\, \mu = 450\right)$

$$= P\left(Z \le \frac{449{,}328-450}{\sigma/\sqrt{20}} \,\middle|\, Z \sim N(0\,;\,1)\right) = P\left(Z \le -0{,}9239 \mid Z \sim N(0\,;\,1)\right) = 0{,}1738.$$

Cette probabilité n'étant pas excessivement faible, on ne peut pas conclure que μ < 450 g. ■

Hypothèse nulle et alternative

Reprenons l'exemple 6.2.1 et supposons que l'échantillon de boîtes n'ait pas encore été prélevé. Nous développerons une procédure dont l'aboutissement sera une conclusion concernant μ, soit si oui ou non μ = 450. Techniquement, on dit qu'on testera une hypothèse, nommée *l'hypothèse null*e est désignée par H_o:

$$\textit{Hypothèse nulle} : H_o : \mu = 450.$$

Cette hypothèse est opposée à une autre, appelée *l'alternative* ou *contre hypothèse* et désignée par H_1.

$$\textit{Alternative} : H_1 : \mu < 450.$$

Remarque *Nous remettons à plus tard l'explication du choix de l'alternative, qui aurait pu être tout autre.*

Nous déciderons laquelle de H_o ou H_1 est vraie après avoir tiré l'échantillon et observé la statistique pertinente $\bar{X}$: une valeur trop petite de $\bar{X}$ mènera au rejet de H_o et à la conclusion que H_1 est vraie. Nous rejetterons H_o si et seulement si

$$\bar{X} \le C$$

pour un certain nombre C.

L'ensemble des valeurs $\bar{X} \le C$ est appelé *région critique* et C est appelé *point critique*.

Définition *Région critique et point critique*

La *région critique* est l'ensemble des valeurs de $\bar{X}$ pour lesquelles on rejettera H_o.
Les *points critiques* sont les limites de la région critique

La région critique $\{\bar{X} \le C\}$ n'est pas entièrement définie, puisque le point critique C n'a pas été fixé. C'est ce qui reste à faire. Considérons les conséquences d'un choix arbitraire, disons $C = 449{,}5$: on rejettera H_o si et seulement si $\bar{X} \le 449{,}5$.

Comment décider si cette règle est bonne? On la jugera bonne si elle contrôle suffisamment le risque de rejeter H_o lorsqu'il se trouve que H_o est vraie. On appelle ce type d'erreur *erreur de première espèce*.

Définition *Erreur de première espèce*

L'*erreur de première espèce* consiste à

rejeter H_o alors que H_o est vraie.

On choisira donc la région critique de telle sorte que la probabilité d'une erreur de première espèce soit suffisamment faible. Dans l'exemple, cette probabilité est calculable une fois la région critique fixée (voir l'exemple 6.2.1):

$$\text{P(Rejeter } H_o \mid H_o) = P(\bar{X} < 449{,}5 \mid \mu = 450) = 0{,}24235.$$

Cette probabilité est appelée la *taille de la région critique*.

Définition *La taille d'une région critique*

La taille d'une région critique ***C*** *est la probabilité de rejet associée à* ***C****.*

Considérons quelques régions critiques de même forme ($\bar{X} \le C$) et évaluons leur taille. Voici 6 régions critiques possibles et la taille de chacune:

	Point critique C	Région critique ***C***: $\{\bar{X} \le C\}$	P(Rejeter H_o | H_o) $= P(\bar{X} \le C \mid \mu = 450)$
1	449,5	$\bar{X} \le 449{,}5$	0,2423
2	449,0	$\bar{X} \le 449{,}0$	0,0811
3	448,5	$\bar{X} \le 448{,}5$	0,0180
4	448,0	$\bar{X} \le 448{,}0$	0,0026
5	447,5	$\bar{X} \le 447{,}5$	0,00024
6	447,0	$\bar{X} \le 447{,}0$	0,00001

La région critique # 1 dans le tableau est de taille 0,2423, une probabilité qui sera, en général, jugée trop grande. Celle de la région critique # 2 aussi, possiblement. On leur préférera les régions critiques # 3—ou même # 4 si une probabilité d'erreur de 0,0180 est elle aussi jugée excessive.

Il appert, dans cette discussion, que la région critique dépendra d'une limite que la probabilité d'une erreur ne doit pas dépasser. Une limite révélée ici implicitement et a posteriori. Il serait raisonnable de fixer ce seuil dès le départ et ensuite de déterminer une région critique—une région critique dont la taille ne dépasse pas ce seuil. Ce seuil, nommé *niveau du test*, est désigné par α.

Définition *Niveau d'un test*

Un test est de niveau α si la taille de la région critique est inférieure ou égale à α.

En général, on détermine une région critique de taille α. Mais on verra dans certaines applications cela n'est pas possible. Dans ce cas, on trouvera la plus grande région critique dont la taille est inférieure ou égale à α.

Dans l'exemple, supposons qu'on fixe $\alpha = 0{,}05$. Nous devons déterminer une région critique de taille inférieure ou égale à α, c'est-à-dire, nous devons déterminer C de telle sorte que $\mathrm{P}(\bar{X} \le C \mid \mu = 450) \le \alpha$. Puisque $\bar{X}$ est une variable continue, le seuil α est atteignable: nous chercherons donc la valeur de C pour laquelle $\mathrm{P}(\bar{X} \le C \mid \mu = 450) = \alpha$.

C'est une équation facilement résolue:

$$\mathrm{P}(\bar{X} \le C \mid \mu = 450) = \alpha \Rightarrow \mathrm{P}\left(\frac{\bar{X} - 450}{\sigma / \sqrt{n}} \le \frac{C - 450}{\sigma / \sqrt{n}} \,\middle|\, \mu = 450\right) = \alpha$$

$$\Rightarrow \mathrm{P}\left(Z \le \frac{C - 450}{3{,}2 / \sqrt{20}} \,\middle|\, Z \sim N(0\,;1)\right) = \alpha \Rightarrow \frac{C - 450}{3{,}2 / \sqrt{20}} = \text{-}z_\alpha$$

$$\Rightarrow \; C = 450 - z_\alpha \frac{3{,}2}{\sqrt{20}}.$$

Si $\alpha = 0{,}05$, $z_\alpha = -1{,}645 \Rightarrow C = 448{,}823$.

Remarques concernant le dernier exemple

1 *En principe, l'hypothèse que* $\mu = 450$ *devrait être rejetée pour toute valeur de* $\bar{X}$ *qui s'éloigne trop de son espérance sous* H_o, *soit* $\mu = 450$. *Donc on devrait rejeter* H_o (*et conclure que* $\mu < 450$) *si* $\bar{X}$ *est excessivement petit ; mais aussi si* $\bar{X}$ *est excessivement* grand (*et dans ce cas conclure que* $\mu > 450$). *En d'autres termes, la région critique, que nous avons choisie de la forme* $\{X \le C\}$, *aurait peut-être dû être de la forme* $\{\bar{X} \le C\} \cup \{\bar{X} \ge D\}$, *où D serait un nombre supérieur* à 450. *Mais le choix que nous avons fait—de nous limiter à une région critique de la forme* $\{\bar{X} \le C\}$*—est légitime. Il correspond à la décision de réagir seulement si on conclut que* $\mu < 450$. *C'est cette décision qu'on exprime lorsqu'on définit la contre hypothèse comme*

$$H_1 : \mu < 450,$$

alors que normalement la négation de H_o *devrait être* : $\mu \ne 450$. *L'énoncé H_1 précise ce qu'on conclut lorsqu'on rejette* H_o.

2 *En posant* $\mu < 450$ *pour contre hypothèse, on se prive de la possibilité de conclure que* $\mu > 450$*, s'il se trouve que la valeur de* $\bar{X}$ *le justifie. Si dans l'exemple* 6.2.1 *ceci ne pose pas de problème, c'est que cette conclusion n'entraînerait aucune action particulière : on n'a rien à faire si le poids net moyen est supérieur au poids affiché. On ne tient pas à détecter cette éventualité.*

3 *Dans l'exemple* 6.2.1, *on aurait aussi pu justifier autrement le choix de* $\mu < 450$ *pour contre hypothèse : en affirmant que l'hypothèse que* $\mu > 450$, *étant contre l'intérêt du fournisseur, est peu vraisemblable et peut donc être exclue d'emblée. Les hypothèses* $H_o : \mu = 450$ *et* $H_1 : \mu < 450$ *seraient alors exhaustives.*

4 *On pourrait se demander pourquoi l'hypothèse nulle dans l'exemple* 6.2.1 *ne serait pas* $H_o : \mu \geq 450$ *plutôt que* $H_o : \mu = 450$. *Ce serait en effet plus naturel* : H_o et H_1 *seraient exhaustives*, H_o *signifiant que les choses peuvent rester telles quelles,* H_1 *exigeant une action. Rien ne s'oppose à ce que* H_o *soit exprimée ainsi, mais alors la taille de la région critique doit être redéfinie. Dans l'exemple* 6.2.1, *la taille de la région critique* $\{\bar{X} \leq 448{,}8229\}$ est $P(\textit{rejeter } H_o \mid H_o) = P(\bar{X} \leq 448{,}8229 \mid \mu = 450) = 0{,}05$ *si* H_o *stipule que* $\mu = 450$. *Mais si* H_o *est exprimée comme* $H_o \geq 450$, *la taille de la région critique est* $P(\textit{rejeter } H_o \mid H_o) = P(\bar{X} \leq 448{,}8229 \mid \mu \geq 450)$, *qui prend plusieurs valeurs dépendantes de la valeur de* μ. *Mais sa valeur maximale* (*sur* $\mu \geq 450$) *est prise à* $\mu = 450$. *Le niveau choisi* α *est dans ce cas une borne supérieure à la probabilité d'une erreur de première espèce.* ■

Types d'erreur

L'erreur de première espèce n'est qu'une de deux erreurs possibles. L'autre erreur possible, l'*erreur de seconde espèce,* consiste à accepter H_o quand H_o est fausse.

Les quatre situations possibles sont schématisées dans le tableau suivant :

		États de la nature	
		H_o vraie	H_o fausse
Décision	On rejette H_o	Erreur de première espèce	Bonne décision
	On accepte H_o	Bonne décision	Erreur de seconde espèce

Remarque sur le terme « accepter »

On se permettra de dire « accepter H_o *», mais il est important de comprendre que cela n'équivaut pas à affirmer «* H_o *est vraie ». « Accepter* H_o *», c'est dire que l'évidence ne permet pas de la rejeter. En revanche, « Rejeter* H_o *», c'est affirmer* H_1. *Non pas que* H_1 *soit incontestablement vraie; elle peut être fausse. Mais la procédure suivie pour y arriver est conçue pour empêcher trop facilement.* ■

Les concepts introduits jusqu'ici seront maintenant appliqués à certains paramètres particuliers : une moyenne μ ; une différence de moyennes $\delta = \mu_1 - \mu_2$; une variance σ^2; et un quotient de variances σ_1^2/σ_2^2 .

6.3 Tests d'hypothèses pour μ (σ connu)

On reprend, dans un cadre plus formel, l'exemple 6.2.1. Le modèle est le suivant. On observe un échantillon aléatoire simple, c'est-à-dire, n variables aléatoires indépendantes $X_1, X_2, ..., X_n$, où $X_i \sim \mathcal{N}(\mu\,;\sigma^2)$. Supposons que la variance σ^2 de la population est connue (ce qui est rare; le cas où σ^2 est inconnue sera traité dans la section 6.4). Le but de l'échantillonnage est de tester une hypothèse, appelée *hypothèse nulle*

$$H_o : \mu = \mu_o \tag{6.3.1}$$

où μ_o est un nombre donné. La décision dépendra naturellement de la valeur de $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$,

puisque $\bar{X}$ est un estimateur de μ. Reste à décider pour quelles valeurs de $\bar{X}$ on rejettera H_o. Cela dépendra de l'alternative H_1, qui, selon du contexte, pourrait être l'une des suivantes:

$$H_1: \mu < \mu_o, \quad H_1: \mu > \mu_o \quad \text{ou } H_1: \mu \neq \mu_o \qquad (6.3.2)$$

Alternative $H_1: \mu < \mu_o$

Rejeter H_o, c'est conclure que $\mu < \mu_o$, ce qu'on fait si $\bar{X}$ est trop petit. Il est plus commode, cependant, d'exprimer la région critique en fonction de la statistique

$$Z = \frac{\bar{X} - \mu_o}{\sigma_{\bar{X}}}, \ (\text{où } \sigma_{\bar{X}} = \sigma / \sqrt{n}) \qquad (6.3.3)$$

et de rejeter H_o si $Z \leq C$, où le point critique C doit satisfaire la condition

$P(Z \leq C \mid H_o) = \alpha$.

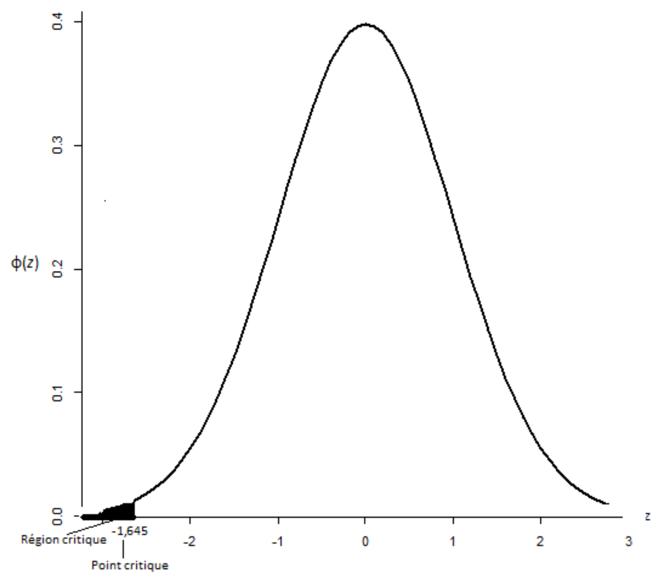
Or si H_o est vraie (et *seulement* si H_o est vraie), $Z \sim N(0 ; 1)$ et donc $C = -z_\alpha$ satisfait cette condition. La règle est donc

$$\textit{Rejeter } H_o \textit{ si } Z \leq -z_\alpha \Leftrightarrow \textit{Rejeter } H_o \text{ si } \bar{X} \leq \mu_o - z_\alpha \sigma_{\bar{X}} \qquad (6.3.4)$$

La figure 6.3.1 illustre la région critique:

Figure 6.3.1

Illustration du test $H_o: \mu = \mu_o$ *contre* $H_1: \mu < \mu_o$



Alternative $H_1: \mu > \mu_o$

Dans ce cas, par un argument parallèle, on trouve la région critique

$$\textit{Rejeter } H_o \textit{ si } Z \geq z_\alpha \Leftrightarrow \textit{Rejeter } H_o \text{ si } \bar{X} \geq \mu_o + z_\alpha \sigma_{\bar{X}} . \qquad (6.3.5)$$

Alternative $H_1: \mu \neq \mu_o$

Dans ce cas, on rejettera H_o si $\bar{X}$ est trop petit *ou* si $\bar{X}$ est trop grand. La région critique est donc

$$|Z| \geq z_{\alpha/2} \qquad (6.3.6)$$

En termes de $\bar{X}$ la règle est

$$\textit{Rejeter } H_o \text{ si } \bar{X} \leq \mu_o - z_{\alpha/2} \sigma_{\bar{X}} \ \textit{ ou } \ \bar{X} \geq \mu_o + z_{\alpha/2} \sigma_{\bar{X}} \qquad (6.3.7)$$

Exemple 6.3.1 *Test unilatéral pour une moyenne, σ connu*

Dans le contexte de l'exemple 6.2.1, considérons l'hypothèse nulle

$$H_o : \mu = 450,$$

à tester contre l'alternative

$$H_1 : \mu < 450.$$

Posons $\alpha = 0{,}05$. Alors $z_\alpha = 1{,}645$, $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{3{,}2}{\sqrt{20}} = 0{,}7155$ et on rejettera H_o si

$$Z = \frac{\bar{X} - 450}{0{,}7155} \le -1{,}645.$$

Cette région critique est illustrée dans la figure 6.3.1.

En termes de $\bar{X}$, cette règle devient :

On rejette H_o *si* $\bar{X} \le 448{,}8229$.

Étant donné la formulation de H_1, rejeter H_o signifie conclure que $\mu < 450$. Autrement dit, si le lot est acceptable ($\mu = 450$), la probabilité de conclure (à tort) qu'il ne l'est pas est de 5 %. Si $\mu > 450$ (« plus qu'acceptable »), la probabilité de conclure que $\mu < 450$ est *inférieure* à 5 %.

Si on avait posé pour alternative l'hypothèse $H_1 : \mu \neq 450$, le point critique à 5 % aurait été $z_{0,025} = 1{,}96$ et la région critique aurait été $|Z| \ge 1{,}96 \Leftrightarrow \bar{X} \le 448{,}60$ ou $\bar{X} \ge 451{,}40$. ■

Autres expressions d'une région critique

La région critique $Z \le -z_\alpha$ peut s'exprimer de plusieurs façons. Par exemple, elle est équivalente à $\bar{X} \le \mu_o - z_\alpha \sigma_{\bar{X}}$. Elle peut aussi s'exprimer en termes de probabilité. Si z est la valeur observée de Z, et $p = P(Z \le z \mid H_o)$, la région critique est $p \le \alpha$.

La valeur p

L'approche formelle devait remplacer les tests informels que nous avons illustrés au début du chapitre (section 6. 1). Ce formalisme est nécessaire pour clarifier les propriétés des procédures proposées. Il permet de quantifier — en termes probabilistes — l'effet d'un changement de région critique ou d'un changement d'effectif.

Mais on perd quelque chose lorsqu'on conclut le test catégoriquement avec un rejet ou un non rejet. Ce qu'on perd, c'est un sens de la confiance qu'on peut accorder à la conclusion. Supposons que la région critique est $\{Z \le -1{,}645\}$ — une région critique de niveau $\alpha = 0{,}05$. Selon la procédure décrite, si on observe $Z = -1{,}7$, on dira qu'on rejette H_o. Si on observe $Z = -3$, on dira également qu'on rejette H_o — rien de plus. Or la différence entre $Z = -1{,}7$ et $Z = -3$ est importante : on rejette avec beaucoup plus de confiance lorsque $Z = -3$.

Pour transmettre cette information, il suffit de citer la probabilité d'une valeur aussi extrême que la valeur observée, c'est-à-dire, de calculer la probabilité $P(Z \le -3)$ sous H_o. On trouve $P(Z \le -3) = 0{,}0014$, alors que $P(Z \le -1{,}7) = 0{,}0446$. Cette probabilité est appelée *valeur p*, (*p-value* en anglais) et parfois « seuil expérimental ». Nous désignerons la *valeur p* par *vp*.

Définition *Valeur p*

Voici deux façons de définir la *valeur p vp* :

- *vp* est la plus petite valeur de α pour laquelle H_o serait rejetée avec les données observées
- *vp* est la probabilité, sous H_o, d'observer un écart aussi grand (ou plus) que celui qui a été observé — donc moins conforme à H_o

Essentiellement, *vp* est la probabilité d'un échantillon moins conforme donc moins conforme à H_o que l'échantillon observé. Plus cette probabilité est faible, plus H_o peut être rejetée avec fermeté.

Notons que les notions de test unilatéral ou bilatéral peuvent aussi être accommodées par la *valeur p*. Si la région critique est $Z > z_\alpha$, $vp = P(Z > z \mid H_o)$. Si la région critique est $|Z| > z_{\alpha/2}$, $vp = P(|Z| > z \mid H_o)$, z étant la valeur observée de Z.

6.4 Tests d'hypothèses pour µ (σ inconnu)

Supposons encore que nous disposons d'un échantillon $X_1, \ldots, X_n$ d'une population $\mathcal{N}(\mu\,;\sigma^2)$, et que l'hypothèse à tester est H_o: $\mu = \mu_o$ contre l'alternative H_1: $\mu < \mu_o$. L'écart-type de la population σ est généralement inconnu et doit donc être estimé. On l'estime par l'écart-type échantillonnal $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$ et on estime $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ par $\hat{\sigma}_{\bar{X}} = \frac{S}{\sqrt{n}}$.

Finalement, on remplace la statistique $Z = \frac{\bar{X} - \mu_o}{\sigma_{\bar{X}}}$ par

$$T = \frac{\bar{X} - \mu_o}{\hat{\sigma}_{\bar{X}}} = \frac{\bar{X} - \mu_o}{S/\sqrt{n}}\,. \tag{6.4.1}$$

La statistique T ne suit pas une loi normale; elle suit une loi de *Student* à n - 1 degrés de liberté. La région critique est alors

$$T \leq -t_{n-1;\alpha}. \tag{6.4.2}$$

Exemple 6.4.1 *Test unilatéral pour une moyenne, σ inconnu*

Dans le contexte de l'exemple 6.3.1, supposons que σ n'est pas connu et considérons les hypothèses

$$H_o : \mu = 450 \quad vs \quad H_1 : \mu < 450$$

On remplace σ par S = 3,297, $\sigma_{\bar{X}}$ par $\hat{\sigma}_{\bar{X}} = \frac{S}{\sqrt{n}} = \frac{3,2975}{\sqrt{20}} = 0,73713$, et Z par $T = \frac{\bar{X} - \mu_o}{\hat{\sigma}_{\bar{X}}}$ = -0,9116. Le point critique à 5 % est $t_{n-1;0,05}$ = -1,73. Puisque T n'est pas inférieur à -1,73, on ne rejette pas H_o. ■

Point critique: $t_{\mathtt{nu;alpha}}$ `=LOI.STUDENT.INVERSE.N(alpha;nu)`
Probabilité: $P(T \leq \mathtt{t} \mid T \sim t_{\mathtt{nu}})$= `LOI.STUDENT.N(t;nu;1)`

Le cas traité dans le dernier exemple et les deux autres possibilités d'alternatives sont résumées dans le tableau ci-dessous. T_{n-1} désigne une variable de loi de *Student* à n-1 degrés de liberté, T est la statistique (6.4.1) et t est la valeur observée de T.

Test de l'hypothèse $\mu = \mu_o$

		H₁		
		$\mu < \mu_o$	$\mu > \mu_o$	$\mu \neq \mu_o$
Région critique en fonction	de T	$T \leq -t_{n-1;\alpha}$	$T \geq t_{n-1;\alpha}$	$\|T\| \geq t_{n-1;\alpha/2}$
	de vp	$P(T_{n-1} \leq t) \leq \alpha$	$P(T_{n-1} \geq t) \leq \alpha$	$P(\|T_{n-1}\| \geq \|t\|) \leq \alpha$

Calcul de la valeur p

Le calcul de la *valeur p* dépend de la forme de la région critique, qui, elle dépend de H_1. Si la région critique est de la forme $Z \leq C$, $vp = P(Z \leq z \mid H_o)$, où z est la valeur observé de Z. Si la région critique est de la forme $Z \geq C$, $vp = P(Z \geq z \mid H_o)$. Finalement, si la région critique est de la forme $|Z| \geq C$, $vp = P(|Z| \geq |z| \mid H_o)$.

6.5 Test sur la différence de deux moyennes

Plusieurs enquêtes et plusieurs expériences scientifiques ont pour but de déterminer s'il y a une différence entre les moyennes de deux populations. Lorsqu’on veut comparer les revenus annuels de deux groupes ethniques, par exemple ; ou évaluer l'effet d'un traitement en comparant un groupe expérimental qui reçoit le traitement, à un groupe témoin, qui reçoit un placebo.

Le modèle mathématique est le suivant : On dispose de deux échantillons indépendants, $X_1, X_2, \ldots, X_{n_1}$ et $Y_1, Y_2, \ldots, Y_{n_2}$ et que :

$$X_i \sim N(\mu_1 ; \sigma_1^2), \quad i = 1, 2, \ldots, n_1 \;; \quad Y_j \sim N(\mu_2 ; \sigma_2^2), \quad j = 1, 2, \ldots, n_2 \tag{6.5.1}$$

Soit $\bar{X}$ et $\bar{Y}$ les moyennes de deux échantillons de taille n_1 et n_2 respectivement, tirés de deux populations de moyennes μ_1 et μ_2 et de variance σ_1^2 et σ_2^2 respectivement. Soit δ la différence entre les deux moyennes:

$$\delta = \mu_1 - \mu_2 \tag{6.5.2}$$

Considérons l'hypothèse

$$H_o : \delta = \delta_o \tag{6.5.3}$$

contre l'une des alternatives

$$H_1 : \delta \neq \delta_o \;;\; H_1^{'} : \delta > \delta_o \;;\; H_1^{''} : \delta < \delta_o \tag{6.5.4}$$

Dans la plupart des applications courantes, $\delta_o = 0$: H_o est l'hypothèse d'égalité des deux moyennes.

Variances connues

Supposons pour le moment que σ_1^2 et σ_2^2 sont connues. Il est clair, intuitivement, que nous devrions baser notre test sur mesure de l'écart entre $\bar{X} - \bar{Y}$ et δ_o, soit $\bar{X} - \bar{Y} - \delta_o$. Cet écart sera normé, comme il est coutume de le faire en statistique, c’est-à-dire, divisé par son écart-type $\sigma_{\bar{X}-\bar{Y}-\delta_o} = \sigma_{\bar{X}-\bar{Y}}$:

$$z_{\bar{X}-\bar{Y}} = \frac{\bar{X} - \bar{Y} - \delta_o}{\sigma_{\bar{X}-\bar{Y}}} \tag{6.5.5}$$

où

$$\sigma_{\bar{X}-\bar{Y}} = \sqrt{\mathrm{Var}(\bar{X} - \bar{Y} - \delta_o)} = \sqrt{\mathrm{Var}(\bar{X}) + \mathrm{Var}(\bar{Y})} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\,. \tag{6.5.6}$$

L'égalité $\mathrm{Var}(\bar{X} - \bar{Y} - \delta_o) = \mathrm{Var}(\bar{X}) + \mathrm{Var}(\bar{Y})$ est permise étant donné l'indépendance des échantillons. Sous H_o cette statistique suit une loi $\mathcal{N}(0\,;\,1)$.

Voici les régions critiques correspondant à trois alternatives

Alternative	$\delta \neq \delta_o$	$\delta > \delta_o$	$\delta > \delta_o$
Région critique	$\lvert Z_{\bar{X}-\bar{Y}} \rvert \geq z_{\alpha/2}$	$Z_{\bar{X}-\bar{Y}} \geq z_\alpha$	$Z_{\bar{X}-\bar{Y}} \leq -z_\alpha$

Remarque *Ce test est valide lorsque la statistique $Z_{\bar{X}-\bar{Y}}$ est de loi normale, ce qui découle de la normalité de la population, une des hypothèses du modèle. L'applicabilité de ce test serait très limitée si la normalité de la population était une hypothèse incontournable. Mais elle ne l'est pas, car par le théorème limite central, on sait $Z_{\bar{X}-\bar{Y}}$ est à peu près de loi normale, à condition que n_1* et *n_2 soient assez grands.*

■

Variances inconnues

En pratique, bien sûr, lorsque les variances σ_1^2 et σ_2^2 ne sont *pas* connues, la variance $\mathrm{Var}(\bar{X} - \bar{Y})$ au dénominateur de la statistique $Z_{\bar{X}-\bar{Y}}$ devra être remplacée par une estimation. Nous présentons ici deux approches, dépendant du contexte et des suppositions qu'on peut faire:

1. n_1 et n_2 sont très grands; ou
2. La population est normale et $\sigma_1 = \sigma_2$.

<u>*Premier cas: n_1 et n_2 sont très grands*</u>

Si n_1 et n_2 *sont très grands*, l'approche décrite ci-dessus s'applique, avec la seule différence que σ_1^2 et σ_2^2 sont remplacés par leurs estimateurs respectifs

$$S_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2 \quad \text{et} \quad S_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2\,, \tag{6.5.7}$$

On remplace alors $Z_{\bar{X}-\bar{Y}}$ par la statistique $\hat{Z}_{\bar{X}-\bar{Y}}$ définie par

$$\hat{Z}_{\bar{X}-\bar{Y}} = \frac{\bar{X} - \bar{Y} - \delta_o}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}\,. \tag{6.5.7}$$

La variable $\hat{Z}_{\bar{X}-\bar{Y}}$ suit une loi à peu près normale, même lorsqu'on remplace σ_1^2 et σ_2^2 par S_1^2 et S_2^2, à condition que les deux échantillons soient assez grands. Donc on suit la même démarche que lorsque σ_1^2 et σ_2^2 sont connues.

<u>*Variances inconnues mais supposées égales*</u>

Si le contexte permet de supposer que les *populations* sont normales (à peu près) et qu'elles ont la *même variance* σ^2, il existe un test exact basé sur la loi de *Student*.

Supposons donc que $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

L'écart-type de $\bar{X} - \bar{Y}$ est alors $\sigma_{\bar{X}-\bar{Y}} = \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}} = \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ et la différence normalisée est

$$Z_{\bar{X}-\bar{Y}} = \frac{\bar{X} - \bar{Y} - \delta_0}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim \mathcal{N}(0\,;\,1). \tag{6.5.9}$$

Il faut maintenant estimer σ. Les statistiques S_1^2 et S_2^2 sont toutes deux des estimateurs du même paramètre σ^2. Un estimateur efficace devra les combiner en un seul estimateur. On peut montrer qu'une combinaison optimale est

$$\hat{\sigma}^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \frac{\sum_{i=1}^{n_1}(X_i - \bar{X})^2 + \sum_{i=1}^{n_2}(Y_i - \bar{Y})^2}{n_1+n_2-2} \tag{6.5.10}$$

On remplacera donc $\sigma_{\bar{X}-\bar{Y}} = \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ par $\hat{\sigma}_{\bar{X}-\bar{Y}} = \hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ et $Z_{\bar{X}-\bar{Y}}$ par

$$T_{\bar{X}-\bar{Y}} = \frac{\bar{X} - \bar{Y} - \delta_0}{\hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{6.5.11}$$

Sous H_0,

$$T_{\bar{X}-\bar{Y}} \sim t_{n_1+n_2-2}\,. \tag{6.5.12}$$

L'approche à partir de là est la même que dans les deux dernières situations, avec la différence que la statistique $T_{\bar{X}-\bar{Y}}$ remplace $Z_{\bar{X}-\bar{Y}}$ et la loi de *Student* à $n_1 + n_2 - 2$ degrés de liberté et la loi de $\mathcal{N}(0\,;\,1)$. Voici donc les régions critiques correspondant aux trois alternatives habituelles.

T_ν désigne une variable de loi de *Student* à ν degrés de liberté, $\nu = n_1 + n_2 - 2$.

Test de l'hypothèse $\delta = \delta_0$

T est la statistique (6.5.11) et t est valeur observée de T,

		H₁		
		$\delta < \delta_0$	$\delta > \delta_0$	$\delta \neq \delta_0$
Région critique en fonction	de T	$T \leq -t_{\nu;\alpha}$	$T \geq t_{\nu;\alpha}$	$\lvert T\rvert \geq t_{\nu;\alpha/2}$
	de *vp*	$P(T_\nu \leq t) \leq \alpha$	$P(T_\nu \geq t) \leq \alpha$	$P(\lvert T_\nu\rvert \geq \lvert t\rvert) \leq \alpha/2$

Exemple 6.5.1 *Test d'égalité de moyennes*, $\sigma_1 = \sigma_2$

On fait passer un examen à un groupe de 22 étudiants qui ont suivi un cours utilisant une nouvelle méthode d'enseignement ; on compare leurs résultats à ceux d'un groupe témoin de 22 étudiants, qui ont suivi le cours traditionnel. Les résultats sont les suivants :

Groupe témoin :
41, 41, 43, 46, 46, 45, 45, 32, 33, 39, 42, 45, 39, 44, 36, 49, 40, 35, 36, 40, 54, 32

Groupe expérimental :
53, 47, 41, 49, 43, 45, 50, 48, 49, 42, 38, 42, 34, 48, 51, 33, 44, 48, 49, 33, 45, 42

Est-ce qu'on peut conclure que la nouvelle méthode est meilleure ? Faire un test bilatéral à 5 %.

Solution Voici un résumé des résultats :

	Moyenne	Écart-type
Groupe témoin	$\bar{X} = 41{,}05$	$S_1 = 5{,}64$
Groupe expérimental	$\bar{Y} = 44{,}27$	$S_2 = 5{,}77$

On devrait normalement se demander si on peut ou non supposer l'égalité des variances, mais le problème ne se pose pas ici, puisque les écarts-types échantillonnaux S_1 et S_2 sont presque égaux, ce qui rend l'hypothèse $\sigma_1 = \sigma_2$ crédible. L'estimation commun est $\hat{\sigma} = 5{,}70$ et $T_{\bar{X}-\bar{Y}} = -1{,}88$. Le point critique est $t_{42\,;0{,}025} = 2{,}02$. On ne peut pas rejeter H_o. La $vp = P(|T_{42}| \geq 1{,}88) = 0{,}067$, d'où on conclut que la différence n'est pas significative au seuil $\alpha = 0{,}05$.

Mais si on avait effectué un test unilatéral (à gauche), la *valeur p* aurait été $P(T_{42} \leq -1{,}88) = 0{,}0337$, et nous aurions alors conclu que la différence est significative. Un test unilatéral, cependant, n'aurait pas été facile à justifier dans ce contexte.

Remarque Le fait que S_1 et S_2 soient si proches est doublement réconfortant. D'abord, il nous conforte dans notre supposition que $\sigma_1 = \sigma_2$. Ensuite parce que si on n'avait pas supposé l'égalité des variances, la statistique de test n'aurait presque pas changé. ■

6.6 Données appariées

Il arrive que les paires $[X_i\,; Y_i]$ soient prises sur un même individu. On parle alors de *données appariées*. Lorsque les données sont appariées, les X_i et les Y_i ne sont pas indépendantes, une des conditions essentielles pour l'application des techniques de la section 6.6. La solution, cependant n'est pas compliquée. Elle est, en fait, déjà connue, comme le montre le prochain exemple.

Certains plans d'expérience visant à comparer deux moyennes ne comportent qu'un groupe de sujets, chacun générant deux valeurs, X_i et Y_i. On parle alors de *données appariées*. L'objectif est encore de comparer $\mu_1 = E(X_i)$ et $\mu_2 = E(Y_i)$, et la statistique pertinente est encore $\bar{D} = \bar{X} - \bar{Y}$. Mais le fait que le couple $[X_i\,; Y_i]$ est observé sur un même individu rend l'hypothèse d'indépendance (entre les X et les Y) peu vraisemblable. Or cette hypothèse est l'une des conditions essentielles pour l'application des techniques présentées à la section dernière. Il se trouve, pourtant, que la solution n'est pas compliquée: elle se réduit à un test sur une moyenne.

Considérons les différences $D_i = X_i - Y_i$. Alors $\delta = E(D_i) = \mu_1 - \mu_2$. L'échantillon est donc constitué de n différences $D_1, D_2, \ldots, D_n$ de moyenne δ, et l'hypothèse à tester est

$$H_o : \delta = \delta_o$$

où δ_o est un nombre fixe (souvent égal à 0 mais pas toujours).

Le test est basé sur la statistique $T = \dfrac{\bar{D} - \delta_o}{S_{\bar{d}}}$, où $\bar{D}$ est la moyenne des différences, $S_{\bar{d}} = S_d/\sqrt{n}$ et S_d est l'écart-type des différences.

Exemple 6.6.1 *Test d'égalité de moyennes, données appariées*

Une expérience est entreprise afin de déterminer si la prise d'un certain médicament contre l'asthme affecte la biodisponibilité d'une certaine composante d'un contraceptif oral. L'étude emploie 22 femmes qui prennent régulièrement le contraceptif. On a mesuré la concentration de cette composante pendant 24 heures à deux occasions : une fois après la prise du médicament (Y) et une fois après la prise d'un placebo (X). Voici les données :

Placebo (x)	Médicament (y)	$x - y$ (d)	Placebo (x)	Médicament (y)	$x - y$ (d)	Placebo (x)	Médicament (y)	$x - y$ (d)
219	216	3	542	178	364	267	218	49
233	251	-18	205	224	-19	344	256	88
232	229	3	143	101	42	366	242	124
181	164	17	282	266	16	194	163	31
224	216	8	249	255	-6	132	127	5
172	134	38	313	401	-88	280	246	34
243	208	35	213	213	0	286	272	14
242	181	61						

Peut-on conclure que la biodisponibilité de cette composante est affectée par le médicament ?

Solution Soit $\mu_1 = E(X_i)$ et $\mu_2 = E(Y_i)$, $i = 1, 2, \ldots, 22$; et $\delta = \mu_1 - \mu_2$. L'hypothèse à tester est $H_o : \delta = 0$. Ce sont des données appariées, et donc les observations X_i et Y_i ne sont pas indépendantes: la biodisponibilité d'un médicament dépend en partie de la personne qui l'ingère, de sorte que la corrélation ρ entre X_i et Y_i doit être positive.

Considérons les couples $D_i = X_i - Y_i$, dont la moyenne est $\delta = \mu_1 - \mu_2$. L'hypothèse à tester est donc $H_o : \delta = 0$. Les deux échantillons on été réduits à un seul et le test est maintenant un test sur une moyenne qu'on peut effectuer à l'aide de la méthode de la section 6.4. Le test est basé sur la statistique $T = \dfrac{\bar{D}}{S_{\bar{d}}}$, où $\bar{D}$ est la moyenne des différences, $S_{\bar{d}} = S_d \big/ \sqrt{22}$ et S_d est l'écart-type des différences. On a : $\bar{D} = 36{,}41$, $S_d = 84{,}07$, $S_{\bar{d}} = 17{,}92$ et $T = 2{,}03$. Le point critique pour un test bilatéral à 5 % est 2,08. Puisque $T \le 2{,}08$, on ne peut pas, strictement parlant, rejeter H_o. La valeur p est 0,055.

Remarque Un test unilatéral à 5 % aurait permis de rejeter H_o si l'alternative avait été $\mu_1 < \mu_2$. La conclusion aurait-elle été légitime ? ■

Remarque *Le dernier exemple illustre bien l'utilité de la notion de valeur p. Il n'est pas raisonnable de donner une réponse catégorique — dans un sens comme dans l'autre — lorsque la valeur de T est si proche du point critique. Et lorsque, comme dans le cas présent, un test unilatéral aurait mené à une conclusion différente* (*on aurait rejeté* H_o *si l'alternative avait été* $\mu_1 < \mu_2$). *Les résultats sont ambigus, et il faut vivre avec. C'est ce qu'on fait lorsqu'on annonce une valeur p de* 0,055, *qui n'est ni trop grande ni trop petite, et qui, de surcroît, aurait été de* 0,028 *dans un test unilatéral.* ■

Problématique des données appariées

Les données ici sont appariées dans le sens que deux valeurs, X et Y, sont observées sur une même unité statistique, ce pourquoi il faut tenir compte d'une éventuelle dépendance entre X et Y. Mais il existe un mode d'appariement dans lequel X et Y proviennent de deux unités distinctes mais qui néanmoins appariées, dans le sens suivant. Une façon de comparer deux traitements, celle discutée dans ce chapitre, consiste à tirer deux échantillons indépendants et de soumettre l'un des deux au traitement (groupe expérimental), l'autre pas (groupe témoin). Mais quand cela est possible, il est préférable de s'assurer que les deux groupes soient comparables par rapport à certains critères pertinents, par exemple, même âge, même sexe, etc. Donc pour chaque membre du groupe témoin on choisira pour le groupe expérimental un individu ayant les mêmes caractéristiques.

Dans ce cas, la méthode décrite ici reste valable (voir les exercices 6.42 et 6.43).

6.7 Test d'hypothèse sur σ

Supposons qu'on veut tester l'hypothèse $H_o : \sigma = \sigma_o$ où σ_o est un nombre positif donné. Il est naturel d'appuyer la décision sur la variance échantillonnale S^2—tout comme une hypothèse sur μ est évaluée à partir de la moyenne échantillonnale $\overline{X}$; et une hypothèse sur une différence $\mu_1 - \mu_2$ est évaluée à partir de la différence entre deux moyennes échantillonnales. On utilisera une fonction croissante de S^2, soit la statistique

$$Q = \frac{(n-1)S^2}{\sigma_o^2} \qquad (6.7.1)$$

qui suit une loi de χ^2_{n-1} sous H_o. La région critique dépend de l'alternative.

Voici les régions critiques correspondant aux alternatives habituelles. χ^2_{n-1} désigne une variable de loi de *khi-deux* à *n*-1 degrés de liberté; Q est la statistique (6.7.1) et q est la valeur observée de Q.

		H_1		
		$\sigma < \sigma_o$	$\sigma > \sigma_o$	$\sigma \neq \sigma_o$
Région critique en fonction	de Q	$Q \leq \chi^2_{n-1;1-\alpha}$	$Q \geq \chi^2_{n-1;\alpha}$	$Q \leq \chi^2_{n-1;1-\alpha/2}$ ou $Q \geq \chi^2_{n-1;\alpha/2}$
	de vp	$P(\chi^2_{n-1} \leq q) \leq \alpha$	$P(\chi^2_{n-1} \geq q) \leq \alpha$	$P(\chi^2_{n-1} \leq q) \leq \alpha/2$ ou $P(\chi^2_{n-1} \geq q) \leq \alpha/2$

Remarque *Soit C_{1g} et C_{1d} les régions critiques pour tester H_o contre $H_g : \sigma < \sigma_o$ et $H_d : \sigma > \sigma_o$, respectivement. Si C_g et C_d sont de tailles α_g et α_d, respectivement, alors $C = C_g \cup C_d$ est une région critique de taille $\alpha = \alpha_g + \alpha_d$ pour tester H_o contre $H_1: \sigma \neq \sigma_o$. Un test bilatéral correspond donc à deux tests unilatéraux, H_o étant alors rejetée si l'un des deux tests unilatéraux mène à un rejet. Le choix $\alpha_g = \alpha_d = \alpha/2$ en est un parmi d'autres.*
En pratique, on effectuera au plus un seul des deux tests quelle que soit l'alternative : s'il se trouve que $S > \sigma_o$, on ne testera pas H_o contre H_g ; et s'il se trouve que $S < \sigma_o$ on ne testera pas H_o contre H_d .

Exemple 6.7.1 *Test sur une variance*

Dans l'exemple 6.4.1, supposons que l'écart-type des poids des boîtes est censé être de 2,5 g. L'échantillon a donné un écart-type estimé S = 3,297. Doit-on conclure que l'écart-type de la population est maintenant supérieur à 2,5 ? Faire un test unilatéral à 5 %.

Solution L'hypothèse nulle est $H_o : \sigma^2 = (2,5)^2$, l'alternative est $H_1 : \sigma^2 > (2,5)^2$. À 5 %, le point critique est $\chi^2_{19;0,05} = 30,144$. Alors que $\frac{(n-1)S^2}{\sigma_o^2} = \frac{(20-1)(3,2965)^2}{(2,5)^2} = 33,04 > 30,14$, ce qui permet de rejeter H_o et de conclure que l'écart-type est en effet supérieur à 2,5. La *valeur p*, $P(\chi^2_{39} \geq 71,0) = 0,024$, confirme qu'il est peu probable d'obtenir une valeur aussi élevée de Q sous l'hypothèse nulle. ■

6.8 Test sur le rapport de deux variances

Supposons qu'on dispose de deux échantillons aléatoires indépendants de tailles n_1 et n_2 issus de deux populations normales de variances σ_1^2 et σ_2^2 et qu'on veuille tester l'hypothèse

$$H_o: \sigma_1^2 = \sigma_2^2 \qquad (6.8.1)$$

contre une alternative bilatérale ou unilatérale.

Comme statistique de test, il est raisonnable d'utiliser le quotient $F = \frac{S_1^2}{S_2^2}$ (ou $\frac{1}{F} = \frac{S_2^2}{S_1^2}$). Si F est trop grand, on rejettera H₀ pour conclure que $\sigma_1 > \sigma_2$. Si F est trop petit (ou $\frac{1}{F} = \frac{S_2^2}{S_1^2}$ est trop grand) on rejettera H₀ pour conclure que $\sigma_2 > \sigma_1$. Pour décider de ce qui est « trop grand » ou « trop petit », il faut connaître la loi de F. Elle suit une loi appelée *loi de Fisher* ou *loi F*, que nous définissons maintenant.

La loi F de Fisher

Comme la loi χ^2 et la loi de *Student*, la loi $\mathcal{F}$ est une des lois fondamentales de la statistique. En voici une définition.

Définition *Loi de Fisher*

Soit $U_1 \sim \chi^2_{\nu_1}$ et $U_2 \sim \chi^2_{\nu_2}$, deux variables aléatoires indépendantes. Alors la variable

$$F = \frac{U_1 / \nu_1}{U_2 / \nu_2} \qquad (6.8.2.)$$

suit une loi appelée *loi de Fisher* à ν_1 et ν_2 degrés de liberté

Les paramètres ν_1 et ν_2 sont appelés « nombre de degrés de liberté du *numérateur* » et « nombre de degrés de liberté du *dénominateur* », respectivement. On écrit $F \sim \mathcal{F}_{\nu_1;\nu_2}$ pour signifier que F suit une loi de Fisher à ν_1 et ν_2 degrés de liberté. Le logiciel Excel calcule des probabilités relatives à la loi $\mathcal{F}$, ce qui permet d'obtenir la *valeur p* d'une statistique F.

La commande

```
=LOI.F.DROITE(f;nu1;nu2)
```

donne la probabilité $P(F \geq f)$ lorsque $F \sim \mathcal{F}_{nu_1;nu_2}$.

Si S_1^2 et S_2^2 sont les variances échantillonnales, alors $(n_1 - 1)\frac{S_1^2}{\sigma_1^2} \sim \chi^2_{n_1-1}$ et $(n_2 - 1)\frac{S_2^2}{\sigma_2^2} \sim \chi^2_{n_2-1}$. D'où

$$\frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} \sim \mathcal{F}_{n_1-1;n_2-1}. \qquad (6.8.3)$$

Donc si H₀ est vraie,

$$F = \frac{S_1^2}{S_2^2} \sim \mathcal{F}_{n_1-1;n_2-1}. \qquad (6.8.4)$$

Voici les régions critiques correspondant aux trois alternatives habituelles. Notez que H₀ est équivalente à l'égalité des écarts-types, $\sigma_1 = \sigma_2$.

Test de l'hypothèse $\sigma_1 = \sigma_2$;

Supposons que $F = \dfrac{S_1^2}{S_2^2} = f$ et $F_{n_1-1;n_2-1}$ une variable de loi de *Fisher* à n_1-1 et n_2 - 1 degrés de liberté.

	H_1		
	$\sigma_1 < \sigma_2$	$\sigma_1 > \sigma_2$	$\sigma_1 \neq \sigma_2$
Région critique en fonction de *vp*	$P(F_{n_1-1;n_2-1} \le f) \le \alpha$	$P(F_{n_1-1;n_2-1} \ge f) \le \alpha$	$P(F_{n_1-1;n_2-1} \le f) \le \alpha/2$ ou $P(F_{n_1-1;n_2-1} \ge f) \le \alpha/2$

Remarque *Un seul calcul est à faire, quelle que soit l'alternative. Supposons que* $S_1 > S_2$*. On calcule* $p = P(F_{n_1-1;n_2-1} \ge q)$ *et on rejette* H_o *si* $p < \alpha$ *ou si* $p < \alpha/2$*, selon que l'alternative est* $\sigma_1 > \sigma_2$ *ou* $\sigma_1 \neq \sigma_2$*. Si l'alternative est* $\sigma_1 < \sigma_2$*, on accepte* H_o *d'emblée, car il n'est pas question de conclure que* $\sigma_1 < \sigma_2$ *lorsque* $S_1 > S_2$*. Évidemment, si* $S_2 > S_1$ *la situation est identique avec des libellés intervertis.* ■

Exemple 6.9.1 *Test sur l'égalité de deux variances*

Les écarts-types de deux échantillons de tailles 30 et 40 tirés de deux populations normales A et B sont, respectivement, 28 et 38. Tester à 5 % l'hypothèse que les variances des populations sont égales contre

a) contre l'hypothèse la population B est plus dispersée.
b) contre l'hypothèse la population A est plus dispersée.
c) contre une hypothèse bilatérale, soit que l'une des deux populations est plus dispersée que l'autre.

Solutions

a) Posons $S_1 = 38$, $S_2 = 28$, $n_1 = 40$ et $n_2 = 30$. La valeur f de la statistique F est $f = \dfrac{38^2}{28^2} = 1{,}84$.

Sous H_o, F est de loi $\mathcal{F}_{39;29}$. On calcule $P(F \ge 1{,}84) = 0{,}045$. Puisque α est fixé à 0,05, on rejette H_o, pour conclure que la population B est plus dispersée.

b) Il n'est pas nécessaire de calculer puisque, ayant observé $S_1 < S_2$, il n'est pas question de conclure que la population A est plus dispersée.

c) Puisque $P(F \ge 1{,}84) = 0{,}045 \ge \alpha/2$, on ne peut pas rejeter H_o. ■

6.9 Fonction de puissance

Jusqu'ici nous avons retenu la probabilité d'une erreur de première espèce comme seule mesure de la validité d'un test. Supposons, par exemple, que l'on compare deux moyennes afin de montrer qu'un nouveau médicament est plus efficace que l'ancien (H_1). Une erreur de première espèce, c'est conclure à tort que le nouveau médicament est plus efficace—une erreur à éviter si elle a des conséquences graves, bien sûr, mais à éviter aussi pour une autre raison: quiconque affirme avoir créé un nouveau médicament a le devoir de le « démontrer ».

À l'issue d'un test, si on rejette H_o, on peut dire, en reconnaissant un certain abus de langage, qu'on a « démontré » H_1. On ne peut pas dire qu'on en est certain, mais on peut dire qu'on en est confiant. Cela dit, il est également important de contrôler l'erreur de seconde espèce. Lorsqu'on conçoit un plan d'expérience, il est important de s'assurer que si H_o est fausse, on puisse la rejeter—que la probabilité de rejeter H_o quand H_o est fausse soit forte. Lorsqu'on compare deux médicaments, et qu'on souhaite démontrer que l'un est supérieur à l'autre, on ne pourra le faire qu'en autant que la différence entre les deux soit assez importante pour être détectée. Une différence réelle, même petite, peut éventuellement être détectable, mais au prix d'un échantillon assez grand. Un test qui permet de détecter de petites différences est dit *puissant*.

Remarque *Afin de désencombrer le langage, nous éviterons dans cette section de parler d'*accepter, *pour ne parler que de* rejeter. *Ainsi quand* H_o *est fausse, on dira « la probabilité de rejeter* H_o *doit être élevée » plutôt que « la probabilité d'accepter* H_o *doit être faible ».* ■

Exemple 6.9.1 *Fonction de puissance d'un test unilatéral sur une moyenne,* σ *connu*

On examine dans le contexte d'un contrôle de la qualité de la production d'un pain français appelé *flûte*, qui doit normalement peser 400 g en moyenne. On doit tester, à partir d'un échantillon de taille $n = 8$, l'hypothèse

$$H_o : \mu = \mu_o \text{ contre } H_1 : \mu < \mu_o.$$

où $\mu_o = 400$.

On admettra que $\sigma = 2{,}5$ est connu et donc que $\sigma_{\bar{X}} = \dfrac{2{,}5}{\sqrt{8}} = 0{,}8839$.

Considérons la région critique

$$\bar{X} \le 398{,}55.$$

On peut maintenant déterminer la probabilité de rejet, laquelle est fonction de la vraie valeur de μ. Nous désignons cette fonction par φ

$$\varphi(\mu) = P(\text{rejeter } H_o \mid \mu)$$

Cette fonction est appelée *fonction de puissance*. Elle peut être calculée pour toute valeur de μ. Par exemple, si $\mu = \mu_o$ (H_o est vraie), alors

$$\varphi(\mu_o) = P(\text{rejeter } H_o \mid \mu_o) = P(\bar{X} \le 398{,}5461 \mid \mu = 400) \approx 0{,}05.$$

C'est la probabilité d'une erreur de première espèce.

Si μ 398,

$$\varphi(398) = P(\bar{X} \le 398{,}5461 \mid \mu = 398) = P\left(\frac{\bar{X} - 398}{\sigma_{\bar{X}}} \le \frac{398{,}5461 - 398}{0{,}8839} \mid \mu\right) = 0{,}7331.$$

Cette probabilité est nettement supérieure à 0,05, et c'est ce qu'on veut: si $\mu = 398$, H_o est fausse et il est bon qu'on la rejette.

Dans le cas présent nous pouvons exprimer la fonction de puissance à l'aide de Φ, la fonction de répartition d'une variable de loi $\mathcal{N}(0\,;\,1)$:

$$\varphi(\mu) = P(\text{rejeter } H_o \mid \mu) = P(\bar{X} \le 398{,}5461 \mid \mu) = P\left(\frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \le \frac{398{,}5461 - \mu}{\sigma_{\bar{X}}} \mid \mu\right)$$

$$= \Phi\left(\frac{398{,}5461 - \mu}{0{,}8839}\right).$$ ■

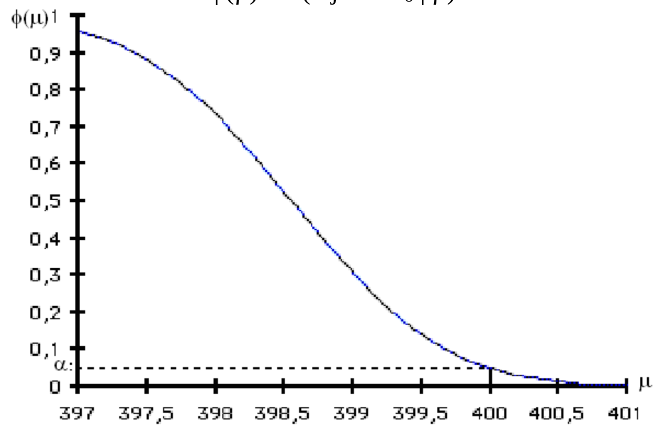
Toute l'information concernant les propriétés d'un test est contenue dans la fonction de puissance: la probabilité de rejeter H_o quand elle est vraie et la probabilité de la rejeter quand elle est fausse.

Plusieurs logiciels (y compris Excel) permettent de calculer ces probabilités, et d'en tracer la courbe. La figure 6.9.1, qui présente un graphique de la fonction de puissance, a été tracée par le logiciel Excel.

Figure 6.9.1

Fonction de puissance pour le test de l'exemple 6.9.1

$\varphi(\mu) = P(\text{rejeter } H_o \mid \mu)$



Remarques à propos de la fonction de puissance **(*Figure* 6.9.1)**

- *La valeur de* φ *au point* $\mu = 400$ *est* 0,05*, la probabilité d'une erreur de première espèce*.
- *Lorsque* $\mu < 400$*, la valeur de* φ *est supérieure à* 0,05*, ce qui est normal* : *lorsque* $\mu < 400$, H_o *est fausse, et il faut bien que la probabilité de la rejeter soit élevée.*
- *Lorsque* μ *n'est que légèrement inférieure à* 400*, la valeur de* φ*, bien que supérieure à* 0,05*, est faible* : *lorsque* H_o *n'est que « un petit peu fausse », il y a peu de chance qu'on la rejette.*
- *Inversement, lorsque* μ *est de beaucoup inférieure à* 400 (H_o *est « très fausse »*)*, la probabilité de rejet est forte* : *l'écart est d'autant plus facile à détecter qu'il est important.*
- *La probabilité d'une erreur de seconde espèce—qui dépend de la valeur de* μ*—est* $1-\varphi(\mu)$*, représentée graphiquement par la distance entre* $\varphi(\mu)$ et 1, $\mu < 400$.
- $\varphi(\mu)$ *est une fonction décroissante, de sorte* $\varphi(\mu) < \varphi(400)$ *pour tout* $\mu > 400$. *On aurait donc pu exprimer l'hypothèse nulle comme* $H_o : \mu \geq 400$ *au lieu de* $H_o : \mu = 400$*. Que pourrait-on dire dans ce cas de la probabilité d'une erreur de première espèce ? L'erreur de première espèce est* $P(\text{Rejeter } H_o \mid H_o) = P(\bar{X} \leq 398{,}55 \mid \mu \geq 400)$. *Cette probabilité dépend de la valeur de* μ*, mais quelle qu'elle soit, si* H_o *est vraie*, $\mu \geq 400$ *et conséquemment*, $P(\bar{X} \leq 398{,}55 \mid \mu \geq 400) \leq 0{,}05$). *La probabilité d'une erreur de première espèce est donc bornée supérieurement par* α

Exemple 6.9.2 *Fonction de puissance d'un test bilatéral sur une moyenne*, σ *connu*

Dans l'exemple 6.9.1 l'alternative

$$H_1 : \mu < 400$$

signifiait qu'il ne fallait rejeter H_o que si le poids moyen des pains se révélait inférieur à 400 g. Ce serait, par exemple, l'attitude d'un inspecteur gouvernemental dont le seul souci serait de protéger le consommateur.

Mais le fabricant serait normalement intéressé à détecter tout écart à la norme, qu'il soit en trop ou en moins. Il faudra donc rejeter H_o si les données révèlent que $\mu > 400$ aussi bien que si elles révèlent que $\mu < 400$. On signifie ceci en posant pour alternative l'hypothèse :

$$H_1 : \mu \neq 400.$$

(H_o reste toujours $H_o : \mu = 400$). On rejette H_o lorsque $\bar{X}$ s'éloigne trop de 400. La région critique est donc de la forme

$$\left|\frac{\bar{X} - 400}{\sigma_{\bar{X}}}\right| \geq z_{\alpha/2}$$

où $\sigma_{\bar{X}}$ = 0,8839. Pour α = 0,05, $z_{\alpha/2}$ = 1,96, et la région critique, en termes de $\bar{X}$ est

$$\bar{X} \geq 401{,}73 \text{ ou si } \bar{X} \leq 398{,}27$$

La fonction de puissance de ce test est

$$\varphi(\mu) = P(\bar{X} \geq 401{,}73 \mid \mu) + P(\bar{X} \leq 398{,}27 \mid \mu) = P\left(\frac{\bar{X}-\mu}{\sigma_{\bar{X}}} \geq \frac{401{,}73-\mu}{\sigma_{\bar{X}}}\right) + P\left(\frac{\bar{X}-\mu}{\sigma_{\bar{X}}} \leq \frac{398{,}27-\mu}{\sigma_{\bar{X}}}\right)$$

$$= 1 - \Phi\left(\frac{401{,}73-\mu}{0{,}8839}\right) + \Phi\left(\frac{398{,}27-\mu}{0{,}8839}\right).$$

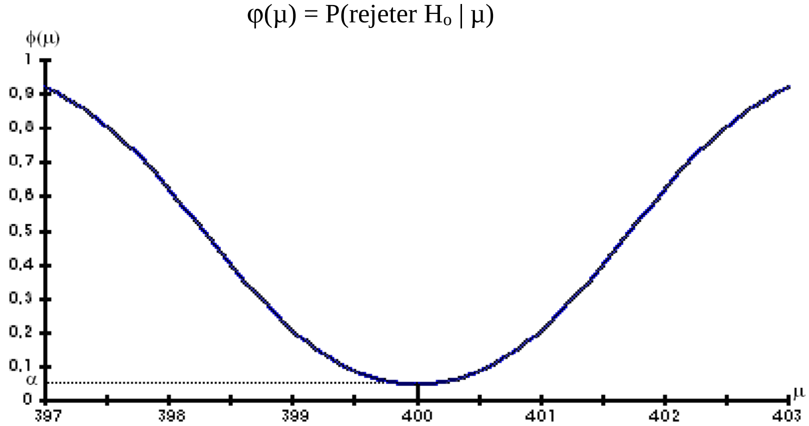
La puissance au point μ = 398, par exemple, est 0,6190. C'est la probabilité de rejeter l'hypothèse que μ = 400 lorsqu'en fait μ = 398. La figure 6.9.2 présente le graphique de cette fonction de puissance.

Comme il le faut, la puissance croît à mesure qu'on s'éloigne de la valeur μ = 400 : « plus H_o est fausse », plus la probabilité de le rejeter est forte. ■

Figure 6.9.2

Fonction de puissance pour le test de l'exemple 6.9.2

$\varphi(\mu)$ = P(rejeter H_o | μ)



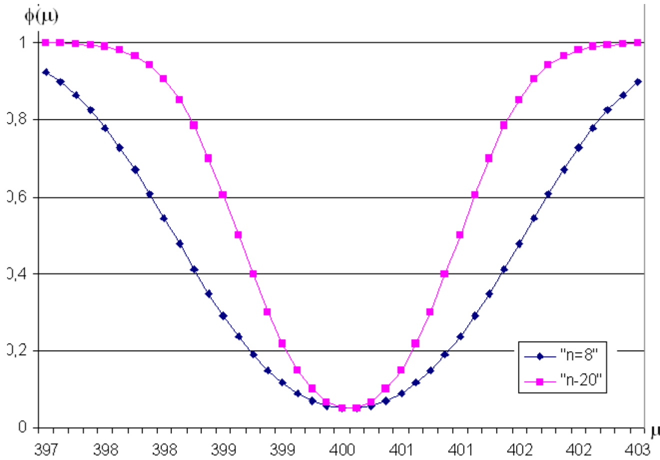
L'effet de la taille de l'échantillon

La notion de fonction de puissance est en particulier utile lorsqu'il existe plus d'une façon possible pour tester une même hypothèse.

Un test est conçu pour contrôler la probabilité d'une erreur de première espèce. Évidemment, on voudrait aussi que la probabilité d'une erreur de seconde espèce soit faible, c'est-à-dire, que la puissance soit élevée. Pour un α donné, on peut augmenter la puissance en prenant un échantillon plus grand. La figure 6.9.3 compare la fonction de puissance de la figure 6.9.2 (pour un test basé sur un échantillon de taille 8) avec la fonction de puissance pour un test de même niveau basé sur un échantillon de taille 20.

Figure 6.9.3

Comparaison de puissances pour un test basé sur un échantillon de taille 8 (*exemple* 6.9 .2) *et pour un test de même niveau basé sur un échantillon de taille* 20



RÉSUMÉ

1 *Région critique* : l'ensemble des valeurs de la statistique pour lesquelles on rejettera H_o.

Erreur de première espèce : Rejeter H_o lorsque H_o est *vraie*.
Erreur de seconde espèce : Accepter H_o lorsque H_o est *fausse.*
Taille de la région critique : Probabilité d'une erreur de première espèce

2 *Test de l'hypothèse* $H_o : \mu = \mu_o$. Le test est basé sur l'écart normalisé entre $\overline{X}$ et μ_o. L'écart-type de $\overline{X}$, $\sigma_{\overline{X}} = \sigma/\sqrt{n}$, est estimé par $\hat{\sigma}_{\overline{X}} = S/\sqrt{n}$ lorsque σ n'est pas connu .

Le tableau suivant donne la statistique de test et la leurs lois sous H_o.

Hypothèse	*Statistique*	*Distribution*
Population normale ou n grand, σ connu	$Z = \dfrac{\overline{X}-\mu_o}{\sigma_{\overline{X}}}$	$Z \sim N(0\,;1)$
n très grand, σ inconnu	$T = \dfrac{\overline{X}-\mu_o}{\hat{\sigma}_{\overline{X}}}$	$T \sim N(0\,;1)$
Population normale	$T = \dfrac{\overline{X}-\mu_o}{\hat{\sigma}_{\overline{X}}}$	$T \sim t_{n-1}$

3 *Test de l'hypothèse* $H_o : \delta = \delta_o$. $\delta = \mu_1 - \mu_2$

	Suppositions		
	σ_1 et σ_2 connus	*σ_1 et σ_2 inconnus*	
	Grands échantillon ou population normale	*Grands échantillons*	*Populations normales* $\sigma_1^2 = \sigma_2^2$
Écart-type de $\overline{X}-\overline{Y}$	$\sigma_{\overline{X}-\overline{Y}} = \sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}$	$\hat{\sigma}_{\overline{X}-\overline{Y}} = \sqrt{\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}}$	$\hat{\sigma} = \sqrt{\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$
Statistique	$Z_{\overline{X}-\overline{Y}} = \dfrac{\overline{X}-\overline{Y}-\delta_o}{\sigma_{\overline{X}-\overline{Y}}}$	$\hat{Z}_{\overline{X}-\overline{Y}} = \dfrac{\overline{X}-\overline{Y}-\delta_o}{\hat{\sigma}_{\overline{X}-\overline{Y}}}$	$T_{\overline{X}-\overline{Y}} = \dfrac{\overline{X}-\overline{Y}-\delta_o}{\hat{\sigma}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$
Loi sous H_o	$Z_{\overline{X}-\overline{Y}} \sim N(0\,;1)$ (*à peu près, si la population n'est pas normale*)	$\hat{Z}_{\overline{X}-\overline{Y}} \sim N(0\,;1)$ (*à peu près*)	$T_{\overline{X}-\overline{Y}} \sim t_{n_1+n_2-2}$

4 *Tests de l'hypothèse* $H_o : \sigma^2 = \sigma_o^2$ La statistique de test est $Q = \dfrac{(n-1)S^2}{\sigma_o^2}$. Lorsque H_o est vraie et la population est normale, $Q \sim \chi^2_{n-1}$.

5 *Test de l'hypothèse* H_o : $\sigma_1^2 = \sigma_2^2$. La statistique est $F = S_1^2/S_2^2$. Sous H_o, $F \sim F_{n_1-1,n_2-1}$

Hypothèse alternative	*Région critique*
$H_1 : \sigma_1^2 > \sigma_2^2$	$\dfrac{S_1^2}{S_2^2} \geq F_{n_1-1;n_2-1;\alpha}$
$H_1 : \sigma_1^2 \neq \sigma_2^2$	$\dfrac{S_1^2}{S_2^2} \geq F_{n_1-1;n_2-1;\alpha/2}$ ou $\dfrac{S_2^2}{S_1^2} \geq F_{n_2-1;n_1-1;\alpha/2}$

Expression générale

La multiplicité des tests dans ce chapitre est à première vue déconcertante. Il est bon de se rappeler que si les détails sont nombreux, les concepts sous-jacents ne le sont pas. Tous les tests traités dans ce chapitre (à l'exception des tests sur les variances), suivent essentiellement la même démarche, une démarche que l'on peut décrire en des termes généraux :

- Une hypothèse H_o porte sur un paramètre θ ;
- L'hypothèse nulle est de la forme $\theta = \theta_o$;
- Le test est basé sur un estimateur $\hat{\theta}$ de θ d'écart-type $\sigma_{\hat{\theta}}$;
- Il existe un estimateur $\hat{\sigma}_{\hat{\theta}}$ de $\sigma_{\hat{\theta}}$;
- La statistique de test est une mesure normalisée de l'écart entre $\hat{\theta}$ et θ, soit $T = \dfrac{\hat{\theta} - \theta_o}{\hat{\sigma}_{\hat{\theta}}}$;
- La région critique est de la forme $|T| \geq A$, ou $T \geq B$, ou $T \leq C$ selon que H_1 est $\theta \neq \theta_o$, $\theta \geq \theta_o$, ou $\theta \leq \theta_o$;
- Les points critiques A, B, ou C sont choisis de telle sorte que $P(|T| \geq A) = \alpha$, $P(T \geq B) = \alpha$, $P(T \leq C) = \alpha$;
- Les points critiques A, B et C dépendent de la loi sous H_o de la statistique T.

Variance des estimateurs :

Paramètre (θ)	Estimateur ($\hat{\theta}$)	Écart-type de l'estimateur $\sigma_{\hat{\theta}}$	Estimateur de $\sigma_{\hat{\theta}}$ ($\hat{\sigma}_{\hat{\theta}}$)
μ	$\bar{X}$	$\sigma_{\bar{X}} = \sigma/\sqrt{n}$	$S/\sqrt{n}$
$\mu_1 - \mu_2$	$\bar{X} - \bar{Y}$	Si $\sigma_1 \neq \sigma_2$: $\sigma_{\bar{X}-\bar{Y}} = \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$	$\hat{\sigma}_{\bar{X}-\bar{Y}} = \sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}$
$\mu_1 - \mu_2$	$\bar{X} - \bar{Y}$	Si $\sigma_1 = \sigma_2$: $\sigma_{\bar{X}-\bar{Y}} = \sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$	$\hat{\sigma}_{\bar{X}-\bar{Y}} = S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$ où $S = \sqrt{\dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$
p	$\hat{p}$ (proportion échantillonnale)	$\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$	$\hat{\sigma}_{\hat{p}} = \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$
$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}$	$\hat{\sigma}_{\hat{p}_1-\hat{p}_2} = \sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$
$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	Sous H_o : $p_1 = p_2 = p$ $\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{p(1-p)}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$	$\hat{\sigma}_{\hat{p}_1-\hat{p}_2} = \sqrt{\hat{p}(1-\hat{p})}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$ où $\hat{p} = \dfrac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1+n_2}$