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** 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** 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 3 000 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** 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. 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). Pour calculer cette probabilité, nous supposerons que la population est de loi $\mathcal{N}(\mu\,;\sigma^2)$ et que σ est connu, σ = 3,2. Alors

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

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

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

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:

Hypothèse nulle : H_o : μ = 450.

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

Alternative : H_1 : μ < 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

L'expression $\{\bar{X} < C\}$ ne définit pas entièrement la région critique, puisque le point critique C n'a pas été fixé; elle affirme seulement que nous rejetterons H_o si $\bar{X}$ est trop petit, donc inférieure ou égale à un certain nombre C (le point critique), un nombre qui reste à déterminer .

Considérons les conséquences d'un choix arbitraire, disons C = 449,5:

Une règle arbitraire : on rejettera H_o si et seulement si $\bar{X} < 449{,}5$.

On se propose de suivre cette règle, sachant qu'elle peut induire en erreur : elle peut mener au rejet de H_o, même si H_o est vraie. On appelle ce type d'erreur *erreur de première espèce*. Quelle elle est la probabilité d'une telle erreur? Elle est calculable (voir l'exemple 6.2.1):

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

Il y a donc une probabilité d'environ 24 % de rejeter H_o à tort. C'est au regard de cette probabilité qu'on décidera d'utiliser ou non la région critique $\{\bar{X} < 449{,}5\}$. On l'utilisera si on considère que cette probabilité est suffisamment petite pour nos besoins.

***Définition** Erreur de première espèce*

L'*erreur de première espèce* consiste à

rejeter H_o alors que H_o est vraie.

La probabilité d'une erreur de première espèce correspondant à la région critique $\{\bar{X} < 449{,}5\}$ est 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 d'une erreur de première espèce associée à **C**.*

Exemple 6.2.2 *Taille d'une région critique donnée*

Quelle est la taille de la région critique $\boldsymbol{C} = \{\bar{X} < 449{,}0\}$?

Solution

La taille de la région critique est la probabilité de rejeter H_o si en fait H_o est vraie.

Donc

$$\text{P(rejeter } H_o \mid H_o \text{ est vrai)} = \mathrm{P}(\bar{X} \le 449{,}0 \mid \mu = 450) = \mathrm{P}\left(\frac{\bar{X} - 450}{\sigma / \sqrt{20}} \le \frac{449{,}0 - 450}{\sigma / \sqrt{20}} \,\middle|\, \mu = 450\right)$$

$$= \mathrm{P}\left(Z \le -1{,}3955 \,\middle|\, Z \sim N(0\,;1)\right) = 0{,}0811.$$

On montre dans le tableau 6.2.1 la taille de quelques régions critiques considérées.

Tableau 6.2.1

Probabilités de rejet pour des régions critiques de la forme C:{ $\bar{X} \le C$}

	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	$\bar{X} < 449{,}5$	0,2423
2	449,0	$\bar{X} < 449{,}0$	0,0811
3	448,5	$\bar{X} < 448{,}5$	0,0180
4	448,0	$\bar{X} < 448{,}0$	0,0026
5	447,5	$\bar{X} < 447{,}5$	0,00024
6	447,0	$\bar{X} < 447{,}0$	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.

Exemple 6.2.3 *Quelle région critique choisir?*

Supposons qu'on se fixe un seuil d'acceptabilité à la taille de la région critique. Parmi les régions critiques présentées au tableau 6.2.1, lesquelles seraient acceptables si on fixe le seuil à

a) 0,10? b) 0,05? c) 0,01?

Réponses

a) Les régions 2 à 6 b) les régions 3 à 6 c) Les régions 4 à 6

Il appert, dans cette discussion, que le choix de la région critique dépendra d'un *seuil*, la limite supérieure à la taille de la région critique, évoquée implicitement ci-dessus. Théoriquement, ce seuil est fixé d'avance et la région critique est déterminée en fonction du seuil choisi.

Exemple 6.2.4 *Région critique respectant un seuil donné*

Dans l'exemple 6.2.1, supposons qu'on fixe le seuil à 0,05. Nous devons déterminer une région critique dont la taille est inférieure ou égale à 0,05.

Il faut donc déterminer C de telle sorte que $P(\bar{X} \le C \mid \mu = 450) \le 0{,}05$.

Puisque $\bar{X}$ est une variable continue, le seuil est atteignable: nous chercherons donc la valeur de C pour laquelle $P(\bar{X} \le C \mid \mu = 450) = 0{,}05$.

C'est une équation facilement résolue:

$$P(\bar{X} \le C \mid \mu = 450) = 0{,}05 \Rightarrow P\left(\frac{\bar{X} - 450}{\sigma / \sqrt{n}} \le \frac{C - 450}{\sigma / \sqrt{n}} \,\middle|\, \mu = 450\right) = 0{,}05$$

$$\Rightarrow P\left(Z \le \frac{C - 450}{3{,}2 / \sqrt{20}} \,\middle|\, Z \sim N(0\,;1)\right) = 0{,}05 \Rightarrow \frac{C - 450}{3{,}2 / \sqrt{20}} = -1{,}645$$

$$\Rightarrow C = 450 - 1{,}645 \frac{3{,}2}{\sqrt{20}} \Rightarrow C = 448{,}823.$$

Définition *Niveau d'un test*
Un test est de niveau α si la région critique est de taille maximale *parmi toutes celles de taille inférieure ou égale à α.*

Remarque *Niveau d'un test et taille d'une région critique*

Dans plusieurs applications, comme dans l'exemple traité jusqu'ici, la taille *de la région critique coïncide exactement avec le* niveau α *du test. C'est le cas dans l'exemple traité jusqu'ici parce que la variable sur laquelle le test est basé, $\bar{X}$, est une variable continue. La définition tarabiscotée ci-dessus est cependant nécessaire afin de tenir compte des situations où la variable du test est* discrète. *Dans ce cas, la région critique ne peut pas être fixée à* exactement α. *Sa taille est celle qui s'en rapproche le plus, tout en demeurant inférieure ou égale à α.*

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 :* $\mu \neq 450$. *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* $H_o : \mu \ge 450$ *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* $\{\bar{X} \le 448{,}8229\}$ est P(*rejeter* $H_o \mid H_o$) = $P(\bar{X} \le 448{,}8229 \mid \mu = 450) = 0{,}05$ *si* H_o *stipule que* $\mu = 450$. *Mais si* H_o *est exprimée comme* $H_o \ge 450$, *la taille de la région critique est* P(*rejeter* $H_o \mid H_o$) = $P(\bar{X} \le 448{,}8229 \mid \mu \ge 450)$, *qui prend plusieurs valeurs dépendantes de la valeur de* μ. *Mais sa valeur maximale* (*sur* $\mu \ge 450$) *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*

$$H_o : \mu = \mu_o,$$

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:

$$H_1\text{: } \mu < \mu_o, \quad H_1\text{: } \mu > \mu_o \quad \text{ou } H_1\text{: } \mu \neq \mu_o$$

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

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

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

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

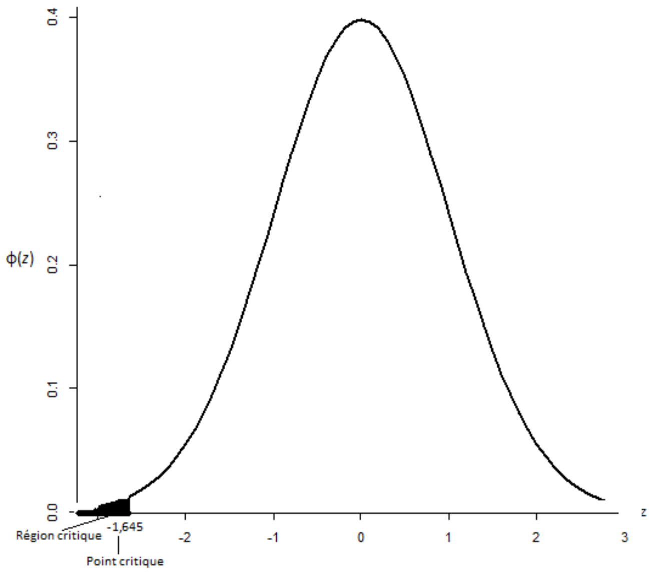
Or si H_o est vraie (et *seulement* si H_o est vraie), $Z \sim \mathcal{N}(0\,;\,1)$ et donc $C = -z_\alpha$ satisfait cette condition. La règle est donc

$$\textit{Rejeter } H_o \textit{ si } Z \leq -z_\alpha \Leftrightarrow \textit{Rejeter } H_o \text{ si } \bar{X} \leq \mu_o - z_\alpha \sigma_{\bar{X}} .$$

La figure 6.3.1 illustre la région critique:

Figure 6.3.1

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



<u>*Alternative* H_1: $\mu > \mu_o$</u>

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

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

<u>*Alternative* H_1: $\mu \neq \mu_o$</u>

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

$$|Z| \geq z_{\alpha/2}$$

En termes de $\bar{X}$ la règle est

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

***Exemple* 6.3.1**

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 α = 0,05. Alors z_α = 1,645, $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{3,2}{\sqrt{20}}$ = 0,7155 et on rejettera H_o si

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

Cette région critique est illustrée dans la figure 6.3.1.

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

$$\textit{On rejette } H_o \textit{ si } \bar{X} \leq 448,8229.$$

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 $|Z| \geq 1{,}96 \Leftrightarrow \bar{X} \leq 448{,}60$ ou $\bar{X} \geq 451{,}40$. ■

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

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 $\{Z \leq -1{,}645\}$ — 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 $P(Z \leq -3)$ sous H_o. On trouve $P(Z \leq -3) = 0{,}0014$, alors que $P(Z \leq -1{,}7) = 0{,}0446$. Cette probabilité est appelée *valeur p*, (*p-value* en anglais) et parfois « seuil expérimental ». Nous désignerons la *valeur p* par *vp*.

Voici une définition du seuil expérimental : la plus petite valeur de α pour laquelle on rejetterait H_o avec les données observées. Ou encore, lorsque la statistique de test est un indice de l'écart entre les données observées et celles attendues sous H_o, la *valeur p* est la probabilité, sous H_o, d'un écart aussi grand (ou plus) que celui qui a été observé — donc moins conforme à H_o. Plus cette probabilité est faible, plus H_o peut être rejetée avec fermeté.

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

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

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

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

***Exemple* 6.4.1**

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

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

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: $t_{\text{nu;alpha}}$ `=LOI.STUDENT.INVERSE.N(alpha;nu)`

Probabilité: $P(T \le t \mid T \sim t_{\text{nu}})$ = `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 suivant:

Test de l'hypothèse $\mu = \mu_o$; $T = \frac{\bar{X} - \mu_o}{\hat{\sigma}_{\bar{X}}}$.

T_{n-1}: une variable de loi de *Student* à n-1 degrés de liberté		H_1		
		$\mu < \mu_o$	$\mu > \mu_o$	$\mu \neq \mu_o$
Région critique	en fonction de T	$T \le -t_{n-1;\alpha}$	$T \ge t_{n-1;\alpha}$	$\lvert T \rvert \ge t_{n-1;\alpha/2}$
	en fonction de vp	$P(T_{n-1} \le t) \le \alpha$	$P(T_{n-1} \ge t) \le \alpha$	$P(\lvert T_{n-1} \rvert \ge \lvert t \rvert) \le \alpha$

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 $Z \le C$, $vp = P(Z \le z \mid H_o)$, où z est la valeur observé de Z. Si la région critique est de la forme $Z \ge C$, $vp = P(Z \ge z \mid H_o)$. Finalement, si la région critique est de la forme $|Z| \ge C$, $vp = P(|Z| \ge |z| \mid H_o)$.

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, $X_1, X_2, ..., X_{n_1}$ et $Y_1, Y_2, ..., Y_{n_2}$ et que :

$$X_i \sim N(\mu_1 ; \sigma_1^2), \quad i = 1, 2, ..., n_1$$

$$Y_j \sim N(\mu_2 ; \sigma_2^2), \quad j = 1, 2, ..., n_2$$

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:

$$\delta = \mu_1 - \mu_2$$

Considérons l'hypothèse

$$H_o : \ \delta = \delta_o$$

contre l'une des alternatives

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

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 $\sigma_{\bar{X}-\bar{Y}-\delta_o} = \sigma_{\bar{X}-\bar{Y}}$:

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

où

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

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 *L'hypothèse de normalité*

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

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

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

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

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 $\sigma_{\bar{X}-\bar{Y}} = \sqrt{\dfrac{\sigma^2}{n_1}+\dfrac{\sigma^2}{n_2}} = \sigma\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}$ et la différence normalisée est

$$z_{\bar{X}-\bar{Y}} = \frac{\bar{X}-\bar{Y}-\delta_o}{\sigma\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} \sim \mathcal{N}(0\,;\,1).$$

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

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

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

$$T_{\bar{X}-\bar{Y}} = \frac{\bar{X}-\bar{Y}-\delta_o}{\hat{\sigma}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$$

Sous H_o,

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

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.

$$\textit{Test de l'hypothèse } \delta = \delta_o\,;\, T = \frac{\bar{X}-\bar{Y}-\delta_o}{\hat{\sigma}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\ ;\ t = \text{valeur observée de } T$$

T_v: une variable de loi de *Student* à $v = n_1 + n_2$ -2 degrés de liberté		H₁ $\delta < \delta_o$	H₁ $\delta > \delta_o$	H₁ $\delta \neq \delta_o$
Région critique	en fonction de T	$T \le -t_{v;\alpha}$	$T \ge t_{v;\alpha}$	$\lvert T\rvert \ge t_{v;\alpha/2}$
	en fonction de *vp*	$P(T_v \le t) \le \alpha$	$P(T_v \ge t) \le \alpha$	$P(\lvert T_v\rvert \ge \lvert t\rvert) \le \alpha/2$

***Exemple* 6.5.1** 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 α = 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 Concernant le dernier exemple
Le fait que S_1 et S_2 sont 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** 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, $S_{\bar{d}} = S_d/\sqrt{22}$ 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 $T \leq 2{,}08$, 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.* ■

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 $\bar{X}$; 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

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

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} : une variable de loi de *khi-deux* à *n*-1 degrés de liberté		H₁ $\sigma < \sigma_o$	$\sigma > \sigma_o$	$\sigma \neq \sigma_o$
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 $Q \geq \chi^2_{n-1;\alpha/2}$
	en fonction 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** 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

$$H_o : \sigma_1^2 = \sigma_2^2$$

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 H_o 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 H_o 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

$$F = \frac{U_1 / \nu_1}{U_2 / \nu_2}$$

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 $(n_1-1)\,S_1^2/\sigma_1^2 \sim \chi^2_{n_1-1}$ et $(n_2-1)\,S_2^2/\sigma_2^2 \sim \chi^2_{n_2-1}$. D'où

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

Donc si H_o est vraie,

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

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

Test de l'hypothèse $\sigma_1 = \sigma_2$; $F = \dfrac{S_1^2}{S_2^2}$, f = valeur observée de F.

$F_{n_1-1;n_2-1}$: une variable de loi de *Fisher* à n_1-1 et n_2 - 1degrés de liberté	H₁		
	$\sigma_1 < \sigma_2$	$\sigma_1 > \sigma_2$	$\sigma_1 \neq \sigma_2$
Région critique en fonction de *vp*	$P(F_{n_1-1;n_2-1} \leq q) \leq \alpha$	$P(F_{n_1-1;n_2-1} \geq q) \leq \alpha$	$P(F_{n_1-1;n_2-1} \leq q) \leq \alpha/2$ ou $P(F_{n_1-1;n_2-1} \geq q) \leq \alpha/2$

Remarque *Un seul calcul est à faire, quelle que soit l'alternative. Supposons que* $S_1 > S_2$. *On calcule* $p = P(F_{n_1-1;n_2-1} \geq q)$ *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 d'égalité de 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 $P(F \geq 1{,}84) = 0{,}045$. 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 $P(F \geq 1{,}84) = 0{,}045 \geq \alpha/2$, 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 *On évitera le terme « accepter »*

*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 sur une moyenne*

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 $\sigma_{\bar{X}} = \frac{2{,}5}{\sqrt{8}} = 0{,}8839$.

Considérons la région critique

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

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

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

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

Si μ 398,

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

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 a 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 $\mathcal{N}(0\,;\,1)$:

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

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

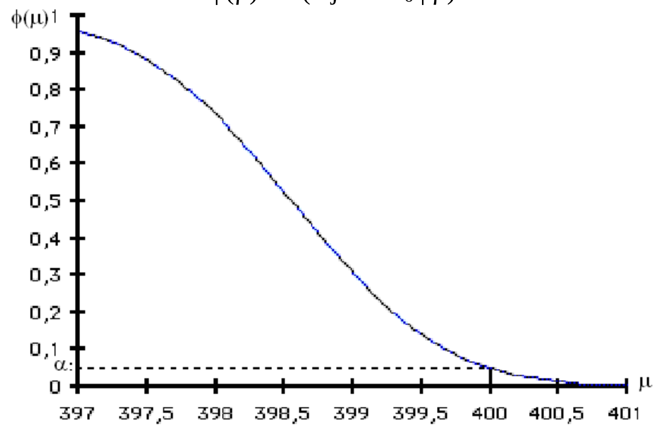
Toute l'information concernant les propriétés d'un test est contenue dans la fonction de puissance: la probabilité de rejeter H_o quand elle est vraie et la probabilité de la rejeter quand elle est fausse.

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(rejeter H_o | μ)



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(Rejeter H_o | H_o) = $P(\overline{X} \leq 398{,}55 \mid \mu \geq 400)$. *Cette probabilité dépend de la valeur de* μ*, mais quelle qu'elle soit, si* H_o *est vraie*, $\mu \geq 400$ *et conséquemment*, $P(\overline{X} \leq 398{,}55 \mid \mu \geq 400) \leq 0{,}05$). *La probabilité d'une erreur de première espèce est donc bornée supérieurement par* α

***Exemple* 6.9.2** *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 $\overline{X}$ 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

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

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

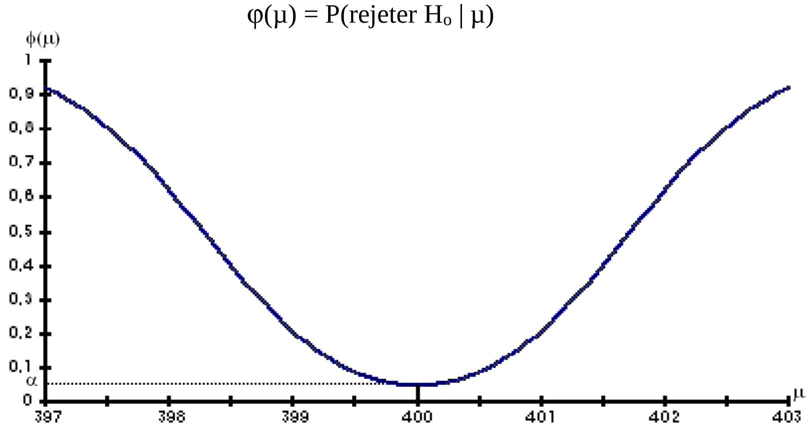
La puissance au point μ = 398, par exemple, est 0,6190. C'est la probabilité de rejeter l'hypothèse que μ = 400 lorsqu'en fait μ = 398. La figure 6.9.2 présente le graphique de cette fonction de puissance.

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

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



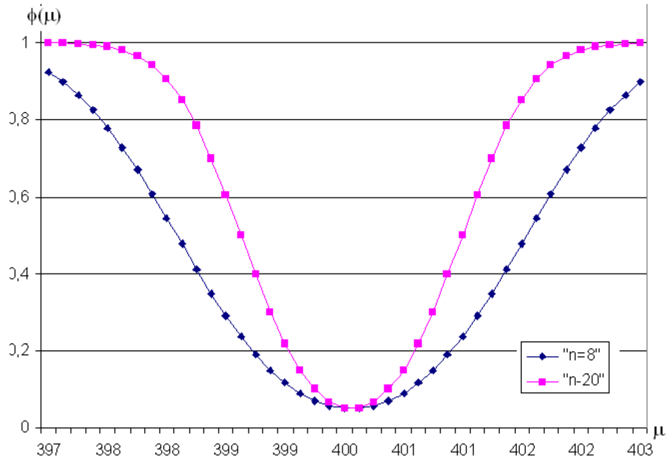
L'effet de la taille de l'échantillon

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 $\bar{X}$ et μ_o. L'écart-type de $\bar{X}$, $\sigma_{\bar{X}} = \sigma/\sqrt{n}$, est estimé par $\hat{\sigma}_{\bar{X}} = S/\sqrt{n}$ 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	$Z = \dfrac{\bar{X} - \mu_o}{\sigma_{\bar{X}}}$	$Z \sim \mathcal{N}(0\,;1)$
n très grand, σ inconnu	$T = \dfrac{\bar{X} - \mu_o}{\hat{\sigma}_{\bar{X}}}$	$T \sim \mathcal{N}(0\,;1)$
Population normale	$T = \dfrac{\bar{X} - \mu_o}{\hat{\sigma}_{\bar{X}}}$	$T \sim t_{n-1}$

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

	Suppositions		
	σ₁ et σ₂ connus	*σ₁ et σ₂ inconnus*	
	Grands échantillon ou population normale	*Grands échantillons*	*Populations normales* $\sigma_1^2 = \sigma_2^2$
Écart-type de $\bar{X} - \bar{Y}$	$\sigma_{\bar{X}-\bar{Y}} = \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$	$\hat{\sigma}_{\bar{X}-\bar{Y}} = \sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}$	$\hat{\sigma} = \sqrt{\dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$
Statistique	$Z_{\bar{X}-\bar{Y}} = \dfrac{\bar{X} - \bar{Y} - \delta_o}{\sigma_{\bar{X}-\bar{Y}}}$	$\hat{Z}_{\bar{X}-\bar{Y}} = \dfrac{\bar{X} - \bar{Y} - \delta_o}{\hat{\sigma}_{\bar{X}-\bar{Y}}}$	$T_{\bar{X}-\bar{Y}} = \dfrac{\bar{X} - \bar{Y} - \delta_o}{\hat{\sigma}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$
Loi sous H_o	$Z_{\bar{X}-\bar{Y}} \sim \mathcal{N}(0\,;1)$ (*à peu près, si la population n'est pas normale*)	$\hat{Z}_{\bar{X}-\bar{Y}} \sim \mathcal{N}(0\,;1)$ (*à peu près*)	$T_{\bar{X}-\bar{Y}} \sim t_{n_1+n_2-2}$

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, $F \sim \mathcal{F}_{n_1-1,n_2-1}$

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

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 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 $\hat{p} = \dfrac{n_1\hat{p}_1+n_2\hat{p}_2}{n_1+n_2}$

Exercices

6.1 [*Statistics in the real world*, Richard J. Larsen et Donna Fox Stroup, MacMillan, 1976, p.59] On a souvent affirmé que la lune a un effet sur la santé mentale de certaines personnes (ce qui explique le terme "lunatique"). Des études récentes ont tenté d'établir un lien entre les phases de la lune et le taux de suicides, la pyromanie, et même l'épilepsie. Une étude plus modeste a été faite dans une clinique psychiatrique en Virginie. On a noté le nombre d'admissions par jour durant la pleine lune pendant 12 mois consécutifs — du mois d'août 1971 au mois de juillet 1972. Voici les nombres de patients admis :

Mois	1	2	3	4	5	6	7	8	9	10	11	12
Nombre d'admissions	5	13	14	12	6	9	13	16	25	13	14	20

On sait que durant le reste de l'année, le nombre d'admissions moyen est de 11,3 par jour. Testez l'hypothèse que la lune n'a pas d'effet sur la santé mentale contre l'alternative selon laquelle la pleine lune provoque la maladie mentale.

6.2 Les données suivantes représentent les poids « avant et après » de 10 individus de même sexe qui ont suivi un régime d'amaigrissement de 5 semaines :

Individus	1	2	3	4	5	6	7	8	9	10
Avant	148	179	125	149	147	151	145	169	138	120
Après	144	162	126	131	132	146	145	152	127	118

Testez à 5% l'hypothèse que le régime n'a pas d'effet contre l'hypothèse qu'il fait réduire le poids.

6.3 [*Statistics in the real world*, Richard J. Larsen et Donna Fox Stroup, MacMillan, 1976, p.160] Une expérience est faite pour savoir si le glaucome (maladie caractérisée par une forte pression intra-oculaire) est généralement accompagné d'autres défauts de l'œil. Plus particulièrement, le but de l'expérience est de savoir si un œil glaucomateux a une cornée plus épaisse que celle d'un œil normal. Les participants à cette expérience souffraient du glaucome à un œil seulement, ce qui permettait de comparer l'œil glaucomateux à l'œil normal. Voici donc les épaisseurs des cornées des 8 patients

Patient	1	2	3	4	5	6	7	8
Œil normale	488	478	480	426	440	410	458	460
Œil glaucomateux	484	478	492	444	436	398	464	476

Qu'en pensez-vous ?

6.4 Dans une conserverie, un protocole de contrôle de la qualité consiste à tirer régulièrement un échantillon de 20 boîtes de conserves et de tester l'hypothèse

$$H_o : \mu = 200 \text{ g contre } \boldsymbol{H_1} : \mu < 200 \text{ g},$$

où μ est le poids moyen de toutes les boîtes de conserves de la population. Vous supposerez que les poids sont distribués normalement, d'écart-type connu $\sigma = 2$. On considère deux tests :

Test 1 : On rejette H_o si $Y \geq C$, où Y est le nombre de boîtes de l'échantillon qui pèsent moins de 200 g.
Test 2 : On rejette H_o si $\overline{X} \leq D$, où $\overline{X}$ est le poids moyen des boîtes de l'échantillon.

a) Déterminez la valeur de C dans le test 1 pour un niveau $\alpha = 0{,}05$. Quelle est la taille de la région critique ? Justifiez le choix de la forme $Y \geq C$ de la région critique.
b) Déterminez les valeurs de la fonction de puissance du test 1 aux points μ = 196 ; 197 ; 198 ; 199 ; 199,25 ; 199,50 ; 199,75 ; 199,8.
c) Déterminez la valeur de D qui donne au test 2 la même taille que le test 1.
d) Déterminez les valeurs de la fonction de puissance aux mêmes points qu'en b) et comparez.

6.5 La durée des ampoules d'un certain type est distribuée normalement d'écart-type 130. Si un échantillon de n = 50 ampoules donne une durée moyenne de 952,2 heures, testez à 5% l'hypothèse

$$H_o : \mu = 1000 \quad \text{contre} \quad H_1 : \mu < 1000.$$

(Trouver d'abord la région critique, c'est-à-dire, trouver le critère avant de l'appliquer au résultat obtenu $\overline{x} = 952{,}2$).

6.6 [*Données du tableau* A.5] Le tableau A.5 présente les températures prises par 130 sujets.

a) Testez l'hypothèse qu'il n'y a pas de différence entre la température moyenne des femmes et celle des hommes.
b) Testez l'hypothèse qu'il y a autant de dispersion entre les températures des femmes qu'entre les températures des hommes.

6.7 Pour tester l'hypothèse $H_o : \mu \leq 180$ contre $H_1 : \mu > 180$, où μ est la moyenne d'une population normale d'écart-type $\sigma = 24$, on prélève un échantillon de taille $n = 64$ et on décide de rejeter H_o si $\overline{X} \geq 186$.

a) Quelle est la probabilité de rejeter H_o si $\mu = 175$?
b) Vous ne pouvez pas calculer la probabilité exacte d'une erreur de première espèce. Mais quelle est la probabilité *maximale* d'une erreur de première espèce ?
c) Quelle est la probabilité maximale d'une erreur de seconde espèce ?

6.8 Déterminez une région critique de taille $\alpha = 0{,}05$ pour tester l'hypothèse $H_o : \mu \geq 1\,800$ contre $H_1 : \mu < 1\,800$ où μ est la moyenne d'une population normale d'écart-type $\sigma = 22$. L'échantillon est de taille $n = 121$. [*N.B.* Dans cet exercice, une région critique de taille α est une région critique telle que la probabilité d'une erreur de première espèce est *au maximum* α.]

6.9 Le poids réel du contenu de certaines boîtes de conserves est distribué normalement d'écart-type $\sigma = 0{,}2$ onces. Vous soupçonnez que le poids moyen μ est inférieur à la moyenne affichée de 8 onces. Un échantillon de taille 25 vous donne une moyenne de 7,95 onces. Déterminez un test unilatéral à 5% pour savoir si vous devriez déposer une plainte contre le fabricant.

6.10 [*Statistics in the real world*, Richard J. Larsen et Donna Fox Stroup, MacMillan, 1976, p.44] Les Grecs employaient souvent dans leur architectures des "rectangles d'or". Ce sont des rectangles de dimensions a et b ($a < b$) telles que $a/b = b/(a + b)$. Dans un rectangle d'or le rapport a/b vaut environ 0,618. Ce nombre a servi d'étalon esthétique non seulement dans l'art de la Grèce antique mais dans tout l'art occidental. Un anthropologue s'est demandé si l'art d'autres peuples non touchés par la civilisation grecque était aussi conforme à cet étalon. Il prélève un échantillon de 20 rectangles trouvés dans des œuvres artisanales des indiens Shoshone. Il calcule le rapport de la largeur sur la longueur pour chacun de ces rectangles. Voici les résultats, multipliés par 100.

69,3 ; 66,2 ; 69,0 ; 57,0 ; 74,9 ; 67,2 ; 62,8 ; 60,9 ; 84,4 ; 60,6
65,4 ; 61,5 ; 66,8 ; 57,6 ; 67,0 ; 60,6 ; 61,1 ; 55,3 ; 93,3 ; 60,1

Tester l'hypothèse qu'en moyenne le rapport de la largeur sur la longueur des rectangles trouvés dans l'artisanat des indiens Shoshone est conforme à l'idéal grec.

6.11 Le fabricant de la boisson *Energo* prétend que sa boisson a un effet énergisant. Afin de mettre cette prétention à l'épreuve, un chercheur a l'intention de faire passer une épreuve de course à 12 sujets. Chaque sujet sera soumis à l'épreuve deux fois : une fois juste après avoir consommé un grand verre d'*Energo*; et une autre fois avec un placebo (même couleur, même goût, mais sans l'ingrédient actif.) Il comptera ensuite le nombre X de sujets qui auront mieux performé sous l'effet de l'*Energo*. Il sait qu'*Energo* ne peut pas nuire et il est disposé à certifier que *Energo* est efficace à condition d'en avoir une preuve solide.

a) La variable X suit une loi
b) Énoncez l'hypothèse nulle H_o dans le langage du chercheur (en termes d'efficacité du produit, pas en termes statistiques) :
c) Énoncez l'hypothèse nulle H_o en termes statistiques (en termes des paramètres de la loi) :
d) Déterminer une région critique de niveau $\alpha = 0{,}01$ (calcul exact, pas d'approximation).

6.12 Dans le contexte du numéro précédent, un autre chercheur effectue la même expérience, mais avec 5 sujets seulement. L'épreuve de course consiste à parcourir une certaine distance, le plus vite possible, et on observe le temps (en secondes) mis à la parcourir. Voici les données :

Temps sous le placebo	22	7	11	15	16
Temps sous l'*Energo*	11	14	4	14	18

a) Définissez le modèle, c'est-à-dire, définissez les variables observées, les suppositions que vous faites concernant leur loi, et les paramètres de ces loi. Finalement, exprimer en termes de ces paramètres les hypothèses

H_o : Energo n'a aucun effet

H_1 : Energo a un effet énergisant

b) Déterminer une région critique à 10 % pour tester H_o versus H_1 :

c) Effectuer le test et tirer votre conclusion

6.13 On tire un échantillon de 46 cadres employés d'une *très* grande compagnie. Voici leurs salaires, classés selon le sexe :

Sexe	Nombre de cadres	Salaire moyen	Écart-type des salaires
Hommes	24	45 880 $	11 093
Femmes	22	37 002 $	10 413

Accusée de discrimination à cause de la différence entre les moyennes des hommes et celles des femmes, la présidente de la compagnie demande à son analyste du service des ressources humaines de lui dire si la différence est statistiquement significative.

a) Énoncez l'hypothèse nulle dans le langage de l'analyste (c'est-à-dire, en termes de discrimination)

b) Définissez le modèle (quelles sont les variables que vous observez? quelle loi suivent-elles?) et énoncez H_o en termes du modèle.

c) Effectuez votre test et exprimez clairement votre conclusion dans le langage de l'analyste. Faites un test bilatéral et posez $\alpha = 5$ %.

6.14 (*Suite du dernier numéro*) Pour être plus sûre, la présidente de la compagnie pose le problème à un deuxième analyste. Celui-ci procède différemment : il compte le nombre de femmes et le nombre d'hommes dont le salaire est supérieur à un certain seuil (36 500 $, jugé comme « moyen »). Voici ce qu'il obtient :

Sexe	Nombre de personnes	Nombre au-dessus du seuil
Hommes	24	16
Femmes	22	7

La question est essentiellement la même : est-ce qu'il y a vraiment une différence de salaire entre hommes et femmes? Décrivez le modèle, effectuez le test, et exprimez votre conclusion clairement. Posez $\alpha = 0{,}05$.

6.15 La teneur en nicotine de deux marques de cigarettes a un écart-type de $\sigma_1 = 1{,}2$ mg pour la première marque et $\sigma_2 = 1{,}4$ mg pour la deuxième. Pour tester l'hypothèse que $\mu_1 - \mu_2 = 2$, on a prélevé un échantillon de chacune des deux marques. Le premier échantillon, de taille 50, avait en moyenne 26,6 *mg* ($\bar{X}$) ; le deuxième, de taille 40, avait en moyenne 23,8 *mg* ($\bar{Y}$).

a) Tester à 5% l'hypothèse $H_o : \mu_1 - \mu_2 = 2$ contre $H_1 : \mu_1 - \mu_2 \neq 2$

b) Tester à 5% l'hypothèse $H_o : \mu_1 - \mu_2 = 2$ contre $H_1 : \mu_1 - \mu_2 > 2$

c) Calculer la puissance au point $\mu_1 - \mu_2 = 2{,}5$ pour le test en a) et le test en b).

6.16 Une enquête menée en 1950 et à nouveau en 1975 a montré qu'en 1950 la taille moyenne de 400 garçons de 10 ans était 135,13cm avec un écart-type de 6,1 cm (c.-à.-d. $\bar{X} = 135{,}13$ et $S_1 = 6{,}1$), alors qu'en 1975 la taille moyenne de 500 garçons de 10 ans était de 136,1 cm avec un écart-type de 6,4 cm (c.-à.-d. $\bar{Y} = 136{,}1$ et $S_2 = 6{,}4$). Tester à 5% l'hypothèse nulle que la population visée a connu un accroissement de taille de 1 cm contre l'alternative que l'accroissement a été inférieur à 1 cm.

6.17 Pour comparer deux marques de peinture, on emploie 4 bidons de chaque marque et on note la superficie qu'ils peuvent couvrir. Les 4 bidons de la première marque ont couvert 47,6 m^2 en moyenne avec un écart-type de 2,9 m^2 ; les 4 bidons de la deuxième marque ont couvert en moyenne 45,7 m^2 avec un écart-type de 2,41 m^2. Tester à 5% l'hypothèse nulle $\mu_1 = \mu_2$ contre l'alternative $\mu_1 - \mu_2 > 0$.

6.18 Pour déterminer si les habitants de deux îles du Pacifique sont de même race, un anthropologue détermine les indices céphaliques de 6 adultes de sexe masculin de chacune des îles. Il obtient les résultats suivants :

$$\bar{X}_1 = 77{,}4,\ \ \bar{X}_2 = 72{,}2$$
$$S_1 = 3{,}3,\ \ S_2 = 2{,}1$$

Tester à 1% l'hypothèse qu'il n'y a pas de différence réelle entre les deux séries. (Supposez que les populations sont normales et qu'elles ont la même variance).

6.19 Si 8 fusées à courte portée ont une erreur moyenne par rapport à la cible de $\bar{X}_1 = 29{,}87\,\text{m}$ avec un écart-type de $S_1 = 2{,}43\,\text{m}$ et si 10 fusées d'une autre sorte ont une erreur moyenne de $\bar{X}_2 = 23{,}16\,\text{m}$ avec un écart-type de $S_2 = 4{,}57\,\text{m}$, tester à 5% l'hypothèse nulle $\mu_1 - \mu_2 = 5$ contre l'alternative $\mu_1 - \mu_2 > 5$. Supposons que les populations sont normales et que leurs variances sont égales.

6.20 Les planchers des hôpitaux devraient-ils être tapissés ? Un tapis propre est-il aussi propre qu'un plancher lavé ? Si l'on pouvait être sûr que non, l'administrateur de l'hôpital disposerait d'un élément d'information important. Dans un hôpital du Montana, une expérience a été faite pour répondre à cette question. Seize chambres de malades ont été testées, dont huit avaient des tapis et huit autres n'en avaient pas. Les chambres étaient comparables autrement. En forçant de l'air à un taux constant à travers un vase de pétri, et en comptant ensuite le nombre de colonies de bactéries qui s'y forment, on a pu mesurer la quantité de bactéries en suspension dans les chambres. Voici les résultats :

No de la chambre avec tapis	#212	#216	#220	#223	#225	#226	#227	#228
Nombre de colonies par m^3	1,10	0,76	0,66	1,21	1,00	0,94	1,36	1,30
No de la chambre sans tapis	#210	#214	#215	#217	#221	#222	#224	#229
Nombre de colonies par m^3	1,12	0,77	0,35	0,67	1,11	1,03	0,94	1,27

Quelle est votre conclusion ? Faites un test unilatéral à 5%.

6.21 Une expérience agricole est menée dans le but de déterminer s'il est rentable d'arroser le maïs d'un engrais chimique. L'expérience comprenait 14 champs de maïs, chacun divisé en deux parties. L'une des parties était arrosée, l'autre pas. La récolte moyenne, en nombre de boisseaux par arpent, est donnée dans le tableau suivant :

Champs	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Parties arrosées	64,3	78,1	93,0	80,7	89,0	79,9	90,6	102,4	70,7	106,1	107,4	74,0	72,6	69,5
Parties non arrosées	78,1	74,4	86,6	79,2	84,7	75,1	87,3	98,8	70,2	101,1	83,4	65,2	68,1	68,4

Testez l'hypothèse que l'arrosage n'améliore pas le rendement.

6.22 [*Statistics in the real world*, Richard J. Larsen et Donna Fox Stroup, MacMillan, 1976, p.107] L'analyse de la teneur en métaux précieux des pièces de monnaie peut fournir une information précieuse aux études de l'histoire politique et culturelle d'une société. (Une pénurie d'argent a déjà incité le gouvernement canadien à frapper des pièces de 10 ¢ et de 25 ¢ à faible teneur en argent). Les données de cet exemple comparent la teneur en argent de deux séries de pièces de monnaie byzantines du XII^e siècle. Ces pièces font partie d'une grande quantité trouvée récemment suite à des fouilles archéologiques à Chypre. La teneur, en pourcentage, de 9 pièces de la première série, est :

No de la pièce	1	2	3	4	5	6	7	8	9
Teneur en %	5,9	6,8	6,4	7,0	6,6	7,7	7,2	6,9	6,2.

Pour la deuxième série, comprenant 7 pièces battues plus tard, les pourcentages sont

No de la pièce	1	2	3	4	5	6	7
Teneur en %	5,3	5,6	5,5	5,1	6,2	5,8	5,8

La différence entre les deux séries est-elle significative ?

6.23 [*Statistics in the real world*, Richard J. Larsen et Donna Fox Stroup, MacMillan, 1976, p.4] Le rôle qu'à joué le célèbre humoriste Mark Twain dans la guerre civile a longtemps été un sujet de controverse. Certains savants ont présenté comme évidence des exploits militaires de Mark Twain une série de 10 lettres publiées dans le New Orleans Daily Crescent durant l'année 1861. Les lettres constituaient une chronique des aventures de l'auteur, un dénommé "Quintus Curtis Snodgrass". Les historiens admettent que les événements décrits dans ces lettres sont authentiques, mais on n'a trouvé aucune trace du dénommé Quintus Curtis Snodgrass. Selon certains critiques, cependant, le style et l'esprit de ces lettres étaient typiques de ceux de Mark Twain. La question est donc : Mark Twain était-il l'auteur des lettres de Quintus Curtis Snodgrass ? Pour répondre à cette question, on a choisi 8 lettres écrites par Mark Twain et on se proposait de les comparer de façon quantitative aux lettres de Quintus Curtis Snodgrass. Un style est caractérisé par plusieurs variables. Pour les fins de cette étude, on en a choisi une seule : la fréquence des mots de 3 lettres. Pour les 8 lettres de Mark Twain, les fréquences sont les suivantes :

No de la lettre	1	2	3	4	5	6	7	8
Fréquence	0,225	0,262	0,217	0,240	0,230	0,229	0,235	0,217

Pour les 10 lettres de Quintus Curtis Snodgrass, les fréquences sont :

No de la pièce	1	2	3	4	5	6	7	8	9	10
Fréquence	0,229	0,205	0,196	0,210	0,202	0,207	0,224	0,223	0,220	0,201

Testez l'hypothèse que les lettres signées « Quintus Curtis Snodgrass » ont été écrites par Mark Twain.

6.24 [*Données du tableau* A.3] Testez l'hypothèse que la moyenne des scores G est la même pour les femmes et pour les hommes. Répétez l'exercice avec les scores V et P. Vous supposerez que l'écart-type des femmes et des hommes est le même pour les trois variables.

6.25 [*Données du tableau* A.3] Testez l'hypothèse que les hommes et les femmes ont le même poids en moyenne ; et testez l'hypothèse que les hommes et les femmes ont la même taille en moyenne. (On sait bien que ces hypothèses sont fausses ; vos tests devraient donc les rejeter).

6.26 [*Données du tableau* A.3] On affirme souvent que pour plusieurs variables, les valeurs sont plus dispersées chez les hommes que chez les femmes. Vérifiez si c'est bien vrai pour les variables dans ce tableau.

6.27 [*Données du tableau* A.4] Testez l'hypothèse que la méthode 2 n'est ni plus ni moins efficace que la méthode traditionnelle

a) en vous basant sur une comparaison des moyennes des pré-tests A1, A2 avec les post-tests B1, B2;
b) en vous basant sur le post-test B3.

6.28 [*Données du tableau* A.5] Y a-t-il une différence entre les hommes et les femmes quant à la température du corps ? Y a-t-il une différence au niveau des battements du coeur ?

6.29 [*Données du tableau* A.3] Testez l'hypothèse que la médiane des différences entre le test verbal (V) et le test global (G) est nulle. Pour ce faire, il suffit de tester l'hypothèse que la proportion des gens pour qui le score verbal est supérieur ou égal au score global est 1/2.

6.30 [*Données du tableau* A.3] Testez l'hypothèse que la grosseur du cerveau des hommes est en moyenne égale à la grosseur du cerveau des femmes. On pourrait considérer cette comparaison inacceptable, puisque les femmes ne sont pas aussi grandes en moyenne que les hommes, et qu'il serait donc normal qu'elles aient un cerveau moins gros. Refaites l'analyse en tâchant de tenir compte de cette critique.

6.31 [*Données du tableau* A.9] Y a-t-il une différence de tailles entre les sopranos et les altos ? Y en a-t-il une entre ténors et basses ?

6.32 [*Données du tableau* A.8] Faites un test d'hypothèse qui permette de confirmer ou d'infirmer l'affirmation suivante : Lorsqu'on croit à l'astrologie, on ne croit pas à la théorie de l'évolution. Il y a plusieurs façons de procéder ; vous devrez préciser la question, rassembler les données et formuler vos hypothèses de façon à pouvoir les tester en utilisant les techniques de ce chapitre.

6.33 [*Données du tableau* A.8] Faites un test d'hypothèse qui permette de confirmer ou d'infirmer l'affirmation suivante : Les gens qui s'opposent à l'avortement sont ceux qui croient à la vie après la mort. Il y a plusieurs façons de procéder ; vous devrez préciser la question, rassembler les données et formuler vos hypothèses de façon à pouvoir les tester en utilisant les techniques de ce chapitre.

6.34 [*Données du tableau* A.3] On veut savoir s'il y a une dépendance entre la taille du cerveau et l'aptitude au test de performance. À cette fin, on partage l'échantillon en deux groupes selon la taille du cerveau :

le premier groupe étant composé de ceux dont le cerveau est de taille inférieure à 90000 ; les autres appartenant au deuxième groupe. Dans chacun de ces groupes on compte le nombre de personnes avec un score supérieur à 110. Utilisez ces données pour tester l'hypothèse qu'il n'y a pas de relation entre la taille du cerveau et le score de performance. Vous devrez d'abord formuler cette hypothèse rigoureusement.

6.34 [*Données du tableau* A.10] Le but des données du tableau est de savoir si le traitement au calcium a pour effet de réduire la pression systolique. Effectuer le test qu'il faut pour arriver à une conclusion. Serait-il acceptable d'effectuer un test basé uniquement sur la pression systolique *à la fin* du traitement ?

6.36 [*Statistics in the real world*, Richard J. Larsen et Donna Fox Stroup, MacMillan, 1976, p.156] Les expériences sur la perception extrasensorielle sont souvent faites à l'aide des "Cartes de Zener". C'est un jeu de 5 cartes distinctes. Récemment, plusieurs recherches ont été faites pour déterminer si l'hypnose n'aurait pas pour effet de faciliter la perception extrasensorielle. Nous décrivons ici une expérience faites avec 15 étudiants. On a demandé à chacun d'eux de deviner l'identité de 200 cartes de Zener. Aux premiers 100 essais, l'étudiant était dans un état normal. Aux 100 essais suivants il était hypnotisé. A chaque essai le "message" était envoyé par un même "émetteur". L'émetteur était lui-même hypnotisé lorsque l'étudiant l'était. On prend note du nombre de réponses correctes parmi 100. Voici les résultats :

Étudiant	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
État normal	18	19	16	21	16	20	20	14	11	22	19	29	16	27	15
Sous hypnose	25	20	26	26	20	23	14	18	18	20	22	27	19	27	21

a) Peut-on conclure que l'hypnose facilite la perception extrasensorielle ? [*Suggestion* : X_i et Y_i sont les résultats du i^e étudiant, alors $D_i = Y_i - X_i$ est une mesure de l'effet de l'hypnose sur les capacités de cet étudiant. Faites votre test à partir des D_i.]

b) Pour les sujets hypnotisés, le nombre moyen de bonnes réponses est 21,7. Est-ce que ceci démontre que des gens ordinaires, sous l'effet de l'hypnose, ont des pouvoirs de perception extrasensorielle ?

6.37 Au numéro précédent, désignez par $Y_1, \ldots, Y_{15}$, les scores des 15 sujets sous hypnose. Supposez que la taille $n = 15$ de l'échantillon est assez grande pour que s'appliquent les théorèmes limites. Montrez que chacune des régions critiques unilatérales ci-dessous est valide, dans le sens qu'elle est à peu près de niveau α. (Notez que sous l'hypothèse H₀ que l'hypnose n'a pas d'effet, $E[Y_i] = 20$ et $Var[Y_i] = 16$.)

a) $\dfrac{\overline{Y} - 20}{S_Y / \sqrt{15}} \geq T_{14}$	b) $\dfrac{\overline{Y} - 20}{4 / \sqrt{15}} \geq z_\alpha$
c) $\displaystyle\sum_{i=1}^{15} \frac{(Y_i - \overline{Y})^2}{16} > \chi^2_{14;\alpha}$	d) $\displaystyle\sum_{i=1}^{15} \frac{(Y_i - 20)^2}{16} > \chi^2_{15;\alpha}$

6.38 [*Statistics in the real world*, Richard J. Larsen et Donna Fox Stroup, MacMillan, 1976, p.166] Les enfants qui ont de graves problèmes d'apprentissage et de comportement ont souvent le même type d'électroencéphalogramme que celui qu'on retrouve chez les enfants souffrant du petit mal épileptique. Ceci a mené à la conjecture que l'éthosuximide, un anticonvulsif qui sert à traiter le petit mal, pourrait faciliter l'apprentissage. Une expérience a été faite récemment pour tester cette théorie. Dix enfants, âgés de 8 à 14 ans ont servi de sujets. L'expérience a duré 6 semaines. Pendant 3 semaines l'enfant recevait un placebo, et pendant 3 autres semaines il était à l'éthosuximide. A la fin de chacune des deux périodes de 3 semaines, l'enfant composait plusieurs parties du test d'intelligence de Wechsler. De façon purement aléatoire on choisissait l'ordre dans lequel les traitements étaient appliqués : certains enfants recevaient l'éthosuximide durant les 3 premières semaines et le placebo ensuite ; et d'autres les recevaient dans l'ordre inverse. Voici les scores pour chaque sujet. L'ordre dans lequel chaque traitement a été donné n'est pas indiqué.

Patient	1	2	3	4	5	6	7	8	9	10
Sous placebo	97	106	106	95	102	111	115	104	90	96
Sous médication	113	113	101	119	111	122	121	106	110	126

Que concluez-vous ?

6.39 Un échantillon de taille 18 est tiré d'une population normale pour tester l'hypothèse $H_o : \sigma = 0{,}6$, contre l'alternative $H_1 : \sigma \neq 0{,}6$. Trouver une région critique de niveau 5%. Quelle serait votre décision si les données conduisent au résultat $S^2 = 0{,}68$. [Répétez avec $H_1 : \sigma > 0{,}6$].

6.40 Un test écrit de conduite est composé par 30 sujets. Les temps mis a compléter le test ont une variance échantillonnale de 6,4 (minutes)^2. Tester à 5% l'hypothèse que $\sigma^2 = 8$ contre l'alternative $\sigma^2 < 8$.

6.41 Les durées de vie de certaines batteries sont censées avoir un écart-type d'au moins 15 heures. Tester à 5 % l'hypothèse $\sigma = 15$ contre l'alternative $\sigma > 15$, sachant que 20 de ces batteries (qui constituent un échantillon aléatoire) ont une variance éch S_2^2 antillonnale de $S^2 = 397$.

6.42 Pour comparer la dispersion dans la résistance à la rupture de deux métaux, on fait une expérience avec $n_1 = 13$ pièces d'un métal et $n_2 = 16$ pièces de l'autre. Voici les résultats :

$$s_1^2 = 19{,}2 \text{ et } s_2^2 = 3{,}5$$

Tester l'hypothèse $\sigma_1^2 = \sigma_2^2$ contre l'alternative $\sigma_1^2 \neq \sigma_2^2$, avec $\alpha = 0{,}1$.

6.43 Voici les scores obtenus par neuf femmes mariées et neuf célibataires dans un test de personnalité.

Célibataires	88	68	77	82	63	80	78	71	72
Mariées	73	77	67	74	74	64	71	71	72

Testez l'hypothèse $\sigma_1^2 = \sigma_2^2$ contre l'alternative $\sigma_1^2 > \sigma_2^2$.

6.44 Une expérience a été faite afin d'établir l'existence d'un phénomène appelé « régression » par lequel un sujet, sous l'effet du stress, aurait tendance à recourir aux choses familières. Rappelons que 12 sujets avaient appris chacun deux façons d'attacher un nœud ; on leur a ensuite demandé, à un moment de stress, de faire un nœud de la façon qu'ils voulaient. L'expérimentateur étudiant a compté le nombre de sujets qui ont recouru à la première méthode apprise, considérant ce nombre comme une variable binomiale avec $n = 12$ et probabilité de succès p. Il a ensuite testé l'hypothèse que $p = 1/2$. Étant donné que 6 étudiants avaient appris la méthode A, puis la méthode B, et les 6 autres dans le sens inverse, cette façon de procéder n'est pas correcte. Qu'est-ce que le test de l'étudiant teste, exactement ? Quelle serait une meilleure façon de procéder ?

6.45 Un étudiant de Pennsylvania State University s'est donné pour projet de maîtrise l'étude de l'effet des boissons au Cola sur ceux qui en consomment en quantités abusives. Il s'est trouvé 57 étudiants qui consommaient quotidiennement de 48 à 111 onces de Cola. Comme groupe témoin, il a pris 112 étudiants qui ne consommaient pas plus de 8 onces par jour. Dans le premier groupe, 33 ont déclaré qu'ils se sentent tendus et nerveux quelques fois. Dans le groupe témoin, le nombre n'était que de 45. Cette différence est-elle significative ?

6.46 [*Statistics in the real world*, Richard J. Larsen et Donna Fox Stroup, MacMillan, 1976, p.205] On estime à environ 46 % le pourcentage des amputés qui succombaient aux infections post chirurgicales lorsqu'on n'employait pas de désinfectant dans les salles d'opérations. Joseph Lister, un chirurgien anglais, s'inspirant de l'œuvre de Louis Pasteur, a effectué 40 amputations en se servant du phénol comme désinfectant. De ces 40, seulement 6 sont morts d'une infection post-opératoire. Peut-on conclure que le phénol est bénéfique ?

6.47 [*Statistics in the real world*, Richard J. Larsen et Donna Fox Stroup, MacMillan, 1976, p.205] Au numéro précédent, le pourcentage de décès pour les patients sans désinfectant a en fait été estimé à partir d'un groupe témoins. Joseph Lister a traité 40 patients au phénol après amputation ; mais il a également pris soin d'observer 35 amputations faites sans désinfection au phénol. Voici les résultats :

		Phénol		
		Non	*Oui*	*Total*
Le patient a survécu	*Non*	16	6	22
	Oui	19	34	53
	Total	35	40	75

Que concluez-vous ?

6.48 La revue Time rapportait dans l'une de ses éditions [30 mars 1987] les résultats d'une étude effectuée par des chercheurs portant sur le taux de mortalité de 92 patients atteints de troubles cardiaques graves. 53 de ces patients possédaient un animal familier (chien, chat, poissons rouges,…); 3 de ceux-ci ont survécu moins d'un an. Parmi les 39 qui ne possédaient pas d'animal familier, 11 sont décédés durant la même période d'un an. Peut-on conclure la compagnie d'un animal familier a un effet sur la probabilité de décès (au cours de la première année) ?

6.49 Une étude des décisions rendues par des jurys dans des cas de vols par effraction où l'accusé était de race noire a révélé les faits suivants : parmi les 28 cas où les victimes étaient de race noire, l'accusé a été trouvé coupable dans 12 cas ; parmi les 36 cas où la victime était de race blanche, l'accusé a été trouvé coupable dans 23 cas. Peut-on conclure que les jurys ont une plus forte tendance à déclarer coupables ceux qui sont accusés d'avoir commis des vols contre des Blancs ?

6.50 Soit α le nombre moyen de fautes de frappe par page dans un gros livre. Comparez les fonctions de puissance correspondant aux deux procédures suivantes pour tester $H_o : \alpha = 1$ contre $H_1 : \alpha > 1$ (montrez que leurs tailles sont à peu près comparables, et que la puissance du deuxième test est supérieure à celle du premier pour α légèrement supérieur à 1. Présentez les fonctions de puissance sous forme graphique :

a) On observe le nombre total X de fautes dans un échantillon aléatoire de 10 pages ; on rejette H_o si $X \geq 17$;
b) On observe le nombre total Y de fautes dans un échantillon aléatoire de 20 pages ; on rejette H_o si $X \geq 30$.

6.51 Un défaut dans la fabrication des bouteilles se présente sous la forme de minuscules bulles dans le verre. Pour comparer deux procédés de fabrication, on prélève un échantillon de 24 bouteilles de chacun des procédés. Soit X le nombre total de bulles dans les 24 bouteilles livrées du premier procédé, et Y le nombre de bulles dans les bouteilles livrées du deuxième procédé. Supposez que X et Y sont de loi de Poisson. Testez l'hypothèse

$$H_o : \lambda_1 = \lambda_2$$

en vous basant sur la distribution conditionnelle de X étant donnée $X + Y = n$.

6.52 Une grande classe est constituée de 98 étudiants nés au Canada et 47 nés ailleurs. On constate que 20 des étudiants nés au Canada et 7 de ceux qui sont nés ailleurs portent des lunettes. La différence entre les deux groupes est-elle significative ?

6.53 On sait que normalement la probabilité d'avoir un garçon est de 0,515. On prélève un échantillon de 125 naissances parmi les membres d'une secte religieuse dont les mœurs sexuelles seraient susceptibles d'influencer le sexe du bébé. On trouve que 68 des nouveau-nés sont des garçons. Peut-on conclure que la probabilité d'avoir un garçon chez le gens de cette secte est différente de 0,515 ?

6.54 Un cardiologue du John Hopkins Medical Institution a rapporté les données suivantes au Congrès de l'American Heart Association à Dallas en 1986. Il avait observé 103 patients qui avaient eu une attaque du cœur et se rétablissaient sans complications. Trente des 103 avaient vécu des épisodes d'« ischémie silencieuse », un arrêt intermittent du débit sanguin au cœur. Une année plus tard, 9 de ceux-ci mouraient d'un arrêt cardiaque. Parmi les 73 qui n'avaient pas connu d'épisode d'ischémie silencieuse, 8 seulement mouraient d'un arrêt du cœur. Peut-on conclure que ceux qui souffrent de ces ischémies silencieuses sont plus vulnérables ? [Source : TIME, December 1,1986].

6.55 [Suite du numéro 6.54] Lors du même congrès, des chercheurs ont présenté les données suivantes sur 98 femmes âgées (plus de 50 ans) dont 51 avaient été victimes d'une attaque du cœur; les 47 autres non. On a constaté que parmi celles qui ont eu une attaque, 12 n'avaient jamais eu d'enfants, alors que parmi les 47 autres, 2 seulement n'avaient jamais eu d'enfants. Considérons un tirage dans la population à laquelle appartiennent ces patientes (des femmes de plus de 50 ans), et définisson les événements A et B comme suit :

A : la personne choisie a une crise cardiaque;
B : la personne choisie n'a pas eu d'enfants.

a) Estimez à partir des données les probabilités $P(B \mid A)$ et $P(B \mid A^c)$.
b) Testez l'hypothèse H_o : $P(B \mid A) = P(B \mid A^c)$.

c) Des données de ce genre sont généralement prélevées afin de tester l'hypothèse que le fait d'avoir eu ou pas des enfants n'influencent pas l'état de santé cardiaque. Or cette hypothèse se traduit plutôt par $P(A \mid B) = P(A \mid B^c)$ et non par $P(B \mid A) = P(B \mid A^c)$. Montrez qu'en fait ces deux hypothèses sont équivalentes.

6.56 On sait par expérience que 60 % des rats infectés d'un certain virus finissent par en mourir. Un chercheur a l'intention d'injecter le virus à un groupe de 20 rats afin de tester l'efficacité d'un certain médicament. Si p est la probabilité de décès d'un rat traité au médicament, il testera $H_o : p = 0,6$ contre $H_1 : p < 0,6$.

a) Supposons qu'en fait $p = 0,55$. Quelle est la probabilité que le chercheur rejette H_o (et donc conclue correctement que son médicament a un effet) ?
b) Le chercheur sait que si son médicament a un effet, il pourra réduire la probabilité de décès de 5 points de pourcentage (donc à 0,55), *pas plus*. Est-ce que l'expérience vaut la peine d'être effectuée. Discuter.

6.57 Supposons que le revenu médian des familles d'une certaine population est de 53 000 $. Dans un échantillon de 250 familles tiré parmi celles dont le chef est une femme on en trouve 24 dont le revenu est supérieur à 53 000 $. Tester l'hypothèse H_o que M, le revenu médian des ménages dont le chef de famille est une femme, est égal à celui des familles de la population entière contre l'hypothèse que M est inférieur à 53 000 $.

6.58 Considérons l'ensemble des maisons de Montréal habitées par leur propriétaire; et supposons que, pour les fins d'une enquête, on divise cet ensemble en blocs de 10 maisons chacun. On prélève au hasard un échantillon de 100 blocs, on interroge tous les propriétaires des blocs sélectionnés, et on note pour chaque bloc la valeur de la variable

X : nombre de propriétaires qui emploient les services d'un jardinier.

On obtient la distribution suivante :

x	0	1	2	3	4	5	6	7	8	
Effectif	25	10	12	9	8	16	10	7	3	100

Nous voulons tester l'hypothèse que 34,5 % des propriétaires utilisent les services d'un jardinier, ce qui est équivalent à l'hypothèse que le nombre moyen de propriétaires qui, dans un bloc, utilisent les services d'un jardinier est 3,45. Tester cette hypothèse

a) En considérant qu'on a prélevé un échantillon de 1000 maisons dans une population de maisons ;
b) En considérant qu'on a prélevé un échantillon de 100 blocs dans une population de blocs.

6.59 Pour connaître l'efficacité d'un certain régime d'amaigrissement, on le fait suivre à un groupe de 6 sujets. On prend leur poids *avant* et *après* le régime. Voici les résultats :

Sujet	1	2	3	4	5	6
Avant	64	54	73	59	64	68
Après	61	55	71	58	61	66

Peut-on conclure que le régime a un effet?

6.60 Dans le cadre d'une étude sur la sécurité routière et la visibilité nocturne des couleurs, on a conçu une expérience dans laquelle deux objets de même dimension, l'un rouge et l'autre jaune, ont été présentés à 218 sujets qui devaient déclarer si oui ou non ils parvenaient à les percevoir. 138 sujets ont vu des deux objets; 25 n'ont vu que l'objet jaune; 11 n'ont vu que l'objet rouge; et les 44 autres n'ont vu aucun des deux objets. Tester l'hypothèse que la probabilité de perception est la même pour les deux couleurs.

6.61 Dans une étude sur la relation entre certains traits de personnalité et des facteurs astrologiques, des chercheurs [Sakofske, Kelly et McKerracher, The Journal of Psychology 110, 275-80, 1982] ont fait compléter un questionnaire (le *Eysenck Personality Questionnaire*] à 241 étudiants néo-élandais. L'hypothèse (avancée antérieurement par des astrologues) que ces chercheurs se proposer de vérifier est que les personnes nées sous un signe positif (Bélier, Balance, Gémeaux, Lion, Verseau, Sagittaire) sont moins introverties que les personnes nées sous un signe négatif (Cancer, Capricorne, Poisson, Scorpion, Taureau, Vierge). Sur l'échelle introversion-extraversion du test, les extravertis ont un score élevé. Les sujets nés sous un signe positif, au nombre de 117, ont obtenu un score moyen de 13,28 avec un écart-type de 4,51. Les 124 sujets nés sous un signe négatif ont eu un score moyen de 14,28 avec un écart-type de 4,41. La différence entre les deux groupes est-elle significative ?

6.62 Pour comparer l'aptitude en mathématiques des Orientaux à celle des Américains de race blanche, un chercheur [Tsang, Sau-Lim, *Journal for Research in Mathematics Education* 15, No. 2, 115-122 (1984)] a prélevé les données suivantes sur les résultats obtenus au *Scholastic Aptitude Test* (SAT) par 20 097 Orientaux et 502 990 Américains de race blanche.

Score au SAT	Orientaux	Blancs
[700 ; 800[	601	22 564
[600 ; 700[	2001	86 521
[500 ; 600[	3190	158 049
[400 ; 500[	2788	151 466
[300 ; 400[	1309	74 498
[200 ; 300[	208	9892
Total	10 097	502 990

a) Tester l'hypothèse que la proportion de personnes d'habileté *supérieure* (ceux dont le score est ≥ 700) est la même parmi les Orientaux et parmi les Blancs.
b) Tester l'hypothèse que le score moyen de Orientaux est égal à celui des Blancs.
c) On aurait pu envisager la solution suivante au problème posé en b) : la moyenne pour les Blancs est de 510,5 et l'écart-type est de 112,5. On teste alors l'hypothèse que la moyenne μ des Orientaux est $\mu = 522{,}0$—un test sur une moyenne, soit avec σ supposé connu et égal à 112,5, soit en estimant σ par l'écart-type des scores des Orientaux. Effectuez le test et discutez de sa validité.
d) Les résultats en b) et c) sont très semblables. Pouvez-vous proposer une explication intuitive de ce fait ?

6.63 Un test d'« estime de soi » donne normalement un score moyen de 40 avec un écart-type (connu) de 16). Une psychologue voudrait vérifier une certaine conjecture selon laquelle les acteurs de théâtre jouissent d'une estime de soi supérieure à la moyenne. Si μ est leur score moyen, la psychologue veut tester l'hypothèse $H_o : \mu = 40$ contre $H_1 : \mu > 40$. Elle a l'espoir de convaincre au plus 8 acteurs à se soumettre au test, et vous demande si, avec un échantillon aussi petit, l'expérience vaut la peine d'être tentée. Selon elle, μ est probablement supérieur à 40, mais pas de beaucoup : peut-être 42, mais pas plus.

a) Si effectivement $\mu = 42$, quelle est la probabilité qu'elle puisse démontrer, avec un test à 5 %, ce qu'elle souhaite démontrer, soit que $\mu > 40$?

b) En supposant que μ est en fait égal à 42, dites-lui combien d'acteurs elle devra recruter pour avoir au moins une chance sur 2 d'affirmer ce qu'elle souhaite affirmer.

6.64 Supposons que la durée (en milliers d'heures) des pièces électroniques d'un certain type est de loi exponentielle de moyenne β. Pour tester l'hypothèse $H_o : \beta = 10$, on tire une pièce et on observe sa durée X. Considérez les deux tests suivants :

TEST 1 : on rejette H_o si $X < 0{,}5$
TEST 2 : on rejette H_o si $X < 0{,}2$

Le but de ce problème est d'étudier les avantages et désavantages relatifs de ces deux tests.

a) Déterminer la probabilité de rejeter H_o si H_o est vraie (i) lorsqu'on utilise le test 1, et (ii) lorsqu'on utilise le test 2. Lequel des deux tests paraît meilleur selon ce critère ?
b) Vérifier les probabilités présentées dans le tableau suivant :

	Probabilité de rejeter H_o	
Valeur de β	Test 1	Test 2
9	0,0540	0,0220
5	0,0952	0,0392
0,5	0,6321	0,3297
0,1	0,9933	0,8647
0,05	0,99995	0,9817

Lequel des deux tests semble meilleur à la lumière de ces calculs ?

c) Supposons qu’on décide qu’on rejettera H_o si $X \leq a$. Quelle doit être la valeur de a pour que la probabilité de rejeter H_o à tort soit de 5 % ?

6.65 Pour étudier l’effet de l’alcool sur les réflexes, on fait passer à 14 sujets un test de dextérité manuelle avant et après avoir consommé 100 ml de vin. Les scores avant et après sont présentés dans le tableau suivant (ce sont les temps de réaction; donc un score élevé dénote un ralentissement des réflexes).

Sujet	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Avant (X)	57	54	62	64	71	65	70	73	68	70	77	74	80	83
Après (Y)	55	60	68	69	70	73	74	74	75	76	76	78	81	90

a) Tester l’hypothèse que l’alcool n’a pas d’effet en utilisant le test basé sur la loi de *Student* .

b) Une autre façon de tester essentiellement la même hypothèse consiste à noter seulement la *direction* du changement, c’est-à-dire, s’il y a eu une baisse ou augmentation du score. Soit U le nombre de changements positifs (augmentation du score) et n le nombre de changements (qui peut être différent du nombre de sujets si certains sujets n’ont pas changé). Alors U est de loi binomiale de paramètres n et p.

   i) Exprimer l’hypothèse que l’alcool n’a pas d’effet en fonction de p ;

   ii) Effectuer un test d’hypothèse en utilisant comme seule information la valeur de U (et non les scores eux-mêmes).

c) Les conclusions en a) et b) sont les mêmes. Mais supposons que le 8^e sujet ait obtenu les scores 75 et 74 au lieu de 73 et 74.

   i) Tester encore l’hypothèse de deux manières ;

   ii) Expliquer les contradictions dans les conclusions.

6.66 Supposons que le nombre d’erreurs typographiques dans une page est de loi de Poisson de moyenne μ. Considérer deux procédures pour tester l’hypothèse $H_o : \mu = 1$.

Procédure 1 : on tire une page au hasard, on rejette H_o si le nombre d’erreurs X est supérieur ou égal à 4.

Procédure 2 : on tire deux pages au hasard ; on rejette H_o si le nombre total d’erreurs Y dans les deux pages est supérieur ou égal à 6.

a) Déterminer la probabilité d’une erreur de première espèce pour chacune des deux procédures.

b) Vérifier les probabilités dans le tableau suivant :

	Probabilité de rejeter H_o	
Valeur de θ	Test 1	Test 2
1,0	0,0190	0,0166
1,2	0,0338	0,0357
1,3	0,0431	0,0490
1,4	0,0537	0,0651

Lequel des deux tests semble meilleur à la lumière de ces calculs ?

c) Supposons qu’on décide de rejeter H_o si $X \geq a$. Quelle doit être la valeur de a si on veut que la probabilité d’une erreur de première espèce soit inférieure ou égale à 1 % ?

Exercices théoriques

6.67 Test de l’égalité de deux proportions avec de petits échantillons Considérez les données suivantes :

Population 1	X_1	$N_1 - X_1$	N_1
Population 2	X_2	$N_2 - X_2$	N_2

Supposons que X_1 et X_2 sont indépendantes et que $X_1 \sim \mathcal{B}(N_1, p_1)$ et $X_2 \sim \mathcal{B}(N_2, p_2)$. Montrez que sous l’hypothèse $H_o : p_1 = p_2$, la distribution conditionnelle de X_1 étant donné $X_1 + X_2 = n$ est $\mathcal{H}(n\ ;\ N_1\ ,\ N_2)$. Montrez comment se servir de ce fait pour tester H_o.

6.68 Une façon de tester l'égalité de deux moyennes consiste à déterminer un intervalle de confiance pour chacune des moyennes et de rejeter l'hypothèse d'égalité si les deux intervalles ne se recoupent pas. Montrer que cette procédure est conservatrice, dans le sens qu'elle donne une probabilité de rejet inférieur au niveau nominal α sous μ_o. Supposez que les variances σ_1^2 et σ_2^2 sont connues.

Montrez que sous μ_o, cette probabilité est supérieure à $P\left\{\left|\overline{X}-\overline{Y}\right|>z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right\}$.

6.69 Soit S_1^2 et S_2^2 les variances de deux échantillons indépendants issus de populations normales de même variance σ^2.

a) Montrez que $aS_1^2 + bS_2^2$ est un estimateur sans biais de σ^2 si et seulement si $a + b = 1$.

b) Montrez que parmi tous les estimateurs sans biais de la forme $aS_1^2 + bS_2^2$, celui qui minimise la variance est $aS_1^2 + (1-a)S_2^2$ où $a = (n_1 - 1)/(n_1 + n_2 - 2)$.

6.70 Un psychologue veut tester l'hypothèse qu'un certain entraînement mental permet d'augmenter le quotient intellectuel. Il considère l'une des deux expériences suivantes :

***Expérience* 1** On tire un échantillon de n sujets expérimentaux qu'on soumet à l'entraînement mental. On observe $Y_1, \ldots, Y_n$, les quotients intellectuels des sujets expérimentaux après l'entraînement. On tire aussi, dans la même population, n sujets témoins à qui on ne fait pas subir l'entraînement. On observe les quotients intellectuels $X_1, \ldots, X_n$ de ces sujets.

***Expérience* 2** On tire un échantillon de n sujets qu'on soumet à l'entraînement mental, mais on observe les quotients intellectuels $X_1, \ldots, X_n$ avant l'entraînement, ainsi que les quotients intellectuels $Y_1,\ldots,Y_n$ après l'entraînement.

Supposez que dans les deux cas l'écart-type des X_i et les Y_i est connu : $\sigma = 10$. Dans l'expérience Z_1, vous supposez que $\text{Cov}(X_i ; Y_i) = 80$, connue également ; dans l'expérience 1 toutes les variables sont indépendantes. Montrez que l'expérience 2 est plus efficace dans le sens que sa puissance est supérieure pour tout point $\delta = \mu_y - \mu_x > 0$. [*Suggestion* : montrez que $\text{Var}\left(\overline{Y}-\overline{X}\right)=\frac{200}{n}$ dans l'expérience 1 et $\text{Var}\left(\overline{Y}-\overline{X}\right)=\frac{40}{n}$ dans l'expérience 2.]

6.71 Pour évaluer l'efficacité du nouveau rasoir SuperMach4, on demande à 50 employés (de sexe masculin) de se raser au bureau : une joue avec le rasoir SuperMach4, l'autre avec l'ancien rasoir, le Superman2. On leur demande ensuite de dire lequel est meilleur. On trouve que 30 ont choisi le SuperMach4. Peut-on conclure qu'il y a une différence perceptible entre les deux rasoirs?

6.72 On conduit une expérience afin d'évaluer l'efficacité d'un certain traitement au calcium pour réduire la pression artérielle. Quinze personnes sont choisies comme sujets. On prend leur pression avant le traitement au calcium; puis on la prend encore après le traitement. Voici les résultats des 14 sujets :

	Sujet													
Pression	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Avant	120	135	145	167	180	155	112	125	138	144	139	125	188	175
Après	101	100	105	168	185	145	106	112	141	120	114	128	185	160

a) Une façon de tester l'hypothèse que le traitement n'a pas d'effet (H_o) consiste à prendre les différences *avant-après* et tester l'hypothèse que l'espérance mathématique (la moyenne) de ces différences est nulle. Effectuez ce test.

b) Mais si on suppose que la distribution des différences est symétrique, il y a une autre façon de faire : on compte le nombre X d'individus dont la pression a *baissé* à la suite du traitement au calcium. Alors $X \sim \mathcal{B}(n ; p)$ avec $n = 14$. Si le traitement n'a pas d'effet, $p = ½$. Tester donc l'hypothèse que le traitement n'a pas d'effet.

6.73 On sait que la teneur en protéines d'une douzaine d'œufs gros format est de 72 grammes, en moyenne, avec un écart-type de $\sigma = 15$. Un chercheur voudrait démontrer qu'un certain supplément de vitamines dans l'alimentation des poules fait croître la quantité de protéines dans les œufs qu'elles produisent. Il possède 9 poules, qu'il nourrira au régime enrichi, puis évaluera la teneur en protéines d'une douzaine d'œufs produite par chaque poule. On sait que la moyenne ne peut pas diminuer sous l'effet du supplément. Le chercheur s'attend en fait à un accroissement de 1 ou 2 grammes dans la moyenne, mais pas plus. Faites les calculs nécessaires pour savoir si l'expérience vaut la peine d'être faite. Il fera son test avec $\alpha = 0{,}05$. Que lui conseilleriez-vous de faire? Vous supposerez que $\sigma = 15$, indépendamment du régime alimentaire. [Ne vous préoccupez pas du fait qu'une douzaine d'œufs comprend 12 œufs: traitez la douzaine comme une unité].

6.74 On tire un échantillon de taille $n = 1$ d'une population de densité $f(x) = \theta e^{-\theta x}$ pour $x > 0$ et zéro ailleurs [Fonction de répartition : $F(x) = 1\text{-}e^{-\theta x}$, $x > 0$].

a) Déterminer une région critique de la forme $X \geq C$ de taille $\alpha = 0{,}1$ pour tester l'hypothèse H_o: $\theta = 2$.

b) Déterminer la taille de la région critique $\{X < 0{,}3)$.

c) Déterminer $\phi(4)$, où ϕ est la fonction de puissance correspondant au test décrit en b).

6.75 On considère trois tests différents de l'hypothèse $H_o : \theta = 2$, tous les trois de niveau $\alpha = 0{,}1$. Voici deux valeurs de la fonction de puissance φ pour chaque test :

Test 1 : $\phi(1) = 0{,}820$; $\phi(3) = 0{,}002$
Test 2 : $\phi(1) = 0{,}001$; $\phi(3) = 0{,}654$
Test 3 : $\phi(1) = 0{,}458$; $\phi(3) = 0{,}521$

Pour chacune des alternatives suivantes, dites lequel des 3 tests vous choisiriez :

Alternative $H_1 : \theta > 2$ Alternative $H_1 : \theta < 2$ Alternative $H_1 : \theta \neq 2$

6.76 Un certain procédé de fabrication produit des pièces dont le taux de défectuosité est stable : en moyenne **20** % des pièces sont défectueuses. Ce taux est bien établi, car on utilise ce procédé de depuis fort longtemps. Un nouveau procédé est maintenant utilisé et pour le comparer à l'ancien on tire un échantillon de **1000** pièces fabriquées selon le deuxième procédé. On y trouve **250** défectueuses. On effectue un test statistique afin de déterminer si cette différence observée est significative. Soit p le taux réel de défectuosité du *nouveau* procédé.

a) Lequel ou lesquels des énoncés suivants pourrai(en)t servir d'hypothèse nulle

E_1 : $p = 25$ % E_2 : $p = 20$ % E_3 $p \neq 25$ % E_4 : $p \neq 20$ %

b) Lequel ou lesquels des énoncés suivants pourrai(en)t servir d'hypothèse nulle?

E_1 : Le taux de défectuosité est le même dans les deux procédés
E_2 : Le taux de défectuosité du deuxième procédé n'est pas affecté par celui du premier
E_3 Le taux de défectuosité est-il plus élevé avec le nouveau procédé?
E_4 : Le taux de défectuosité du deuxième procédé est supérieur à celui du premier.

c) Supposons que la p-valeur est 0,001. Lequel ou lesquels des énoncés suivants est (sont) justifié(s)?

E_1 : La probabilité que p soit égale à 25 % est 0,001
E_2 : On peut conclure avec confiance que le taux de défectuosité est avec le nouveau procédé est de 25 %.
E_3 : On ne peut pas conclure avec confiance que le nouveau procédé est différent de l'ancien
E_4 : On peut conclure avec confiance que le taux de défectuosité est avec le nouveau procédé est différent de l'ancien.

d) Supposons que $vp = 0{,}6723$. Lequel ou lesquels des énoncés suivants est (sont) justifié(s)?

E_2 : On ne peut pas conclure avec confiance que $p \neq 20$ %
E_3 : On peut conclure avec confiance que $p = 25$ %
E_4 : On ne peut pas conclure avec confiance que les taux de défectuosité des deux procédés sont différents.