

Chapitre 7

Tests d'hypothèses Données qualitatives

7.1 Introduction

Au chapitre 6, les observations d'un échantillon X_1, X_2, \dots, X_n représentent des quantités mesurées — des poids, des longueurs, des résultats de tests psychométriques, etc. C'étaient donc des données *quantitatives*. Nous passons maintenant aux données dites *qualitatives* ou *catégorielles*: chaque élément de l'échantillon est *catégorisé*, et les observations seront des décomptes: le nombre d'observations appartenant à telle ou telle catégorie. Les observations seront de loi hypergéométrique, binomiale ou multinomiale; et les paramètres seront des *proportions*.

Les deux premiers tests sont une version qualitative de tests sur une moyenne: on teste l'hypothèse qu'une proportion (ou probabilité) p est égale à une valeur donnée p_0 ; et puis que la différence $\delta = p_1 - p_2$ entre deux proportions (ou probabilités) est égale à une valeur donnée δ_0 . Ces mêmes tests sont ensuite généralisés à des tests concernant une série de proportions (*test d'ajustement*) et une série de différences de proportions (*test d'indépendance*). Finalement, on reviendra à la différence entre deux proportions dans le cas de données appariées.

L'exemple suivant illustre une approche informelle à un test concernant une proportion:

Exemple 7.1.1 Un test informel

Une compagnie ayant conclu une entente d'achat d'un bloc de 80 appartements se propose d'examiner l'état des appartements à partir d'un échantillon de 20 unités tirées au hasard. Le vendeur affirme que 52 des unités sont dans un état impeccable (ne nécessitent pas de réparations majeures). L'acheteur a le droit d'annuler le contrat s'il est établi (à un certain degré de confiance que le nombre d'unités parfaites est inférieur à 52. L'échantillon est tiré, les appartements examinés et le nombre d'unités en état impeccable est compté: 3 seulement. Peut-on affirmer avec confiance que le nombre d'unités impeccables dans la population est inférieur à 52?

Solution La population est de taille $N = 80$ et l'échantillon de taille $n = 12$. Le nombre X d'unités impeccables dans l'échantillon est une variable de loi hypergéométrique $H(n; N_1, N_2)$, où N_1 est le nombre d'appartements impeccables et $N_2 = N - N_1$. Selon l'affirmation du vendeur, $N_1 = 52$, ce qui voudrait dire que la proportion p d'appartements impeccables dans la population est $p = 0,65$. Dans l'échantillon, l'observation $X = 3$ correspond à une proportion observée de $3/12 = 0,25$. Utilisant la fonction de probabilité de la loi hypergéométrique, on évalue la probabilité d'une valeur de X aussi petite que 3: $P(X \leq 3 \mid N_1 = 52) = 0,0028$, une valeur extrêmement petite qui permet d'affirmer avec confiance que $N_1 < 52$. ■

L'exemple ci-dessus est un test d'hypothèse sur le paramètre N_1 , mais il est équivalent à un test sur une proportion, soit $p = \frac{N_1}{N}$. L'hypothèse nulle est

$$H_0 : N_1 = 52 \text{ ou } p = \frac{52}{80} \text{ et l'alternative est } H_1 : N_1 < 52 \text{ ou } p < \frac{52}{80} .$$

Ce qui est calculé est la *valeur p* = 0,0028. On peut effectuer le test de façon plus formelle, une approche qui pourrait être stipulé dans l'entente.

Calcul par Excel: $P(X \leq x \mid n = n; N_1 = N_1; N = N) = \text{=LOI.HYPERGEOMETRIQUE.N}(x; n; N_1; N; 1) .$

Exemple 7.1.2 Suite de l'exemple 7.1.1

Par exemple, il pourrait être entendu, avant le tirage de l'échantillon, que le contrat peut être annulé seulement s'il est établi à un niveau de 5 % que $N_1 < 52$. Déterminer une région critique de niveau $\alpha = 0,05$ en termes de X

Solution Une région critique de la forme $X \leq C$ doit satisfaire la condition $P(X \leq C | N_2 = 52) \leq 0,05$. On choisira la plus grande région critique qui satisfait cette condition. Étant donné que X est une variable discrète, on peut procéder par tâtonnement: on détermine la taille de plusieurs régions critiques. Les voici pour $C = 0$ à 5.

C	0	1	2	3	4	5
$P(X \leq C N_2 = 52)$	0,000000505	0,000019	0,000308	0,002841	0,0168	0,0679

La région critique est donc $X \leq 4$. Sa taille est 0,0168, nettement inférieure au seuil $\alpha = 0,05$. Ceci est dû au fait que X est une variable discrète: la région critique $X \leq 5$ est de taille supérieure au seuil. Plus n est grand, cependant, plus on peut s'approcher du niveau α ■

Remarque Une « proportion » p n'est pas toujours un quotient concret comme dans les exemples précédents. Il représente plutôt une probabilité. Par exemple, la proportion p de pièces défectueuses produites par un procédé manufacturier est en fait la probabilité qu'une pièce soit défectueuse, une mesure de la qualité du processus de fabrication. C'est le cas de l'exemple suivant, où la population est essentiellement infinie, ce qui justifie l'emploi de la loi binomiale plutôt que la loi hypergéométrique. ■

Calcul par Excel de la probabilité

$$P(X \leq x | n = n; p = p; N = N) : \boxed{=LOI.BINOMIALE.N(x; n; p; 1)}$$

Exemple 7.1.3 Une hypothèse sur une proportion testé à l'aide de la loi binomiale

Un procédé de fabrication de boulons est considéré satisfaisant si le taux de défectuosité est de 2 % (ou moins, bien sûr). Un inspecteur prélève un échantillon de 200 boulons pour s'assurer que le procédé est bien sous contrôle. Soit X le nombre de boulons défectueux qu'on trouvera dans l'échantillon. Déterminer une région critique de niveau $\alpha = 0,05$ pour tester l'hypothèse

$$H_0 : p = 0,02$$

Solution $X \sim \mathcal{B}(200 ; p)$. Si H_0 est vraie, $E(X) = 200(0,02) = 4$. On rejettera H_0 si $X \geq C$, où C est le plus petit nombre supérieur à 4 pour lequel

$$P(X \geq C | p = 0,02) \leq 0,05.$$

Le tableau suivant présente la probabilité $P(X \geq C | p = 0,02)$ pour quelques valeurs de C .

C	5	6	7	8	9	10	11
$P(X \geq C p = 0,02)$	0,3712	0,2133	0,1086	0,0493	0,0202	0,0075	0,0025

La région critique de niveau 0,05 est donc $\{X \geq 8\}$. Elle est de taille 0,0493. ■

Dans les derniers exemples, les tests ont été exécutés à partir de la distribution exacte de l'observation X , soit la loi hypergéométrique (si la population est finie) soit la loi binomial (si la population est infinie, ou très grande). Si la taille de la population N est très grande, la loi hypergéométrique peut être approchée par la loi binomiale. Et si la taille de l'échantillon est grande (sans trop approcher N), la loi binomiale peut être approchée par la loi normale. Dans la plupart des applications réelles, ces conditions sont réunies: la population est très grande par rapport à l'échantillon (sinon infinie) et l'échantillon est assez grand pour que la loi normale soit utilisée pour approcher la loi binomiale.

7.2 Tests sur une proportion - Grands échantillons

L'exemple 7.1.3 traite d'un test d'hypothèse concernant une proportion. Lorsque n est grand, nous pouvons nous servir de la loi normale comme approximation de la distribution de $\hat{p} = X/n$. La loi de X , une loi $\mathcal{B}(n; p)$, peut être approchée par une loi $\mathcal{N}(np; np(1-p))$, et donc

$$\hat{p} \sim \mathcal{N}\left(p; \frac{p(1-p)}{n}\right)$$

Considérons l'hypothèse

$$H_0 : p = p_0$$

et l'alternative

$$H_1 : p \neq p_0.$$

La région critique peut s'exprimer en fonction de la statistique centrée-réduite

$$Z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}.$$

Lorsque H_0 est vraie, Z suit à peu près une loi $\mathcal{N}(0; 1)$. On rejette donc H_0 si

$$\frac{|\hat{p} - p_0|}{\sqrt{\frac{p_0(1-p_0)}{n}}} > z_{\alpha/2}.$$

Remarque Une autre expression pour l'écart-type

Le dénominateur de Z est l'écart-type de $\sigma_{\hat{p}}$ de \hat{p} lorsque H_0 est vraie. Mais on pourrait remplacer le dénominateur par une estimation de l'écart-type et baser le test sur la statistique

$$\hat{Z} = \frac{\hat{p} - p_0}{\hat{\sigma}_{\hat{p}}} = \frac{\hat{p} - p_0}{\sqrt{\hat{p}(1-\hat{p})/n}}$$

qui, elle aussi, suit à peu près une loi $\mathcal{N}(0; 1)$ lorsque n est grand et donc la région critique est $|\hat{Z}| > z_{\alpha/2}$:

$$\text{On rejette } H_0 \text{ si } \left| \frac{\hat{p} - p_0}{\sqrt{\hat{p}(1-\hat{p})/n}} \right| > z_{\alpha/2}.$$

Les deux tests ne sont pas équivalents et peuvent théoriquement mener à des conclusions différentes. Mais cette éventualité est peu probable lorsque n est grand. ■

Pour des tests unilatéraux, on a les régions critiques suivantes :

$$Z > z_{\alpha} \quad \text{si l'alternative est } H_1' : p > p_0$$

$$Z < -z_{\alpha} \quad \text{si l'alternative est } H_1'' : p < p_0$$

Exemple 7.2.1 Un test unilatéral

Un grossiste conclut avec son fournisseur de noix une entente selon laquelle 30 % des noix de chaque lot livré doivent être des noix de cajou. Le grossiste prévoit prélever un échantillon de 100 noix à chaque livraison afin de s'assurer que la proportion de noix de cajou ne soit pas inférieure à 30 %.

Lors d'une livraison, on trouve seulement 20 noix de cajou dans un échantillon de 100 noix. Déterminer un test unilatéral à 5 % et effectuer le test.

Solution Si p est la proportion de noix de cajou dans le lot, l'hypothèse nulle est $H_0 : p = 0,3$, l'alternative

$H_1 : p < 0,3$. La région critique est $Z < -1,645$, où $Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$.

$$Z = \frac{0,2 - 0,3}{\sqrt{\frac{0,3(1-0,3)}{100}}} = -2,18 < -1,645. \text{ Donc on rejette } H_0. \quad \blacksquare$$

Exemple 7.2.2 Comparaison des deux estimations de l'écart-type

Dans l'exemple 7.2.1, comparer $\hat{Z} = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$ à $Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ et les valeurs p correspondantes

(telles qu'approchées par la loi normale). Examiner la qualité des approximations normales en comparant les valeurs p d'un test basé sur \hat{Z} , d'un test basé sur Z et d'un test exact basé sur la loi binomiale à la valeur p d'un test basé sur $X =$ nombre de noix de cajou.

Solution $\hat{Z} = -2,5$, la valeur p est $P(\hat{Z} \leq -2,5 \mid H_0) = 0,0062$; $Z = -2,18$ et la valeur p est $P(Z \leq -2,18 \mid H_0) = 0,0145$.

X est de loi $\mathcal{B}(100; 0,3)$ sous H_0 et la valeur p exacte est $P(X \leq 20 \mid H_0) = 0,0165$. On constate que le test basé sur Z est ici plus précis que celui basé sur \hat{Z} . \blacksquare

7.3 Tests sur la différence entre deux proportions

Considérons deux variables binomiales indépendantes, X_1 et X_2 , où $X_1 \sim \mathcal{B}(n_1; p_1)$ et $X_2 \sim \mathcal{B}(n_2; p_2)$. La statistique pour tester l'hypothèse

$$H_0 : p_1 = p_2$$

contre

$$H_1 : p_1 \neq p_2$$

serait, idéalement,

$$\frac{\hat{p}_1 - \hat{p}_2}{\sigma_{\hat{p}_1 - \hat{p}_2}} \quad (7.3.1)$$

où

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \quad (7.3.2)$$

On estimerait donc $\sigma_{\hat{p}_1 - \hat{p}_2}$ par

$$\hat{\sigma}_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \quad (7.3.3)$$

Cependant, si on tient compte de H_0 ($p_1 = p_2 = p$ disons), alors

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}} = \sqrt{p(1-p)} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad (7.3.4)$$

et alors on estimera $\sigma_{\hat{p}_1 - \hat{p}_2}$ par

$$\hat{\sigma}_{\hat{p}_1 - \hat{p}_2} = \sqrt{\hat{p}(1 - \hat{p})} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad (7.3.5)$$

où \hat{p} est une estimation de la proportion commune aux deux populations.

Si $p_1 = p_2$, alors $X + Y$ est une variable de loi $\mathcal{B}(n_1 + n_2; p)$. p est alors estimée par

$$\hat{p} = \frac{X + Y}{n_1 + n_2} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2} \quad (7.3.6)$$

La statistique

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\hat{\sigma}_{\hat{p}_1 - \hat{p}_2}}$$

suit à peu près une loi $N(0; 1)$ sous H_0 , quelle que soit la façon dont $\sigma_{\hat{p}_1 - \hat{p}_2}$ est estimée. On rejette H_0 si $|Z| > z_{\alpha/2}$. Pour des tests unilatéraux, on a les régions critiques suivantes :

$$Z > z_{\alpha} \quad \text{si l'alternative est } H_1' : p_1 > p_2$$

$$Z < -z_{\alpha} \quad \text{si l'alternative est } H_1'' : p_1 < p_2.$$

Dans la suite de ce manuel, le dénominateur de Z sera (7.3.5).

Exemple 7.3.1 Test d'égalité de deux proportions

À l'occasion d'un sondage d'opinion mené auprès d'étudiants de l'UQÀM, on s'est intéressé à savoir si l'opposition à un mariage interconfessionnel émane d'une opposition généralisée au mariage à l'extérieur du groupe. Ce qui aurait pour implication que ceux qui s'opposent au mariage interconfessionnel tendent à s'opposer également au mariage interracial. Parmi les 90 répondants, 33 se sont dits opposés au mariage inter-religieux, contre 57 qui ne s'y opposaient pas. Dans le premier groupe, 22 s'opposaient aussi au mariage interracial, contre 10 dans le deuxième groupe. Peut-on conclure que ceux qui s'opposent au mariage inter-religieux sont plus nombreux à s'opposer aussi au mariage interracial ?

Solution On a deux échantillons, de tailles $n_1 = 33$ et $n_2 = 57$. Soit X_1 et X_2 le nombre de personnes qui s'opposent au mariage interracial dans le premier et deuxième groupe, respectivement.

$$X_1 \sim \mathcal{B}(33; p_1), X_2 \sim \mathcal{B}(57; p_2).$$

L'hypothèse nulle H_0 (que les deux groupes ne diffèrent pas quant à leur attitude sur le mariage interracial) est $H_0 : p_1 = p_2$.

On prendra pour alternative l'hypothèse $H_1 : p_1 \neq p_2$, même si on soupçonne fortement que p_1 ne peut pas être inférieur à p_2 : on le soupçonne, mais on s'abstient de l'affirmer d'avance.

On calcule : $\hat{p}_1 = 22/33 = 0,6667$ et $\hat{p}_2 = 10/57 = 0,1754$; $\hat{p} = \frac{22+10}{33+57} = 0,3556$;

$$\hat{\sigma}_{\hat{p}_1 - \hat{p}_2} = \sqrt{\hat{p}(1 - \hat{p})} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 0,1047; Z = 4,69.$$

Il est certain, donc, que l'attitude envers le mariage interracial est liée à l'attitude envers le mariage interconfessionnel. ■

Remarque Si au dernier exemple on avait estimé $\sigma_{\hat{p}_1 - \hat{p}_2}$ par la formule avec

$$\hat{\sigma}_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} = \sqrt{\frac{(22/33)(11/33)}{33} + \frac{(10/57)(47/57)}{57}} = 0,0963,$$

on aurait eu $Z = 5,10$, ce qui n'aurait pas changé la conclusion. ■

7.4 Tests d'ajustement

Le test de l'hypothèse $H_0 : p = p_0$ est basé sur une observation X de loi binomiale de paramètres n et p . Ce test est utile lorsque la variable observée est dichotomique: une variable qualitative qui n'a que deux modalités. Mais une variable qualitative a souvent plus de deux modalités, donc plusieurs proportions à tester—et qu'on souhaite tester simultanément. On effectue alors un test dit d'*ajustement*.

Un test d'ajustement est un test basé sur un vecteur $\mathbf{X} = [X_1; \dots; X_k]$ de loi multinomiale de paramètres n et $[p_1; \dots; p_k]$. L'hypothèse à tester porte simultanément sur toutes les composantes du vecteur de probabilités. Le cas particulier $k = 2$ est familier : c'est le test sur p discuté à la section 7.2.

Exemple 7.4.1 Test sur une proportion vu comme un test d'ajustement

Si on observe 100 naissances pour tester l'hypothèse que la probabilité d'avoir une fille est 0,5, les données peuvent être présentées sous la forme d'une distribution comme celle-ci :

	Filles	Garçons	Total
Effectifs	$X_1 = 48$	$X_2 = 52$	100

Bien qu'il y ait deux variables, X_1 et X_2 , dans le décor, l'une d'elles est redondante : si X_1 est connu, X_2 est totalement déterminé. X_1 suit une loi binomiale de paramètres $n = 100$ et p et l'hypothèse est $H_0 : p = 1/2$. La statistique de test, $Z = \frac{X_1/100 - 1/2}{\sqrt{\frac{(1/2)(1-1/2)}{100}}} = \frac{X_1 - 50}{\sqrt{100(1/2)(1-1/2)}}$, mesure

l'écart entre X_1 et 50, son espérance sous H_0 . On rejette H_0 si cet écart, en valeur absolue, est trop grand, donc si $|Z| > z_{\alpha/2}$ ou, ce qui est équivalent, si $Z^2 > z_{\alpha/2}^2$.

Le test est basé sur le fait que la statistique Z est à peu près de loi $\mathcal{N}(0; 1)$, ce qui entraîne que Z^2 est de loi χ_1^2 (à peu près). Donc la règle qui consiste à rejeter H_0 si $Z^2 > \chi_{1,\alpha}^2$ est un test de niveau α .

Les deux critères de rejet, $|Z| > z_{\alpha/2}$ et $Z^2 > \chi_{1,\alpha}^2$ sont équivalents car $z_{\alpha/2}^2 = \chi_{1,\alpha}^2$. ■

Nous commençons par décrire la procédure, sans justification, à l'aide d'un exemple, après quoi nous développerons la théorie et le modèle sous-jacents.

Hypothèse nulle, distributions observée et théorique — un exemple

Considérons les données suivantes sur un échantillon de 780 suicides prélevé afin de déterminer si le jour de la semaine a une influence sur les suicides. L'hypothèse à tester est

H_0 : La fréquence des suicides est la même pour tous les jours de la semaine.

On classe les suicides en 4 catégories, selon le moment de la semaine où le suicide a eu lieu : le début de la semaine (lundi), le milieu de la semaine (mardi à jeudi), la fin de la semaine (vendredi), et le week-end (samedi-dimanche). :

	Jour				Total
	Lundi	Mardi-jeudi	Vendredi	Samedi-dimanche	
Effectif observé	110	320	100	250	780

En termes quantitatifs, H_0 s'exprime comme ceci :

H_0 : La fréquence des suicides est la même pour tous les jours de la semaine

La distribution ci-dessus, appelée *distribution observée*, doit être comparée à la distribution *attendue* sous H_0 . Il est clair que si H_0 est vraie, les 780 suicides devraient se répartir de façon proportionnelle au nombre de jours : 1/7 le lundi ; 3/7 les mardi-jeudi ; 1/7 le vendredi ; et 2/7 les samedi-dimanche. Le tableau suivant présente cette distribution, appelée *distribution théorique* :

	Jour				Total
	Lundi	Mardi-jeudi	Vendredi	Samedi-dimanche	
Effectif théorique	$\frac{1}{7}780 = 111,4$	$\frac{3}{7}780 = 334,3$	$\frac{1}{7}780 = 111,4$	$\frac{2}{7}780 = 222,9$	780

Nous comparerons donc la distribution observée à la distribution théorique. Si l'écart est important, on rejettera H_0 .

Remarque *Le même raisonnement est à la base de tout test statistique. Une valeur postulée du paramètre donne lieu à une valeur attendue de l'observation. Dans l'exemple 7.1.1, la valeur postulée $p = 1/2$ entraîne une valeur attendue de X_1 , soit 50. La statistique Z mesure l'écart entre la valeur observée X_1 et la valeur attendue. L'hypothèse $p = 1/2$ est rejetée si l'écart est trop important. C'est encore ce que nous ferons dans ce chapitre, sauf qu'il faut maintenant mesurer l'écart entre un vecteur d'observations et un vecteur de valeurs attendues.* ■

Nous commençons donc par définir une mesure de l'écart entre une série d'effectifs et une autre.

Écart entre les effectifs observés et théoriques

L'écart entre les deux tableaux est mesuré par une quantité χ^2 et définie par

$$\chi^2 = \sum_i \frac{(O_i - T_i)^2}{T_i},$$

où les O_i sont les effectifs observés et les T_i sont les effectifs théoriques.

Dans l'exemple, on a

$$\chi^2 = \frac{(110-111,4)^2}{111,4} + \frac{(320-334,3)^2}{334,3} + \frac{(100-111,4)^2}{111,4} + \frac{(250-222,9)^2}{222,9} = 5,09.$$

Région critique

La région critique est exprimée en fonction de χ^2 . On rejettera H_0 si χ^2 est *trop grand* car un grand χ^2 reflète une grande différence entre les deux tableaux. La région critique sera donc de la forme

$$\chi^2 \geq C.$$

La constante C doit être choisie de telle sorte que la probabilité d'une erreur de première espèce soit égale (à peu près) à un certain nombre α . Il faut donc que

$$P(\chi^2 \geq C \mid H_0) = \alpha.$$

Pour déterminer la valeur de C , il faut connaître la distribution de χ^2 sous H_0 .

Théorème 7.4.1 *Distribution de la statistique χ^2*

Lorsque H_0 est vraie, la statistique χ^2 suit à peu près une loi χ^2_v , où

$$v = \text{nombre de cases} - 1$$

Notation Pour tout nombre positif $\beta < 1$, on désigne par $\chi_{v;\beta}^2$ le point tel que

$$P(\chi_v^2 \geq \chi_{v;\beta}^2) = \beta$$

où χ_v^2 est une variable aléatoire de loi khi-deux à v degrés de liberté.

Donc la région critique est :

$$\chi^2 \geq \chi_{v;\alpha}^2$$

où $\chi_{v;\alpha}^2$ le point tel que $P(\chi_v^2 \geq \chi_{v;\alpha}^2) = \alpha$ si χ_v^2 est une variable aléatoire de loi khi-deux à v degrés de liberté.

Dans l'exemple, $v = 3$ et si on prend $\alpha = 0,05$, le point critique est $\chi_{3;0,05}^2 = 7,8147$.

Calcul par Excel: $\chi_{nu;alpha}^2$ =LOI.KHIDEUX.INVERSE.DROITE(alpha;nu).

Conclusion Puisque $\chi^2 = 5,09 < 7,8147$, on ne rejette pas H_0 : on ne peut pas affirmer que le taux de suicide varie selon le jour de la semaine. L'écart entre les deux séries d'effectifs (observés et théoriques) pourrait bien être dû au hasard tout seul.

Le modèle

Ce n'est pas tout vecteur de k effectifs qui se prête à un test du khi-deux. Il importe de s'assurer que le vecteur des effectifs observés est de loi multinomiale, de la statistique χ^2 en dépend.

On suppose donc que les valeurs observées constituent un vecteur

$$\mathbf{X} = [X_1; X_2; \dots; X_k] \sim \mathcal{M}(n; p_1; p_2; \dots; p_k).$$

L'hypothèse nulle est de la forme

$$H_0 : p_1 = p_{1o}, p_2 = p_{2o}, \dots, p_k = p_{ko}.$$

où $p_{1o}, p_{2o}, \dots, p_{ko}$ sont des nombres positifs tels que $\sum_i p_{io} = 1$.

Les effectifs théoriques T_i sont les *espérances* des X_i sous H_0 , soit

$$T_i = E(X_i | H_0) = np_{io}$$

La statistique χ^2 est

$$\chi^2 = \sum_{i=1}^k \frac{(X_i - np_{io})^2}{np_{io}}$$

Dans ces conditions, si l'hypothèse nulle est vraie, alors χ^2 suit à peu près une loi χ_v^2 où $v = k - 1$.

Remarque Le cas $k = 2$

Lorsque $k = 2$, le test d'ajustement se réduit à un test concernant une proportion p . La loi multinomiale est alors une loi binomiale, l'hypothèse $H_0 : p_1 = p_{1o}, p_2 = p_{2o}$ est l'hypothèse $H_0 : p = p_o$, l'énoncé $p_2 = p_{2o}$ (ou $1 - p = 1 - p_o$), étant redondant. ■

Remarque L'hypothèse d'une loi multinomiale

Il faut savoir que les deux conditions qui justifient la loi multinomiale — indépendance et probabilités fixes — sont facilement violées. Dans l'exemple des suicides, deux observations qui font partie

d'un pacte de suicide ne sont pas indépendantes. Dans le même exemple, si l'échantillon de 780 suicides est constitué de quatre échantillons provenant de quatre pays différents (ou de quatre époques différentes), l'hypothèse que les probabilités p_i sont les mêmes à chaque tirage n'est pas nécessairement vérifiée.

Un autre exemple. Supposons qu'on tire n bulletins de votes dans une élection contestée par quatre partis. Soit $\mathbf{X} = [X_1 ; X_2 ; X_3 ; X_4]$ la distribution des votes ($X_i =$ nombre de votes pour le parti i). Dans quelles conditions ces variables sont-elles de loi multinomiale ? Voici quelques situations où l'hypothèse d'indépendance est douteuse :

- Les résultats de $n = 200$ bulletins tirés dans une population de 300 bulletins ne sont pas indépendants (à moins que les tirages se fassent avec remise, ou que la population comprenne 10 000 bulletins et non 300, rendant les tirages pratiquement indépendants).
- L'hypothèse que les quatre probabilités ne changent pas d'une épreuve à l'autre n'est probablement pas vérifiée si l'échantillon est tiré de quatre circonscriptions différentes.
- Supposons qu'on suit la procédure suivante: on sélectionne au hasard 10 circonscriptions, puis on tire 20 bulletins dans chacune d'elles. Ici l'hypothèse d'indépendance ainsi que l'hypothèse que les probabilités sont les mêmes à chaque tirage sont toutes deux violées.

Encore un exemple. On tire un échantillon (à partir des dossiers de la police) de n personnes blessées dans des accidents de voiture. On classe les personnes selon le jour de la semaine où l'accident a eu lieu, ce qui donne les effectifs O_1, \dots, O_7 . L'indépendance des épreuves est-elle respectée ? Certainement pas si, par exemple, 10 des personnes choisies ont été impliquées dans un même carambolage. Si les O_i comptent les accidents plutôt que les personnes impliquées, le modèle multinomial s'appliquerait sans doute bien mieux. ■

7.5 Tests d'indépendance

Si l'hypothèse d'ajustement généralise l'hypothèse $p = p_0$, un test d'indépendance généralise le test d'égalité de deux proportions. Nous présenterons un cas particulier avant de décrire la procédure formellement. Considérons les données suivantes sur deux variables, la *scolarité* S et l'*attitude* face à l'avortement A :

		A : Attitude face à l'avortement			
		Pour	Mixte	Contre	
S : Scolarité	≤ 8	31	23	56	110
	9 — 12	171	89	177	437
	> 12	116	39	74	229
Total		318	151	307	776

Ces données ont été recueillies afin de déterminer s'il y a une relation entre les deux variables. Le tableau suivant présente la distribution conditionnelle de l'attitude étant donné chaque niveau de scolarité :

		A : Attitude face à l'avortement			
		Pour	Mixte	Contre	
S : Scolarité	≤ 8	0,28	0,21	0,51	1
	9 — 12	0,39	0,20	0,41	1
	> 12	0,51	0,17	0,32	1

Une certaine dépendance se manifeste clairement dans ces distributions conditionnelles. On y voit, par exemple, que la proportion de «pour» augmente avec le niveau de scolarité. Il s'agit maintenant de savoir si cette dépendance, évidente au niveau de l'échantillon, existe aussi au niveau de la population. On doit *tester* l'hypothèse d'indépendance.

La logique du test d'indépendance est, à quelques détails techniques près, identique à celle d'un test d'ajustement : On énonce H_0 , on construit un tableau d'effectifs théoriques, on mesure l'écart entre les effectifs observés et les effectifs théoriques et on rejette H_0 si l'écart est trop grand.

Hypothèse nulle

L'hypothèse nulle est

H_0 : les variables *scolarité* et *attitude face à l'avortement* sont indépendantes.

Calcul des effectifs théoriques

Nous devons déterminer les effectifs théoriques sous l'hypothèse d'indépendance. Ce sont les effectifs auxquels on doit s'attendre lorsque S et A sont indépendantes. Considérons la distribution marginale de la variable *Attitude face à l'avortement*, dans le tableau suivant :

		A : Attitude face à l'avortement			
		Pour	Mixte	Contre	
S : Scolarité	≤ 8				110
	9 - 12				437
	> 12				229
Marginale		$\frac{318}{776}$	$\frac{151}{776}$	$\frac{307}{776}$	1

Si les deux variables sont indépendantes (H_0 est vraie), les distributions conditionnelles de A , étant donné les valeurs de S devraient être les mêmes, et donc les mêmes que la marginale, c'est-à-dire $\frac{318}{776}$, $\frac{151}{776}$, et $\frac{307}{776}$. On répartit donc l'effectif total de chaque groupe de scolarité (110, 437, et 229) selon ces fréquences :

		A : Attitude face à l'avortement			
		Pour	Mixte	Contre	
S : Scolarité	≤ 8	$\frac{318}{776} \times 110 = 45,08$	$\frac{151}{776} \times 110 = 21,40$	$\frac{307}{776} \times 110 = 53,52$	110
	9 - 12	$\frac{318}{776} \times 437 = 179,08$	$\frac{151}{776} \times 437 = 85,03$	$\frac{307}{776} \times 437 = 172,89$	437
	> 12	$\frac{318}{776} \times 229 = 93,84$	$\frac{151}{776} \times 229 = 44,56$	$\frac{307}{776} \times 229 = 90,60$	229
Total		$\frac{318}{776} \times 776$	$\frac{151}{776} \times 776$	$\frac{307}{776} \times 776$	776

Mesure de l'écart entre les effectifs observés et les effectifs théoriques

On mesure l'écart entre les effectifs observés et les effectifs théoriques exactement comme avec un test d'ajustement :

$$\begin{aligned}
 \chi^2 &= \frac{(31 - 45,08)^2}{45,08} + \frac{(23 - 21,40)^2}{21,40} + \frac{(56 - 53,52)^2}{53,52} \\
 &+ \frac{(171 - 179,08)^2}{179,08} + \frac{(89 - 85,03)^2}{85,03} + \frac{(177 - 172,89)^2}{172,89} \\
 &+ \frac{(116 - 93,84)^2}{93,84} + \frac{(39 - 44,56)^2}{44,56} + \frac{(74 - 90,60)^2}{90,60} \\
 &= 17,71
 \end{aligned}$$

Le point critique

On rejettera H_0 si $\chi^2 \geq C$, où C est choisi de telle sorte que $P(\chi^2 \geq C) = \alpha$. Le théorème suivant permet de déterminer C .

Théorème 7.5.1 *Si les effectifs théoriques sont assez grands, la statistique χ^2 suit à peu près une loi χ^2_v sous H_0 . Le nombre de degrés de liberté est*

$$v = (\ell - 1) \times (c - 1)$$

où ℓ est le nombre de lignes du tableau et c est le nombre de colonnes.

Le point critique est donc $\chi^2_{(\ell-1)(c-1)}$ et la région critique est

$$\chi^2 > \chi^2_{(\ell-1)(c-1); \alpha}$$

Ici $(\ell-1)(c-1) = (3-1)(3-1) = 4$ et $\chi^2_{4;0.05} = 9,4877$. Puisque $\chi^2 = 17,71 \geq 9,4877$, nous rejetons H_0 à 5 %. Nous concluons que dans la population il y a vraiment une dépendance entre la scolarité et l'attitude face à l'avortement.

Le modèle

Dans un tableau de dimension $\ell \times c$, les effectifs théoriques

					Sommes
	X_{11}	X_{12}	...	X_{1c}	$X_{1\cdot}$
	X_{21}	X_{22}	...	X_{2c}	$X_{2\cdot}$
	\vdots	\vdots	\ddots	\vdots	\vdots
	$X_{\ell 1}$	$X_{\ell 2}$...	$X_{\ell c}$	$X_{\ell \cdot}$
Sommes	$X_{\cdot 1}$	$X_{\cdot 2}$...	$X_{\cdot c}$	n

suivent conjointement une loi multinomiale. Dans l'exemple, les X_{ij} sont les effectifs du tableau original. Les paramètres de la loi multinomiale sont n , le nombre total d'observations, et l'ensemble des probabilités

					Sommes
	p_{11}	p_{12}	...	p_{1c}	$p_{1\cdot}$
	p_{21}	p_{22}	...	p_{2c}	$p_{2\cdot}$
	\vdots	\vdots	\ddots	\vdots	\vdots
	$p_{\ell 1}$	$p_{\ell 2}$...	$p_{\ell c}$	$p_{\ell \cdot}$
Sommes	$p_{\cdot 1}$	$p_{\cdot 2}$...	$p_{\cdot c}$	n

Le tableau suivant présente les 9 paramètres relatifs à l'exemple, ainsi que certaines fonctions de ces probabilités, soit les sommes des lignes $p_{1\cdot}$, $p_{2\cdot}$, $p_{3\cdot}$ et les sommes des colonnes $p_{\cdot 1}$, $p_{\cdot 2}$, $p_{\cdot 3}$:

		A : Attitude face à l'avortement			
		Pour	Mixte	Contre	
S : Scolarité	≤ 8	p_{11}	p_{12}	p_{13}	$p_{1\cdot}$
	$9 - 12$	p_{21}	p_{22}	p_{23}	$p_{2\cdot}$
	> 12	p_{31}	p_{32}	p_{33}	$p_{3\cdot}$
Total		$p_{\cdot 1}$	$p_{\cdot 2}$	$p_{\cdot 3}$	1

L'hypothèse nulle, en fonction de ces paramètres, est

$$H_0 : p_{ij} = p_{i\cdot} \times p_{\cdot j} \text{ pour tout } i, j$$

Les fréquences théoriques seraient donc

		A : Attitude face à l'avortement			
		Pour	Mixte	Contre	
S : Scolarité	≤ 8	$p_{1 \cdot} p_{\cdot 1}$	$p_{1 \cdot} p_{\cdot 2}$	$p_{1 \cdot} p_{\cdot 3}$	$p_{1 \cdot}$
	9 — 12	$p_{2 \cdot} p_{\cdot 1}$	$p_{2 \cdot} p_{\cdot 2}$	$p_{2 \cdot} p_{\cdot 3}$	$p_{2 \cdot}$
	> 12	$p_{3 \cdot} p_{\cdot 1}$	$p_{3 \cdot} p_{\cdot 2}$	$p_{3 \cdot} p_{\cdot 3}$	$p_{3 \cdot}$
Total		$p_{\cdot 1}$	$p_{\cdot 2}$	$p_{\cdot 3}$	1

Puisque, sous H_0 , $E(X_{ij}) = np_{i \cdot} p_{\cdot j}$, la statistique χ^2 devrait normalement être

$\sum_i \sum_j \frac{(X_{ij} - np_{i \cdot} p_{\cdot j})^2}{np_{i \cdot} p_{\cdot j}}$. Mais étant donné que les $p_{i \cdot}$ et les $p_{\cdot j}$ ne sont pas connus, elles seront estimées par

$$\hat{p}_{i \cdot} = \frac{1}{n} \sum_j X_{ij} \quad \hat{p}_{\cdot j} = \frac{1}{n} \sum_i X_{ij}$$

La statistique χ^2 est donc

$$\chi^2 = \sum_{i=1}^l \sum_{j=1}^c \frac{(X_{ij} - n\hat{p}_{i \cdot} \hat{p}_{\cdot j})^2}{n\hat{p}_{i \cdot} \hat{p}_{\cdot j}}$$

Étant donné le théorème 7.4.2, la région critique est

$$\chi^2 > \chi_{(l-1)(c-1); \alpha}^2$$

Dans l'exemple, puisque $\chi^2 = 17,7 > \chi_{4; 0,05}^2 = 9,4877$, nous rejetons H_0 à 5%. Nous concluons qu'il y a vraiment une dépendance entre la scolarité et l'attitude face à l'avortement.

Remarque La notion d'indépendance

En probabilités, l'indépendance entre deux événements A et B est définie par $P(A \cap B) = P(A)P(B)$. C'est encore le sens dans lequel nous l'entendons ici. Par exemple, si A est l'évènement « la scolarité du répondant est ≤ 8 » et B l'évènement « le répondant est pour l'avortement », alors, dans la notation du dernier tableau, $P(A \cap B) = p_{11}$, $P(A) = p_{1 \cdot}$ et $P(B) = p_{\cdot 1}$. L'hypothèse d'indépendance stipule que $p_{11} = p_{1 \cdot} \times p_{\cdot 1}$. ■

Remarque Une autre modélisation

Dans l'exemple de cette section, le chercheur a choisi 776 sujets qu'il a ensuite classés selon la scolarité et l'attitude. Par conséquent, tous les effectifs observés sont aléatoires, y compris les effectifs des marges; seul $n = 776$ est fixe. Mais il existe des situations où les effectifs des marges sont fixés. Dans ce cas, la description des données comme un vecteur de loi multinomiale n'est pas conforme à la réalité. Le tableau suivant présente des données recueillies dans le but de savoir si la durée d'une hospitalisation (pour une certaine maladie) varie d'un hôpital à l'autre :

		Durée de l'hospitalisation			
		1-2 jours	3-5 jours	plus de 5 jours	
Hôpital	1	40	20	40	100
	2	50	35	65	150
	3	95	45	60	200

Visiblement, les effectifs de la marge de droite sont fixes : on a décidé, avant de prélever les données, qu'on prendrait respectivement 100, 150 et 200 cas dans les trois hôpitaux. Ce ne sont pas des variables aléatoires.

Dans le tableau suivant on nomme les variables observées :

		Durée de l'hospitalisation			
		1-2 jours	3-5 jours	5 jours+	
Hôpital	1	X_1	Y_1	Z_1	n_1
	2	X_2	Y_2	Z_2	n_2
	3	X_3	Y_3	Z_3	n_3

Les 9 variables aléatoires dans le tableau ne suivent pas une loi multinomiale. Nos observations sont constituées de 3 vecteurs de loi multinomiale. Le modèle est plutôt celui-ci :

$$\begin{aligned}
 [X_1 ; Y_1 ; Z_1] &\sim \mathcal{M}(n_1 ; p_1 ; r_1 ; s_1) ; \\
 [X_2 ; Y_2 ; Z_2] &\sim \mathcal{M}(n_2 ; p_2 ; r_2 ; s_2) ; \\
 [X_3 ; Y_3 ; Z_3] &\sim \mathcal{M}(n_3 ; p_3 ; r_3 ; s_3) .
 \end{aligned}$$

L'hypothèse nulle est

$$H_0 : [p_1 ; r_1 ; s_1] = [p_2 ; r_2 ; s_2] = [p_3 ; r_3 ; s_3] \quad \blacksquare$$

Malgré la différence entre ce modèle (trois multinomiales) et celui de l'exemple précédent (une seule multinomiale), la procédure reste la même : le même test est valable dans les deux cas.

Remarque Le cas particulier d'un tableau 2x2 est un test d'égalité de deux proportions. La formulation habituelle du test d'égalité de proportions est celle-ci : on observe deux variables aléatoires indépendantes, $X_1 \sim \mathcal{B}(n_1 ; p_1)$ et $X_2 \sim \mathcal{B}(n_2 ; p_2)$, et on teste l'hypothèse $H_0 : p_1 = p_2$, un test essentiellement équivalent à un test d'indépendance. Mais le déroulement de l'expérience ne mène pas toujours à cette modélisation. Par exemple, supposons qu'on veuille tester l'hypothèse que la proportion des personnes favorables à une certaine loi (permettant le mariage gai, par exemple) est la même chez les femmes et les hommes. Les données prennent la forme suivante :

	Pour projet de loi	Contre le projet de loi	Total
Femmes	X_1	$Y_1 = n_1 - X_1$	$n_1 = X_1 + Y_1$
Hommes	X_2	$Y_2 = n_2 - X_2$	$n_2 = X_2 + Y_2$
	$X_1 + X_2$	$Y_1 + Y_2$	$n = n_1 + n_2$

Si le nombre de femmes n_1 et le nombre d'hommes n_2 sont fixes, alors effectivement le modèle de deux binomiales indépendantes s'impose directement. Mais il est possible que le tirage soit fait au hasard dans la population entière. Dans ce cas, nos observations constituent un vecteur $[X_1 ; Y_1 ; X_2 ; Y_2]$ de loi $\mathcal{M}(n ; p_{11} ; p_{12} ; p_{21} ; p_{22})$. La raison pour laquelle on peut néanmoins continuer à agir comme si on avait affaire à deux variables indépendantes de loi binomiale, c'est que les variables X_1 et X_2 sont conditionnellement indépendantes et de lois binomiales. C'est-à-dire, étant donné $X_1 + Y_1 = n_1$ et $X_2 + Y_2 = n_2$:

$$\begin{aligned}
 X_1 | X_1 + Y_1 = n_1 &\sim \mathcal{B}(n_1 ; p_1) \text{ et } X_2 | X_2 + Y_2 = n_2 \sim \mathcal{B}(n_2 ; p_2), \text{ où} \\
 p_1 &= \frac{p_{11}}{p_{11} + p_{12}} \text{ et } p_2 = \frac{p_{21}}{p_{21} + p_{22}} .
 \end{aligned} \quad \blacksquare$$

7.6 Données appariées (test de McNemar)

Le conditionnement décrit dans la dernière remarque est utile aussi dans un autre contexte, que nous présentons ici. Il s'agit d'un test d'égalité de proportions lorsque les données sont appariées. Cette situation est tout à fait parallèle au test d'égalité de moyennes dans le cas de données appariées. L'hypothèse à tester est $p_1 = p_2$ lorsque les données sont appariées, et donc les estimateurs \hat{p}_1 et \hat{p}_2 ne sont pas indépendants. La solution, illustrée dans l'exemple suivant, est parfois appelé *test de McNemar*.

On doit évaluer l'efficacité d'une pilule conçue pour réduire les risques d'une crise de panique chez les personnes qui en sont susceptibles. Une façon de procéder est la suivante : on soumet deux groupes — un groupe expérimental qui reçoit le traitement et un groupe témoin qui reçoit un placebo — à une situation de stress susceptible de provoquer une crise. Les variables X_1 et X_2 (nombre de réactions de panique enregistrées dans les deux groupes), sont indépendantes, de loi binomiale de probabilités p_1 et p_2 respectivement. Les techniques de la section 7.3 peuvent alors être utilisées pour tester l'hypothèse $p_1 = p_2$.

Mais un autre plan d'expérience consiste à placer chaque individu dans une situation de stress *deux fois*, l'une avec le traitement, l'autre avec le placebo. On observe encore les variables X_1 et X_2 , elles sont de loi binomiale de paramètres p_1 et p_2 , et c'est encore l'hypothèse $p_1 = p_2$ qu'il s'agit de tester. Mais on ne peut pas procéder comme d'habitude parce que X_1 et X_2 ne sont pas indépendantes. Nous observons en fait quatre variables aléatoires, disons $X_{11}, X_{12}, X_{21}, X_{22}$, de loi multinomiale de paramètres n (le nombre de sujets) et p_{11}, p_{12}, p_{21} et p_{22} , comme dans la tableau suivant :

		Sous traitement		
		Réaction	Pas de réaction	
Sous placebo	Réaction	X_{11} p_{11}	X_{12} p_{12}	$X_1 = (X_{11} + X_{12})$ p_1
	Pas de réaction	X_{21} p_{21}	X_{22} p_{22}	
		$X_2 = (X_{11} + X_{21})$ p_2	$n - X_2$	n

L'hypothèse à tester est

$$H_0: p_1 = p_2 \Leftrightarrow p_{11} + p_{12} = p_{11} + p_{21} \Leftrightarrow p_{12} = p_{21}$$

On peut tester cette hypothèse en considérant la distribution *conditionnelle* de X_{12} étant donné la valeur observée de $X_{12} + X_{21}$:

$$X_{12} \mid X_{12} + X_{21} = m \sim \mathcal{B}\left(m; \frac{p_{12}}{p_{12} + p_{21}}\right)$$

H_0 est équivalente à $H_0: \frac{p_{12}}{p_{12} + p_{21}} = 1/2$. Il s'agit donc d'un test classique sur une proportion.

Exemple 7.6.1 Comparaison de deux moyennes – données appariées

Considérons donc les données suivantes sur un échantillon de 150 personnes soumises deux fois à une situation de stress, l'une sous l'effet d'un médicament, l'autre sous l'effet d'un placebo.

		Sous traitement		Total
		Réaction de panique	Pas de réaction de panique	
Sous placebo	Réaction de panique	12	28	40
	Pas de réaction de panique	13	47	60
		25	75	100

On considère les 41 (13+28) sujets dont le comportement diffère sous le traitement et sous le placebo. On teste l'hypothèse que parmi ceux-ci, la probabilité que la réaction de panique se produise sous l'effet du traitement est 1/2. La variable observée X suit conditionnellement une loi $\mathcal{B}(41; p)$, $H_0: p =$

$$\frac{1}{2} \hat{p} = \frac{28}{41} \text{ et } Z = \frac{28/41 - 1/2}{\frac{1}{2} \sqrt{\frac{1}{41}}} = 2,34, \text{ ce qui est hautement significatif, le seuil expérimental (pour}$$

un test bilatéral) étant 0,019. Parmi ceux dont la réponse diffère sous l'effet du médicament, ceux qui ont mieux réagi sous l'effet du médicament que sous l'effet du placebo sont plus nombreux (28/41) que les autres. On conclut que le médicament est bénéfique. ■

7.7 Sommes de khi-deux

Considérons le tableau suivant, qui donne la distribution des quotients intellectuels (QI) d'un échantillon d'étudiants pris dans chacune de quatre universités :

	Université				Toutes
	A	B	C	D	
QI < 115	12	30	30	46	118
115 ≤ QI < 125	47	49	34	18	148
QI ≥ 125	41	21	36	36	134
Effectif total	100	100	100	100	400

Les quatre colonnes sont des vecteurs

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}; \mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}; \mathbf{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix}; \mathbf{W} = \begin{bmatrix} W_1 \\ W_2 \\ W_3 \end{bmatrix},$$

chacun de loi multinomiale de paramètres $n = 100$ et

$$\mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}; \mathbf{q} = \begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix}; \mathbf{r} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix}; \text{ et } \mathbf{s} = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix},$$

respectivement. L'hypothèse d'indépendance stipule que ces quatre vecteurs sont égaux ($\mathbf{p} = \mathbf{q} = \mathbf{r} = \mathbf{s}$), mais ne spécifie par leur valeur. Il peut arriver, cependant, qu'on veuille tester une hypothèse de la forme :

$$\mathbf{p} = \mathbf{q} = \mathbf{r} = \mathbf{s} = \mathbf{p}_o$$

où \mathbf{p}_o est un vecteur donné. Par exemple,

$$H_o : \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

(Pourquoi 1/3? Ces probabilités correspondent à l'hypothèse que, dans chaque université, les QI sont de loi normale de moyenne 120 et d'écart-type 12, ces paramètres étant une norme validée au cours des années sur un très grand nombre d'individus, d'âge et d'éducation comparables à ceux des étudiants de ces universités).

On peut tester cette hypothèse en effectuant un test d'uniformité dans chaque université séparément. On obtient les résultats suivants :

	Université			
	A	B	C	D
χ^2	21,02	35,06	0,56	12,08
Valeur p	0,00003	2×10^{-8}	0,756	0,0024

Chacune des valeurs de χ^2 est de loi L'hypothèse d'uniformité est acceptée à l'université C et rejetée aux trois autres. Mais là nous nous retrouvons avec quatre conclusions alors que nous n'avons qu'une seule hypothèse :

$$H_0 : p = p_0 \text{ et } q = p_0 \text{ et } r = p_0 \text{ et } s = p_0, \text{ où } p_0 = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

Rejeter cette hypothèse, c'est conclure qu'au moins l'une de ces affirmations est fausse. Une statistique appropriée est simplement la somme des valeurs de χ^2 :

$$\chi^2 = 21,02 + 35,06 + 0,56 + 12,08 = 68,72.$$

Les quatre échantillons sont indépendants et sous H_0 chacune des valeurs de χ^2 est de loi χ^2 (à peu près) à 2 degrés de liberté. Leur somme est donc de loi χ^2 à 8 degrés de liberté. La valeur observée $\chi^2 = 68,72$ correspond à une valeur p de $8,8 \times 10^{-12}$. Donc on rejette l'hypothèse, pour conclure qu'il n'est pas vrai que la distribution des QI est normale (de moyenne 120 et d'écart-type 12) dans chacune des universités.

Remarque On pourrait songer à l'approche suivante. On combine les données des quatre universités en une seule série, soit $U = X + Y + Z + W = [118; 148; 134]$. Sous H_0 , ces effectifs sont de loi multinomiale de paramètres $n = 400$ et probabilités $[1/3; 1/3; 1/3]$. Un test d'ajustement donne une valeur de χ^2 de 3,38 à 2 degrés de liberté. La valeur p étant 0,18, on ne peut pas rejeter H_0 .

Le premier rejette H_0 , alors que le deuxième, basé sur U , ne rejette pas H_0 . Auquel faut-il croire?

Strictement parlant, les deux tests sont corrects dans le sens suivant. Dans les deux cas, la probabilité d'une erreur de première espèce est d'environ 5 %, et dans ce sens, les deux tests sont corrects. C'est au niveau de la puissance que les deux tests diffèrent—la capacité du test de rejeter H_0 lorsqu'elle est fausse. Le deuxième test, qui repose entièrement sur le vecteur U , est insensible aux différences entre X , Y , Z et W , et donc ne contient aucune information sur les différences entre p , q , r , et s . Il accepte H_0 parce que les composantes de U sont assez proches les unes des autres. Or il y a d'importantes différences entre X , Y , Z et W . Le deuxième test détecte ces différences et conclut à des différences entre p , q , r , et s . ■

7.8 Comparaisons multiples

Dans certaines études scientifiques, une erreur de première espèce constitue ce qu'on appelle une *fausse découverte*. Une question courante en génétique, par exemple, est de savoir si un gène donné est lié à une certaine maladie. H_0 est l'hypothèse que le gène n'est pas lié à la maladie. Si on rejette H_0 on affirme avoir fait une découverte. Si l'hypothèse est vraie, c'est une *fausse découverte*. La probabilité d'une fausse découverte est α , le niveau du test. On se protège donc contre une fausse découverte en prenant α petit.

Une difficulté surgit, cependant, lorsqu'un laboratoire, à la recherche d'un gène pouvant être lié à une même maladie considère un grand nombre n de gènes possibles, donc n tests d'hypothèses. Si le risque d'une fausse découverte est maintenu à α à chacun des tests, le risque d'au moins une fausse découverte au cours des n tests croît avec n . Plus précisément, si n est assez grand, il est très probable de rejeter au moins une des hypothèses, mêmes si elles sont toutes vraies. Le défi est donc celui-ci:

Lorsqu'on teste n hypothèses, comment contrôler la probabilité de rejeter au moins l'une d'elles si elles sont toutes vraies?

La méthode de Bonferroni

Soit $H_{01}, H_{02}, \dots, H_{0n}$ n hypothèses, et H_0 la conjonction des n hypothèses $H_{01}, H_{02}, \dots, H_{0n}$. On désigne H_0 par

$$H_0 = H_{01} \wedge H_{02} \wedge \dots \wedge H_{0n}.$$

(H_0 est vraie si les n hypothèses sont toutes vraies).

Soit R_1, R_2, \dots, R_n n régions critiques correspondant aux n hypothèses, et R_0 une région critique pour H_0 . Le problème est de s'assurer que $P(R_i | H_{0i})$ assez petit pour que $P(R_0 | H_0) \leq \alpha$.

On vérifie aisément que si les R_i sont tels que $P(R_i | H_{0i}) \leq \alpha/n$ pour tout i , et

$$R_0 = R_1 \cup R_2 \cup \dots \cup R_n,$$

alors $P(R_0 | H_0) \leq \alpha$. Ceci découle du fait que la probabilité de l'union de n événements est inférieure ou égale à la somme de leurs probabilités:

$$P(R_0 | H_0) = P(R_1 \cup R_2 \cup \dots \cup R_n) \leq P(R_1) + P(R_2) + \dots + P(R_n).$$

Si les hypothèses sont toutes vraies, chacun des événements R_i est de probabilité α/n (ou inférieures à α/n), et par conséquent, $P(R_1) + P(R_2) + \dots + P(R_n) \leq n(\alpha/n) = \alpha$.

La méthode de Bonferroni consiste donc à effectuer chaque test au niveau α/n .

Exemple 7.8.1 Comparaison de deux vecteurs de probabilité

Les chercheurs d'un laboratoire s'intéressent aux causes génétiques d'une certaine maladie. Certaines considérations théoriques leur fait penser que 4 gènes particuliers (les gènes g_1, g_2, g_3 et g_4) pourraient être en cause. Afin de confirmer cette hypothèse, on rassemble deux groupes de personnes. Le groupe 1 (le groupe «expérimental») est constitué de $n_1 = 70$ sujets souffrant de la maladie en question. Le groupe 2 (le groupe témoin), de $n_2 = 200$ personnes n'en est pas atteint. On compte alors le nombre de personnes— X_i dans le groupe 1 et Y_i dans le groupe 2—qui portent le gène g_i .

$$X_i \sim \mathcal{B}(70 ; p_{1i}) \text{ et } Y_i \sim \mathcal{B}(200 ; p_{2i}) .$$

Si on conclut que $p_{1i} > p_{2i}$, on est en mesure d'affirmer (du moins provisoirement) que le gène g_i est en partie responsable de la maladie.

On a 4 hypothèses à tester:

$$H_{0i} : p_{1i} = p_{2i}, \text{ contre } H_{1i} : p_{1i} > p_{2i}, i = 1, 2, 3, 4.$$

(Le choix d'une alternative unilatérale pourrait, dépendant du contexte, être contesté, mais nous le supposons justifiable ici).

Voici les résultats:

Nombre de sujets possédant le gène	Gène (g_i)			
	g_1	g_2	g_3	g_4
Dans le groupe 1 (X_i)	12	23	31	31
Dans le groupe 2 (Y_i)	25	39	59	77

Posons $\alpha = 0,05$. On effectue chaque test au niveau $\alpha/4 = 0,0125$. Les statistiques de test

$$\text{sont } Z_i = \frac{\hat{p}_{1i} - \hat{p}_{2i}}{\hat{p}_i(1 - \hat{p}_i)\sqrt{\frac{1}{70} + \frac{1}{200}}}, \text{ où } \hat{p}_i = \frac{70\hat{p}_{1i} + 200\hat{p}_{2i}}{70 + 200} .$$

Les régions critiques sont $Z_i \geq z_{0,0125} = 2,24$.

Voici les calculs:

	Gène g ₁	Gène g ₂	Gène g ₃	Gène g ₄
\hat{p}_{1i}	0,1714	0,3286	0,4429	0,4429
\hat{p}_{2i}	0,1250	0,1950	0,2950	0,3850
\hat{p}_i	0,1370	0,2296	0,3333	0,4000
z_i	0,972	2,287	2,2586	0,8504
Valeur p	0,166	0,011	0,012	0,198

Puisque $z_2 > 2,24$ et $z_3 > 2,24$, on peut rejeter les hypothèses H_{02} et H_{03} et conclure donc que $p_{12} > p_{22}$ et $p_{13} > p_{23}$: les gènes g_2 et g_3 pourraient être liés à la maladie.

Remarque Le but de cet exemple est d'illustrer la méthode de Bonferroni. L'analyse qui y est faite est possiblement simpliste dans la mesure où elle ne tient pas compte des interactions possibles entre les gènes.

Remarque Au dernier exemple, les régions critiques sont exprimées en fonction des valeurs Z_i . Il est souvent plus commode de s'exprimer en fonction des valeurs p . La région critique $\{Z \geq z_{0,0125}\}$ est équivalente à : $\{\text{valeur } p > 0,0125\}$.

Procédure de Hochberg

[Yosef Hochberg, A sharper Bonferroni procedure for multiple tests of significance, *Biometrika* (1988), **75**, 4, pp. 800-2]

Un des inconvénients majeurs de la méthode de Bonferroni est qu'elle oblige à réduire la taille de la région critique à α/n , ce qui a pour conséquence que le rejet d'une hypothèse devient très difficile si n est grand. Certains proposent de prendre α plus grand—jusqu'à 30 %, s'il le faut. Ce n'est pas déraisonnable, si on en accepte le risque: annoncer des découvertes qui n'en sont pas.

Mais avant de recourir à cette solution, on présente la procédure de Hochberg, qui accroît quelque peu la probabilité de rejeter des hypothèses tout en maintenant la propriété essentielle de la méthode de Bonferroni. Nous décrivons la méthode ici, suivie d'une justification heuristique, exprimée dans le contexte de l'exemple 7.8.1.

On exprime les régions critiques en termes des *valeurs p*, soit $\{p \leq \alpha\}$ pour un test d'hypothèse de niveau α .

Description de la procédure:

Soit p_1, p_2, \dots, p_n les *valeurs p* correspondant aux hypothèses $H_{01}, H_{02}, \dots, H_{0n}$; $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(n)}$ les statistiques d'ordre et $H_{(01)}, H_{(02)}, \dots, H_{(0n)}$ les hypothèses correspondantes.

On calcule les quantités

$$z_1 = np_{(1)}; z_2 = (n - 1)p_{(2)}; z_3 = (n - 2)p_{(3)}; \dots; z_n = p_{(n)};$$

On rejette toute hypothèse $H_{(0i)}$ pour laquelle $z_i \leq \alpha$.

Justification de la procédure de Hochberg

Par comparaison, la méthode de Bonferroni rejette toute hypothèse $H_{(0i)}$ pour laquelle $np_{(i)} \leq \alpha$. Il est donc possible que l'approche de Hochberg permette de rejeter une hypothèse qui ne pourrait être rejetée par la méthode de Bonferroni. C'est le cas de l'hypothèse $H_{(03)}$, par exemple, si $(n - 2)p_{(3)} \leq \alpha < np_{(3)}$.

Cette méthode possède la propriété voulue, soit que $P(R_0 | H_0) \leq \alpha$, et elle répond aussi à un souci de cohérence. Par exemple, si on teste deux hypothèses, soit H_{01} et H_{02} , ainsi que $H_0 = H_{01} \wedge H_{02}$, il serait incohérent de rejeter H_{01} et d'accepter H_0 . Il serait raisonnable alors de rejeter H_{01} seulement si on rejette également H_0 . Si R_1, R_2 et $R_0 = R_1 \cup R_2$ sont des régions critiques,

chacune de taille α , alors pour éviter toute incohérence, on définit de nouvelles régions critiques $R_1^* = R_1 \cap R_0$, $R_2^* = R_2 \cap R_0$ et $R_0^* = R_1^* \cup R_2^*$. Puisque $R_1^* \subset R_0$ et $R_2^* \subset R_0$ et $R_0^* \subset R_0$ sous H_0 , $P(R_1^* | H_0) \leq \alpha$, $P(R_2^* | H_0) \leq \alpha$ et $P(R_0^* | H_0) \leq \alpha$.

En général, la procédure se définit comme suit.

Considérons les hypothèses $H_{01}, H_{02}, \dots, H_{0n}$, ainsi que toutes les conjonctions possibles, telles, par exemple, $H_{01} \wedge H_{02}$, $H_{02} \wedge H_{03}$, $H_{01} \wedge H_{02} \wedge H_{03}$, $H_{01} \wedge H_{02} \wedge \dots \wedge H_{0n}$. Supposons qu'il existe une région critique de niveau α pour chacune de ces hypothèses, et que dans le cas d'une conjonction de 2 hypothèses ou plus, la région critique est définie par l'approche de Bonferroni. Le test de Hochberg consiste à rejeter H_{0i} si et seulement si on rejette H_{0i} ainsi que toute conjonction d'hypothèses qui comprend H_{0i} . Cette procédure mène à la procédure de Hochberg telle que décrite ci-dessus.

On illustre dans le cas $n = 3$. Soit $p_{(1)} \leq p_{(2)} \leq p_{(3)}$ les valeurs p ordonnées et $H_{(01)}, H_{(02)}, H_{(03)}$ les hypothèses correspondantes. Voici les conditions dans lesquelles on rejette $H_{(01)}$:

$H_{(01)}$ et toutes les conjonctions la contenant	Région critique (selon Bonferroni)	Expression simplifiée
$H_{(01)}$	$p_{(1)} \leq \alpha$	$p_{(1)} \leq \alpha$
$H_{(01)} \wedge H_{(02)}$	$p_{(1)} \leq \alpha/2$ ou $p_{(2)} \leq \alpha/2$	$p_{(1)} \leq \alpha/2$
$H_{(01)} \wedge H_{(03)}$	$p_{(1)} \leq \alpha/2$ ou $p_{(3)} \leq \alpha/2$	$p_{(1)} \leq \alpha/2$
$H_{(01)} \wedge H_{(02)} \wedge H_{(03)}$	$p_{(1)} \leq \alpha/3$ ou $p_{(2)} \leq \alpha/3$ ou $p_{(3)} \leq \alpha/3$	$p_{(1)} \leq \alpha/3$

On rejette $H_{(01)}$ si $p_{(1)} \leq \alpha$ et $p_{(1)} \leq \alpha/2$ et $p_{(1)} \leq \alpha/3$. Donc on rejette $H_{(01)}$ si $p_{(1)} \leq \alpha/3$.

Voici les conditions dans lesquelles on rejette $H_{(02)}$:

$H_{(01)}$ et toutes les conjonctions la contenant	Région critique (selon Bonferroni)	Expression simplifiée
$H_{(02)}$	$p_{(2)} \leq \alpha$	$p_{(2)} \leq \alpha$
$H_{(01)} \wedge H_{(02)}$	$p_{(1)} \leq \alpha/2$ ou $p_{(2)} \leq \alpha/2$	$p_{(1)} \leq \alpha/2$
$H_{(02)} \wedge H_{(03)}$	$p_{(2)} \leq \alpha/2$ ou $p_{(3)} \leq \alpha/2$	$p_{(2)} \leq \alpha/2$
$H_{(01)} \wedge H_{(02)} \wedge H_{(03)}$	$p_{(1)} \leq \alpha/3$ ou $p_{(2)} \leq \alpha/3$ ou $p_{(3)} \leq \alpha/3$	$p_{(1)} \leq \alpha/3$

On rejette $H_{(02)}$ si $p_{(2)} \leq \alpha$ et $p_{(1)} \leq \alpha/2$ et $p_{(2)} \leq \alpha/2$ et $p_{(1)} \leq \alpha/3$, donc si $p_{(1)} \leq \alpha/3$ et $p_{(2)} \leq \alpha/2$.

Voici les conditions dans lesquelles on rejette $H_{(03)}$:

$H_{(03)}$ et toutes les conjonctions la contenant	Région critique (selon Bonferroni)	Expression simplifiée
$H_{(03)}$	$p_{(3)} \leq \alpha$	$p_{(3)} \leq \alpha$
$H_{(01)} \wedge H_{(03)}$	$p_{(1)} \leq \alpha/2$ ou $p_{(3)} \leq \alpha/2$	$p_{(1)} \leq \alpha/2$
$H_{(02)} \wedge H_{(03)}$	$p_{(2)} \leq \alpha/2$ ou $p_{(3)} \leq \alpha/2$	$p_{(2)} \leq \alpha/2$
$H_{(01)} \wedge H_{(02)} \wedge H_{(03)}$	$p_{(1)} \leq \alpha/3$ ou $p_{(2)} \leq \alpha/3$ ou $p_{(3)} \leq \alpha/3$	$p_{(1)} \leq \alpha/3$

On rejette $H_{(03)}$ si $p_{(3)} \leq \alpha$ et $p_{(1)} \leq \alpha/2$ et $p_{(2)} \leq \alpha/2$ et $p_{(1)} \leq \alpha/3$, donc si $p_{(1)} \leq \alpha/3$, $p_{(2)} \leq \alpha/2$ et $p_{(3)} \leq \alpha$.

Pour illustrer la différence entre cette méthode et la méthode de Bonferroni, considérons les conditions qui permettent de rejeter $H_{(01)}$, $H_{(02)}$ et $H_{(03)}$. Selon la méthode modifiée, on rejette $H_{(01)}$, $H_{(02)}$ et $H_{(03)}$ si

$$p_{(1)} \leq \frac{\alpha}{n}, p_{(2)} \leq \frac{\alpha}{n-1} \text{ et } p_{(3)} \leq \frac{\alpha}{n-2},$$

alors que par la méthode de Bonferroni ces mêmes trois hypothèses ne sont rejetées que lorsque

$$p_{(1)} \leq \frac{\alpha}{n}, p_{(2)} \leq \frac{\alpha}{n} \text{ et } p_{(3)} \leq \frac{\alpha}{n}.$$

Exemple 7.8.2 Méthode de Hochberg appliquée au contexte de l'exemple 7.8.1.

Les valeurs p ordonnées sont 0,011 0,012 0,165 0,198

Les valeurs $(n - i + 1)p(i)$ sont : 0,044 0,036 0,331 0,198.

Si le niveau souhaité est $\alpha = 0,05$, on rejette les hypothèses $H_{(01)}$ et $H_{(02)}$, soit H_{02} et H_{03} .

6.9 Trois problèmes résolus pour le chapitre 7

Problème 1

Pour des raisons d'économie, un chocolatier compte remplacer le chocolat traditionnel T par un nouveau chocolat N (même formule que pour T, avec un ingrédient exotique en moins) et se demande si la différence est vraiment perceptible. Il fait tester les deux chocolats par un expert en lui demandant d'identifier 12 chocolats, dont certains sont traditionnels (T), d'autres nouveaux (N). Posons comme hypothèse nulle l'hypothèse

H_0 : La différence entre les deux sortes est imperceptible

a) Considérons le plan suivant :

Plan A : On lui fait goûter 6 chocolats T et 6 N, sans lui dire combien il y en a de chaque.

Supposons que l'expert réussit à identifier 10 des chocolats. Peut-on rejeter H_0 ?

Réponse

Si X le nombre de succès, nous allons calculer $P(X \geq 10)$ sous H_0 . Dans ce plan, on suppose que les essais sont indépendants et que la probabilité de succès est $\frac{1}{2}$ à chaque essai. Donc X est de loi $\mathcal{B}(10 ; \frac{1}{2})$, et $P(X \geq 10 | p = \frac{1}{2}) = 0,01928711$. Cette probabilité étant plutôt faible, nous rejetons H_0 : le sujet semble bien avoir certaines capacités.

b) Considérons le plan suivant :

Plan B : On lui fait goûter 6 chocolats T et 6 N, mais il sait qu'il y en a 6 de chaque

Supposons que l'expert réussit à identifier 10 des chocolats. Peut-on rejeter H_0 ?

Réponse

$X = 2Y$, où Y , le nombre de chocolats T identifiés comme T, est de loi hypergéométrique de paramètres $n = 6, N_1 = N_2 = 6$.

		Le chocolat est		
		T	N	
Le chocolat est déclaré	T	Y	$6-Y$	6
	N	$6-Y$	Y	6
		6	6	12

$P(X \geq 10) = P(Y \geq 5) = 0,0400$.

c) Considérons le plan suivant :

Plan C : On lui fait goûter 4 chocolats T et 8 N, et il sait qu'il y en a 4 T et 8 N.

Supposons que l'expert réussit à identifier 10 des chocolats. Peut-on rejeter H_0 ?

Réponse

$X = 2Y + 4$, où Y est de loi hypergéométrique de paramètres $n = 4, N_1 = 4, N_2 = 8$.

		Le chocolat est		
		T	N	
Le chocolat est déclaré	T	Y	4-Y	4
	N	4-Y	4+Y	8
		4	8	12

$P(X \geq 10) = P(Y \geq 3) = 0,0667$

Le test A étant le plus exigeant des trois, le fait d’avoir réussi 10 fois est particulièrement significatif, ce qui l’est moins dans le cas des tests B et C, plus faciles.

Problème 2 [Suite]

Supposons que l’expérience décrite au dernier numéro n’a pas encore été faite et qu’on doit la planifier, c’est-à-dire, déterminer une région critique. Celle-ci sera exprimée en fonction de X, le nombre de succès, et la région critique sera de la forme $X \geq a$. On décide qu’on rejettera H_0 si le nombre de succès est assez élevé, donc si X est supérieur ou égal à un certain nombre a. On choisira a de telle sorte que $P(X \geq a/H_0) \leq \alpha$. Fixons $\alpha = 0,05$ Déterminer a pour chacun des plans, prenant pour seuil $\alpha = 0,05$.

Plan A

Le nombre de succès X est de loi $\mathcal{B}(12 ; 1/2)$. Voici les valeurs de $P(X \geq a | H_0)$ pour différentes valeurs de a :

a	$P(X \geq a)$
12	0,0002
11	0,0032
10	0,0193
9	0,0730
8	0,1938
7	0,3872
6	0,6128

Si $X \geq 10$, on rejettera l’hypothèse. La région critique est de taille 0,0193.

Plan B

Le nombre de succès X est égal à 2Y, où Y est de loi hypergéométrique de paramètres $n = 6, N_1 = N_2 = 6$.

Nombre de succès $X = 2Y$	Nombre de T déclarés T (Y)	$P(X \geq a)$
12	6	0,00108
10	5	0,0400
8	4	0,2835
6	3	0,7165
4	2	0,9600
2	1	0,9989
0	0	1

On ne rejettera H_0 que si le sujet réussi à identifier au moins 10 des chocolats. La région critique est de taille 0,0400.

Plan C

Le nombre de succès X est égal $2Y+4$, où Y est de loi hypergéométrique de paramètres $n = 4, N_1 = 4, N_2 = 8$.

Nombre de succès $a = 2y+4$	Nombre de T déclarés T (y)	$P(X \geq a)$
12	4	0,0020
10	3	0,0667
8	2	0,4061
6	1	0,8586
4	0	1

On ne rejettera H_0 que si le sujet réussit à identifier tous les chocolats. La région critique est de taille 0,0020.

Problème 3

Une personne prétend être capable de distinguer 15 marques de bière. Pour le prouver, il se soumet à un test qui lui demande de les identifier (il prend une gorgée de chaque sorte). Considérons l'hypothèse

H_0 : Cette personne est totalement incapable de distinguer une bière de l'autre.

Soit X le nombre de marques de bière correctement identifiées. Voici la distribution de X sous l'hypothèse H_0 :

a	$P(X = a)$	$P(X \geq a)$	a	$P(X = a)$	$P(X \geq a)$	a	$P(X = a)$	$P(X \geq a)$
0	0,36788	1	6	0,00051	0,00059	11	0,00000	0,00000
1	0,36788	0,63212	7	0,00007	0,00008	12	0,00000	0,00000
2	0,18394	0,26424	8	0,00001	0,00001	13	0,00000	0,00000
3	0,06131	0,08030	9	0,00000	0,00000	14	0,00000	0,00000
4	0,01533	0,01899	10	0,00000	0,00000	15	0,00000	0,00000
5	0,00307	0,00366						

Si X est assez élevée on rejettera H_0 .

Réponse À un seuil $\alpha = 0,05$, on rejettera H_0 s'il identifie correctement 4 marques ou plus.

Résumé

Expression générale

La multiplicité des tests dans ce chapitre et le précédent 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 manuel (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_0 porte sur un paramètre θ ;
- L'hypothèse nulle est de la forme $\theta = \theta_0$;
- 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_0 , soit $T = \frac{\hat{\theta} - \theta_0}{\hat{\sigma}_{\hat{\theta}}}$;
- La région critique est de la forme $|T| \geq A$, ou $T \geq B$, ou $T \leq C$ selon que H_1 est $\theta \neq \theta_0$, $\theta \geq \theta_0$, ou $\theta \leq \theta_0$;
- 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 $\sigma_1 \neq \sigma_2$: $\sigma_{\bar{X}-\bar{Y}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$\hat{\sigma}_{\bar{X}-\bar{Y}} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$
$\mu_1 - \mu_2$	$\bar{X} - \bar{Y}$	Si $\sigma_1 = \sigma_2$: $\sigma_{\bar{X}-\bar{Y}} = \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$	$\hat{\sigma}_{\bar{X}-\bar{Y}} = S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ où $S = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}}$
p	\hat{p} (proportion échantillonnale)	$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$	$\hat{\sigma}_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$	$\hat{\sigma}_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$
$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	Sous $H_0 : p_1 = p_2 = p$ $\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{p(1-p)} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$	$\hat{\sigma}_{\hat{p}_1-\hat{p}_2} = \sqrt{\hat{p}(1-\hat{p})} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ où $\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$

Exercices

7.1 Voici la répartition, selon le mois, de 1 855 décès choisis au hasard parmi tous les décès qui ont eu lieu en Amérique du Nord en 1976.

Mois	Jan	Fév.	Mars	Avril	Mai	Juin	Juil.	Août	Sept.	Oct.	Nov.	Déc.	Total
Effectif	157	142	158	152	156	153	158	159	153	157	152	158	1855

- a) Peut-on attribuer les différences uniquement au hasard ?
- b) Supposons que ces données vous aient été fournies par un assistant en qui vous n’avez pas trop confiance. Les données ont-elles l’air d’avoir été manipulées ?

7.2 Dans une étude célèbre, des données ont été prélevées sur 6 587 suicides en France. Voici la distribution des suicides selon le jour de la semaine :

Jour	Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi	Dimanche	Total
Effectif	1001	1035	982	1033	905	737	894	6587

- a) Testez l’hypothèse que les suicides se répartissent uniformément sur les jours de la semaine.
- b) Selon une certaine conjecture, les taux de suicides diminuent à l’approche d’un week-end. Plus précisément, le taux quotidien global pour l’ensemble des jours du vendredi au dimanche est inférieur au taux quotidien global pour l’ensemble des jours du lundi au jeudi. Les données appuient-elles cette conjecture ?
- c) Testez l’hypothèse que chacun des jours du lundi au jeudi a le même taux de suicides.
- d) Testez l’hypothèse que chacun des jours du vendredi au dimanche a le même taux de suicides.
- e) Essayez de résumer en une phrase ou deux l’ensemble des conclusions tirées ci-dessus.

7.3 Un sociologue s’intéresse à la relation entre la couleur de la peau et la mobilité professionnelle. Il prélève un échantillon de 94 personnes de peau pâle, 175 personnes de peau moyenne, et 80 personnes de peau brune. Il construit une mesure de mobilité à l’aide de quoi il classifie ses sujets selon leur mobilité. Voici les résultats.

Mobilité	Couleur de la peau			Total
	Pâle	Moyenne	Brune	
Grande	35	84	51	170
Faible	59	91	29	179
Total	94	175	80	349

Tester l’hypothèse qu’il n’y a aucune relation entre la couleur de la peau et la mobilité professionnelle avec $\alpha = 0,05$.

7.4 Dans une étude sur l’absentéisme scolaire, une des causes d’absence est d’ordre non volontaire : décès, maladie, fatigue excessive, traitement médical. Parmi les 642 répondants à un questionnaire sur les motifs d’absence, 200 se classent dans cette catégorie. On a classifié ces 200 répondants selon le sexe et la cause.

	Décès	Maladie	Fatigue	Traitement médical
Garçons	4	39	25	18
Filles	7	56	32	19

Ces types de causes d’absence sont-ils liés au sexe ?

7.5 Un sociologue a fait subir un test d’estime de soi à 3 750 américains de 3 religions différentes. Voici les résultats

		Religion			Total
		Protestante	Catholique	Juive	
Estime de soi	Forte	1191	988	435	2614
	Faible	517	493	126	1136
		1708	1 481	561	3750

- a) Calculez, pour chaque religion, le pourcentage des sujets qui ont une forte estime de soi.

- b) Les différences entre ces pourcentages sont-elles significatives ?
 - c) La différence entre les catholiques et les protestants est-elle significative ?
 - d) En mettant les protestants et les catholiques ensemble, trouvez-vous ceux-ci différents des juifs ?
 - e) Résumez vos conclusions.
- 7.6 Les données du numéro précédent sont celles de la population générale. Les données dans le tableau suivant portent sur une sous-population : ceux qui n'ont pas complété l'école secondaire.

		Religion			Total
		Protestante	Catholique	Juive	
Estime de soi	Forte	330	234	89	653
	Faible	152	91	37	280
		482	325	126	933

Les conclusions sont-elles les mêmes ici ?

- 7.7 Une équipe de chercheurs dispose de données sur une population d'accidents : l'ensemble de tous les accidents qui ont eu lieu au Québec en 1980 et qui ont entraîné des blessures corporelles. Pour la plupart des variables, il était aisé d'obtenir les données pour la population entière. Pour certaines autres variables, comme celles qui identifiaient la position exacte du véhicule lors de l'accident, il était difficile d'en déterminer les valeurs et on ne pouvait se permettre de le faire pour la population entière. Il a donc fallu prélever un échantillon pour étudier la distribution de ces variables-là. On a prélevé un échantillon de 600 accidents. Malheureusement, l'échantillon n'a pas été prélevé de façon purement aléatoire, ce qui faisait douter de sa représentativité. Pour étudier la question de la représentativité, on a choisi une variable particulière, la variable "gravité de la blessure", dont on connaissait la distribution pour la population entière ainsi que pour l'échantillon. Formulez convenablement l'hypothèse que l'échantillon est représentatif et testez-la. Les distributions sont :

	Blessure				Total
	mortelle	très grave	grave	pas grave	
Fréquence dans la population	0,20	0,30	0,30	0,20	1
Fréquence dans l'échantillon	0,10	0,30	0,40	0,20	1

- 7.8 Déterminez s'il y a un lien entre le *type* de tumeur cérébrale, et son *site*. On distingue trois types de tumeurs

A_1 : tumeurs bénignes ; A_2 : tumeurs malignes ; A_3 : autres tumeurs

et les sites sont

B_1 : lobes frontaux ; B_2 : lobes temporaux ; B_3 : autres régions cérébrales

Les données sont présentées dans le tableau suivant :

	B_1	B_2	B_3	
A_1	23	21	34	78
A_2	9	4	24	37
A_3	6	3	17	26
	38	28	75	141

- 7.9 [Données du tableau A.8] Faites un test permettant de confirmer ou d'infirmer l'affirmation suivante : il y a une relation entre le fait de croire en l'astrologie et le fait de croire en l'évolution. Il y a plusieurs façons de procéder. Vous devez exprimer l'hypothèse nulle formellement. Vous pouvez décider de grouper les catégories, mais justifiez votre décision.
- 7.10 [Données du tableau A.8] Faites un test permettant de confirmer ou d'infirmer l'affirmation suivante : Il y a une relation entre le fait de croire en l'astrologie et le fait de croire que la longueur de la ligne de vie sur la paume contient de l'information sur la durée de vie. Il y a plusieurs façons de procéder. Vous devez exprimer l'hypothèse nulle formellement. Vous pouvez décider de grouper les catégories, mais justifiez votre décision.

- 7.11 [Données du tableau A.8] Faites un test permettant de confirmer ou d’infirmer l’affirmation suivante : les gens qui pratiquent leur religion préfèrent épouser des coreligionnaires. Il y a plusieurs façons de procéder. Vous devez exprimer l’hypothèse nulle formellement. Vous pouvez décider de grouper les catégories, mais justifiez votre décision.
- 7.12 Un psychologue prétend pouvoir décider si un sujet est schizophrène ou pas en examinant son Électroencéphalogramme (EEG). Pour vérifier cette prétention, on lui présente l’EEG de 50 sujets dont 40 sont en fait des schizophrènes. Se basant sur l’EEG, le psychologue fait un simple diagnostic : schizophrène ou normal. On constate que 34 des 50 diagnostics étaient corrects : 32 correctement classifiés comme schizophrènes et 2 correctement classifiés comme normaux. Deux étudiants *A* et *B* se proposent de tester l’hypothèse que l’EEG ne fournit au psychologue aucune information sur la schizophrénie.
- A - L’étudiant *A* traite le nombre de diagnostics corrects comme une variable binomiale avec $n = 50$ et teste l’hypothèse que $p = 1/2$ (Ayant observé $X = 34$ succès il rejette cette hypothèse et conclut que l’EEG est un bon indice de la schizophrénie)
- B- L’étudiant *B* dresse les tableaux suivants :

État du patient	Effectifs observés			État du patient	Effectifs théoriques		
	Classification du psychologue				Classification du psychologue		
	Correcte	Incorrecte			Correcte	Incorrecte	
Schizo.	32	8	40	Schizo.	27,2	12,8	40
Normal	2	8	10	Normal	6,8	3,2	10
	34	16	50		34	16	50

Il effectue ensuite le test d’indépendance classique. Dites pourquoi les deux étudiants se trompent et faites le test correctement.

- 7.13 On examine les deux premières questions d’un test psychologique en étudiant les résultats obtenus par 80 sujets sur ces deux questions. Parmi ces 80 sujets, 30 ont répondu correctement aux deux questions, 21 ont répondu incorrectement aux deux questions, 9 ont répondu correctement à la première et incorrectement à la deuxième, et 20 ont répondu incorrectement à la première et correctement à la deuxième. Trois hypothèses sont formulées ci-dessous. Pour chacune d’elles, dites si vous pouvez la tester avec les données à votre disposition. Si oui, testez-la, après l’avoir formulée correctement. Sinon, expliquez pourquoi les techniques que vous connaissez ne s’appliquent pas.

H_A : La moitié de la population est capable de répondre à la question 1.

H_B : Les deux questions mesurent des choses différentes ; elles n’ont rien à voir l’une avec l’autre ; elles sont indépendantes ; on peut être bon dans l’une et pas dans l’autre.

H_C : Les deux questions sont de même difficulté.

- 7.14 Pour savoir si une certaine maladie est aussi fréquente chez les hommes que chez les femmes, un chercheur prélève un échantillon de 450 hommes et 450 femmes. Il trouve que 100 hommes et 85 femmes sont atteints de la maladie. Testez l’hypothèse que les hommes et les femmes sont également vulnérables à la maladie.

Pour tester la même hypothèse, un autre chercheur procède différemment. Il prélève un échantillon de 900 personnes atteintes de la maladie. Parmi celles-ci, il trouve qu’il y a 420 femmes et 480 hommes. Testez l’hypothèse avec ces données.

La deuxième expérience fournit moins d’information que la première. Comment se fait-il qu’on peut s’en tirer quand même ?

- 7.15 Pour savoir s’il y a de la discrimination contre les femmes, on prélève les données suivantes sur 1 200 demandes d’admission aux deux universités d’une ville.

		Résultat de la demande		
		Acceptée	Refusée	Total
Sexe de l'individu	Masculin	416	184	600
	Féminin	184	416	600
Total		600	600	1 200

Peut-on conclure qu'il y a de la discrimination contre les femmes ? Voici les données sur les mêmes 1 200 candidatures, réparties cette fois selon l'université à laquelle la demande est adressée.

Université A

		Résultat de la demande		
		Acceptée	Refusée	Total
Sexe de l'individu	Masculin	8	92	100
	Féminin	92	408	500
Total		100	500	600

Université B

		Résultat de la demande		
		Acceptée	Refusée	Total
Sexe de l'individu	Masculin	408	92	500
	Féminin	92	8	100
Total		500	100	600

Vos conclusions changent-elles ?

7.16 On classe 661 finissants d'une université d'après les secteurs d'emploi et la discipline. Voici les données :

	Enseignement	Fonction publique	P.M.E.	Grandes corporations	
Génie	25	40	10	80	155
Sciences physiques	38	40	25	75	178
Sciences humaines	51	45	15	78	189
Lettres	65	50	4	20	139
Total	179	175	54	253	661

Y a-t-il une relation entre la discipline et le secteur d'emploi ?

7.17 Le tableau ci-dessous donne, pour une population de 87 627 mariages qui ont eu lieu au Canada en 1974, la distribution de la religion de l'époux et de la religion de l'épouse.

		Religion de l'épouse				
		Anglicane	Baptiste	Catholique romaine	Église Unie	
Religion de l'époux	Anglicane	7504	701	4731	4 680	17 616
	Baptiste	710	2222	992	1 050	4 974
	Catholique romaine	4589	914	26919	6 044	38 466
	Église Unie	4247	932	6143	15 249	26 571
Total		17 050	4 769	38 785	27 023	87 627

- Dressez un tableau qui expose clairement la forte dépendance entre ces deux variables.
- Testez l'hypothèse que 50% des femmes Baptistes épousent des coreligionnaires.
- Testez l'hypothèse que 50% des mariages entre catholiques et baptistes sont des mariages entre un homme catholique et une femme baptiste.
- Testez l'hypothèse que la probabilité qu'un homme épouse une coreligionnaire est la même pour les 4 religions représentées.

7.18 La distribution suivante a été dressée par Haberman (1978) à partir de données fournies par le National Opinion Research Center de l'Université de Chicago. Les variables sont le nombre d'années de scolarité (X) et l'attitude face à l'avortement (Y).

		Attitude			
		Pour	Mixte	Contre	
Scolarité	≤ 8	31	23	56	110
	9 — 12	171	89	177	437
	> 12	116	39	74	229
Total		318	151	307	776

- a) Testez l'hypothèse que la scolarité et l'attitude sont indépendantes.
- b) Les données ci-dessous portent sur deux groupes : des catholiques et des protestants du nord des États-Unis. Voici la distribution pour chaque groupe :

		Attitude : Catholiques			
		Pour	Mixte	Contre	
Scolarité	≤ 8	8	10	24	42
	9 — 12	65	39	89	193
	> 12	37	18	43	98
Total		110	67	156	333

		Attitude : Protestants			
		Pour	Mixte	Contre	
Scolarité	≤ 8	23	13	32	68
	9 — 12	106	50	88	244
	> 12	79	21	31	131
Total		208	84	151	443

La conclusion en (a) doit-elle être modifiée ?

7.19 Il y a quelques années, un comité fédéral de 5 personnes, chargé de prendre des décisions sur la télédiffusion, a statué sur une demande particulière en provenance de la Province de Québec. La décision prise par le comité était nettement défavorable au Québec. Les journalistes n'ont pas manqué de signaler que les deux membres francophones du comité s'étaient opposés à la décision, alors que les 3 membres anglophones l'avaient appuyée. Les journalistes insinuaient qu'il ne s'agissait pas là d'une simple coïncidence, mais bien de préjugés de la part des uns ou des autres. Ils faisaient donc un test d'hypothèse informel. Le but de cet exercice est de développer une façon plus formelle de le faire. Soit X_1 le nombre de francophones qui voteraient en faveur de la proposition et X_2 le nombre d'anglophones en faveur. Énoncez H_0 et montrez que la distribution conditionnelle de X_1 étant donnée $X_1 + X_2 = m$ est une variable de loi $\mathcal{H}(m ; n_1 + n_2)$ sous H_0 .

7.20 Il existe une méthode qu'on propose aux couples pour augmenter leurs chances d'avoir un enfant du sexe de leur choix. Voici les données sur 48 couples qui ont essayé la méthode.

		Les parents ont eu		Total
		une fille	un garçon	
Les parents voulaient	une fille	19	3	22
	un garçon	3	23	26
Total		22	26	48

La question posée est de savoir si la méthode en question est efficace. Certaines des approches décrites ci-dessous sont erronées car elles ne testent pas l'hypothèse qu'on veut vraiment tester. Dites ce que chacune des approches ci-dessous teste. Exprimez les hypothèses en termes des paramètres de la loi multinomiale $\mathcal{M}(48 ; p_1 ; p_2 ; p_3 ; p_4)$ dont les quatre effectifs sont une réalisation.

- a) *Première approche* On se base sur le fait que la méthode «a réussi» pour 32 des 48 couples et on teste l'hypothèse que la probabilité de réussite est $\frac{1}{2}$.

- b) *Deuxième approche* On effectue un test d'indépendance sur le tableau ci-dessus, tel quel.
- c) *Troisième approche* On teste l'hypothèse que la probabilité de succès pour un couple qui veut un garçon est $\frac{1}{2}$ et que la probabilité de succès pour un couple qui veut une fille est $\frac{1}{2}$ —un test du khi-deux à deux degrés de liberté.
- d) *Quatrième approche* On effectue un test d'indépendance sur un tableau constitué comme ceci :

		Les parents		Total
		Ont eu ce qu'ils voulaient	N'ont pas eu ce qu'ils voulaient	
Les parents voulaient	une fille	19	3	22
	un garçon	23	3	26
Total		42	6	48

- 7.21 Une usine fabrique des vis qu'elle met dans des boîtes de 100. Soit X le nombre de vis défectueuses dans une boîte. Supposons qu'une longue expérience du procédé de fabrication a permis de dresser la fonction de masse suivante :

x	0	1	2	3	4	5	6	Total
Fréquence	0,13	0,27	0,27	0,18	0,09	0,04	0,02	1

Un nouveau procédé est considéré, et pour le tester on l'emploie à la fabrication de 100 boîtes de vis. On obtient la distribution suivante :

x	0	1	2	3	4	5	6	Total
Fréquence	30	21	13	15	12	5	4	100

Le patron de l'usine demande à deux statisticiens de tester l'hypothèse nulle selon laquelle «le nouveau procédé est pareil à l'ancien». Le statisticien A fait un test basé sur la proportion p de vis défectueuses parmi les 10 000 de l'échantillon. Le statisticien B fait un test basé sur le nombre moyen \bar{X} de vis défectueuses par boîte. Effectuez les deux tests et commentez.

- 7.22 Pour comparer deux marques de bière, on fait une expérience avec 100 amateurs de chaque marque. Chaque groupe affirme connaître la différence entre les deux et préférer nettement la leur. On demande à chaque sujet d'identifier sa préférence, après avoir goûté les deux. Voici les résultats :

		Habituellement boivent		Total
		A	B	
Ont préféré	A	65	45	110
	B	35	55	90
Total		100	100	200

- a) Il n'y a pas de différence discernable entre les deux bières
 - b) Les buveurs de la bière B ne peuvent pas distinguer entre les deux bières
 - c) Les fidèles de la bière A discriminent mieux entre les deux marques que ceux de la bière B.
- 7.23 Pour comparer l'aptitude en mathématiques des Orientaux à celle des Américains de race blanche, Tsang [The Mathematics Education of Asian Americans, *Journal for Research in Mathematics Education* 15, No. 2, 115-122] a examiné les résultats de 10 097 étudiants orientaux au test de mathématiques du SAT [*Scholastic Aptitude Test*]. La distribution des scores est donnée dans le tableau suivant, qui présente également la distribution, en fréquences, des scores des Américains de race blanche.

Score	Orientaux (effectif)	Américains blancs (fréquence)
700 – 800	601	0,045
600 - 690	2001	0,172
500 – 590	3190	0,314
400 – 490	2788	0,301
300 – 390	1309	0,148
200 - 290	208	0,020
	10 097	1

Y a-t-il une différence significative entre les Orientaux et les Américains de race blanche ?

- 7.24 Au numéro précédent, les fréquences concernant les Américains de race blanche sont présentées comme si elles étaient fixes et connues pour la population entière, ce qui n'est pas en fait le cas. Elles sont basées sur un échantillon de 502 990 étudiants dont les effectifs, dans l'ordre des lignes du tableau ci-dessus, sont : 22 564 ; 86 521 ; 158 049 ; 151 466 ; 74 498 ; 9892. Utilisez ces données pour répondre à nouveau à la question posée au numéro précédent. Une discussion s'impose sur la relation entre les deux tests : a) Est-ce qu'on teste exactement la même chose dans les deux cas ? b) Aurait-on pu prévoir que la conclusion n'aurait pas changé. c) Pourquoi la différence entre les valeurs de khi-deux calculées ici et au numéro précédent est-elle si minuscule ?
- 7.25 Aux deux numéros précédents, on aurait pu traiter les scores comme des variables quantitatives. Voici, approximativement, la moyenne et l'écart-type des scores pour les deux groupes :

Groupe	Moyenne	Écart-type
Orientaux	$\bar{X} = 522,0$	$S = 115,7$
Américains blancs	510,5	112,5

- a) Testez l'hypothèse que les moyennes des populations respectives sont égales.
- b) Considérez que les résultats obtenus par les Orientaux proviennent d'une population de moyenne μ et testez l'hypothèse que $\mu = 510,5$ (la moyenne des blancs). Ici on voit peut-être mieux que dans les problèmes précédents pourquoi cette approche est fautive.
- c) Confirmez, cependant, que la différence entre les tests en a) et b) est minuscule et expliquez pourquoi.

Exercices théoriques

- 7.26 Après avoir administré un test d'aptitude à un groupe de 100 sujets, on se propose de tester l'hypothèse que les questions 1 et 2 sont de même difficulté, c'est-à-dire, que la probabilité qu'un sujet réponde correctement est la même pour les deux questions. Supposons que les résultats obtenus par les 100 sujets sur les deux questions sont les suivants :

		Question 2		
		Correcte	Incorrecte	Total
Question 1	Correcte	22	35	57
	Incorrecte	11	32	43
Total		33	67	100

Les taux de succès aux questions 1 et 2 sont, respectivement, de 57% et de 33%. On ne peut pas faire un test d'égalité de proportions pour vérifier si la différence est significative, car il ne s'agit pas de deux échantillons indépendants : ce sont les mêmes sujets qui ont répondu aux deux questions. Désignons les 4 observations par

X_{11}	X_{12}	$X_{1.}$
X_{21}	X_{22}	$X_{2.}$
$X_{.1}$	$X_{.2}$	n

Ces quatre variables aléatoires suivent conjointement une loi $\mathcal{M}(n ; p_{11}, p_{12}, p_{21}, p_{22})$.

- a) Montrez que la distribution conditionnelle de X_{12} étant donné $X_{12} + X_{22} = m$ est $\mathcal{B}(m ; p_{12}/(p_{12} + p_{21}))$.
- b) Formulez l’hypothèse nulle en fonction des paramètres de la distribution conditionnelle de X_{12}
- c) Effectuez le test à partir de la distribution conditionnelle de X_{12} .

7.27 L’objectif de cet exercice est de montrer l’équivalence entre un certain test d’ajustement et un test sur une proportion. Considérons les données suivantes sur 120 cas de suicide :

Date du suicide	1 décembre – 31 mars	1 avril — 30 novembre
Effectif	50	70

L’hypothèse nulle est que les suicides ne sont pas affectés par la saison. Soit X le nombre de suicides parmi 120 qui ont eu lieu entre le 1^{er} décembre et le 31 mars. Alors $X \sim \mathcal{B}(120 ; p)$, et l’hypothèse à tester est $H_0 : p = 1/3$ contre l’alternative $H_1 : p \neq 1/3$. Le test peut être basé sur la *statistique* $Z =$

$$\frac{X - n/3}{\sqrt{n(1/3)(2/3)}}$$

qui suit, sous H_0 , une loi $\mathcal{N}(0 ; 1)$. La région critique est $|Z| > z_{\alpha/2}$. Mais le test peut

aussi bien être basé sur Z^2 , et puisque $Z^2 \sim \chi^2_1$, on a aussi la région critique $Z^2 > \chi^2_{1;\alpha}$.

- a) Vérifier numériquement que les deux tests sont équivalents.
- b) Vérifiez numériquement que $Z^2 = \sum_{i=1}^2 \frac{(O_i - T_i)^2}{T_i}$, la statistique utilisée dans ce chapitre pour faire un test d’ajustement.
- c) Vérifiez en général que pour tester l’hypothèse $H_0 : p = p_0$, les régions critiques

$$|Z| > 1,96 \text{ avec } Z = \frac{X - np_0}{\sqrt{np_0(1 - p_0)}}$$

et

$$\sum_{i=1}^2 \frac{(O_i - T_i)^2}{T_i} > \chi^2_{1;\alpha}$$

sont équivalents.

7.28 Supposons qu’on dispose de données provenant de 3 études indépendantes sur les suicides. La première étude, basée sur 120 cas de suicide, rapporte que 50 de ces suicides ont lieu entre le 1^{er} décembre et le 31 mars. La deuxième étude, basée sur 80 cas, rapporte que 15 de ceux-ci ont eu lieu entre le 1^{er} janvier et la fin février. La troisième étude enfin, basée sur 90 cas, rapporte que 27 ont eu lieu entre 1^{er} janvier et le 31 mars. Vous voulez savoir si la saison froide a un effet sur le taux de suicide. Faites un test avec $\alpha = 0,01$, en employant l’ensemble des données à votre disposition que vous pouvez disposer comme suit

	Nombre de suicides en hiver sur un total de
1 ^{ère} étude	50	120
2 ^e étude	15	80
3 ^e étude	27	90

L’hypothèse qu’on veut tester est l’hypothèse que les suicides ne sont pas affectés par la saison.

- a) En un premier temps, testez cette hypothèse trois fois, une fois dans le cadre de chaque étude.
- b) Maintenant montrez qu’on peut tester l’hypothèse globalement, en mettant ensemble les ressources des trois études. Il s’agit d’effectuer un test basé sur la somme des trois khi-deux calculés séparément : quelle est la région critique ? Justifiez votre procédure.

7.29 Montrez que dans un tableau 2×2 , la statistique χ^2 pour tester l’indépendance est égale à Z^2 , où

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

La notation est définie ci-dessous :

X_1	$n_1 - X_1$	n_1
X_2	$n_2 - X_2$	n_2
$n \hat{p}$	$n(1 - \hat{p})$	n

$$\hat{p}_1 = X_1/n_1, \hat{p}_2 = X_2/n_2, \hat{p} = \frac{X_1 + X_2}{n}, n = n_1 + n_2.$$

7.30 De très nombreuses études ont été effectuées dans le but de démontrer que les premiers-nés d'une famille sont d'un caractère distinct, plus conservateurs que la norme. L'une de ces études est basée sur le fait que lors de la condamnation de Louis XVI, les députés de la Convention nationale qui étaient en faveur de l'exécution étaient beaucoup plus souvent premiers-nés que les autres. Voici les données :

	En faveur de la guillotine	Contre la guillotine	
Premiers-nés	182	67	249
Non premiers-nés	179	293	472
	361	360	721

- Montrez qu'un test du khi-deux (ou autre test) rejette effectivement l'hypothèse H_0 qu'il n'y a pas de relation entre l'ordre de naissance et l'attitude face au traitement à accorder à Louis XVI.
- Maintenant on vous propose de montrer que cette dépendance pourrait n'être qu'un artifice, dû, par exemple, au fait que les gens susceptibles de favoriser la guillotine auraient en moyenne moins d'enfants (bien qu'on serait mal pris s'il fallait expliquer pourquoi ce serait le cas). Supposons les distributions suivantes :

Personnes en faveur de la guillotine (on traite les enfants uniques comme des premiers-nés) :

Nombre d'enfants	1	2	3	4	
Proportion	0,3	0,1	0,1	0,5	1

Personnes contre la guillotine :

Nombre d'enfants	4	5	6	
Proportion	0,1	0,4	0,5	1

Déterminer les effectifs théoriques. Pourquoi est-ce que ceux-ci contredisent la conclusion en a) ?

- Le numéro précédent n'est qu'un exemple des nombreuses études destinées à tester H_0 . Puisqu'à chaque étude il y a une probabilité α de rejeter H_0 à tort, il n'est pas certain que le rejet observé dans l'exemple n'est pas un de ces rejets erronés. Il faut donc analyser l'ensemble des études du genre pour en arriver à un bilan global (on appelle cela une «méta-analyse»). C'est ce qu'a fait un chercheur, à partir d'un grand nombre d'études. Parmi les 196 études qu'il a examinées, il y en avait 72 qui rejetaient H_0 (donc confirmaient l'existence d'un lien). Supposez que toutes les études ont été faites à un niveau $\alpha = 0,05$ et dites si l'ensemble des études permet de rejeter H_0 .
- Au numéro précédent, supposons qu'on ait retenu la valeur p de chacun des 196 tests, donc 196 nombres compris dans l'intervalle $[0 ; 1]$, et que leur moyenne est de 0,07 avec leur écart type de 0,04. Sachant que sous H_0 ces nombres sont de loi uniforme sur $[0 ; 1]$, testez H_0 .
- Une usine fabrique des vis qu'elle met dans des boîtes de 100. Soit X le nombre de vis défectueuses dans une boîte. Pour comparer deux procédés de fabrication, P_1 et P_2 , on les emploie à la fabrication de 100 boîtes chacun. Voici la distribution des X pour chacun des deux procédés :

x	0	1	2	3	4	5	6	Total
-----	---	---	---	---	---	---	---	-------

Effectif, procédé P ₁	13	27	27	18	9	4	2	100
Effectif, procédé P ₂	30	21	13	15	12	5	4	100

Le patron de l'usine demande à trois statisticiens de tester l'hypothèse, formulée vaguement comme «les deux procédés sont pareils». Le statisticien A effectue un test d'indépendance à partir du tableau ci-dessus, tel quel. Le statisticien B considère qu'il a deux échantillons de 10 000 vis chacun et fait un test d'égalité de proportions. Le statisticien C, enfin, fait un test basé sur le nombre moyen de vis défectueuses par boîte. Effectuez les trois tests et commentez. Sont-ils tous valides ?