

Chapitre 6 : Tests d'hypothèses

Les statistiques peuvent être une aide à la décision permettant de choisir entre deux hypothèses. Par exemple, on teste un nouveau traitement anti-cancéreux sur un échantillon de souris. Avant traitement, le taux de développement de la maladie p est de 20%. On veut savoir si le taux observé après traitement est significativement différent ou non. Les différences observées pouvant être dues au seul hasard, il s'agit de donner un sens précis à "significativement différent".

1 Principe général d'un test.

Hypothèses statistiques. On émet deux hypothèses :

- une "hypothèse nulle", notée H_0 : elle doit toujours se traduire mathématiquement par une *égalité* sur un paramètre p .
- une hypothèse alternative, notée H_1 : elle doit être incompatible avec H_0 (elles ne peuvent être vraies ensemble). Elle se traduit par des signes \neq (test bilatéral), $<$ ou $>$ (test unilatéral).

Exemple du traitement médical :

- H_0 : le traitement est inactif. Elle se traduit par l'égalité $p = 0,2$ (le taux n'a pas varié).
- H_1 : le traitement est actif. Il se traduit par $p \neq 0,2$. On aurait aussi pu considérer $p < 0,2$, ou encore $p > 0,2$: le médicament agit en aggravant les choses.

Attention ! on ne teste que les hypothèses nulles, le but étant d'essayer de montrer qu'elles sont fausses. Le test n'a que deux issues : ou bien H_0 est fausse (on la rejette et accepte son contraire : "le traitement est actif" avec un certain risque d'erreur fixé à l'avance), ou bien le test est non concluant (on ne peut conclure d'après le test si le médicament est actif ou non).

Le principe d'un test d'hypothèses est analogue à un raisonnement par l'absurde :

1. on suppose que H_0 est vraie (*Exemple* : $p = 0,2$) ;
2. sous cette hypothèse, on cherche alors un intervalle I_α –dit intervalle de pari– autour de p tel que P_n (la variable aléatoire estimant p sur les échantillons de taille n) appartienne à I_α avec un risque d'erreur α fixé à l'avance : $p(P_n \in I_\alpha) = 1 - \alpha$. Pour $\alpha = 0,05$, la signification de cet intervalle de pari est donc la suivante :

si H_0 est vraie, la valeur observée du paramètre sur un échantillon a 95 chances sur 100 d'appartenir à l'intervalle de pari $I_{0,05}$.

3. on regarde sur un échantillon donné la valeur "observée" p_n de P_n .
 - (i) si p_n n'appartient pas à I_α (dans le cas bilatéral), on obtient une contradiction (avec un certain risque d'erreur). **On rejette donc H_0** –ce qui revient à accepter son contraire– avec un risque de se tromper α . (*Sur l'exemple : on conclut que le médicament agit, avec un risque d'erreur de 5%*).

Le complémentaire de l'intervalle de pari I_α s'appelle donc "zone de rejet". Dans le cas d'un test unilatéral, la zone de rejet se limite à une des deux composantes du complémentaire de I_α (voir TD).

- (ii) si p_n appartient à I_α , il n'y a aucune contradiction et le test ne prouve rien. La seule chose que l'on peut dire est : **Le test ne permet pas de rejeter H_0** . *Sur l'exemple, on peut seulement dire : l'activité du médicament n'est pas démontrée par ce test.*

Dans la pratique, dans la situation (ii-), on utilise si on peut un test plus puissant, avec un échantillon plus gros. S'il reste non concluant, on accepte alors l'hypothèse H_0 , non pas pour des raisons mathématiques, mais "par défaut" ou "par sécurité" : on ne met pas sur le marché un médicament dont on n'a pas pu prouver l'efficacité.

La méthode de calcul de I_α et donc des zones de rejet sera à préciser dans chaque situation : on remplacera P_n par une variable aléatoire centrée réduite Z_n qui, si H_0 est vraie et si l'échantillon est suffisamment grand, suivra une loi connue par des tables (loi normale, lois du χ^2 ...).

Le risque α s'appelle **risque de première espèce**. C'est la probabilité de rejeter H_0 alors qu'elle était vraie. On le fixe au départ de la procédure :

- pas trop grand. Sinon le test a un trop grand risque d'erreur.
- pas trop petit. Sinon avec un risque voisin de 0, I_α sera très grand et on sera presque toujours dans le cas (ii-), et le test ne sera presque jamais concluant.

Le bon compromis assez généralement adopté est $\alpha = 0,05$.

Le risque de deuxième espèce est la probabilité β de ne pas conclure alors que H_0 était fausse, et aurait dû être rejetée. Il est en général difficile à évaluer. Retenons seulement qu'il diminue avec la taille n de l'échantillon. La puissance du test est le nombre $1 - \beta$ et augmente avec la taille de l'échantillon.

2 Test de comparaison de proportions.

Comparaison d'une proportion théorique avec une proportion observée.

Nous décrivons ce test sur l'exemple précédent de l'efficacité d'un traitement. Sans le traitement, on sait que la proportion théorique de souris malades est de $p = 0,2$. Sur un échantillon de 100 souris traitées, la proportion observée p_n est de 0,17.

1. On pose :
 H_0 : le traitement est inactif. La proportion dans la population traitée reste $p = 0,2$.
 H_1 : le traitement est actif.
2. On suppose H_0 vraie. De plus n est suffisamment grand : $100 > 30$. On sait alors que la variable aléatoire

$$Z_n = \frac{P_n - p}{\sqrt{p(1-p)/n}}$$

suit approximativement la loi normale centrée réduite (c'est une conséquence du théorème central limite). Au risque $\alpha = 0,05$, l'intervalle de pari est l'intervalle $[-z, +z]$ tel que

$$p(|Z_n| < z) = 1 - 0,05.$$

D'après la table de $\mathcal{N}(0, 1)$, $z = 1,96$ (voir ch. 5). Donc, $I_{0,05} = [-1,96, +1,96]$. La signification de cet intervalle est ici la suivante : *si le traitement est inactif, la valeur observée a 95 chances sur 100 d'appartenir à cet intervalle.*

3. On calcule la valeur observée z_n de Z_n sur l'échantillon considéré :

$$z_n = 0,03/\sqrt{0,2 \times 0,8/100} = 0,75.$$

Puisque cette valeur est dans l'intervalle de pari, on ne peut pas conclure sur l'efficacité ou l'inefficacité du traitement.

Exercice 1. *Puisque le test n'est pas concluant, on le refait sur un échantillon de 1000 souris traitées. On observe à nouveau un taux de souris malades de 0,17. Quelle est la conclusion du test (au même risque d'erreur 5%) ?*

Degré de signification d'un test. Lorsqu'un test a permis de conclure au rejet de H_0 (et seulement dans ce cas !), on veut savoir si la conclusion est acquise par une marge confortable, ou au contraire de manière "limite". On veut donc connaître jusqu'à quelle valeur du risque α on aurait continué à accepter cette conclusion :

*On appelle degré de signification du test le plus petit α
tel que I_α ne contienne pas la valeur observée.*

La confiance en la conclusion du test sera d'autant meilleure que ce degré de signification sera proche de zéro.

Exercice 2. *Calculer le degré de signification du test précédent.*

Comparaison de deux proportions observées.

On veut comparer non plus une proportion théorique à une proportion observée, mais deux proportions observées p_A et p_B sur deux échantillons distincts A et B d'effectifs n_A et n_B . L'hypothèse à tester est

$$H_0 : p_A = p_B.$$

La stratégie étant identique, il suffit de préciser la variable aléatoire Z à utiliser pour obtenir l'intervalle de pari I_α . On pose

$$Z = \frac{P_A - P_B}{\sigma(P_A - P_B)}.$$

où P_A et P_B sont les variables aléatoires : proportion sur les échantillons de tailles n_A et n_B . La difficulté est ici de calculer l'écart-type $\sigma(P_A - P_B)$. On considère que les deux échantillons définissent des variables indépendantes. En conséquence,

$$\sigma^2(P_A - P_B) = \sigma^2(P_A) + \sigma^2(P_B) = \frac{pq}{n_A} + \frac{pq}{n_B}$$

où p est la proportion théorique et $q = 1 - p$. Cette proportion étant inconnue, on l'approxime par la moyenne pondérée des deux proportions observées :

$$p \simeq \frac{n_A p_A + n_B p_B}{n_A + n_B}.$$

On trouve l'intervalle de pari en utilisant alors le fait suivant : si H_0 est vraie, et si les deux échantillons sont suffisamment grands (dans la pratique : $n_A p$, $n_A q$, $n_B p$ et $n_B q$ doivent être supérieurs à 5), la variable Z ainsi obtenue suit approximativement la loi normale centrée réduite.

3 Tests de comparaison de deux moyennes

A nouveau, il suffit de préciser la variable aléatoire Z à utiliser (et sa loi) pour obtenir l'intervalle de pari sous l'hypothèse : " H_0 est vraie" (pas de différence significative entre les deux moyennes).

Comparaison d'une moyenne théorique μ à une moyenne observée m_n sur un échantillon de taille n . Pour tester l'égalité de ces deux moyennes, on utilise la variable aléatoire :

$$Z_n = \frac{M_n - \mu}{\sigma/\sqrt{n}}$$

qui, si H_0 est vraie, et si $n > 30$, suit approximativement la loi normale centrée réduite. Ici, M_n désigne l'estimateur "moyenne d'un échantillon de taille n " et σ l'écart-type théorique du caractère. S'il est inconnu, on remplace σ/\sqrt{n} par $\Sigma_n/\sqrt{n-1}$ ($=\Sigma_{n-1}/\sqrt{n}$), où Σ_n est la variable aléatoire : "écart-type d'un échantillon de taille n ".

Comparaison de deux moyennes observées sur deux échantillons de taille n_A et n_B . Soit M_A et M_B les variables aléatoires "moyennes observées sur les échantillons de taille n_A et n_B ", et S_A et S_B les variables aléatoires "écarts-type estimés sur les échantillons de taille n_A et n_B ". Si H_0 est vraie et si $n_A \geq 30$ et $n_B \geq 30$ la loi

$$Z = \frac{M_A - M_B}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}}$$

suit approximativement la loi normale centrée réduite.

4 Comparaison de deux distributions statistiques.

On considère un paramètre prenant k valeurs (k est parfois appelé la *modalité*). On connaît une répartition théorique des fréquences de ces valeurs : p_1, \dots, p_k . Sur un échantillon de taille n , on observe des effectifs o_1, \dots, o_k . On veut tester l'hypothèse :

H_0 : les deux répartitions théoriques et observées coïncident.

Attention : les tests de comparaisons de distributions comparent des *effectifs*, non des proportions. On calcule donc les effectifs théoriques attendus sur l'échantillon de taille n : $n_i = np_i$. On les compare aux effectifs observés o_i à l'aide de la variable aléatoire :

$$Z = \sum_{i=1}^k \frac{(o_i - n_i)^2}{n_i}.$$

où les O_i sont les variables aléatoires prenant les valeurs o_i sur un échantillon donné. On admet ici le résultat suivant :

si H_0 est vraie et si les effectifs théoriques vérifient $n_i \geq 5$, alors Z suit une loi connue :
la loi du χ^2 à $\nu = k - 1$ degrés de liberté (ou loi de Pearson).

Loi du χ^2 à ν degrés de liberté : cette loi est la somme $X_1^2 + \dots + X_\nu^2$ des carrés de ν variables indépendantes qui suivent toutes la loi normale centrée réduite. Elle permet de généraliser l'utilisation de la loi normale pour deux proportions $p, q = 1 - p$ à une répartition en plusieurs

proportions $p_1, \dots, p_{\nu+1}$. Les tables du χ^2 (voir à la fin de ce chapitre) donnent pour chaque valeur de ν et pour un risque d'erreur α le nombre z_α tel que

$$p(\chi_\nu^2 \geq z_\alpha) = \alpha$$

ou ce qui est équivalent, z_α tel que $p(\chi_\nu^2 \leq z_\alpha) = 1 - \alpha$. L'intervalle de pari est ici $[0, z_\alpha]$.

Si la valeur calculée de Z **dépasse** ce seuil z_α , on considère qu'on est dans la zone de rejet et **on refuse** H_0 avec un risque d'erreur α : les deux distributions sont distinctes. Dans le cas contraire, le test n'a pas permis d'observer de différences entre les deux distributions (mais un test plus puissant pourrait éventuellement le faire). En l'absence d'indication contraire, on conclut que les deux distributions coïncident.

Exercice 3. Les résultats à une élection nationale opposant 4 partis ont donné les proportions :

$p_1 = 10\%$, $p_2 = 5\%$, $p_3 = 65\%$, $p_4 = 20\%$.

Dans un village de 200 habitants, le nombre de voix obtenues par chaque parti a été de :

$o_1 = 9$, $o_2 = 5$, $o_3 = 141$, $o_4 = 45$.

On veut tester l'hypothèse : ce village est représentatif des tendances nationales. Quelle est votre conclusion ?

(Méthode : on calculera les effectifs théoriques, puis la valeur prise par Z sur l'échantillon, puis ν et le seuil $z_{0,05}$ et on conclura.

Remarques.

- Ce test permet de vérifier si une distribution statistique suit une loi connue (loi normale, de Poisson, binomiale etc...). Cette loi hypothétique donne sur un échantillon des effectifs théoriques que l'on compare aux effectifs observés.

- lorsque la condition $o_i \geq 5$ n'est pas vérifiée pour tout i , on peut faire des regroupements de classe pour l'obtenir.

5 Test d'indépendance.

On veut tester l'indépendance de deux caractères statistiques X et Y d'une population. On pose :

H_0 : les deux caractères sont indépendants.

On dispose d'un tableau des effectifs observés o_{ij} , $i = 1, \dots, k$, $j = 1, \dots, l$ sur un échantillon. A partir des effectifs marginaux observés, on construit un second tableau des effectifs théoriques n_{ij} (pas nécessairement entiers!) obtenus sous l'hypothèse d'indépendance en utilisant la relation :

$$n_{ij} = \frac{n_{i.} \times n_{.j}}{n}.$$

On compare ces deux tableaux en utilisant la variable aléatoire

$$Z = \sum_{i,j} \frac{(n_{ij} - O_{ij})^2}{n_{ij}}.$$

où les O_{ij} sont les variables aléatoires prenant les valeurs o_{ij} sur un échantillon donné. On détermine l'intervalle de pari en utilisant le résultat suivant :

si H_0 est vraie et si les effectifs théoriques vérifient $n_{ij} \geq 5$, alors Z suit la loi du χ^2 à $\nu = (k-1)(l-1)$ degrés de liberté.

On conclut comme précédemment : si la valeur calculée de l'échantillon dépasse le seuil z_α donné par la table de χ^2_ν , on rejette H_0 avec un risque d'erreur α : les deux caractères sont liés. Dans le cas contraire, le test ne peut conclure au rejet de H_0 .

Exemple. Les caractères "couleur des yeux / couleurs des cheveux" observés sur un échantillon de 124 personnes donnent le tableau suivant :

Yeux \ cheveux	blonds	bruns	roux	noirs	total
bleus	25	9	7	3	44
gris	13	17	7	10	47
marrons	7	13	5	8	33
total	45	39	19	21	124

Exercice 4. Complétez le tableau des effectifs théoriques ci-dessous sous l'hypothèse d'indépendance (les effectifs obtenus ne sont plus nécessairement des entiers ; arrondir au centième).

Yeux \ cheveux	blonds	bruns	roux	noirs	total
bleus					44
gris					47
marrons					33
total	45	39	19	21	124

Exercice 5. Calculez :

- la valeur de Z prise sur l'échantillon ;
- le nombre de degré de liberté ν de sa loi ;
- le seuil z_α de l'intervalle de pari au risque 0,05 ;
- conclure le test.

Test sur le chapitre 6

1. On veut tester une hypothèse H_0 face à l'hypothèse alternative H_1 . Qu'appelle-t-on risque de première espèce ? risque de deuxième espèce ?
2. On veut comparer une proportion observée p_n sur un échantillon de taille n , à une proportion théorique p . Décrire la méthode et la variable aléatoire utilisée.
3. Même question pour la comparaison d'une moyenne théorique et d'une moyenne observée.
4. Décrire la méthode permettant de tester l'indépendance de deux variables aléatoires.

Chapitre 6 : Travaux dirigés

1. (*Comparaison d'une moyenne avec une moyenne théorique.*) Une machine fabrique des pièces dont la longueur théorique est $\mu = 100$ cm. Pour vérifier si la machine est bien réglée, on prélève un échantillon de 50 pièces, sur lequel on mesure la longueur moyenne $m = 99$ cm avec un écart-type $\sigma = 2$ cm. La machine est-elle bien réglée ? (on souhaite une réponse avec un risque d'erreur de 1%).

On suivra la méthodologie suivante :

- écrire l'hypothèse H_0 à tester ainsi que son hypothèse alternative.
- écrire la variable aléatoire Z à considérer (voir le cours), l'intervalle de pari et la zone de rejet au risque donné.
- évaluer la valeur z de Z sur l'échantillon considéré.
- conclure le test et répondre à la question posée.

2. (*Un test de comparaison de moyennes unilatéral.*) Dans un atelier de réparation de téléviseurs, on a noté les temps de réparation (en mn) sur 50 téléviseurs de marque A et sur 40 téléviseurs de marque B. Sur ces deux populations de téléviseurs, on a obtenu les résultats suivants :

Durée des interventions	[0,20[[20,40[[40,60[[60,80[[80,100[[100,120[
Nombre d'interventions sur A	1	6	11	15	10	7
Nombre d'interventions sur B	3	7	12	9	6	3

(facultatif) Vérifier que les moyennes et écart-types de ces deux échantillons sont : $m_A = 69,2$, $m_B = 58,5$, $\sigma_A = 25,6$, $\sigma_B = 29,49$.

On veut savoir si la relation $m_A > m_B$ constatée est significative. Pour cela on teste les hypothèses : $H_0 : m_A = m_B$ contre $H_1 : m_A > m_B$ au risque de 5%. On considère la variable Z du ch 6, §3 (comparaison de deux moyennes observées). Si H_0 est vraie Z suit approximativement $\mathcal{N}(0, 1)$. L'intervalle de pari est ici $] -\infty, z_\alpha]$ et la zone de rejet $Z > z_\alpha$ de sorte que $p(Z > z_\alpha) = 5\%$.

- (a) Déterminer cette zone de rejet.
 - (b) Calculer la valeur z de Z prise sur les observations, et conclure le test.
 - (c) Quelle conclusion obtient-on avec un test bilatéral au même risque d'erreur ?
3. Une fabrique interroge 1000 automobilistes auxquels elle a vendu deux types de pneus distincts. On veut savoir si les qualités d'adhérence du véhicule sont liées à ce choix de pneus. Le questionnaire donne les résultats suivants :

Type de pneus / Adhérence	très bonne	correcte	médiocre
Type 1	300	150	90
Type 2	240	120	100

Tester l'indépendance de ces deux distributions au seuil de 5% (voir ch 6, §5).

4. (*Examen 2006*) Un dispositif de veille informatique d'un réseau compte le nombre d'incidents survenus pendant une période de 50 jours. Soit n_i le nombre de jours avec x_i incidents. On obtient la distribution suivante :

$x_i :$	0	1	2	3	4
$n_i :$	21	18	7	3	1

- (a) i. Déterminer le nombre moyen d'incidents par jour.
 ii. Déterminer l'écart-type de cette distribution.
- (b) On veut approximer cette distribution par une variable aléatoire X qui à chaque journée associe le nombre d'incidents survenus. On fait l'hypothèse que cette variable aléatoire X suit une loi de Poisson :

$$p(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

de paramètre $\lambda = 0.9$.

- i. Calculer la loi de probabilité de cette distribution théorique pour les valeurs $k = 0, 1, 2, 3, 4$.
- ii. En déduire les effectifs théoriques t_i correspondants, pour un effectif total de 50 jours (on gardera des valeurs décimales avec une précision de 0,01).
- iii. Avec cette loi théorique, déterminer la probabilité qu'il y ait au plus un accident dans une journée donnée.
- (c) On veut vérifier si cette hypothèse d'ajustement par une distribution de Poisson est valide à l'aide d'un test statistique de type χ^2 .
- i. Pour chacune des deux distributions (observée et théorique), on décide de regrouper les trois dernières classes en une seule. Pour quelle raison doit-on le faire ? Ecrire les deux distributions ainsi obtenues.
- ii. Quelle quantité doit-on calculer pour vérifier la conformité de ces deux distributions ? Quelle est sa valeur ?
- iii. Sous l'hypothèse H_0 : "les deux distributions sont identiques", quelle loi suit cette quantité ? (préciser le nombre de degrés de libertés de cette loi). Peut-on accepter H_0 au seuil d'erreur de 0.05 ?
-

Tables du Chi-Deux

Pour chaque valeur de probabilité α (par exemple 0,05 ou 0,01) et chaque degré de liberté ν de 1 à 30, la table donne le seuil z_α tel que

$$p(\chi_\nu^2 > z_\alpha) = \alpha.$$

$\nu \backslash \alpha$	0,99	0,975	0,95	0,05	0,025	0,01
1	0.000	0.001	0.004	3.841	5.024	6.635
2	0.020	0.051	0.103	5.991	7.378	9.210
3	0.115	0.216	0.352	7.815	9.348	11.345
4	0.297	0.484	0.711	9.488	11.143	13.277
5	0.554	0.831	1.145	11.070	12.833	15.086
6	0.872	1.237	1.635	12.592	14.449	16.812
7	1.239	1.690	2.167	14.067	16.013	18.475
8	1.646	2.180	2.733	15.507	17.535	20.090
9	2.088	2.700	3.325	16.919	19.023	21.666
10	2.558	3.247	3.940	18.307	20.483	23.209
11	3.053	3.816	4.575	19.675	21.920	24.725
12	3.571	4.404	5.226	21.026	23.337	26.217
13	4.107	5.009	5.892	22.362	24.736	27.688
14	4.660	5.629	6.571	23.685	26.119	29.141
15	5.812	6.908	7.962	26.296	28.845	32.000
17	6.408	7.564	8.672	27.587	30.191	33.409
18	7.015	8.231	9.390	28.869	31.526	34.805
19	7.633	8.907	10.117	30.144	32.852	36.191
20	8.260	9.591	10.851	31.410	34.170	37.566
21	8.897	10.283	11.591	32.671	35.479	38.932
22	9.542	10.982	12.338	33.924	36.781	40.289
23	10.196	11.689	13.091	35.172	38.076	41.638
24	10.856	12.401	13.848	36.415	39.364	42.980
25	11.524	13.120	14.611	37.652	40.646	44.314
26	12.198	13.844	15.379	38.885	41.923	45.642
27	12.879	14.573	16.151	40.113	43.195	46.963
28	13.565	15.308	16.928	41.337	44.461	48.278
29	14.256	16.047	17.708	42.557	45.722	49.588
30	14.953	16.791	18.493	43.773	46.979	50.892