

Principe des tests statistiques – Comparaison d’une moyenne et d’une proportion à une norme – Comparaison de deux moyennes avec échantillons indépendants ou appariés

Exercice 1

Soit X une variable aléatoire et x une réalisation de X . On veut construire un test basé sur la valeur de x . On pose que sous H_0 , X est distribué uniformément sur $[0 ; 1]$, et que sous H_1 , X est distribué uniformément sur $[0 ; 2]$. On dessinera les densités de probabilités de X sous H_0 et sous H_1 pour répondre aux questions.

Q1. On définit la région de rejet de H_0 par $\{x : x < 0,25\}$. Cochez LA proposition exacte.

- a- le risque de première espèce est la probabilité de rejeter H_1 à tort
- b- le risque de première espèce est égal à 25%
- c- la puissance est la probabilité de rejeter H_0 à tort
- d- la puissance est égale à 25%
- e- le risque de deuxième espèce est égal à 75%

Q2. On cherche maintenant à définir une région de rejet de H_0 de seuil 5%. Cochez la ou les propositions exactes.

- a- tout intervalle de longueur 0,05 sur $[0 ; 1]$ est une région de rejet de H_0 au seuil 5%
- b- on ne peut pas calculer la puissance du test qui aurait pour région de rejet de H_0 un intervalle de longueur 0,05 sur $[0 ; 1]$
- c- tous les intervalles de longueur 0,05 sur $[0 ; 1]$ sont équivalents du point de vue de la puissance
- d- Tout intervalle de longueur 0,05 sur $[1 ; 2]$ est une région de rejet de H_0 au seuil de 5%
- e- La puissance d’un test de seuil 5% qui comprend un intervalle de longueur 0.05 sur $[0 ; 1]$ et l’intervalle $[1 ; 2]$ est de 0,525

Exercice 2

Des études antérieures ont montré que la prévalence d’une maladie M en 2000 était de 30%. On se propose de réestimer cette prévalence en 2014 à partir d’un échantillon. On trouve alors 40% de malades sur cet échantillon.

Q3. Cochez la ou les propositions exactes.

- a- si la prévalence de la maladie M dans la population est inchangée, son intervalle de pari à 95% est $[27\% ; 33\%]$ (arrondi à 1% près)
- b- à partir de cet intervalle de pari à 95%, on peut conclure que la prévalence de la maladie n’a pas changé
- c- on ne peut pas calculer cet intervalle de pari
- d- on ne peut rien conclure quant à l’hypothèse que la prévalence de la maladie n’a pas changé
- e- si on avait pu calculer l’intervalle de pari de la prévalence, on aurait pu faire le test

Q4. La taille de l'échantillon sélectionné était de 240 personnes. Cochez la ou les propositions exactes.

- a- l'intervalle de confiance à 95% de la prévalence en 2014 est [24% ; 36%] (arrondi à 1% près)
- b- l'intervalle de confiance à 95% de la prévalence en 2014 est [34% ; 46%] (arrondi à 1% près)
- c- pour estimer la prévalence de la maladie en 2014 à 5% près, et dans le cas où cette prévalence est de l'ordre de 40%, il aurait fallu inclure entre 175 et 200 sujets
- d- pour estimer la prévalence de la maladie en 2014 à 5% près, et dans le cas où cette prévalence est de l'ordre de 40%, il aurait fallu inclure entre 275 et 300 sujets
- e- pour estimer la prévalence de la maladie en 2014 à 5% près, et dans le cas où cette prévalence est de l'ordre de 40%, il aurait fallu inclure entre 350 et 425 sujets

Q5. Pour la même taille d'échantillon $n = 240$ on donne $\sqrt{\frac{0,21}{240}} = 0,03$ Cochez LA proposition exacte.

- a- l'hypothèse alternative testée est $\{\pi = 30\%\}$
- b- si la prévalence de la maladie M dans la population est la même que celle de 2000, l'intervalle de pari à 95% de la prévalence observée est [31% ; 34%] (arrondi à 1% près)
- c- la région de non rejet de l'hypothèse nulle au risque 5% est définie par $\{p \notin [24\% ; 36\%]\}$
- d- 37% appartient à la région de non rejet au risque 5%
- e- 37% n'appartient pas à la région de rejet au risque 1%

Q6. Cochez la ou les propositions exactes.

- a- on conclut au seuil 5% que la prévalence de la maladie a changé
- b- on conclut au seuil 5% que la prévalence de la maladie n'a pas changé
- c- on conclut au seuil 1% que la prévalence de la maladie a changé
- d- on conclut au seuil 1% que la prévalence de la maladie n'a pas changé
- e- aucune de ces réponses

Q7. Un étudiant en médecine décide de refaire cette estimation sur un nouvel échantillon de 70 personnes. Il retrouve 40% de malades sur son échantillon. On donne $\sqrt{0,3} = 0,55$

Cochez la ou les propositions exactes.

- a- on conclut au seuil 5% que la prévalence de la maladie a changé
- b- on conclut au seuil 5% que la prévalence de la maladie n'a pas changé
- c- on conclut au seuil 1% que la prévalence de la maladie a changé
- d- on conclut au seuil 1% que la prévalence de la maladie n'a pas changé
- e- la conclusion du test change avec ce nouvel échantillon

Exercice 3

On suspecte qu'une maladie M touche plus les hommes que les femmes et on souhaite utiliser un test classique permettant de répondre à cette question reposant sur le calcul d'un paramètre de test z . Pour cela, on constitue un échantillon représentatif de 100 malades et on trouve 70 hommes. Soit π , la proportion d'hommes chez les malades.

Q8. Cochez la ou les propositions exactes.

- a- l'estimation ponctuelle de la proportion d'homme malade est 0,7
- b- l'hypothèse nulle qu'il faut tester est $\{p = 0,50\}$ où p est la proportion d'hommes observée
- c- l'hypothèse nulle qu'il faut tester est $\{\pi \neq 0,50\}$
- d- sous H_0 , la statistique du test suit une loi normale d'espérance 0 et de variance 1
- e- $\{z : |z| > 1,96\}$ est la zone de rejet de H_0 pour un test au seuil 5%

Q9. Cochez la ou les propositions exactes.

- a- en faisant le test, on trouve $|z| = 4$
- b- il n'est pas possible de rejeter l'hypothèse selon laquelle cette maladie touche autant les hommes que les femmes (au risque 5%)
- c- en appliquant la théorie des tests, on conclura que la maladie touche plus les hommes que les femmes (au risque 5%)
- d- si on fixe à zéro le risque de rejeter à tort H_0 , il n'est alors jamais possible de rejeter H_0
- e- en appliquant la théorie des tests, on conclura que la maladie touche plus les hommes que les femmes (au risque 1%)

Q10. Cochez la ou les propositions exactes.

- a- Le degré de signification est la plus petite valeur du risque de première espèce
- b- Le degré de signification quantifie la force de conviction à ne pas rejeter H_0
- c- Comme le risque de première espèce, le degré de signification est défini a posteriori
- d- Un degré de signification peut être supérieur à 5%
- e- le degré de signification est inférieur à 10^{-4}

Exercice 4

On veut comparer l'efficacité de deux traitements antihypertenseur A et B. Chez 40 sujets atteints d'hypertension traités par A, on observe une diminution moyenne de la pression artérielle systolique de 26 mmHg avec une variance de 130 mmHg². Chez 40 autres sujets atteints d'hypertension traités par B, on observe une diminution moyenne de la pression artérielle systolique de 30 mmHg avec une variance de 230 mmHg².

Q11. Cochez LA proposition exacte.

- a- au risque 5%, on rejette l'hypothèse nulle d'égalité des diminutions moyennes de la pression artérielle systolique dans les deux populations
- b- au risque 1%, on ne rejette pas l'hypothèse nulle d'égalité des diminutions moyennes de la pression artérielle systolique dans les deux populations
- c- la puissance du test est comprise entre 40% et 50 %
- d- la puissance du test est comprise entre 70% et 80 %
- e- le risque de deuxième espèce est compris entre 20% et 30%

Q12. On pense en fait qu'il existe une différence de 6 mmHg entre les diminutions moyennes de la pression artérielle systolique de deux traitements. Cochez la ou les propositions exactes.

- a- la puissance est la probabilité de ne pas rejeter H_0 alors que H_0 est vraie
- b- sous H_1 , la statistique du test suit approximativement une loi normale de moyenne 2 et de variance 1
- c- la puissance du test vaut 69% (à 2% près)
- d- s'il existe vraiment une différence de 6 mmHg, il y a 50% de risque de ne pas la mettre en évidence (à 2% près)
- e- si on avait voulu montrer une différence de seulement 4 mmHg entre les deux traitements, avec le même nombre de sujets, la puissance aurait été plus petite

En fait on apprend qu'il s'agit des 40 mêmes sujets qui ont été traités par A et par B dans un ordre tiré au sort. On note \bar{D} et s^2 la moyenne observée et la variance observée des différences de pression artérielle systolique (traitement B – traitement A). On note $\mu = (\mu_B - \mu_A)$ la moyenne théorique des différences de pression artérielle systolique. On trouve sur ces données que s^2 est de 160 mmHg².

Q13. Cochez LA proposition exacte.

- a- la moyenne observée des différences de pression artérielle systolique m_d est égale à 4
- b- la moyenne observée des différences de pression artérielle systolique m_d est égale à -4
- c- l'hypothèse H_0 testée est $D = 0$
- d- l'hypothèse H_0 testée est $\mu = 4$
- e- on va effectuer un test de comparaison de moyenne sur échantillons indépendants

Q14. Cochez la ou les propositions exactes.

- a- les moyennes de pression artérielle systolique entre les deux traitements ne diffèrent pas au seuil 5%
- b- au seuil 5%, on rejette l'hypothèse nulle d'égalité des moyennes de pression artérielle systolique entre les deux traitements
- c- au seuil 5%, la diminution de la pression artérielle avec le traitement B est supérieur à celle avec le traitement A.
- d- le test revient à la comparaison d'une moyenne observée à une moyenne théorique égale à zéro
- e- en faisant le test on trouve $|z| = 2$

Q15. Cochez la ou les propositions exactes.

- a- Les conclusions des tests sont identiques pour les 2 échantillons indépendants et pour l'échantillon apparié
- b- Pour l'échantillon apparié, les moyennes de la diminution de pression artérielle des traitements A et B sont indépendantes
- c- On a plus de puissance sans prendre en compte l'appariement
- d- On a moins de puissance sans prendre en compte l'appariement
- e- Le degré de signification pour le test de comparaison de moyenne sur échantillons appariés est inférieur à 4%

Comparaison de deux proportions avec échantillons indépendants – Tests du Chi2 – Tests non paramétriques

Exercice 1

Dans une enquête étiologique sur le cancer broncho-pulmonaire, on a interrogé des sujets atteints de ce cancer et des témoins non cancéreux sur leur consommation de tabac.

On trouve :

- sur 300 sujets atteints de cancer (groupe C) : 3 non fumeurs
- sur 300 témoins (groupe T) : 30 non fumeurs

Q1. Cochez la ou les propositions exactes.

- a- l'estimation du pourcentage de non fumeurs chez les témoins est 10%
- b- l'intervalle de confiance à 95% du pourcentage de non fumeurs chez les témoins est [15% ; 20%] (arrondi à 1% près)
- c- l'estimation du pourcentage de non fumeurs chez les sujets atteints de cancer est 1%
- d- l'intervalle de confiance à 95% du pourcentage de non fumeurs chez les sujets atteints de cancer est [0% ; 2%] (arrondi à 1% près)
- e- l'estimation de la proportion de non fumeurs commune aux deux groupes est 5,5%

Q2. On souhaite comparer la proportion de non fumeurs entre les deux populations. On donne

$$\sqrt{0,055 \times 0,945 \times \left(\frac{1}{300} + \frac{1}{300}\right)} \approx 0,018$$

Cochez LA proposition exacte.

- a- pour tester si la différence entre les deux proportions est significative, on testera l'hypothèse nulle $\pi_C \neq \pi_T$
- b- pour tester si la différence entre les deux proportions est significative, on testera l'hypothèse nulle $p_C \neq p_T$
- c- $|z| > 1,96$
- d- on ne peut pas conclure car $n_{p_C} < 5$ où p_C est la proportion observée de non fumeurs chez les sujets atteints de cancer et n_C le nombre de sujets atteints de cancer dans l'échantillon
- e- on conclut au seuil 5% qu'il y a une proportion identique de non fumeurs chez les sujets atteints de cancer broncho-pulmonaire et les témoins

Exercice 2

Pour les patients épileptiques d'un hôpital, le nombre de crises d'épilepsie par an et les proportions connues sont donnés dans le tableau ci-dessous :

Nombres de crises d'épilepsie	0	1	2	3
Proportions	0,20	0,35	0,25	0,20

On évalue un nouveau médicament pendant un an sur un échantillon représentatif de 100 patients épileptiques de cet hôpital. On observe 0 crise d'épilepsie chez 30 patients, 1 crise chez 35 patients, 2 crises chez 25 patients et 3 crises chez 10 patients. On se demande si la répartition des crises d'épilepsie dans l'échantillon est comparable à celle connue pour les patients de cet hôpital.

Q3. Cochez la ou les propositions exactes.

- a- les conditions de validité du test sont vérifiées car $n \geq 30$
- b- on effectue un test du χ^2 d'indépendance
- c- on rejette H_0 au risque 5%
- d- on ne rejette pas H_0 au risque 5%
- e- le degré de signification du test est inférieur à 2%

Exercice 3

Une étude a été menée pour étudier une éventuelle relation entre l'obésité et le cancer du sein chez la femme jeune. 1805 sujets ont été inclus dont 845 patientes atteintes de cancer du sein (Cas) et 960 non atteintes (Témoins). Les résultats suivants ont été observés :

IMC*	Cas	Témoins	Total
<23,3	550	576	1126
23,3 - 27,0	141	190	331
>27,0	154	194	348
Total	845	960	1805

*IMC : Indice de Masse Corporelle (poids en kg / taille en m^2)

On trouve $K = 5,16$

Q4. Cochez la ou les propositions exactes.

- a- au risque 5%, on conclut qu'il existe une liaison entre l'IMC et la survenue d'un cancer du sein
- b- l'indépendance entre les deux variables peut être testée à partir d'un test du χ^2 d'indépendance à deux degrés de liberté (ddl)
- c- si l'on utilise un test du χ^2 d'indépendance au risque 5% pour tester l'indépendance entre les deux variables, il faudra comparer la valeur obtenue de la statistique à 3,84
- d- au risque de 5%, les résultats de l'étude ne permettent pas de conclure que l'IMC et la survenue d'un cancer du sein sont liés
- e- au risque 5%, l'IMC et la survenue du cancer du sein sont indépendants

Exercice 4

Une étude rétrospective a pour but d'étudier les facteurs associés à la survenue de complications à la suite d'une opération chirurgicale chez 360 malades. La fréquence des complications selon la durée de l'opération est donnée dans le tableau suivant :

	≤ 60 mn	60-90 mn	90-120 mn	>120 mn
Nombre de complications	7	19	26	8
Nombre de non complications	113	100	66	21

Q5. Cochez LA proposition exacte.

- a- la comparaison du nombre de complications entre les différents groupes (de durée) peut se faire par le test de l'écart réduit z
- b- la comparaison du nombre de complications entre les différents groupes peut se faire par un test du χ^2 d'indépendance à 4 ddl
- c- si la proportion de complications est la même quelle que soit la durée de l'intervention, le nombre attendu de complications dans le groupe de durée ≤ 60 mn (C_{11}) vaut 25
- d- on conclut au seuil 5% que la fréquence des complications diffère selon la durée de l'opération
- e- on ne peut pas conclure car les conditions de validité du test ne sont pas remplies

On décide de réunir les deux dernières classes et de refaire le test. On trouve $K = 21,7$.

Q6. Cochez la ou les propositions exactes.

- a- au seuil 5%, on rejette l'hypothèse nulle
- b- au seuil 1%, on rejette l'hypothèse nulle
- c- le degré de signification du test est inférieur à 1%
- d- on ne rejette pas l'hypothèse nulle au seuil 5%
- e- on peut maintenant conclure que la fréquence des complications diffère selon la durée de l'opération au risque 5%

Exercice 5

Q7. On souhaite comparer l'efficacité de 2 traitements A et B contre le cancer. On a obtenu les résultats suivant dans 2 groupes de 4 malades :

	Traitement A	Traitement B
Efficace	1	3
Non efficace	3	1

Cochez LA proposition exacte.

- a- on doit utiliser un test exact de Fisher car au moins un effectif attendu est < 5
- b- on peut utiliser un test du χ^2 d'indépendance à 1 ddl
- c- on rejette H_0 au risque 5%
- d- il y a deux tableaux 2×2 possibles avec les mêmes marges
- e- le degré de signification du test exact de Fisher est inférieur à 0,05

Exercice 6

On veut mettre en évidence une prise de poids dans un groupe de 10 patients 6 mois après l'arrêt du tabac. Le tableau ci-dessous montre les variations de poids de chaque patient après 6 mois de sevrage tabagique :

Différence de poids en Kg
+ 6.0
+ 1.3
+ 10.2
+ 23.9
+ 3.1
+ 6.8
- 1.5
- 14.7
- 3.3
+ 11.1

Q8. Cochez LA proposition exacte.

- a- on effectue un test z de comparaison de 2 moyennes pour échantillons appariés
- b- on effectue un test de Wilcoxon pour 2 échantillons indépendants
- c- la somme des rangs des différences positives est égale à 40 et la somme de tous les rangs est égale à 50
- d- on ne rejette pas H_0 au risque 5%
- e- au risque 5%, la valeur critique donnée par la table est égale à 53

Exercice 7

On veut comparer une série de 10 patients infectés (atteints de méningite) et une série de 10 patients sains dont on a étudié le nombre de cellules par mm^3 à l'examen du liquide céphalo-rachidien (LCR). Les résultats figurent dans le tableau ci-dessous :

Nombre de cellules par mm^3 de LCR	
Patients Sains	Patients infectés
0	37
14	42
18	12
0	32
31	105
0	84
0	15
0	47
11	51
0	65

Q9. Cochez la ou les propositions exactes.

- a- on effectue un test de Wilcoxon pour 2 échantillons indépendants
- b- la somme des rangs des patients sains est égale à 60
- c- la somme des rangs des patients infectés est égale à 150
- d- sous H_0 , la variable « nombre de cellules par mm^3 de LCR » a la même distribution dans les 2 populations dont sont issus les 2 échantillons
- e- on ne rejette pas H_0 au risque 5%

Exercice 8

On souhaite étudier le profil de résistance à un antibiotique des bactéries *E. coli* responsables de septicémies acquises dans 3 services d'un hôpital (chirurgie, réanimation et long séjour). Pour ce faire, les résultats des antibiogrammes (technique de laboratoire visant à tester la résistance d'une bactérie vis-à-vis d'un antibiotique) de 200 patients de chaque service qui ont eu une septicémie à *E. coli* entre 2009 et 2011 sont extraits de la base de données du service de bactériologie. L'automate qui effectue les antibiogrammes peut rendre les 4 résultats suivants : sensible, intermédiaire, résistant ou indéterminé.

Les résultats sur les 600 patients des 3 services étaient les suivants :

	Sensible	Intermédiaire	Résistant	Indéterminé	Total
Chirurgie	124	52	19	5	200
Réanimation	119	50	28	3	200
Long séjour	121	49	23	7	200
Total	364	151	70	15	600

On donne la valeur du paramètre de test $K = 3,5$

Q10. Cochez la ou les propositions exactes.

- a - sous l'hypothèse nulle, les profils de résistance d'*E. coli* dans les trois services sont homogènes
- b - on ne peut pas faire le test car dans une des cases on observe un effectif strictement inférieur à 5
- c - on effectue un test du χ^2 d'homogénéité à 6 ddl
- d - on rejette l'hypothèse nulle au risque 1%
- e - on rejette l'hypothèse nulle au risque 5%

Exercice 9

Le tableau ci-dessous donne le classement de 200 naissances en fonction de la parité et du poids du nouveau-né. On se demande s'il existe une relation entre ces 2 variables.

Poids	Primipares	Multipares	Total
< 3 kg	21	19	40
Entre 3 et 4 kg	61	65	126
> 4 kg	8	26	34
Total	90	110	200

Q11. On trouve $K = 7,8$.

Cochez LA proposition exacte.

- a- l'indépendance entre les deux variables peut être testée à partir d'un test du χ^2 d'indépendance à 3 ddl
- b- si les deux variables sont liées, le nombre attendu de nouveau-nés de moins de 3kg chez les primipares vaut 18
- c- au seuil 5%, on conclut que les deux variables sont significativement liées
- d- au seuil 5%, on conclut que les deux variables sont significativement indépendantes
- e- au seuil 5%, on conclut qu'on ne peut pas rejeter la liaison entre les deux variable