

TESTS STATISTIQUES

UE6

2022 - 2023

Que pensez-vous de ces affirmations ?

- A. Les tests statistiques mettant en œuvre deux caractères qualitatifs s'effectuent par un test non paramétrique
- B. Le nombre de degrés de liberté correspond au nombre de termes dépendants
- C. Un tableau de contingence peut permettre de présenter les résultats de deux caractères quantitatifs
- D. α est appelé le seuil de sécurité (confiance)
- E. La formule du χ^2 sans correction de Yates est légitime quelle que soit la taille de l'échantillon

On a relevé, pour 160 familles de 4 enfants, le nombre de garçons par famille. On obtient la distribution suivante :

Nombre de garçons	0	1	2	3	4
Nombre de familles	11	37	57	43	12

On fait l'hypothèse que la probabilité d'avoir un garçon est égale à 0,5. L'adéquation à une loi de probabilité est recherchée. Un test est choisi et donne comme valeur 1,1. Tous les effectifs théoriques sont supérieurs à 5. Au seuil de sécurité de 95 %, on souhaite savoir si c'est vrai.

- A. Le test choisi est un test d'homogénéité
- B. La loi de probabilité choisie est la loi de Poisson
- C. Le test choisi est un test de conformité
- D. Le nombre de degrés de liberté est 4
- E. En utilisant la table appropriée, on peut dire que l'hypothèse émise est acceptée

Le taux de glycémie a été mesuré chez 80 enfants prématurés. On considère qu'un enfant est hypoglycémique si sa glycémie est égale ou inférieure à 30 cg/l et qu'il est normo-glycémique si sa glycémie est supérieure à 30 cg/l.

On constate que sur 55 enfants dont la glycémie est normale, 24 sont des garçons, alors que sur 25 enfants dont la glycémie est abaissée, on dénombre 9 filles. On demande de déterminer, à l'aide d'un test approprié, si les filles seraient moins sujettes que les garçons à présenter une hypoglycémie, au coefficient de sécurité de 95%.

Que pensez-vous de ces affirmations ?

- A. Le problème à étudier est la comparaison de deux distributions
- B. Le test choisi sera un χ^2 d'homogénéité
- C. Le nombre de degrés de liberté est 1
- D. On choisira la formule du χ^2 avec correction de Yates
- E. L'effectif théorique du nombre de filles hypoglycémiques est 25/2

Quelles sont les affirmations exactes ?

- A. Il faut réaliser un test de comparaison de moyennes sur échantillons indépendants
- B. Il faut réaliser un test de comparaison de moyennes sur échantillons appariés
- C. Il faut réaliser un test de comparaison de variances sur échantillons indépendants
- D. L'hypothèse nulle du test statistique est que la moyenne de B varie dans la population avant et après traitement
- E. Les conditions d'applications sont remplies pour effectuer le test.

Quelles sont les affirmations exactes ? On peut conclure que la variation de B après traitement est significativement différente de :

- A. 0
- B. +1
- C. +1,5
- D. +0,5
- E. -1

Cochez, par Vrai ou Faux, chacune des propositions suivantes :

- A. Un test d'hypothèse paramétrique suppose que l'on connaît la distribution de la variable étudiée dans la population
- B. Si on cherche à montrer une différence entre deux variances, l'hypothèse alternative H_1 est celle que l'on espère vérifier comme étant vraie
- C. Dans le cas d'un test bilatéral où l'on compare deux moyennes de population μ_1 et μ_2 , on aura :
 - $H_0 : \mu_1 = \mu_2$
 - $H_1 : \mu_1 > \mu_2$ ou $\mu_1 < \mu_2$
- D. Les hypothèses nulles et alternatives doivent être définies avant de commencer le test
- E. Le risque de première espèce correspond au risque de rejeter H_0 alors qu'elle est vraie

Le tableau suivant donne la répartition d'un groupe d'enfants par taille :

Taille (cm)	Effectif
[80 ; 90 [3
[90 ; 95 [15
[95 ; 100 [22
[100 ; 105 [18
[105 ; 110 [12
[110 ; 120 [5

Tous les effectifs théoriques sont supérieurs à 5 sauf pour les 2 intervalles extrêmes. L'adéquation à une loi de probabilité est recherchée.

Que pensez-vous de ces affirmations ?

- A. Un test d'adéquation à une loi Gaussienne paraît cohérent
- B. Le nombre de degrés de liberté est 3
- C. Le test choisi traite d'un problème de conformité
- D. Au coefficient de sécurité de 99%, le seuil de décision lu dans la table est de 6,63
- E. Le test choisi est un Chi2 d'homogénéité

Le tableau suivant donne, pour des enfants, la correspondance entre la taille des amygdales et la présence ou l'absence de streptocoques hémolytiques au prélèvement pharyngien.

		Porteurs de germes		Totaux
		Oui	Non	
Amygdales	non hypertrophiées	19	496	515
	hypertrophiées	30	570	600
	très hypertrophiées	23	262	285
Totaux		72	1328	1400

On veut savoir s'il existe un lien entre la taille des amygdales et la présence de germes (au seuil de risque de 5 %)

Que pensez-vous de ces affirmations ?

- A. Le problème à étudier est la comparaison de deux pourcentages

- B.** Le test choisi sera un Chi2 d'homogénéité
 - C.** On choisira la formule du Chi2 avec correction de Yates
 - D.** Le nombre de degrés de liberté est 1
 - E.** L'effectif théorique du nombre de porteurs de germes à amygdales hypertrophiées est 216/7
-

Sur un échantillon représentatif de 100 personnes atteintes d'une maladie M en Meurthe-et-Moselle, on a observé 38 femmes. On se demande si ce pourcentage de femmes diffère de la valeur 30% observée lors d'une étude sur un échantillon de 900 personnes atteintes de la maladie M dans la région Grand-Est. On choisit de réaliser un test statistique du χ^2 au risque 5%. La valeur prise par la statistique du test est égale à environ 2,70.

Indiquez, pour chaque assertion, si elle est vraie ou fausse.

- A. Il s'agit de comparer un pourcentage observé à une valeur théorique
- B. Le pourcentage de femmes atteintes par la maladie M attendu si l'hypothèse nulle est vraie est 30,8%
- C. Les conditions d'utilisation du test sont satisfaites car le plus petit des effectifs attendus sous l'hypothèse nulle est 30,8
- D. La valeur prise par la statistique est inférieure à la valeur seuil lue dans la table du χ^2 de Pearson
- E. On montre une différence significative entre le pourcentage de femmes atteintes par la maladie M en Meurthe-et-Moselle et dans la région Grand-Est

Que pensez-vous de ces affirmations ?

- A. Un test paramétrique nécessite la connaissance de la loi de distribution de la variable étudiée
- B. Une étude statistique utilise généralement les données de toute la population et permet de formuler des lois générales relatives à la population
- C. Lors d'une étude statistique, le choix du test oriente vers la formulation précise de l'hypothèse nulle
- D. Dans une adéquation à une loi normale, pour calculer les effectifs théoriques en se servant des tables statistiques, la recherche de la valeur de l'écart réduit au centre des classes doit être préalablement effectuée
- E. Une étude portant sur des échantillons appariés ne pourra s'effectuer que par un test non paramétrique

La quantité de poussière dans l'atmosphère peut être estimée à l'aide d'un ultramicroscope. Un très petit volume d'air est éclairé par un rayon lumineux et l'observateur compte le nombre de particules vues. En répétant un grand nombre de fois l'expérience, on détermine la quantité de poussière par centimètre cube. 300 mesures sont effectuées. Les résultats suivants ont été obtenus :

Nombre de particules	0	1	2	3	4	5	6
Fréquences observées	38	76	89	54	20	19	4

Tous les effectifs théoriques sont supérieurs à 5 sauf pour la case correspondant à 6 particules.

Que pensez-vous de ces affirmations ?

- A. L'hypothèse du test choisi est celle d'une adéquation à une loi de Poisson
- B. Le nombre de degrés de liberté est 4
- C. Le test choisi traite d'un problème de conformité
- D. Au seuil de risque de 1%, la valeur lue dans la table est de 7,78
- E. Le test choisi est un χ^2 d'homogénéité

Le taux de glycémie a été mesuré chez 80 enfants prématurés. On considère qu'un enfant est hypoglycémique si sa glycémie est égale ou inférieure à 30 cg/l et qu'il est normo-glycémique si sa glycémie est supérieure à 30 cg/l. On dénombre, sur 55 enfants dont la glycémie est normale, 31 filles et 24 garçons ; alors que, sur 25 enfants dont la glycémie est abaissée, on dénombre 9 filles et 16 garçons. On demande de déterminer, à l'aide d'un test approprié, si les filles seraient moins sujettes que les garçons à présenter une hypoglycémie, au coefficient de sécurité de 95 %.

Que pensez-vous de ces affirmations ?

- A. Le problème à étudier est la comparaison de deux distributions théoriques
- B. On peut réaliser un test du χ^2 d'homogénéité
- C. Le nombre de degrés de liberté est 1
- D. On choisira la formule du χ^2 avec correction de Yates
- E. L'effectif théorique du nombre de filles hypoglycémiques est : 25/2

On réalise une étude pour comparer l'efficacité de deux anti-hypertenseurs A et B. 28 patients reçoivent le traitement A, 32 patients le traitement B. On juge l'efficacité de chaque traitement en comparant, pour chaque patient, la différence de pression artérielle systolique entre le début et la fin de l'étude. On sait que cette variable suit une loi normale.

Les variances des différences sont identiques dans les deux groupes.

Que pensez-vous de ces affirmations ?

- A. Pour savoir si l'efficacité des traitements est différente, on réalise un test de comparaison de moyennes sur échantillons appariés
 - B. Pour savoir si l'efficacité des traitements est différente, on réalise un test de comparaison de moyennes sur échantillons indépendants
 - C. Pour savoir si l'efficacité des traitements est différente, on peut réaliser un test de l'écart réduit.
 - D. Pour savoir si l'efficacité des traitements est différente, on peut réaliser un test de Student.
 - E. Pour savoir si la pression artérielle varie entre le début et la fin de l'étude dans le groupe A, on réalise un test de comparaison de moyennes sur échantillons appariés
-

On réalise une étude pour comparer l'efficacité de deux anti-cancéreux A et B. 50 patients reçoivent le traitement A, 50 patients le traitement B. On juge l'efficacité de chaque traitement en comparant, pour chaque patient, la différence D de volume tumoral entre le début et la fin de l'étude. On sait que cette variable suit une loi normale. Les variances des différences sont identiques dans les deux groupes. On a les résultats suivants :

Groupe A : moyenne observée :

$$\overline{D}_A = -50 \text{ mm}^3$$

Groupe B : moyenne observée :

$$\overline{D}_B = -40 \text{ mm}^3$$

Variance commune estimée : 100 mm^6

On réalise le test statistique le plus adapté pour savoir si l'efficacité des traitements est différente.

La statistique calculée vaut :

$$\text{A- } \frac{|-50 - (-40)|}{\sqrt{\frac{10}{50}}}$$

$$\text{B- } \frac{|-50 - (-40)|}{\sqrt{\frac{100}{50}}}$$

$$\text{C- } \frac{|-50 - (-40)|}{\sqrt{\frac{200}{50}}}$$

$$\text{D- } \frac{|-50 - (-40)|}{\sqrt{\frac{100}{100}}}$$

$$\text{E- } \frac{|-50 - (-40)|}{\sqrt{\frac{20}{50}}}$$

2022 - 2023

A est vrai

- A. Vrai :** Le test du χ^2 est non paramétrique et s'applique sur des comparaisons de distributions de caractères qualitatifs.
- B. Faux :** Cela correspond au nombre de termes INdépendants.
- C. Faux :** Deux caractères qualitatifs.
- D. Faux :** Alpha est le risque de première espèce. Le seuil de sécurité est $1-\alpha$.
- E. Faux :** Elle ne peut pas s'appliquer pour des échantillons dont l'effectif est inférieur à 30.

C et E sont vrais

- A. Faux :** C'est un test de conformité (on cherche une adéquation à une loi théorique).
- B. Faux :** La distribution est symétrique et concerne une variable discrète. La loi de probabilité choisie est la loi binomiale.
- C. Vrai :** Cf. A.
- D. Vrai :** Tous les effectifs théoriques sont supérieurs à 5 donc on ne regroupe les classes, $n = 5$. Ici on a une adéquation à une loi binomiale mais dans l'énoncé, on nous donne la probabilité d'avoir un garçon ce qui correspond au paramètre de la loi binomiale. Puisque ce paramètre nous est donné, il n'y aura pas besoin de l'estimer avec les données de l'échantillon donc $r = 0$. On a donc $n - 1 - r = 5 - 1 - 0 = 4$ ddl.
- E. Vrai :** Pour 4 ddl et un risque α à 0,05, la valeur lue dans la table du χ^2 est 9,49. $1,1 < 9,49$, la statistique du test est inférieure à la valeur seuil donc nous acceptons H_0 , la distribution est conforme.

B, C et E sont vrais

- A. Faux :** Nous cherchons à comparer la distribution d'un caractère dans 2 populations différentes, pas une comparaison de distribution.
- B. Vrai :** Selon l'énoncé : "On demande de déterminer, à l'aide d'un test approprié, si les filles seraient moins sujettes que les garçons à présenter une hypoglycémie" : on cherche donc à comparer la prévalence d'un caractère (hypoglycémie) entre 2 populations (filles et garçons), nous effectuons un χ^2 d'homogénéité.
- C. Vrai :** En effectuant un tableau de contingence, nous aurions deux lignes et deux colonnes.
- $(l-1) \times (c-1) = (2 - 1) \times (2 - 1) = 1 \times 1 = 1$.
- D. Faux :** Pour utiliser la correction de continuité de Yates, il faut réunir ces 3 conditions :

1) Un effectif total entre 30 et 50 (ce n'est pas le cas ici donc on ne peut pas utiliser la correction de continuité de Yates)

2) Un tableau de contingence 2×2 (c'est bien le cas ici)

3) Au moins 1 effectif théorique inférieur à 5 (ce n'est pas le cas ici, l'effectif théorique le plus petit est de $25 \times 40 / 80 = 12,5$. On ne peut pas utiliser la correction de continuité de Yates).

E. Vrai : Voici le tableau des effectifs observés et effectifs théoriques (les effectifs théoriques sont entre parenthèses) :

	Garçon	Fille	Total
Normal	24 ($55 \times 40 / 80 = 55 / 2 = 27,5$)	31 ($55 \times 40 / 80 = 55 / 2 = 27,5$)	55
Hypo	16 ($25 \times 40 / 80 = 25 / 2 = 12,5$)	9 ($25 \times 40 / 80 = 25 / 2 = 12,5$)	25
Total	40	40	80

2021 - 2022

B et E justes

A. Faux : On a 2 moyennes d'un même échantillon mais obtenu à des temps différents. On compare alors des moyennes d'un même échantillon, c'est-à-dire qu'elles sont appariées.

B. Vrai : Voir A

C. Faux : Voir A

D. Faux : L'hypothèse nulle du test statistique est que la moyenne de B sont les mêmes dans la population avant et après traitement.

E. Vrai : Oui les conditions sont respectées.

D vraie

A. Faux : On a une différence de 2 ui entre les deux moyenne. Pour trouver la variation significative, on doit faire Différence / racine(Variance) = 2 / racine(25) = 2/5 = 0,5

B. Faux : Voir A

C. Faux : Voir A

D. Vrai : Voir A

E. Faux : Voir A

A, B, et E justes

A. Vrai : Pour les tests paramétriques, il faut connaître la distribution de la variable étudiée.

B. Vrai : En effet l'hypothèse H1 cherche à prouver la différence.

C. Faux : Dans les tests bilatéraux, on compare la différence et non pas la grandeur pour l'hypothèse H1.

D. Faux : Les hypothèses nulles et alternatives sont définies après avoir choisi le test mais avant d'avoir commencé les calculs.

E. Vrai : Il correspond à α .

A, C et D sont vraies

A. Vrai : on voit qu'on a une distribution continue et symétrique, donc on fait bien un test d'adéquation à la loi gaussienne (loi normale).

B. Faux : On nous dit dans l'énoncé que « Tous les effectifs théoriques sont supérieurs à 5 sauf pour les 2 intervalles extrêmes » donc il faut regrouper ces 2 intervalles avec les 2 intervalles adjacents. On a donc 4 classes ($n = 4$). On est ici dans le cas d'une adéquation à une loi normale donc $r = 2$. On a donc $n - 1 - r = 4 - 1 - 2 = 1$ ddl.

C. Vrai : On cherche à savoir si les valeurs présentées se distribuent comme une loi normale, c'est donc de la conformité. Dans l'homogénéité, il y a plusieurs échantillons (pas le cas ici)

D. Vrai : On doit donc lire la table du Chi 2. Comment savoir si on doit faire un test du Chi 2 ? Si dans un énoncé il y a un/des tableaux et des effectifs théoriques, c'est probablement un chi2.

On parle d'un coefficient de 99% de sécurité, soit un risque de 1% = 0,01 \Rightarrow On lit dans la colonne 0,01.

On a vu précédemment qu'il y avait 1 degrés de liberté, on lit dans la ligne 1 ddl de la table du Chi 2. Le croisement entre ligne et colonne donne bien 6,63

E. Faux : cf. C, on fait un chi2 de conformité

E est vrai

A. Faux : Le problème à étudier est l'indépendance de deux critères.

B. Faux : On cherche ici une association entre 2 caractères qualitatifs. On va alors faire un test de Chi2 d'indépendance.

C. Faux : La correction de Yates n'est valable que pour les tableaux en 2X2.

D. Faux : Le degré de liberté est trouvé par la formule : (nombre de colonne - 1) x (nombre de ligne - 1). On obtient dans ce cas ddl = 1 x 2 = 2.

E. Vrai : Pour trouver l'effectif théorique il faut multiplier les valeurs totales de la ligne et de la colonne concerné puis diviser le résultat obtenu par l'effectif total. Dans le cas présent on obtient le calcul : $(72 \times 600) / 1400 = 216 / 7$.

2020 - 2021

B, C et D sont justes.

A. Faux : on cherche à comparer deux valeurs expérimentales entre elles. On a 2 échantillons : un de 100 personnes et un de 900 personnes.

B. Vrai : si l'hypothèse nulle est vraie, on s'attend à ce que le pourcentage observé ne soit pas significativement différent du pourcentage commun calculé dans l'item C.

C. Vrai : les conditions d'utilisation du test sont satisfaites : le plus petit effectif attendu calculé est de $((n_1 * p_1) + (n_2 * p_2)) / (n_1 + n_2) = ((100 * 0,38) + (900 * 0,30)) / (100 + 900) = (38 + 270) / 1000 = 308 / 1000 = 0,308 = 30,8\%$.

D. Vrai : au risque de 5%, on obtient une valeur seuil de 3,84. La valeur du test étant égale à 2,70, elle est inférieure au seuil.

E. Faux : la valeur prise par la statistique étant inférieure à la valeur seuil, on ne rejette pas l'hypothèse nulle : les pourcentages ne sont pas significativement différents.

A et C sont justes.

A. Faux : Les tests paramétriques n'exigent aucune connaissance sur la loi de probabilités. Vrai Un test paramétrique, à l'inverse d'un test non paramétrique exige d'avoir des connaissances sur la distribution de la variable étudiée.

B. Faux : Une étude statistique utilise les données d'un échantillon plus ou moins grand afin de formuler des lois générales relatives à la population.

C. Vrai : L'hypothèse est imposée par construction du test.

D. Faux : C'est aux limites des classes

E. Faux : Elle peut aussi s'effectuer par le test de student qui est un test paramétrique.

A, B et C sont justes.

A. Vrai : on répète un grand nombre de fois (300) une mesure, avec une distribution discrète et asymétrique. On va ici chercher à savoir si la répartition obtenue suit une loi de Poisson.

B. Vrai : L'effectif de la case correspondant à 6 particules étant inférieur à 5, on va la regrouper avec la case correspondant à 5 particules. On a donc 6 classes ($n = 6$). De plus, comme on doit estimer le paramètre de la loi de Poisson par calcul, $r = 1$. On obtient donc $n - 1 - r = 6 - 1 - 1 = 4$ degrés de liberté.

C. Vrai : On cherche ici à comparer une distribution expérimentale à une distribution théorique : c'est bien un problème de conformité.

D. Faux : attention au piège, la valeur donnée correspond à un risque de 10% ! À 1% et un nombre de degrés de liberté de 4, on obtient une valeur de 13,28.

E. Faux : Dans un test du χ^2 d'homogénéité, on compare deux distributions expérimentales pour savoir si elles proviennent de la même population. Ici on ne dispose que d'une seule distribution donc on utilise un test d'adéquation / de conformité.

B, C et E sont vrais

A. Faux : Nous cherchons à comparer la distribution d'un caractère dans 2 populations différentes, pas une comparaison de distribution.

B. Vrai : Selon l'énoncé : "On demande de déterminer, à l'aide d'un test approprié, si les filles seraient moins sujettes que les garçons à présenter une hypoglycémie" : on cherche donc à comparer la prévalence d'un caractère (hypoglycémie) entre 2 populations (filles et garçons), nous effectuons un χ^2 d'homogénéité.

C. Vrai : En effectuant un tableau de contingence, nous aurions deux lignes et deux colonnes. $(l-1) \times (c-1) = (2-1) \times (2-1) = 1 \times 1 = 1$.

D. Faux : Pour utiliser la correction de continuité de Yates, il faut réunir ces 3 conditions :

1) Un effectif total entre 30 et 50 (ce n'est pas le cas ici donc on ne peut pas utiliser la correction de continuité de Yates)

2) Un tableau de contingence 2×2 (c'est bien le cas ici)

3) Au moins 1 effectif théorique inférieur à 5 (ce n'est pas le cas ici, l'effectif théorique le plus petit est de $25 \times 40 / 80 = 12,5$. On ne peut pas utiliser la correction de continuité de Yates).

E. Vrai : Voici le tableau des effectifs observés et effectifs théoriques (les effectifs théoriques sont entre parenthèses) :

	Garçon	Fille	Total
Normal	24 ($55 \times 40 / 80 = 55 / 2 = 27,5$)	31 ($55 \times 40 / 80 = 55 / 2 = 27,5$)	55
Hypo	16 ($25 \times 40 / 80 = 25 / 2 = 12,5$)	9 ($25 \times 40 / 80 = 25 / 2 = 12,5$)	25
Total	40	40	80

B, D et E sont justes.

A. Faux : Il s'agit de deux échantillons indépendants car on compare l'évolution de 2 valeurs prises à des moments différents dans 2 échantillons distincts.

B. Vrai

C. Faux : Le test de l'écart réduit est utilisé pour les comparaisons de moyennes sur échantillons indépendants avec n supérieur ou égal à 30 pour les deux échantillons.

D. Vrai : les conditions d'applications sont bien remplies car selon l'énoncé « On sait que cette variable suit une loi normale. »

E. Vrai : Ici le problème n'est pas le même que dans l'item A et B, on cherche à comparer deux échantillons appariés : le groupe A avant le traitement et le groupe A après le traitement.

C est juste.

A. Faux : cf. réponse C.

B. Faux : cf. réponse C.

C. Vrai : on compare deux moyennes observées sur des échantillons indépendants. On utilise le test de Student :

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_c^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Ainsi, on obtient T = :

$$\frac{-50 - (-40)}{\sqrt{100 \left(\frac{1}{50} + \frac{1}{50} \right)}} = \frac{-50 - (-40)}{\sqrt{100 \times \frac{2}{50}}} = \frac{-50 - (-40)}{\sqrt{\frac{200}{50}}}$$

D. Faux : cf. réponse C.

E. Faux : cf. réponse C.