

**Corrigés des exercices n°8, n°9 et n°10**

Exercice n° 8 :

a) Le khi-deux des effectifs de ce tableau (n = 200) est égal à 5,314. Que peut-il en conclure?

On peut répondre à la question sans faire aucun calcul. Le tableau proposé comporte 3 lignes et 2 colonnes, le nombre de degrés de liberté (d.d.l.) est donc de :

$$\text{d.d.l.} = (\text{nombre de lignes} - 1) \times (\text{nombre de colonnes} - 1) = (3 - 1) \times (2 - 1) = 2 \times 1 = 2$$

On lit sur la table du  $X^2$  la valeur attendue pour d.d.l. = 2 au seuil de 0.05 (cad avec un risque de 5 chances sur 100 de se tromper. La valeur lue est de 5,991.

Dans l'énoncé de cet exercice, on nous dit que le khi-deux est égal à 5,314.

Le  $X^2$  étant < à 5,991 (5,314 < 5,991), nous ne pouvons donc rien conclure (et surtout pas qu'il n'existe pas de liaison statistiquement significative entre ces deux variables<sup>1</sup>).

b) S'il comparait les enfants uniques aux aînés, que pourrait-il dire ?

Le tableau étant présenté sous forme de pourcentages, nous devons reconstituer les effectifs<sup>2</sup>.

	Note < à la moyenne	Note > à la moyenne
<b>Enfant unique</b>	15 %	10 %
<b>Aîné</b>	10 %	15 %
<b>Autre cas</b>	21 %	29 %

Si n = 200, alors 15 % =  $\frac{15 \times 200}{100} = 30$  etc.

On obtient le tableau suivant :

	Note < à la moyenne	Note > à la moyenne	Total :
<b>Enfant unique</b>	30	20	<b>50</b>
<b>Aîné</b>	20	30	<b>50</b>
<b>Autre cas</b>	42	58	<b>100</b>
<b>Total :</b>	<b>92</b>	<b>108</b>	<b>200</b>

Pour comparer les enfants uniques aux aînés, on prend en compte le tableau d'effectifs observés suivant :

	Note < à la moyenne	Note > à la moyenne	Total :
<b>Enfant unique</b>	30	20	<b>50</b>
<b>Aîné</b>	20	30	<b>50</b>
<b>Total :</b>	<b>50</b>	<b>50</b>	<b>100</b>

Calcul des valeurs théoriques de chaque case (a', b', c' et d') :

$$a' = \frac{C1 \times L1}{N} \quad b' = \frac{C2 \times L1}{N} \quad c' = \frac{C1 \times L2}{N} \quad d' = \frac{C2 \times L2}{N}$$

Ici C1 = 50 ; C2 = 50 ; L1 = 50 ; L2 = 50

On calcule donc les valeurs des 4 fractions : exemple :  $a' = \frac{50 \times 50}{100} = \frac{2500}{100} = 25,0$  etc.

Nb : le hasard fait ici que tous les effectifs théoriques sont égaux.

<sup>1</sup> Cf le cours ou le polycopé « Le test du khi-deux », dernier paragraphe en bas de la page 2.

<sup>2</sup> On ne calcule jamais un  $X^2$  avec des pourcentages. Idem que ci-dessus, cf. Remarques importantes n°1 du polycopé p2.

On obtient ainsi le tableau d'effectifs théoriques suivant :

	Note < à la moyenne	Note > à la moyenne	Total :
Enfant unique	25,0	25,0	<b>50</b>
Aîné	25,0	25,0	<b>50</b>
Total :	<b>50</b>	<b>50</b>	<b>100</b>

Formule du  $X^2$  : Somme des [ effectifs (Observés - Théoriques)<sup>2</sup> ]  
effectifs théoriques

$$\begin{aligned} \text{cad : } X^2 &= \frac{(a - a')^2}{a'} + \frac{(b - b')^2}{b'} + \frac{(c - c')^2}{c'} + \frac{(d - d')^2}{d'} \\ &= \frac{(30 - 25,0)^2}{25,0} + \frac{(20 - 25,0)^2}{25,0} + \frac{(20 - 25,0)^2}{25,0} + \frac{(30 - 25,0)^2}{25,0} \\ &= \frac{5,0^2}{25,5} + \frac{5,0^2}{25,5} + \frac{5,0^2}{25,5} + \frac{5,0^2}{25,5} = \frac{25,0}{25,0} + \frac{25,0}{25,0} + \frac{25,0}{25,0} + \frac{25,0}{25,0} \\ &= 1,0 + 1,0 + 1,0 + 1,0 = 4,0 \end{aligned}$$

Le khi deux de ce tableau est de 4,0.

On détermine le nombre de degrés de liberté :

$$\text{d.d.l.} = (\text{nombre de lignes} - 1) \times (\text{nombre de colonnes} - 1) = (2 - 1) \times (2 - 1) = 1 \times 1 = 1$$

On lit sur la table du  $X^2$  la valeur attendue pour d.d.l. = 1 au seuil de 0.05 (cad avec un risque de 5 chances sur 100 de se tromper. La valeur lue est de 3,841.

Le  $X^2$  étant  $\geq$  à 3,841 ( $4,0 \geq 3,841$ ), nous pouvons conclure qu'il existe une liaison statistiquement significative entre ces deux variables.

Pour déterminer le sens de cette liaison, nous effectuons le tableau des signes :

	Note < à la moyenne	Note > à la moyenne
Enfant unique	30 - 25,0 => +	20 - 25,0 => -
Aîné	20 - 25,0 => -	30 - 25,0 => +

Conclusion : Nous pouvons conclure qu'il y a bien un lien statistiquement significatif entre la position dans la famille et les performances scolaires : les enfants uniques obtiennent plus souvent des notes inférieures à la moyenne par rapport aux aînés.

Ou : (ce qui revient au même) les enfants en position d'aîné obtiennent plus souvent des notes supérieures à la moyenne comparativement aux enfants uniques.

c) Comment peut-il interpréter son résultat en terme de recherche [cad quelle(s) hypothèse(s) pourrait-il donc avancer] ?

(On attend ici des hypothèses, des pistes de recherche et non des certitudes). Quelques exemples :

- Les enfants uniques ont de moins bonnes conditions éducatives que les aînés (pas de stimulation ou d'émulation vis à vis de la fratrie)
- Les aînés font l'objet de davantage de vigilance que les enfants uniques
- L'aîné a davantage de pression par rapport à l'école (il doit être un « exemple »)
- (complémentairement) Les enfants uniques sont plus facilement « négligés » que les aînés
- etc.

Exercice n° 9 :

a) Le khi-deux obtenu en opposant les pratiquants d'un sport (individuel ou collectif) aux non pratiquants en fonction du milieu de résidence (rural ou urbain) est de 3,07. Que pouvez-vous en conclure (et pourquoi) ?

Il s'agissait en fait ici d'un khi deux calculé sur un tableau à 4 cases. Il fallait regrouper les pratiquants aux non pratiquants. On obtenait ainsi le tableau suivant :

	<b>Ruraux</b>	<b>Urbains</b>	<b>Total :</b>
<b>Pratiquants d'un sport (individuel ou collectif)</b>	46	75	<b>121</b>
<b>Aucun sport</b>	8	28	<b>36</b>
<b>Total :</b>	<b>54</b>	<b>103</b>	<b>157</b>

On détermine quel était le nombre de degrés de liberté :

$$d.d.l. = (\text{nombre de lignes} - 1) \times (\text{nombre de colonnes} - 1) = (2 - 1) \times (2 - 1) = 1 \times 1 = 1$$

Sur la table du  $X^2$  la valeur attendue pour  $d.d.l. = 1$  au seuil de 0.05 (cad avec un risque de 5 chances sur 100 de se tromper) était de 3,841.

Le  $X^2$  étant  $\leq$  à 3,841 ( $3,07 \leq 3,841$ ), nous ne pouvons rien conclure (et surtout pas qu'il n'existe pas de liaison statistiquement significative entre ces deux variables<sup>1</sup>).

b) Pour les 121 personnes qui pratiquent, y-a-t-il un lien entre le type de sport pratiqué et le milieu de résidence ?

Pour répondre à la question, il fallait calculer un khi deux. Le tableau correspondant à l'énoncé de la question était le suivant (tableau d'effectifs observés) :

	<b>Ruraux</b>	<b>Urbains</b>	<b>Total :</b>
<b>Sport individuel</b>	37	46	<b>83</b>
<b>Sport collectif</b>	9	29	<b>38</b>
<b>Total :</b>	<b>46</b>	<b>75</b>	<b>121</b>

On calcule ensuite le tableau d'effectifs théoriques.

Calcul des valeurs théoriques de chaque case (a', b', c' et d') :

$$a' = \frac{C1 \times L1}{N} \quad b' = \frac{C2 \times L1}{N} \quad c' = \frac{C1 \times L2}{N} \quad d' = \frac{C2 \times L2}{N}$$

Ici  $C1 = 46$  ;  $C2 = 75$  ;  $L1 = 83$  ;  $L2 = 38$

On calcule donc les valeurs des 4 fractions : exemple :  $a' = \frac{46 \times 83}{121} = \frac{3818}{121} = 31,55$  etc.

On obtient ainsi le tableau d'effectifs théoriques suivant :

	<b>Ruraux</b>	<b>Urbains</b>	<b>Total :</b>
<b>Sport individuel</b>	31,55	51,45	<b>83</b>
<b>Sport collectif</b>	14,45	23,55	<b>38</b>
<b>Total :</b>	<b>46</b>	<b>75</b>	<b>121</b>

<sup>1</sup> Cf le cours ou le polycopé « Le test du khi-deux », dernier paragraphe en bas de la page 2.

Formule du  $X^2$  : 
$$\text{Somme des } \left[ \frac{\text{effectifs (Observés - Théoriques)}^2}{\text{effectifs théoriques}} \right]$$

$$\begin{aligned} \text{cad : } X^2 &= \frac{(a - a')^2}{a'} + \frac{(b - b')^2}{b'} + \frac{(c - c')^2}{c'} + \frac{(d - d')^2}{d'} \\ &= \frac{(37,0 - 31,55)^2}{31,55} + \frac{(46,0 - 51,45)^2}{51,45} + \frac{(9,0 - 14,45)^2}{14,45} + \frac{(29,0 - 23,55)^2}{23,55} \\ &= \frac{(5,45)^2}{31,55} + \frac{(-5,45)^2}{51,45} + \frac{(-5,45)^2}{14,45} + \frac{(5,55)^2}{23,55} = \frac{29,70}{31,55} + \frac{29,70}{51,45} + \frac{29,70}{14,45} + \frac{29,70}{23,55} \\ &= 0,94 + 0,58 + 2,06 + 1,26 = 4,84 \end{aligned}$$

Le khi deux de ce tableau est de 4,84.

On détermine le nombre de degrés de liberté :

$$\text{d.d.l.} = (\text{nombre de lignes} - 1) \times (\text{nombre de colonnes} - 1) = (2 - 1) \times (2 - 1) = 1 \times 1 = 1$$

On lit sur la table du  $X^2$  la valeur attendue pour d.d.l. = 1 au seuil de 0.05 (cad avec un risque de 5 chances sur 100 de se tromper. La valeur lue est de 3,841.

Le  $X^2$  étant  $\geq$  à 3,841 ( $4,84 \geq 3,841$ ), nous pouvons conclure qu'il existe une liaison statistiquement significative entre ces deux variables.

Pour déterminer le sens de cette liaison, nous effectuons le tableau des signes :

	<b>Ruraux</b>	<b>Urbains</b>
<b>Sport individuel</b>	37 – 31,55 => +	46 – 51,45 => -
<b>Sport collectif</b>	9 – 14,45 => -	29 – 23,55 => +

Conclusion : Nous pouvons conclure qu'il y a bien un lien statistiquement significatif entre le lieu de résidence et le type de sport pratiqué : les personnes résidant en milieu rural pratiquent davantage un sport individuel par rapport à celles qui résident en milieu urbain qui pratiquent davantage un sport collectif<sup>1</sup>.

c) Que pouvez-vous conclure en clair des résultats à ces deux questions ?

S'il n'est pas possible d'affirmer que le lieu de résidence a une influence sur le fait de pratiquer ou non un sport (cf. conclusion de la question a), nous pouvons néanmoins dire parmi les sportifs, les urbains pratiquent davantage des sports collectifs que les ruraux.

\* \* \*

<sup>1</sup> Rappel : le cas est totalement fictif.

Exercice n° 10 :

a) Peut-il dire que le Groupe 1 soit statistiquement meilleur que le Groupe 2 ?

Il s'agit ici de comparer deux moyennes observées sur des petits effectifs (puisque nous avons les deux écarts-types et que les deux effectifs  $n_A$  ou  $n_B$  sont  $\leq 30$ ).

Formule : on calcule une note t telle que  $t = \frac{\bar{m}_A - \bar{m}_B}{\sqrt{\frac{\sigma^2}{n_A} + \frac{\sigma^2}{n_B}}}$

Il faut donc calculer la variance commune  $\sigma^2$

Formule :  $\sigma^2 = \frac{[(n_A - 1) \cdot \sigma_A^2] + [(n_B - 1) \cdot \sigma_B^2]}{n_A + n_B - 2} = \frac{[(4 - 1) \times 1,29^2] + [(4 - 1) \times 0,82^2]}{4 + 4 - 2}$

$$\sigma^2 = \frac{[3 \times 1,66] + [3 \times 0,67]}{6} = \frac{4,99 + 2,02}{6} = \frac{7,01}{6} = 1,17$$

Calcul de la note t =  $\frac{\bar{m}_A - \bar{m}_B}{\sqrt{\frac{\sigma^2}{n_A} + \frac{\sigma^2}{n_B}}} = \frac{9,5 - 7,6}{\sqrt{\frac{1,17}{4} + \frac{1,17}{4}}} = \frac{1,90}{\sqrt{0,29 + 0,29}} = \frac{1,90}{\sqrt{0,58}} = \frac{1,90}{0,76} = \mathbf{2,50}$

On lire |t| sur la table de t : la valeur trouvée doit être  $\geq$  à la valeur lue pour d.d.l. =  $n_A + n_B - 2$

Ici d.d.l. =  $4 + 4 - 2 = 6$

La valeur lue pour d.d.l. = 6 au risque de 5% (cad colonne 0.05) est de 2,447.

$|t| = |2,50| = 2,50 \geq 2,447$

Conclusion : nous pouvons dire que la différence entre les deux moyennes est statistiquement significative : les notes du groupe 1 sont significativement supérieures à celles du groupe 2.

b) La moyenne nationale au BEPC est de 7,40. Peut-il dire que la moyenne du Groupe 1 soit meilleure que la moyenne nationale ?

Il s'agit ici de comparer une moyenne observée (celle du groupe 1) et une moyenne théorique (la moyenne nationale au BEPC) sur des petits effectifs (puisque  $n_A$  est  $\leq 30$ ).

Formule : on calcule une note t telle que  $t = \frac{(\bar{m}_A - \bar{m}_T) \times \sqrt{n_A}}{\sigma_A}$

Application :  $t = \frac{(\bar{m}_A - \bar{m}_T) \times \sqrt{n_A}}{\sigma_A} = \frac{(9,50 - 7,40) \times \sqrt{4}}{1,29} = \frac{2,10 \times 2,0}{1,29} = \frac{4,20}{1,29} = \mathbf{3,26}$

On lire |t| sur la table de t : la valeur trouvée doit être  $\geq$  à la valeur lue pour d.d.l. =  $n_A - 1$

Ici d.d.l. =  $4 - 1 = 3$

La valeur lue pour d.d.l. = 3 au risque de 5% (cad colonne 0.05) est de 3,182.

$|t| = |3,26| = 3,26 \geq 3,182$

Conclusion : nous pouvons dire que la différence entre les deux moyennes est statistiquement significative : la moyenne du groupe 1 est significativement supérieure à la moyenne nationale.

c) Si le Groupe 2 avait comporté 64 personnes (au lieu de 4, avec la même moyenne et le même écart-type), aurait-il pu affirmer que la moyenne du Groupe 2 était meilleure que la moyenne nationale (7,40) ?

L'exercice est le même que le b) ci-dessus avec des valeurs différentes mais sur un grand échantillon  $n_A \geq 30$ . Il s'agissait de comparer une moyenne observée (celle du groupe 2) et une moyenne théorique (la moyenne nationale au BEPC) sur des grands effectifs

Formule : on calcule une note  $\mathcal{E}$  telle que  $\mathcal{E} = \frac{(\bar{m}_A - \bar{m}_T) \times \sqrt{n_A}}{\sigma_A}$

$$\text{Application : } \mathcal{E} = \frac{(7,60 - 7,40) \times \sqrt{64}}{0,82} = \frac{0,20 \times 8,0}{0,82} = \frac{1,60}{0,82} = \mathbf{1,95}$$

Par définition, la différence est significative si  $|\mathcal{E}| \geq 1,96$

Ici,  $1,95 \leq 1,96$  (je sais, c'est un peu juste mais si vous ne faites pas d'erreur de calcul, vous devez trouver cette valeur).

Conclusion : nous pouvons dire que la différence entre les deux moyennes n'est pas statistiquement significative : la moyenne du groupe 2 n'est pas significativement supérieure à la moyenne nationale.

\* \* \*