Correction - Exercices de révision

EXERCICE 1

Dans une enquête effectuée en 2003, des étudiants de diverses filières des universités alsaciennes ont été interrogés.
Nous avons croisé la question « Avez-vous voté aux dernières élections étudiantes avec leur intérêt politique, ce qui nous donne le tableau 1 Effectuer le test du Chi-deux sur ce tableau. Qu'en déduisez-vous ?
Sur les deuxième tableau, dites ce que vous testez et interprétez les résultats trouvés à l'aide des % lignes ou colonnes quand le test est significatif.

Tableau 1 : Avez-vous voté aux dernières élections étudiantes ? / Etes-vous intéressé par la politique

% ligne % colonne	Pas du tout	Un peu	moyen	Assez	beaucoup	Total
Oui	39 9,8% 36,45%	88 22,1% 44,2%	114 28,6% 52,8%	109 27,4% 55,6%	48 12,1% 57,8%	398 49,7%
Non	68 16,9% 63,55%	111 27,5% 55,8%	102 25,3% 47,2%	87 21,6% 44,4%	35 8,7% 42,2%	403 50,3%
Total	107 13,4%	199 24,8%	216 27,0%	196 24,5%	83 10,4%	801

Tableau 2 : Avez-vous voté aux dernières élections étudiantes ? / Etes-vous inscrit sur les listes électorales

% ligne % colonne	inscrit	Pas inscrit	Total
Oui	325 81,7% 51,7%	73 18,3% 42,4%	398 49,7%
non	304 75,4% 48,3%%	99 12,4% 57,6%	403 50,3%
Total	629 78,5%	172 21,5%	801

Nombre de degré de liberté=1 $x^2 = 4,6$ prob= 0,032

Population= {étudiants des 3 universités strasbourgeoises}
A chaque individu, nous pouvons associer le couple de variables aléatoires (X, Y)
X : Avez-vous voté aux dernières élections étudiantes ?
Il existe 2 modalités de réponses : x_1 = oui x_2 = non
Y : Etes-vous intéressé par la politique ?
5 modalités de réponses : y_1 = pas du tout y_2 = un peu y_3 = moyennement y_4 = assez y_5 = beaucoup

Problème posé : est ce que le fait de voter aux dernières élections étudiantes et l'intérêt porté à la politique de manière générale sont liés ?

Posons l'hypothèse suivante :

H_0 : les variables X et Y sont indépendantes

Ce qui signifie qu'il n'y a aucun lien entre le fait de voter aux dernières élections étudiantes et l'intérêt porté à la politique

Résolution du problème : On ne peut interroger la population toute entière, on va interroger un échantillon de taille 801

1) Cadre théorique : avant d'interroger les étudiants, on pose la théorie comme suit :

N_{ij} = nombre d'étudiants ayant voté aux dernières élections étudiantes selon la modalité x_i de X et qui sont intéressés par la politique selon la modalité y_j de Y dans des échantillons de taille 801

Les N_{ij} sont des variables aléatoires au nombre de 10=5*2

Si H_0 est vrai, c'est-à-dire s'il y a indépendance entre X et Y, alors ces variables aléatoires devraient être les mêmes que les variables aléatoires C_{ij}

$C_{ij} = \dfrac{N_{i.} \times N_{.j}}{n}$ = nombre d'étudiants ayant voté aux dernières élections étudiantes selon la modalité x_i de X et qui sont intéressés par la politique selon la modalité yj de Y dans des 801 échantillons si H0 est vrai

Appliquons le théorème :

Si H_0 est vrai et que « n est grand », alors la variable aléatoire

$\chi^2 = \sum_{i=1}^{i=2} \sum_{j=1}^{j=5} \dfrac{(N_{ij} - C_{ij})^2}{C_{ij}}$ suit une loi du Chi-Deux à (2-1)*(5-1)=4 degrés de liberté

2) Situation après collecte de l'échantillon :

n_{ij} = nombre d'étudiants ayant voté aux dernières élections étudiantes selon la modalité x_i de X et qui sont intéressés par la politique selon la modalité y_j de Y dans l'échantillon recueilli de taille 801.

Les n_{ij} sont des nombres au nombre de 10=5*2

Si H_0 est vrai, c'est-à-dire qu'il y a indépendance entre X et Y, alors ces nombres ne devraient pas s'éloigner des nombres c_{ij} calculés comme suit

$c_{ij} = \dfrac{n_{i.} \times n_{.j}}{n}$ = nombre d'étudiants de l'échantillon qui auraient voté aux dernières élections étudiantes selon la modalité x_i de X et qui sont intéressés par la politique selon la modalité y_j de Y si H_0 est vrai

n_{ij} c_{ij} Contribution au Chi-deux	Pas du tout	Un peu	moyen	Assez	beaucoup	Total
Oui	39 53,2 0,98	88 98,9 1,20	114 107,3 0,41	109 97,4 1,38	48 41,2 1,11	398
Non	68 53,8 3,73	111 100,1 1,18	102 108,7 0,41	87 98,6 1,37	35 41,8 1,09	403
Total	107	199	216	196	83	801

Exemple : chiffre 1 : n_{11} = 39 étudiants ont voté aux dernières élections étudiantes et ne sont pas du tout intéressés par la politique.

Chiffre 2 : $c_{11} = \dfrac{107 \times 398}{801} = 53{,}2$ = on devrait observer 53,2 étudiants ayant voté aux dernières élections étudiantes et n'étant pas du tout intéressés par la politique s'il y a indépendance entre les deux variables.

Chiffre 3 = $\dfrac{(n_{11} - c_{11})^2}{c_{11}} = \dfrac{(39 - 53{,}2)^2}{53{,}2} = 0{,}98$ = contribution de la cellule au calcul du Chi-deux observé.

On peut appliquer le test car au moins 80% des c_{ij} sont supérieurs à 5, d'ailleurs tous les c_{ij} le sont : le test est valide

TEST :

1) ε=5% ➔ table du chi-deux t= 9,49
2) P(χ^2 <9,49)= 95%

Si il y a indépendance, nous avons 95% de probabilité pour que la valeur calculée soit inférieure à 9,49

3) $\chi^2 = \sum_{i=1}^{i=2} \sum_{j=1}^{j=5} \frac{(n_{ij} - c_{ij})^2}{c_{ij}}$ = 0,98 +1,20+0,41+1,38+1,11+3,73+1,18+0,41+1,37+1,09 = 12,86

$\chi^2 \geq 9{,}49 \Rightarrow$ L'hypothèse H_0 est rejetée au risque de 5%

J'ai 5% de probabilité de me tromper en déclarant qu'il existe un lien entre le vote aux élections étudiantes et l'intérêt porté à la politique.

Quels liens ? Pourcentage ligne ou pourcentage colonne ?
A priori il semble logique que l'intérêt porté à la politique engendre un vote plus important aux élections étudiantes. Il est plus judicieux d'interpréter les % colonnes.

Tableau 1 : Avez-vous voté aux dernières élections étudiantes ? **/** Etes-vous intéressé par la politique

% colonne	Pas du tout	Un peu	moyen	Assez	beaucoup
Oui	36,45%	44,2%	52,8%	55,6%	57,8%
Non	63,55%	55,8%	47,2%	44,4%	42,2%
Total	100%	100%	100%	100%	100%

Plus l'intérêt porté à la politique est important et plus le vote aux élections étudiantes est fort, allant de 36,4% pour ceux qui ne s'y intéressent pas du tout à 57,8% pour ceux qui s'y intéressent beaucoup.

Mais on peut quand même remarquer que même dans ce cas, ils ne sont qu'une petite majorité à voter aux élections étudiantes et que les pourcentages bougent peu à partir d'un intérêt moyen portée à la politique.

Tableau 2 : La population est la même mais on veut voir s'il existe un lien entre le fait d'être inscrit sur les listes électorales et celui de voter au dernières élections étudiantes.
Chacune des deux variables possèdent seulement 2 modalités, ce qui donne un tableau 2x2
Posons l'hypothèse d'indépendance entre ces deux variables.
L'échantillon est le même mais cette fois ci la variable $\chi^2 = \sum_{i=1}^{i=2} \sum_{j=1}^{j=5} \frac{(N_{ij} - C_{ij})^2}{C_{ij}}$ suit un chi-deux à 1 degré de liberté.
Les effectifs théoriques $\mathbf{c_{ij}}$ sont tous supérieurs à 5 puisque le plus faible qui est $c_{12} = \frac{172 \times 398}{801} = 85{,}5$ donc on peut appliquer le test :
Les résultats nous indiquent que la réalisation de cette variable est égale à 4,6, ce qui nous amène à un rejet de l'hypothèse au seuil de 5% puisque la valeur critique trouvée dans la table du chi-deux pour 1 degré de liberté et un risque de 5% est de 3,84.
Nous ne rejetons pas au risque de 1% puisque la valeur critique est alors de 6,63.

Prob=0,032=3,2% nous indique que nous pouvons aller jusqu'à un risque de 3,2% de rejeter l'hypothèse, en dessous de ce risque nous ne pouvons pas rejeter H_0.

Pourcentage colonne : Remarquons que 51,7% des étudiants inscrits sur les listes électorales ont voté aux dernières élections étudiantes alors que 42,4% de ceux qui ne sont pas inscrits l'on fait.

EXERCICE 2

En 1995/96, une enquête portant sur l'âge du personnel soignant a été réalisée. Les questionnaires avaient été envoyés par courrier et 534 sont revenus. Dans l'échantillon, sur les 521 personnes qui ont donné leur âge (13 non réponses) on avait trouvé un âge moyen de 36,8 ans avec un écart-type observé de 8,3 ans.

1) **L'âge moyen du personnel soignant en 2005 est de 40,9 ans. Peut-on démontrer au risque de 5 % que le personnel soignant a vieilli en 10 ans ?**

P = Ensemble du personnel soignant en 1995/96

X = Age d'un personnel soignant en 1995/96

$\mathcal{L}(X) = \mathcal{L}(\mu\,;\sigma)$ avec **μ et σ inconnus**

μ = Age moyen de l'ensemble du personnel soignant en 1995/96 = **μ inconnu**

$H_0 : \mu = \mu_0 = 40{,}9$ / $\overline{H_0} : \mu \neq \mu_0$

RESOLUTION : On tire un échantillon de taille 521 : n=521

<u>A priori,</u>

$X_1, X_2, \ldots, X_{521}$, sont des variables aléatoires **indépendantes** de loi $\mathcal{L}(\mu\,;\sigma)$ (taux de sondage < 1/10)

$\overline{X} = \dfrac{\sum X_i}{521}$ = **estimateur de μ = âge moyen dans un échantillon quelconque de 521 personnels soignants**

$\mathcal{L}(\overline{X})$ n'est pas connue car **σ inconnue** donc

$S_c^2 = \dfrac{1}{n-1}\sum(X_i - \overline{X})^2$ = **estimateur de σ^2**

<u>Sous H_0 :</u>

$\mathcal{L}_0\left(\dfrac{\overline{X}-\mu_0}{Sc/\sqrt{n}}\right) = T_{n-1} = T_{520ddl}$ Loi de Student

Mais n est grand (>30), donc :

$\dfrac{\sigma^2}{\sqrt{n}} \approx \dfrac{s_c^2}{\sqrt{n}} \approx \dfrac{s^2}{\sqrt{n}}$

Et $\mathcal{L}_0(\overline{X}) \approx \mathcal{N}\left(\mu_0, \dfrac{s}{\sqrt{n}}\right) = \mathcal{N}\left(40{,}9; \dfrac{8{,}3}{\sqrt{521}}\right) = \mathcal{N}(40{,}9; 0{,}36)$

i) α= 95% $\Rightarrow$ $t_{0,05}$ = 1,96 par lecture de la Table 2

ii) $I_A^{5\%} = \left[\mu_0 - t\times\frac{s}{\sqrt{n}} ; \mu_0 + t\times\frac{s}{\sqrt{n}}\right]$

$= [40,9 – 1,96 \times 0,36 ; 40,9 + 1,96 \times 0,36]$

$= [40,2 ; 41,6]$

$P\left(\overline{X} \in [40,2 ; 41,6]\right) = 95\ \%$

Si je prends un échantillon de 521 personnels soignants et que je calcule l'âge moyen dans cet échantillon, j'ai 95 % de probabilité, **si H_0 est vraie** (c'est-à-dire si le personnel soignant a le même âge en 1995/96 et 2005), pour que l'âge moyen se situe entre 40,2 ans et 41,6 ans.

iii) **<u>A posteriori,</u>** je tire **un** échantillon de 521 personnels soignants

$\overline{x}$ = 36,8 ans : **une** estimation de μ = âge moyen du personnel soignant **observé** dans mon échantillon de taille 521

$\overline{x} \notin I_A \Rightarrow$ Rejet de H_0 :

Le personnel soignant a vieilli depuis 10 ans et j'ai 5 % de risque de me tromper en affirmant cela.

La population totale en France a vieilli entre 1995 et 2005, donc la population du personnel soignant vieillit aussi.

De plus, la population du personnel soignant est surtout composée de femmes, or elles arrêtent peut-être moins souvent de travailler aujourd'hui (après avoir eu 2 enfants par exemple…) par rapport à 1995 ?

2) Quel test aurait-il fallu utiliser si l'échantillon n'avait été que de 25 personnels soignants ? Pourquoi et quelle hypothèse faut-il alors avancer ? Quelle différence avec le 1) ?

Si l'échantillon n'avait été que de 25 personnels soignants, il aurait fallu utiliser le test 2, soit le test de Student car l'échantillon aurait une petite taille (n<30).

Ce test se fait sous la condition de **l'hypothèse de nécessité** soit remplie : $\mathcal{L}(X) = \mathcal{N}(\mu, \sigma)$

$\mathcal{L}(\overline{X})$ n'est alors plus approximativement égale à une loi normale, il faut utiliser la loi de Student pour la variable

$$\mathcal{L}_0\left(\frac{\overline{X} - 40,9}{S_c / \sqrt{n}}\right) = \mathcal{T}_{n-1} = \mathcal{T}_{24ddl}$$

Le test est le même : $H_0 : \mu = \mu_0 = 40,9$ / $\overline{H_0} : \mu \neq \mu_0$

Mais sa résolution est différente :

i) α= 95% $\Rightarrow$ $t_{n-1,\ 1-\alpha} = t_{24\ ;\ 0{,}05}$ = 2,064 valeur lue dans la Table 5

ii) $I_A^{5\%} = \left[\mu_0 - t \times \frac{s_c}{\sqrt{n}} ; \mu_0 + t \times \frac{s_c}{\sqrt{n}}\right]$

$s_c = s\sqrt{n/(n-1)} = 8{,}3\sqrt{25/24} = 8{,}5$

$$\begin{aligned} I_A^{5\%} &= \left[40{,}9 - 2{,}064 \times \frac{8{,}5}{\sqrt{25}} ; 40{,}9 + 2{,}064 \times \frac{8{,}5}{\sqrt{25}}\right] \\ &= [40{,}9 - 2{,}064 \times 1{,}7 ; 40{,}9 + 2{,}064 \times 1{,}7] \\ &= [37{,}4\ ;\ 44{,}4] \end{aligned}$$

iii) Si $\bar{x} \notin I_A \Rightarrow$ Rejet de H_0 avec un risque de 5 % de se tromper
Si $\bar{x} \in I_A \Rightarrow$ On accepte H_0 jusqu'à nouvel ordre.

Ici, $\bar{x}$ = 36,8 ans
$\bar{x} \notin I_A$ donc, Rejet de H_0 : Le personnel soignant a vieilli en 10 ans et j'ai 5 % de risque de me tromper en affirmant cela.

EXERCICE 3

Les bières servies dans ma brasserie préférée doivent contenir 25 cl de bière. Le contenu d'une bière devrait suivre une loi normale d'espérance 25 cl et d'écart-type 3 cl. Depuis le changement de propriétaire de ma brasserie, j'ai l'impression que ma chope de bière est moins remplie.
Pour vérifier, j'ai acheté 15 bières (que j'ai bues accompagné des mes amis !) et j'ai mesuré le contenu en bière de mes chopes. J'ai observé un contenu moyen de 22 cl.

3) Sachant que je connais l'écart-type du contenu de l'ensemble des bières de ma brasserie, peut-on démontrer au risque de 5 % que le nouveau propriétaire est malhonnête ?

P = Ensemble des bières de ma brasserie

X = Contenu en bière d'une chope

$\mathcal{L}(X) = \mathcal{N}(\mu ; 3)$ avec **μ inconnu** et **σ connu**

μ = Contenu moyen en bière de l'ensemble des chopes de ma brasserie = **μ inconnu**

$H_0 : \mu = \mu_0 = 25$ / $\overline{H_0} : \mu \neq \mu_0$

RESOLUTION : On tire un échantillon de taille 15 : n=15

<u>A priori,</u>

$X_1, X_2, \ldots, X_{15}$, sont des variables aléatoires **indépendantes** de loi normale $\mathcal{N}(\mu ; 3)$

$\overline{X} = \frac{\sum X_i}{15}$ = **estimateur de μ**

$\mathcal{L}(\overline{X}) = \mathcal{N}(\mu; \frac{\sigma}{\sqrt{n}})$,

<u>Sous H_0 :</u>

$\mathcal{L}_0(\overline{X}) = \mathcal{N}(\mu_0; \frac{\sigma}{\sqrt{n}}) = \mathcal{N}(25; \frac{3}{\sqrt{15}}) = \mathcal{N}(25; 0{,}77)$

i) $\varepsilon = 5\% \Rightarrow t_{0,05} = 1{,}96$

ii) $I_A^{5\%} = \left[\mu_0 - t \times \frac{\sigma}{\sqrt{n}} ; \mu_0 + t \times \frac{\sigma}{\sqrt{n}}\right]$

$= [25 - 1{,}96 \times 0{,}77 ; 25 + 1{,}96 \times 0{,}77]$

$= [23{,}5 ; 26{,}5]$

$P(\overline{X} \in I_A) = 95\ \%$

Si je prends un échantillon de 15 chopes et que je calcule le contenu moyen en bière de mon échantillon, j'ai 95 % de probabilité, **si H_0 est vraie**, pour que le contenu moyen en bière des 15 chopes soit situé entre 23,5 cl et 26,5 cl.

iii) **A posteriori,** je tire **un** échantillon de 15 bières

$\overline{x} = 22$: **une** estimation de μ = contenu moyen **observé** dans 15 chopes de bière

$\overline{x} \notin I_A \Rightarrow$ Rejet de H_0 :
Le nouveau propriétaire de ma brasserie est malhonnête et j'ai 5 % de risque de me tromper en affirmant cela.

4) **Lors d'une autre soirée, j'apprends que le propriétaire a acheté de nouvelles machines à bière et que l'écart-type du contenu de l'ensemble des bières de ma brasserie n'est plus le même. J'achète à nouveau 15 bières et j'ai mesuré le contenu en bière de mes chopes. J'observe un contenu moyen de 22cl avec un écart-type de 4 cl.**

P = Ensemble des bières de ma brasserie

X = Contenu en bière d'une chope

$\mathcal{L}(X) = \mathcal{N}(\mu\ ;\ \sigma)$ avec **μ et σ inconnus** (**Hypothèse de nécessité** : il faut la loi normale car n est petit)

μ = Contenu moyen en bière de l'ensemble des chopes de ma brasserie = **μ inconnu**

$H_0 : \mu = \mu_0 = 25$ / $\overline{H_0} : \mu \neq \mu_0$

RESOLUTION : On tire un échantillon de taille 15 : n=15

A priori,

$X_1, X_2, \ldots, X_{15}$, sont des variables aléatoires **indépendantes** de loi normale $\mathcal{N}(\mu\ ;\ \sigma)$ avec

$\overline{X} = \dfrac{\sum X_i}{15}$ = **estimateur de μ**

$L(\overline{X})$ n'est pas connue car n est petit et **σ inconnue**

$S_c^2 = \dfrac{1}{n-1}\sum(X_i - \overline{X})^2$ = **estimateur de σ^2**

$\mathcal{L}_0\left(\dfrac{\overline{X} - \mu_0}{Sc/\sqrt{n}}\right) = T_{n-1} = T_{14ddl}$ Loi de Student

α= 95% $\Rightarrow$ $t_{n-1,\,1-\alpha} = t_{14\,;\,0,05}$ = 2,145 par lecture de la Table 5

$$I_A^{5\%} = \left[\mu_0 - t\times\frac{S_c}{\sqrt{n}} ; \mu_0 + t\times\frac{S_c}{\sqrt{n}}\right]$$

<u>A posteriori,</u> je tire **un** échantillon de 15 bières

$s_c = s\sqrt{\dfrac{n}{n-1}} = 4\sqrt{\dfrac{15}{14}} =$ 4,14 = estimation de **σ**

$$I_A^{5\%} = \left[\mu_0 - t\times\frac{s_c}{\sqrt{n}} ; \mu_0 + t\times\frac{s_c}{\sqrt{n}}\right]$$

$$= \left[25 - 2{,}145\times\frac{4{,}14}{\sqrt{15}} ; 25 + 2{,}145\times\frac{4{,}14}{\sqrt{15}}\right]$$

$$= [25 - 2{,}145\times 1{,}07 \; ; \; 25 + 2{,}145\times 1{,}07]$$

= [22,7 ; 27,3]

Si je prends un échantillon de 15 chopes et que je calcule le contenu moyen en bière de mon échantillon, j'ai 95 % de probabilité, **si H_0 est vraie**, pour que le contenu moyen en bière des 15 chopes soit situé entre 22,7 cl et 27,3 cl.

$\bar{x}$ = 22 : **une** estimation de μ = contenu moyen **observé** dans 15 chopes de bière

Soit $\bar{x} \notin I_A \Rightarrow$ Rejet de H_0 :
Le nouveau propriétaire de ma brasserie est malhonnête et j'ai 5 % de risque de me tromper en affirmant cela.