Correction exercices 1 à 3

Tests relevant de la statistique du X^2

1. Je suis malaaddeeee…

Les résultats de l'évolution d'une maladie M, à la suite de l'emploi de l'un ou l'autre des traitements A et B, figurent dans le tableau d'effectifs ci-dessous:

	Guérison	Amélioration	Etat stationnaire
A	280	210	110
B	220	90	90

Peut on dire que les traitements A et B sont différents (au seuil $\alpha = 0.05$) ?

1. Je suis malaaddeeee...

Solution :

H_0 : Homogénéité des effectifs entre échantillons

H_1 : Hétérogénéité des effectifs entre échantillons

	Guérison	Amélioration	Etat stationnaire	Marge colonne
A	280 **300**	210 **180**	110 **120**	600
B	220 **200**	90 **120**	90 **80**	400
Marge ligne	500	300	200	1000

$$\chi^2_{obs} = \frac{(280-300)^2}{300} + \frac{(210-180)^2}{180} + \frac{(110-120)^2}{120} + \cdots + \frac{(90-80)^2}{80} = 17.9$$

Pour α = 0.05 et ddl = 2 ➔ $\chi^2_{théo} = 5.99$ (13.82 pour α = 0.01)

$\chi^2_{obs} > \chi^2_{théo}$, **on rejette H_0** et on accepte H_1 : les traitements ont des effets différents

Effectuer ce test sous R

Code : On utilisera la fonction « **chisq.test** », dans laquelle on renseignera :

Tableau de type « matrix » (2 L * 3 C)

Les effectifs observés

Comparaison par lignes (échantillons)

```
chisq.test(matrix(c(280,210,110,220,90,90), ncol=3, byrow=TRUE))
```

Sortie :

```
        Pearson's Chi-squared test

data:  matrix(c(280, 210, 110, 220, 90, 90), ncol = 3, byrow = TRUE)
X-squared = 17.917, df = 2, p-value = 0.0001287
```

χ^2_{obs}

ddl

p-value hautement significative (< 0.001) → rejet de H0

Effectuer ce test sous R

NB: On peut également utiliser le test exact de Fisher (petits échantillons)

→ Fonction « fisher.test », syntaxe similaire :

```
> fisher.test(matrix(c(280,210,110,220,90,90), ncol=3, byrow=TRUE))

        Fisher's Exact Test for Count Data

data:  matrix(c(280, 210, 110, 220, 90, 90), ncol = 3, byrow = TRUE)
p-value = 0.0001086
alternative hypothesis: two.sided
```

p-value hautement significative (< 0.001), rejet de H0

2. Les chveux dans les yeux...

Sur un échantillon de 6800 individus, on a relevé la couleur des yeux et cheveux de chaque personnes. Les effectifs sont :

	Blonds	Bruns	Roux	Noirs	Marge C
Marrons	115	438	16	288	857
Bleus	1768	807	47	189	2811
Gris ou Verts	946	1387	53	746	3132
Marge L	2829	2632	116	1223	6800

La distribution de la couleur des cheveux varie elle en fonction de la couleur des yeux ?

2. Les chveux dans les yeux...

Sur un échantillon de 6800 individus, on a relevé la couleur des yeux et cheveux de chaque personnes. Les effectifs sont :

	Blonds	Bruns	Roux	Noirs	Marge C
Marrons	115	438	16	288	857
Bleus	1768	807	47	189	2811
Gris ou Verts	946	1387	53	746	3132
Marge L	2829	2632	116	1223	6800

La distribution de la couleur des cheveux varie elle en fonction de la couleur des yeux ?

2 VA qualitatives nominales (non contrôlées par l'investigateur), mesurées simultanément sur 1 seul échantillon d'individus → test d'indépendance

2. Les chveux dans les yeux...

Solution :

H_0 : le fait de connaître la couleur des cheveux ne permet pas d'aider à deviner la couleur des yeux et inversement

H_1 : il y a un lien entre couleur des cheveux et des yeux

	Blonds	Bruns	Roux	Noirs	Marge C
Marrons	115 **356.53**	438 **331.71**	16 **14.62**	288 **154.13**	857
Bleus	1768 **1169.46**	807 **1088.02**	47 **47.95**	189 **505.57**	2811
Gris ou Verts	946 **1303.00**	1387 **1212.27**	53 **53.43**	746 **563.30**	3132
Marge L	2829	2632	116	1223	6800

$X^2_{obs} = 1073{,}89$ $\quad$ $X^2_{Théo} = 12.592$ (α = 0.05, ddl= (4-1)*(3-1)= 6)

$X^2_{obs} >> X^2_{Théo}$, et p-value << 0.001

➔ *Ces deux variables sont liées (indépendance rejetée)*

Effectuer ce test sous R

Code :

```
mat_ChevYeux<-matrix(c(115,438,16,288,
                       1768,807,47,189,
                       946,1387,53,746),
                     ncol = 4, byrow = T)
# rajouter noms lignes et colonnes
colnames(mat_ChevYeux)<-c("Blond","Brun","Roux","Noir")
row.names(mat_ChevYeux)<-c("Marron","Bleu","Gris/Vert")
# rajouter les marges L et C
addmargins(mat_ChevYeux)

## le test
Khi2_chevYeux<-chisq.test(mat_ChevYeux)
Khi2_chevYeux
#résultats attendus/théoriques
round(Khi2_chevYeux$expected,2)
```

Sortie :

```
        Pearson's Chi-squared test

data:  matrix(c(115, 438, 16, 288, 1768, 807, 47, 189,
946, 1387, 53,     746), ncol = 4, byrow = T)
X-squared = 1073.5, df = 6, p-value < 2.2e-16
```

3. Qui vole un boeuf...

Le bœuf musqué (*Ovibos moscatus*) est infecté par des nématodes dont on connait trois variétés différentes. Pour vérifier si l'importance relative de ces trois variétés varie d'une région à l'autre on a dénombré des individus appartenant aux trois variétés dans les excréments de bœufs musqués de trois régions différentes :

	Variété I	Variété II	Variété III
Alaska	7	42	40
T NW	12	54	16
Iles B	6	33	33

L'importance relative des trois variétés de nématodes varie-t-elle entre les trois régions ?

3. Qui vole un boeuf...

Le bœuf musqué (*Ovibos moscatus*) est infecté par des nématodes dont on connait trois variétés différentes. Pour vérifier si l'importance relative de ces trois variétés varie d'une région à l'autre on a dénombré des individus appartenant aux trois variétés dans les excréments de bœufs musqués de trois régions différentes :

	Variété I	Variété II	Variété III
Alaska	7	42	40
T NW	12	54	16
Iles B	6	33	33

L'importance relative des trois variétés de nématodes varie-t-elle entre les trois régions ?

3 échantillons indépendants (facteur qualitatif contrôlé/fixé = localité)

3 modalités (facteur qualitatif aléatoire = variétés nématodes)

Comparaison d'effectifs entre échantillons → Chi^2 homogénéité avec H1 bilatéral

Effectuer ce test sous R

```
chisq.test(matrix(c(7,42, 40, 12, 54, 16, 6,33,33), ncol=3, byrow = TRUE))
```

```
        Pearson's Chi-squared test

data:  matrix(c(7, 42, 40, 12, 54, 16, 6, 33, 33), ncol = 3, byrow = TRUE)
X-squared = 16, df = 4, p-value = 0.003019
```

ddl = (3-1)* (3-1) =4

P-value< 0.01, les **différences observées sont très significatives** et ne sont pas dues seulement aux fluctuations d'échantillonnage (lien avec localité des groupes comparés)