

Tests du χ^2

Outils Mathématiques & Statistique Décisionnelle

Licence 3 Management, Economie de la firme, CCF

Faculté d'économie, gestion & AES

Université de Bordeaux - Collège DSPEG

Automne 2024

A. Lourme

alexandre.lourme@u-bordeaux.fr

Outline

- Adéquation à une loi probabilité (goodness-of-fit test)
- Indépendance de variables qualitatives (contingency table test)

Le dé est-il truqué ?

On lance cent fois le même dé :

face	1	2	3	4	5	6
effectif observé	7	18	26	15	18	16
fréquence observée	0,07	0,18	0,26	0,15	0,18	0,16

Le dé est-il truqué ?

On lance cent fois le même dé :

face	1	2	3	4	5	6
effectif observé	7	18	26	15	18	16
fréquence observée	0,07	0,18	0,26	0,15	0,18	0,16

Supposons le dé équilibré :

face	1	2	3	4	5	6
fréquence théorique	1/6	1/6	1/6	1/6	1/6	1/6

Le dé est-il truqué ?

On lance cent fois le même dé :

face	1	2	3	4	5	6
effectif observé	7	18	26	15	18	16
fréquence observée	0,07	0,18	0,26	0,15	0,18	0,16

Supposons le dé équilibré :

face	1	2	3	4	5	6
fréquence théorique	1/6	1/6	1/6	1/6	1/6	1/6

Comment interpréter l'écart entre fréquences observées et théoriques ?

(i) l'hypothèse d'un dé équilibré est incorrecte : le dé est truqué

(ii) l'hypothèse d'un dé équilibré est correcte, mais la fréquence d'apparition de chaque face est soumise aux fluctuations d'échantillonnage

Le dé est-il truqué ?

On lance cent fois le même dé :

face	1	2	3	4	5	6
effectif observé	7	18	26	15	18	16
fréquence observée	0,07	0,18	0,26	0,15	0,18	0,16

Supposons le dé équilibré :

face	1	2	3	4	5	6
fréquence théorique	1/6	1/6	1/6	1/6	1/6	1/6

Comment interpréter l'écart entre fréquences observées et théoriques ?

(i) l'hypothèse d'un dé équilibré est incorrecte : le dé est truqué

(ii) l'hypothèse d'un dé équilibré est correcte, mais la fréquence d'apparition de chaque face est soumise aux fluctuations d'échantillonnage

Comment choisir entre (i) et (ii) ?

- mesurer l'écart entre fréquences observées et théoriques
- rejeter (ii) au profit de (i) si cet écart dépasse un certain seuil

Test du χ^2 d'adéquation

On considère r valeurs/classes/modalités v_i ($i = 1, \dots, r$) ; chacune des n unités d'un échantillon aléatoire est affectée à v_i avec probabilité p_i ($\sum_{i=1}^r p_i = 1$).

Test du χ^2 d'adéquation

On considère r valeurs/classes/modalités v_i ($i = 1, \dots, r$) ; chacune des n unités d'un échantillon aléatoire est affectée à v_i avec probabilité p_i ($\sum_{i=1}^r p_i = 1$).

Le nombre N_i d'unités affectées à v_i ($\sum_{i=1}^r N_i = n$) est aléatoire et $N = (N_1, \dots, N_r)$ est distribué selon une loi multinomiale .

Test du χ^2 d'adéquation

On considère r valeurs/classes/modalités v_i ($i = 1, \dots, r$) ; chacune des n unités d'un échantillon aléatoire est affectée à v_i avec probabilité p_i ($\sum_{i=1}^r p_i = 1$).

Le nombre N_i d'unités affectées à v_i ($\sum_{i=1}^r N_i = n$) est aléatoire et $N = (N_1, \dots, N_r)$ est distribué selon une loi multinomiale .

valeur/classe/modalité	v_1	v_2	$\dots$	v_r
effectif observé	N_1	N_2	$\dots$	N_r
effectif théorique	np_1	np_2	$\dots$	np_r

Test du χ^2 d'adéquation

On considère r valeurs/classes/modalités v_i ($i = 1, \dots, r$) ; chacune des n unités d'un échantillon aléatoire est affectée à v_i avec probabilité p_i ($\sum_{i=1}^r p_i = 1$).

Le nombre N_i d'unités affectées à v_i ($\sum_{i=1}^r N_i = n$) est aléatoire et $N = (N_1, \dots, N_r)$ est distribué selon une loi multinomiale .

valeur/classe/modalité	v_1	v_2	$\dots$	v_r
effectif observé	N_1	N_2	$\dots$	N_r
effectif théorique	np_1	np_2	$\dots$	np_r

On en déduit :

$$D = \sum_{i=1}^r (N_i - np_i)^2 / (np_i) \underset{\text{ap.}}{\sim} \chi_{r-1}^2 \quad (1)$$

Test du χ^2 d'adéquation

On considère r valeurs/classes/modalités v_i ($i = 1, \dots, r$) ; chacune des n unités d'un échantillon aléatoire est affectée à v_i avec probabilité p_i ($\sum_{i=1}^r p_i = 1$).

Le nombre N_i d'unités affectées à v_i ($\sum_{i=1}^r N_i = n$) est aléatoire et $N = (N_1, \dots, N_r)$ est distribué selon une loi multinomiale .

valeur/classe/modalité	v_1	v_2	$\dots$	v_r
effectif observé	N_1	N_2	$\dots$	N_r
effectif théorique	np_1	np_2	$\dots$	np_r

On en déduit :

$$D = \sum_{i=1}^r (N_i - np_i)^2 / (np_i) \underset{\text{ap.}}{\sim} \chi_{r-1}^2 \quad (1)$$

(1) permet de définir un test (asymptotique) du χ^2 comparant $(p_1, \dots, p_r)$ à une valeur de référence $(p_1^0, \dots, p_r^0)$:

Test du χ^2 d'adéquation

On considère r valeurs/classes/modalités v_i ($i = 1, \dots, r$) ; chacune des n unités d'un échantillon aléatoire est affectée à v_i avec probabilité p_i ($\sum_{i=1}^r p_i = 1$).

Le nombre N_i d'unités affectées à v_i ($\sum_{i=1}^r N_i = n$) est aléatoire et $N = (N_1, \dots, N_r)$ est distribué selon une loi multinomiale .

valeur/classe/modalité	v_1	v_2	$\dots$	v_r
effectif observé	N_1	N_2	$\dots$	N_r
effectif théorique	np_1	np_2	$\dots$	np_r

On en déduit :

$$D = \sum_{i=1}^r (N_i - np_i)^2 / (np_i) \underset{\text{ap.}}{\sim} \chi_{r-1}^2 \quad (1)$$

(1) permet de définir un test (asymptotique) du χ^2 comparant $(p_1, \dots, p_r)$ à une valeur de référence $(p_1^0, \dots, p_r^0)$:

- hypothèses : $\mathcal{H}_0 : \forall i, p_i = p_i^0$ vs $\mathcal{H}_1 : \exists i, p_i \neq p_i^0$

Test du χ^2 d'adéquation

On considère r valeurs/classes/modalités v_i ($i = 1, \dots, r$) ; chacune des n unités d'un échantillon aléatoire est affectée à v_i avec probabilité p_i ($\sum_{i=1}^r p_i = 1$).

Le nombre N_i d'unités affectées à v_i ($\sum_{i=1}^r N_i = n$) est aléatoire et $N = (N_1, \dots, N_r)$ est distribué selon une loi multinomiale .

valeur/classe/modalité	v_1	v_2	$\dots$	v_r
effectif observé	N_1	N_2	$\dots$	N_r
effectif théorique	np_1	np_2	$\dots$	np_r

On en déduit :

$$D = \sum_{i=1}^r (N_i - np_i)^2 / (np_i) \underset{\text{ap.}}{\sim} \chi_{r-1}^2 \quad (1)$$

(1) permet de définir un test (asymptotique) du χ^2 comparant $(p_1, \dots, p_r)$ à une valeur de référence $(p_1^0, \dots, p_r^0)$:

- hypothèses : $\mathcal{H}_0 : \forall i, p_i = p_i^0$ vs $\mathcal{H}_1 : \exists i, p_i \neq p_i^0$
- statistique de test : $X = \sum_{i=1}^r (N_i - np_i^0)^2 / (np_i^0) \underset{\text{ap.}}{\sim} \chi_{r-1}^2$ sous $\mathcal{H}_0$

Test du χ^2 d'adéquation

On considère r valeurs/classes/modalités v_i ($i = 1, \dots, r$) ; chacune des n unités d'un échantillon aléatoire est affectée à v_i avec probabilité p_i ($\sum_{i=1}^r p_i = 1$).

Le nombre N_i d'unités affectées à v_i ($\sum_{i=1}^r N_i = n$) est aléatoire et $N = (N_1, \dots, N_r)$ est distribué selon une loi multinomiale .

valeur/classe/modalité	v_1	v_2	$\dots$	v_r
effectif observé	N_1	N_2	$\dots$	N_r
effectif théorique	np_1	np_2	$\dots$	np_r

On en déduit :

$$D = \sum_{i=1}^r (N_i - np_i)^2 / (np_i) \underset{\text{ap.}}{\sim} \chi_{r-1}^2 \quad (1)$$

(1) permet de définir un test (asymptotique) du χ^2 comparant $(p_1, \dots, p_r)$ à une valeur de référence $(p_1^0, \dots, p_r^0)$:

- hypothèses : $\mathcal{H}_0 : \forall i, p_i = p_i^0$ vs $\mathcal{H}_1 : \exists i, p_i \neq p_i^0$
- statistique de test : $X = \sum_{i=1}^r (N_i - np_i^0)^2 / (np_i^0) \underset{\text{ap.}}{\sim} \chi_{r-1}^2$ sous $\mathcal{H}_0$
- zone de rejet de $\mathcal{H}_0$ au seuil α : $X > \chi_{r-1; 1-\alpha}^2$

Test du χ^2 d'adéquation

On considère r valeurs/classes/modalités v_i ($i = 1, \dots, r$) ; chacune des n unités d'un échantillon aléatoire est affectée à v_i avec probabilité p_i ($\sum_{i=1}^r p_i = 1$).

Le nombre N_i d'unités affectées à v_i ($\sum_{i=1}^r N_i = n$) est aléatoire et $N = (N_1, \dots, N_r)$ est distribué selon une loi multinomiale .

valeur/classe/modalité	v_1	v_2	$\dots$	v_r
effectif observé	N_1	N_2	$\dots$	N_r
effectif théorique	np_1	np_2	$\dots$	np_r

On en déduit :

$$D = \sum_{i=1}^r (N_i - np_i)^2 / (np_i) \underset{\text{ap.}}{\sim} \chi_{r-1}^2 \quad (1)$$

(1) permet de définir un test (asymptotique) du χ^2 comparant $(p_1, \dots, p_r)$ à une valeur de référence $(p_1^0, \dots, p_r^0)$:

- hypothèses : $\mathcal{H}_0 : \forall i, p_i = p_i^0$ vs $\mathcal{H}_1 : \exists i, p_i \neq p_i^0$
- statistique de test : $X = \sum_{i=1}^r (N_i - np_i^0)^2 / (np_i^0) \underset{\text{ap.}}{\sim} \chi_{r-1}^2$ sous $\mathcal{H}_0$
- zone de rejet de $\mathcal{H}_0$ au seuil α : $X > \chi_{r-1; 1-\alpha}^2$
- p -valeur du test : $p\text{-val.} = \mathbb{P}(\chi_{r-1}^2 > X_{\text{obs}})$

Exercice 1. Le dé est-il truqué ?

face	1	2	3	4	5	6
effectif observé	7	18	26	15	18	16

Exercice 1. Le dé est-il truqué ?

face	1	2	3	4	5	6
effectif observé	7	18	26	15	18	16
effectif théorique	100/6	100/6	100/6	100/6	100/6	100/6

Exercice 1. Le dé est-il truqué ?

face	1	2	3	4	5	6
effectif observé	7	18	26	15	18	16
effectif théorique	100/6	100/6	100/6	100/6	100/6	100/6

Chacun des 100 lancers ($n = 100$) fera apparaître l'une des six faces ($r = 6$). N_i : nombre **aléatoire** d'apparitions de la face i au cours des 100 lancers.

Exercice 1. Le dé est-il truqué ?

face	1	2	3	4	5	6
effectif observé	7	18	26	15	18	16
effectif théorique	100/6	100/6	100/6	100/6	100/6	100/6

Chacun des 100 lancers ($n = 100$) fera apparaître l'une des six faces ($r = 6$). N_i : nombre **aléatoire** d'apparitions de la face i au cours des 100 lancers.

Si le dé est équilibré la probabilité p_i d'apparition de la face i est $p_i^0 = 1/6$ pour tout i .

Exercice 1. Le dé est-il truqué ?

face	1	2	3	4	5	6
effectif observé	7	18	26	15	18	16
effectif théorique	100/6	100/6	100/6	100/6	100/6	100/6

Chacun des 100 lancers ($n = 100$) fera apparaître l'une des six faces ($r = 6$). N_i : nombre **aléatoire** d'apparitions de la face i au cours des 100 lancers.

Si le dé est équilibré la probabilité p_i d'apparition de la face i est $p_i^0 = 1/6$ pour tout i .

- Hyp : $\mathcal{H}_0 : \forall i, p_i = p_i^0 = 1/6$ (dé équilibré) vs $\mathcal{H}_1 : \exists i, p_i \neq p_i^0 = 1/6$ (dé truqué)

Exercice 1. Le dé est-il truqué ?

face	1	2	3	4	5	6
effectif observé	7	18	26	15	18	16
effectif théorique	100/6	100/6	100/6	100/6	100/6	100/6

Chacun des 100 lancers ($n = 100$) fera apparaître l'une des six faces ($r = 6$). N_i : nombre **aléatoire** d'apparitions de la face i au cours des 100 lancers.

Si le dé est équilibré la probabilité p_i d'apparition de la face i est $p_i^0 = 1/6$ pour tout i .

- Hyp : $\mathcal{H}_0 : \forall i, p_i = p_i^0 = 1/6$ (dé équilibré) vs $\mathcal{H}_1 : \exists i, p_i \neq p_i^0 = 1/6$ (dé truqué)
- SdT : $X = \sum_{i=1}^6 (N_i - 100/6)^2 / (100/6) \underset{\text{ap.}}{\sim} \chi_5^2$ sous $\mathcal{H}_0$

Exercice 1. Le dé est-il truqué ?

face	1	2	3	4	5	6
effectif observé	7	18	26	15	18	16
effectif théorique	100/6	100/6	100/6	100/6	100/6	100/6

Chacun des 100 lancers ($n = 100$) fera apparaître l'une des six faces ($r = 6$). N_i : nombre **aléatoire** d'apparitions de la face i au cours des 100 lancers.

Si le dé est équilibré la probabilité p_i d'apparition de la face i est $p_i^0 = 1/6$ pour tout i .

- Hyp : $\mathcal{H}_0 : \forall i, p_i = p_i^0 = 1/6$ (dé équilibré) vs $\mathcal{H}_1 : \exists i, p_i \neq p_i^0 = 1/6$ (dé truqué)
- SdT : $X = \sum_{i=1}^6 (N_i - 100/6)^2 / (100/6) \underset{\text{ap.}}{\sim} \chi_5^2$ sous $\mathcal{H}_0$
- ZdR : $\mathcal{H}_0$ rejetée au seuil 1% ($\alpha = 0,01$) ssi $X > \chi_{5,0,99}^2$

Exercice 1. Le dé est-il truqué ?

face	1	2	3	4	5	6
effectif observé	7	18	26	15	18	16
effectif théorique	100/6	100/6	100/6	100/6	100/6	100/6

Chacun des 100 lancers ($n = 100$) fera apparaître l'une des six faces ($r = 6$). N_i : nombre **aléatoire** d'apparitions de la face i au cours des 100 lancers.

Si le dé est équilibré la probabilité p_i d'apparition de la face i est $p_i^0 = 1/6$ pour tout i .

- Hyp : $\mathcal{H}_0 : \forall i, p_i = p_i^0 = 1/6$ (dé équilibré) vs $\mathcal{H}_1 : \exists i, p_i \neq p_i^0 = 1/6$ (dé truqué)
- SdT : $X = \sum_{i=1}^6 (N_i - 100/6)^2 / (100/6) \underset{\text{ap.}}{\sim} \chi_5^2$ sous $\mathcal{H}_0$
- ZdR : $\mathcal{H}_0$ rejetée au seuil 1% ($\alpha = 0,01$) ssi $X > \chi_{5,0,99}^2$
- Déc : $X_{obs} = (7 - 100/6)^2 / (100/6) + \dots + (16 - 100/6)^2 / (100/6) \approx 11$
 $X_{obs} \leq \chi_{5,0,99}^2 \approx 15$; on ne rejette pas $\mathcal{H}_0$ au seuil 1%.

Exercice 1. Le dé est-il truqué ?

face	1	2	3	4	5	6
effectif observé	7	18	26	15	18	16
effectif théorique	100/6	100/6	100/6	100/6	100/6	100/6

Chacun des 100 lancers ($n = 100$) fera apparaître l'une des six faces ($r = 6$). N_i : nombre **aléatoire** d'apparitions de la face i au cours des 100 lancers.

Si le dé est équilibré la probabilité p_i d'apparition de la face i est $p_i^0 = 1/6$ pour tout i .

• Hyp : $\mathcal{H}_0 : \forall i, p_i = p_i^0 = 1/6$ (dé équilibré) vs $\mathcal{H}_1 : \exists i, p_i \neq p_i^0 = 1/6$ (dé truqué)

• SdT : $X = \sum_{i=1}^6 (N_i - 100/6)^2 / (100/6) \underset{\text{ap.}}{\sim} \chi_5^2$ sous $\mathcal{H}_0$

• ZdR : $\mathcal{H}_0$ rejetée au seuil 1% ($\alpha = 0,01$) ssi $X > \chi_{5,0,99}^2$

• Déc : $X_{obs} = (7 - 100/6)^2 / (100/6) + \dots + (16 - 100/6)^2 / (100/6) \approx 11$

$X_{obs} \leq \chi_{5,0,99}^2 \approx 15$; on ne rejette pas $\mathcal{H}_0$ au seuil 1%.

• p -valeur : $p\text{-val.} = \mathbb{P}(\chi_5^2 > X_{obs}) = 1 - \mathbb{P}(\chi_5^2 \leq 11) \approx 0,05 > \alpha = 0,01$

Tant que le seuil de significativité est inférieur à 5% on ne rejette pas $\mathcal{H}_0$.

Même exercice sous R

face	1	2	3	4	5	6
effectif observé	7	18	26	15	18	16

Le dé est-il truqué ?

Même exercice sous R

face	1	2	3	4	5	6
effectif observé	7	18	26	15	18	16

Le dé est-il truqué ?

`observe = c(7,18,26,15,18,16)` : effectifs observés.

Même exercice sous R

face	1	2	3	4	5	6
effectif observé	7	18	26	15	18	16

Le dé est-il truqué ?

`observe = c(7,18,26,15,18,16)` : effectifs observés.

`theorique = c(1/6,1/6,1/6,1/6,1/6,1/6)` : fréquences théoriques.

Même exercice sous R

face	1	2	3	4	5	6
effectif observé	7	18	26	15	18	16

Le dé est-il truqué ?

`observe = c(7,18,26,15,18,16)` : effectifs observés.

`theorique = c(1/6,1/6,1/6,1/6,1/6,1/6)` : fréquences théoriques.

`mytest = chisq.test(observe,p=theorique)` : le test du χ^2 d'adéquation.

Même exercice sous R

face	1	2	3	4	5	6
effectif observé	7	18	26	15	18	16

Le dé est-il truqué ?

`observe = c(7,18,26,15,18,16)` : effectifs observés.

`theorique = c(1/6,1/6,1/6,1/6,1/6,1/6)` : fréquences théoriques.

`mytest = chisq.test(observe,p=theorique)` : le test du χ^2 d'adéquation.

`mytest$statistic` : la valeur observée de la statistique de test (≈ 11).

Même exercice sous R

face	1	2	3	4	5	6
effectif observé	7	18	26	15	18	16

Le dé est-il truqué ?

`observe = c(7,18,26,15,18,16)` : effectifs observés.

`theorique = c(1/6,1/6,1/6,1/6,1/6,1/6)` : fréquences théoriques.

`mytest = chisq.test(observe,p=theorique)` : le test du χ^2 d'adéquation.

`mytest$statistic` : la valeur observée de la statistique de test (≈ 11).

`mytest$p.value` : la p -valeur du test ($\approx 0,05$).

Même exercice sous R

face	1	2	3	4	5	6
effectif observé	7	18	26	15	18	16

Le dé est-il truqué ?

`observe = c(7,18,26,15,18,16)` : effectifs observés.

`theorique = c(1/6,1/6,1/6,1/6,1/6,1/6)` : fréquences théoriques.

`mytest = chisq.test(observe,p=theorique)` : le test du χ^2 d'adéquation.

`mytest$statistic` : la valeur observée de la statistique de test (≈ 11).

`mytest$p.value` : la p -valeur du test ($\approx 0,05$).

`qchisq(0.99,5)` : la valeur critique associée au seuil 1% (≈ 15).

Même exercice sous R

face	1	2	3	4	5	6
effectif observé	7	18	26	15	18	16

Le dé est-il truqué ?

`observe = c(7,18,26,15,18,16)` : effectifs observés.

`theorique = c(1/6,1/6,1/6,1/6,1/6,1/6)` : fréquences théoriques.

`mytest = chisq.test(observe,p=theorique)` : le test du χ^2 d'adéquation.

`mytest$statistic` : la valeur observée de la statistique de test (≈ 11).

`mytest$p.value` : la p -valeur du test ($\approx 0,05$).

`qchisq(0.99,5)` : la valeur critique associée au seuil 1% (≈ 15).

La valeur observée de la statistique de test (11) est inférieure à la valeur critique (15) ; de façon équivalente, la p -valeur (5%) est supérieure au seuil de significativité (1%) ; au seuil 1% on ne rejette pas que le dé soit équilibré.

Exercice 2. Adéquation à une loi exponentielle

durée de vie	[0; 1000]	]1000; 2000]	]2000; 3000]	]3000; 4000]	]4000; 5000]	]5000; 6000]
eff. observé	19	13	11	4	2	1

La durée de vie des ampoules provient-elle de $\mathcal{E}(1/1000)$?

Exercice 2. Adéquation à une loi exponentielle

durée de vie	[0; 1000]	]1000; 2000]	]2000; 3000]	]3000; 4000]	]4000; 5000]	]5000; 6000]
eff. observé	19	13	11	4	2	1

La durée de vie des ampoules provient-elle de $\mathcal{E}(1/1000)$?

Six classes ($r = 6$) ; on y répartit la durée de vie d'un échantillon aléatoire de 50 ampoules ($n = 50$). N_i : nombre aléatoire d'ampoules dans la classe i .

Exercice 2. Adéquation à une loi exponentielle

durée de vie	[0; 1000]	]1000; 2000]	]2000; 3000]	]3000; 4000]	]4000; 5000]	]5000; 6000]
eff. observé	19	13	11	4	2	1

La durée de vie des ampoules provient-elle de $\mathcal{E}(1/1000)$?

Six classes ($r = 6$) ; on y répartit la durée de vie d'un échantillon aléatoire de 50 ampoules ($n = 50$). N_i : nombre aléatoire d'ampoules dans la classe i .

Si la durée de vie des ampoules est distribuée selon $\mathcal{E}(1/1000)$, la probabilité p_i qu'une ampoule meure entre $(i - 1) \times 1000$ et $i \times 1000$ est égale à : $p_i^0 = e^{-i+1} - e^{-i}$.

Exercice 2. Adéquation à une loi exponentielle

durée de vie	[0; 1000]	]1000; 2000]	]2000; 3000]	]3000; 4000]	]4000; 5000]	]5000; 6000]
eff. observé	19	13	11	4	2	1
eff. théorique	31,6	11,6	4,3	1,6	0,6	0,2

La durée de vie des ampoules provient-elle de $\mathcal{E}(1/1000)$?

Six classes ($r = 6$) ; on y répartit la durée de vie d'un échantillon aléatoire de 50 ampoules ($n = 50$). N_i : nombre aléatoire d'ampoules dans la classe i .

Si la durée de vie des ampoules est distribuée selon $\mathcal{E}(1/1000)$, la probabilité p_i qu'une ampoule meure entre $(i - 1) \times 1000$ et $i \times 1000$ est égale à : $p_i^0 = e^{-i+1} - e^{-i}$.

Exercice 2. Adéquation à une loi exponentielle

durée de vie	[0; 1000]	]1000; 2000]	]2000; 3000]	]3000; 4000]	]4000; 5000]	]5000; 6000]
eff. observé	19	13	11	4	2	1
eff. théorique	31,6	11,6	4,3	1,6	0,6	0,2

La durée de vie des ampoules provient-elle de $\mathcal{E}(1/1000)$?

Six classes ($r = 6$) ; on y répartit la durée de vie d'un échantillon aléatoire de 50 ampoules ($n = 50$). N_i : nombre aléatoire d'ampoules dans la classe i .

Si la durée de vie des ampoules est distribuée selon $\mathcal{E}(1/1000)$, la probabilité p_i qu'une ampoule meure entre $(i-1) \times 1000$ et $i \times 1000$ est égale à : $p_i^0 = e^{-i+1} - e^{-i}$.

- Hyp : $\mathcal{H}_0 : \forall i, p_i = p_i^0 = e^{-i+1} - e^{-i}$ vs $\mathcal{H}_1 : \exists i, p_i \neq p_i^0 = e^{-i+1} - e^{-i}$

Exercice 2. Adéquation à une loi exponentielle

durée de vie	[0; 1000]	]1000; 2000]	]2000; 3000]	]3000; 4000]	]4000; 5000]	]5000; 6000]
eff. observé	19	13	11	4	2	1
eff. théorique	31,6	11,6	4,3	1,6	0,6	0,2

La durée de vie des ampoules provient-elle de $\mathcal{E}(1/1000)$?

Six classes ($r = 6$) ; on y répartit la durée de vie d'un échantillon aléatoire de 50 ampoules ($n = 50$). N_i : nombre aléatoire d'ampoules dans la classe i .

Si la durée de vie des ampoules est distribuée selon $\mathcal{E}(1/1000)$, la probabilité p_i qu'une ampoule meure entre $(i-1) \times 1000$ et $i \times 1000$ est égale à : $p_i^0 = e^{-i+1} - e^{-i}$.

- Hyp : $\mathcal{H}_0 : \forall i, p_i = p_i^0 = e^{-i+1} - e^{-i}$ vs $\mathcal{H}_1 : \exists i, p_i \neq p_i^0 = e^{-i+1} - e^{-i}$
- SdT : $X = \sum_{i=1}^6 (N_i - 50 \times (e^{-i+1} - e^{-i}))^2 / (50 \times (e^{-i+1} - e^{-i})) \underset{\text{ap.}}{\sim} \chi_5^2$ sous $\mathcal{H}_0$

Exercice 2. Adéquation à une loi exponentielle

durée de vie	[0; 1000]	]1000; 2000]	]2000; 3000]	]3000; 4000]	]4000; 5000]	]5000; 6000]
eff. observé	19	13	11	4	2	1
eff. théorique	31,6	11,6	4,3	1,6	0,6	0,2

La durée de vie des ampoules provient-elle de $\mathcal{E}(1/1000)$?

Six classes ($r = 6$) ; on y répartit la durée de vie d'un échantillon aléatoire de 50 ampoules ($n = 50$). N_i : nombre aléatoire d'ampoules dans la classe i .

Si la durée de vie des ampoules est distribuée selon $\mathcal{E}(1/1000)$, la probabilité p_i qu'une ampoule meure entre $(i-1) \times 1000$ et $i \times 1000$ est égale à : $p_i^0 = e^{-i+1} - e^{-i}$.

- Hyp : $\mathcal{H}_0 : \forall i, p_i = p_i^0 = e^{-i+1} - e^{-i}$ vs $\mathcal{H}_1 : \exists i, p_i \neq p_i^0 = e^{-i+1} - e^{-i}$
- SdT : $X = \sum_{i=1}^6 (N_i - 50 \times (e^{-i+1} - e^{-i}))^2 / (50 \times (e^{-i+1} - e^{-i})) \underset{\text{ap.}}{\sim} \chi_5^2$ sous $\mathcal{H}_0$
- ZdR : $\mathcal{H}_0$ rejetée au seuil 10% ($\alpha = 0,1$) ssi $X > \chi_{5,0,9}^2$

Exercice 2. Adéquation à une loi exponentielle

durée de vie	[0; 1000]	]1000; 2000]	]2000; 3000]	]3000; 4000]	]4000; 5000]	]5000; 6000]
eff. observé	19	13	11	4	2	1
eff. théorique	31,6	11,6	4,3	1,6	0,6	0,2

La durée de vie des ampoules provient-elle de $\mathcal{E}(1/1000)$?

Six classes ($r = 6$) ; on y répartit la durée de vie d'un échantillon aléatoire de 50 ampoules ($n = 50$). N_i : nombre aléatoire d'ampoules dans la classe i .

Si la durée de vie des ampoules est distribuée selon $\mathcal{E}(1/1000)$, la probabilité p_i qu'une ampoule meure entre $(i-1) \times 1000$ et $i \times 1000$ est égale à : $p_i^0 = e^{-i+1} - e^{-i}$.

- Hyp : $\mathcal{H}_0 : \forall i, p_i = p_i^0 = e^{-i+1} - e^{-i}$ vs $\mathcal{H}_1 : \exists i, p_i \neq p_i^0 = e^{-i+1} - e^{-i}$
- SdT : $X = \sum_{i=1}^6 (N_i - 50 \times (e^{-i+1} - e^{-i}))^2 / (50 \times (e^{-i+1} - e^{-i})) \underset{\text{ap.}}{\sim} \chi_5^2$ sous $\mathcal{H}_0$
- ZdR : $\mathcal{H}_0$ rejetée au seuil 10% ($\alpha = 0,1$) ssi $X > \chi_{5,0,9}^2$
- Déc : $X_{obs} = (19 - 31,6)^2 / 31,6 + \dots + (1 - 0,2)^2 / 0,2 \approx 25,7$
 $X_{obs} > \chi_{5,0,90}^2 \approx 9,2$; on rejette $\mathcal{H}_0$ au seuil 10%.

Exercice 2. Adéquation à une loi exponentielle

durée de vie	[0; 1000]	]1000; 2000]	]2000; 3000]	]3000; 4000]	]4000; 5000]	]5000; 6000]
eff. observé	19	13	11	4	2	1
eff. théorique	31,6	11,6	4,3	1,6	0,6	0,2

La durée de vie des ampoules provient-elle de $\mathcal{E}(1/1000)$?

Six classes ($r = 6$) ; on y répartit la durée de vie d'un échantillon aléatoire de 50 ampoules ($n = 50$). N_i : nombre aléatoire d'ampoules dans la classe i .

Si la durée de vie des ampoules est distribuée selon $\mathcal{E}(1/1000)$, la probabilité p_i qu'une ampoule meure entre $(i - 1) \times 1000$ et $i \times 1000$ est égale à : $p_i^0 = e^{-i+1} - e^{-i}$.

- Hyp : $\mathcal{H}_0 : \forall i, p_i = p_i^0 = e^{-i+1} - e^{-i}$ vs $\mathcal{H}_1 : \exists i, p_i \neq p_i^0 = e^{-i+1} - e^{-i}$
- SdT : $X = \sum_{i=1}^6 (N_i - 50 \times (e^{-i+1} - e^{-i}))^2 / (50 \times (e^{-i+1} - e^{-i})) \underset{\text{ap.}}{\sim} \chi_5^2$ sous $\mathcal{H}_0$
- ZdR : $\mathcal{H}_0$ rejetée au seuil 10% ($\alpha = 0,1$) ssi $X > \chi_{5,0,9}^2$
- Déc : $X_{obs} = (19 - 31,6)^2 / 31,6 + \dots + (1 - 0,2)^2 / 0,2 \approx 25,7$
 $X_{obs} > \chi_{5,0,90}^2 \approx 9,2$; on rejette $\mathcal{H}_0$ au seuil 10%.
- p -valeur : $p\text{-val.} = \mathbb{P}(\chi_5^2 > X_{obs}) = 1 - \mathbb{P}(\chi_5^2 \leq X_{obs}) \approx 0,0001$.

Tant que le seuil de significativité est supérieur à 0,01% on rejette $\mathcal{H}_0$.

Outline

- Adéquation à une loi probabilité (goodness-of-fit test)
- Indépendance de variables qualitatives (contingency table test)

Sexe & couleur des yeux : deux variables indépendantes ?

Un échantillon de 100 français :

		<i>couleur des yeux</i>		
		<i>vert</i>	<i>bleu</i>	<i>marron</i>
<i>sexe</i>	<i>garçon</i>	10	11	28
	<i>filles</i>	7	18	26

Tableau 1. effectifs conjoints

Sexe & couleur des yeux : deux variables indépendantes ?

Un échantillon de 100 français :

		<i>couleur des yeux</i>			total
		<i>vert</i>	<i>bleu</i>	<i>marron</i>	
sexe	garçon	10	11	28	49
	filles	7	18	26	51
	total	17	29	54	100

Tableau 1. effectifs conjoints

Sexe & couleur des yeux : deux variables indépendantes ?

Un échantillon de 100 français :

		couleur des yeux			total
		vert	bleu	marron	
sexe	garçon	10	11	28	49
	filles	7	18	26	51
total		17	29	54	100

Tableau 1. effectifs conjoints

		couleur des yeux			total
		vert	bleu	marron	
sexe	garçon	0, 10	0, 11	0, 28	0, 49
	filles	0, 07	0, 18	0, 26	0, 51
total		0, 17	0, 29	0, 54	1

Tableau 2. fréquences conjointes observées

Sexe & couleur des yeux : deux variables indépendantes ?

Un échantillon de 100 français :

		couleur des yeux			total
		vert	bleu	marron	
sexe	garçon	10	11	28	49
	filles	7	18	26	51
	total	17	29	54	100

Tableau 1. effectifs conjoints

		couleur des yeux			total
		vert	bleu	marron	
sexe	garçon	0, 10	0, 11	0, 28	0, 49
	filles	0, 07	0, 18	0, 26	0, 51
	total	0, 17	0, 29	0, 54	1

Tableau 2. fréquences conjointes observées

		couleur des yeux			total
		vert	bleu	marron	
sexe	garçon	0, 0833	0, 1421	0, 2646	0, 49
	filles	0, 0867	0, 1479	0, 2754	0, 51
	total	0, 17	0, 29	0, 54	1

Tableau 3. fréquences conjointes théoriques

Sexe & couleur des yeux : deux variables indépendantes ?

Un échantillon de 100 français :

		couleur des yeux			total
		vert	bleu	marron	
sexe	garçon	10	11	28	49
	filles	7	18	26	51
	total	17	29	54	100

Tableau 1. effectifs conjoints

		couleur des yeux			total
		vert	bleu	marron	
sexe	garçon	0, 10	0, 11	0, 28	0, 49
	filles	0, 07	0, 18	0, 26	0, 51
	total	0, 17	0, 29	0, 54	1

Tableau 2. fréquences conjoints observées

		couleur des yeux			total
		vert	bleu	marron	
sexe	garçon	0, 0833	0, 1421	0, 2646	0, 49
	filles	0, 0867	0, 1479	0, 2754	0, 51
	total	0, 17	0, 29	0, 54	1

Tableau 3. fréquences conjoints théoriques

Si les variables sont indépendantes l'écart entre fréquences observées et fréquences théoriques tend à disparaître lorsque la taille de l'échantillon augmente.

Test du χ^2 d'indépendance

U et V sont deux variables qualitatives ; chacune des n unités d'un échantillon aléatoire prend une des r valeurs $u_1, \dots, u_r$ de U et une des s valeurs $v_1, \dots, v_s$ de V .

Test du χ^2 d'indépendance

U et V sont deux variables qualitatives ; chacune des n unités d'un échantillon aléatoire prend une des r valeurs $u_1, \dots, u_r$ de U et une des s valeurs $v_1, \dots, v_s$ de V .

		V				
		v_1	v_2		v_s	total
U	u_1	$N_{1,1}$	$N_{1,2}$	$\dots$	$N_{1,s}$	$N_{1,\bullet}$
	u_2	$N_{2,1}$	$N_{2,2}$	$\dots$	$N_{2,s}$	$N_{2,\bullet}$
	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
	u_r	$N_{r,1}$	$N_{r,2}$	$\dots$	$N_{r,s}$	$N_{r,\bullet}$
total		$N_{\bullet,1}$	$N_{\bullet,2}$		$N_{\bullet,s}$	n

Test du χ^2 d'indépendance

U et V sont deux variables qualitatives ; chacune des n unités d'un échantillon aléatoire prend une des r valeurs $u_1, \dots, u_r$ de U et une des s valeurs $v_1, \dots, v_s$ de V .

		V				
		v_1	v_2	...	v_s	total
U	u_1	$N_{1,1}$	$N_{1,2}$	...	$N_{1,s}$	$N_{1,\bullet}$
	u_2	$N_{2,1}$	$N_{2,2}$	...	$N_{2,s}$	$N_{2,\bullet}$
	...	...	...	...	...	...
	u_r	$N_{r,1}$	$N_{r,2}$	...	$N_{r,s}$	$N_{r,\bullet}$
total		$N_{\bullet,1}$	$N_{\bullet,2}$		$N_{\bullet,s}$	n

$N_{i,j}$: nombre **aléatoire** d'unités prenant les valeurs u_i et v_j

$N_{i,\bullet} = \sum_{j=1}^s N_{i,j}$: nombre aléatoire d'unités prenant la valeur u_i de U

$N_{\bullet,j} = \sum_{i=1}^r N_{i,j}$: nombre aléatoire d'unités prenant la valeur v_j de V

$N_{i,j}^* = N_{i,\bullet} \times N_{\bullet,j} / n$: nbre d'unités prenant les valeurs u_i et v_j en cas d'indépendance

$N_{i,j}$ est un effectif observé ; $N_{i,j}^*$ est un effectif théorique sous hypothèse d'indépendance.

Test du χ^2 d'indépendance

U et V sont deux variables qualitatives ; chacune des n unités d'un échantillon aléatoire prend une des r valeurs $u_1, \dots, u_r$ de U et une des s valeurs $v_1, \dots, v_s$ de V .

		V				
		v_1	v_2	...	v_s	total
U	u_1	$N_{1,1}$	$N_{1,2}$	...	$N_{1,s}$	$N_{1,\bullet}$
	u_2	$N_{2,1}$	$N_{2,2}$	...	$N_{2,s}$	$N_{2,\bullet}$
	...	...	...	...	...	...
	u_r	$N_{r,1}$	$N_{r,2}$	...	$N_{r,s}$	$N_{r,\bullet}$
total		$N_{\bullet,1}$	$N_{\bullet,2}$		$N_{\bullet,s}$	n

$N_{i,j}$: nombre **aléatoire** d'unités prenant les valeurs u_i et v_j

$N_{i,\bullet} = \sum_{j=1}^s N_{i,j}$: nombre aléatoire d'unités prenant la valeur u_i de U

$N_{\bullet,j} = \sum_{i=1}^r N_{i,j}$: nombre aléatoire d'unités prenant la valeur v_j de V

$N_{i,j}^* = N_{i,\bullet} \times N_{\bullet,j} / n$: nbre d'unités prenant les valeurs u_i et v_j en cas d'indépendance

$N_{i,j}$ est un effectif observé ; $N_{i,j}^*$ est un effectif théorique sous hypothèse d'indépendance.

• hypothèses : $\mathcal{H}_0$: U et V indépendantes vs $\mathcal{H}_1$: U et V non indépendantes

Test du χ^2 d'indépendance

U et V sont deux variables qualitatives ; chacune des n unités d'un échantillon aléatoire prend une des r valeurs $u_1, \dots, u_r$ de U et une des s valeurs $v_1, \dots, v_s$ de V .

		V				
		v_1	v_2	...	v_s	total
U	u_1	$N_{1,1}$	$N_{1,2}$	...	$N_{1,s}$	$N_{1,\bullet}$
	u_2	$N_{2,1}$	$N_{2,2}$	...	$N_{2,s}$	$N_{2,\bullet}$
	...	...	...	...	...	...
	u_r	$N_{r,1}$	$N_{r,2}$	...	$N_{r,s}$	$N_{r,\bullet}$
total		$N_{\bullet,1}$	$N_{\bullet,2}$		$N_{\bullet,s}$	n

$N_{i,j}$: nombre **aléatoire** d'unités prenant les valeurs u_i et v_j

$N_{i,\bullet} = \sum_{j=1}^s N_{i,j}$: nombre aléatoire d'unités prenant la valeur u_i de U

$N_{\bullet,j} = \sum_{i=1}^r N_{i,j}$: nombre aléatoire d'unités prenant la valeur v_j de V

$N_{i,j}^* = N_{i,\bullet} \times N_{\bullet,j} / n$: nbre d'unités prenant les valeurs u_i et v_j en cas d'indépendance

$N_{i,j}$ est un effectif observé ; $N_{i,j}^*$ est un effectif théorique sous hypothèse d'indépendance.

• hypothèses : $\mathcal{H}_0$: U et V indépendantes vs $\mathcal{H}_1$: U et V non indépendantes

• statistique de test : $X = \sum_{i=1}^r \sum_{j=1}^s \left(N_{i,j} - N_{i,j}^* \right)^2 / N_{i,j}^* \underset{\text{ap.}}{\sim} \chi^2_{(r-1) \times (s-1)}$ sous $\mathcal{H}_0$

Test du χ^2 d'indépendance

U et V sont deux variables qualitatives ; chacune des n unités d'un échantillon aléatoire prend une des r valeurs $u_1, \dots, u_r$ de U et une des s valeurs $v_1, \dots, v_s$ de V .

		V				
		v_1	v_2	...	v_s	total
U	u_1	$N_{1,1}$	$N_{1,2}$	...	$N_{1,s}$	$N_{1,\bullet}$
	u_2	$N_{2,1}$	$N_{2,2}$	...	$N_{2,s}$	$N_{2,\bullet}$
	...	...	...	...	...	...
	u_r	$N_{r,1}$	$N_{r,2}$	...	$N_{r,s}$	$N_{r,\bullet}$
total		$N_{\bullet,1}$	$N_{\bullet,2}$		$N_{\bullet,s}$	n

$N_{i,j}$: nombre **aléatoire** d'unités prenant les valeurs u_i et v_j

$N_{i,\bullet} = \sum_{j=1}^s N_{i,j}$: nombre aléatoire d'unités prenant la valeur u_i de U

$N_{\bullet,j} = \sum_{i=1}^r N_{i,j}$: nombre aléatoire d'unités prenant la valeur v_j de V

$N_{i,j}^* = N_{i,\bullet} \times N_{\bullet,j} / n$: nbre d'unités prenant les valeurs u_i et v_j en cas d'indépendance

$N_{i,j}$ est un effectif observé ; $N_{i,j}^*$ est un effectif théorique sous hypothèse d'indépendance.

- hypothèses : $\mathcal{H}_0$: U et V indépendantes vs $\mathcal{H}_1$: U et V non indépendantes
- statistique de test : $X = \sum_{i=1}^r \sum_{j=1}^s \left(N_{i,j} - N_{i,j}^* \right)^2 / N_{i,j}^* \underset{\text{ap.}}{\sim} \chi^2_{(r-1) \times (s-1)}$ sous $\mathcal{H}_0$
- zone de rejet de $\mathcal{H}_0$ au seuil α : $X_{\text{obs}} > \chi^2_{(r-1) \times (s-1); 1-\alpha}$

Test du χ^2 d'indépendance

U et V sont deux variables qualitatives ; chacune des n unités d'un échantillon aléatoire prend une des r valeurs $u_1, \dots, u_r$ de U et une des s valeurs $v_1, \dots, v_s$ de V .

		V				
		v_1	v_2	...	v_s	total
U	u_1	$N_{1,1}$	$N_{1,2}$	...	$N_{1,s}$	$N_{1,\bullet}$
	u_2	$N_{2,1}$	$N_{2,2}$	...	$N_{2,s}$	$N_{2,\bullet}$
	...	...	...	...	...	...
	u_r	$N_{r,1}$	$N_{r,2}$	...	$N_{r,s}$	$N_{r,\bullet}$
total		$N_{\bullet,1}$	$N_{\bullet,2}$		$N_{\bullet,s}$	n

$N_{i,j}$: nombre **aléatoire** d'unités prenant les valeurs u_i et v_j

$N_{i,\bullet} = \sum_{j=1}^s N_{i,j}$: nombre aléatoire d'unités prenant la valeur u_i de U

$N_{\bullet,j} = \sum_{i=1}^r N_{i,j}$: nombre aléatoire d'unités prenant la valeur v_j de V

$N_{i,j}^* = N_{i,\bullet} \times N_{\bullet,j} / n$: nbre d'unités prenant les valeurs u_i et v_j en cas d'indépendance

$N_{i,j}$ est un effectif observé ; $N_{i,j}^*$ est un effectif théorique sous hypothèse d'indépendance.

- hypothèses : $\mathcal{H}_0$: U et V indépendantes vs $\mathcal{H}_1$: U et V non indépendantes
- statistique de test : $X = \sum_{i=1}^r \sum_{j=1}^s \left(N_{i,j} - N_{i,j}^* \right)^2 / N_{i,j}^* \underset{\text{ap.}}{\sim} \chi^2_{(r-1) \times (s-1)}$ sous $\mathcal{H}_0$
- zone de rejet de $\mathcal{H}_0$ au seuil α : $X_{obs} > \chi^2_{(r-1) \times (s-1); 1-\alpha}$
- p -valeur du test : $p\text{-val.} = \mathbb{P}(\chi^2_{(r-1) \times (s-1)} > X_{obs})$

Exercice 3. Sexe & couleur des yeux : indépendantes ?

		<i>couleur des yeux</i>			total
		<i>vert</i>	<i>bleu</i>	<i>marron</i>	
sexe	garçon	10	11	28	49
	filles	7	18	26	51
	total	17	29	54	100

Tableau 2. *effectifs conjoints observés*

Exercice 3. Sexe & couleur des yeux : indépendantes ?

		couleur des yeux			total
		vert	bleu	marron	
sexe	garçon	10	11	28	49
	filles	7	18	26	51
	total	17	29	54	100

Tableau 2. effectifs conjoints observés

Chaque sujet d'un échantillon **aléatoire** de $n = 100$ français est un garçon ($i = 1$)/une fille ($i = 2$) et ses yeux sont verts ($j = 1$)/bleus ($j = 2$)/marron ($j = 3$).

Exercice 3. Sexe & couleur des yeux : indépendantes ?

		couleur des yeux			total
		vert	bleu	marron	
sexe	garçon	10	11	28	49
	filles	7	18	26	51
	total	17	29	54	100

Tableau 2. effectifs conjoints observés

Chaque sujet d'un échantillon **aléatoire** de $n = 100$ français est un garçon ($i = 1$)/une fille ($i = 2$) et ses yeux sont verts ($j = 1$)/bleus ($j = 2$)/marron ($j = 3$).

$N_{i,j}$: nombre de garçons /filles aux yeux verts/bleus/marron.

Exercice 3. Sexe & couleur des yeux : indépendantes ?

		couleur des yeux			
		vert	bleu	marron	total
sexe	garçon	10	11	28	49
	filles	7	18	26	51
	total	17	29	54	100

Tableau 2. *effectifs conjoints observés*

Chaque sujet d'un échantillon **aléatoire** de $n = 100$ français est un garçon ($i = 1$)/une fille ($i = 2$) et ses yeux sont verts ($j = 1$)/bleus ($j = 2$)/marron ($j = 3$).

$N_{i,j}$: nombre de garçons /filles aux yeux verts/bleus/marron.

$N_{i,\bullet}$: nbre de garçons/filles ; $N_{\bullet,j}$: nbre de sujets aux yeux verts/bleus/marron.

Exercice 3. Sexe & couleur des yeux : indépendantes ?

		couleur des yeux			total
		vert	bleu	marron	
sexe	garçon	10	11	28	49
	filles	7	18	26	51
	total	17	29	54	100

Tableau 2. effectifs conjoints observés

		couleur des yeux			total
		vert	bleu	marron	
sexe	garçon	8, 33	14, 21	26, 46	49
	filles	8, 67	14, 79	27, 54	51
	total	17	29	54	100

Tableau 3. effectifs conjoints théoriques

Chaque sujet d'un échantillon **aléatoire** de $n = 100$ français est un garçon ($i = 1$)/une fille ($i = 2$) et ses yeux sont verts ($j = 1$)/bleus ($j = 2$)/marron ($j = 3$).

$N_{i,j}$: nombre de garçons /filles aux yeux verts/bleus/marron.

$N_{i,\bullet}$: nbre de garçons/filles ; $N_{\bullet,j}$: nbre de sujets aux yeux verts/bleus/marron.

$N_{i,j}^* = N_{i,\bullet} \times N_{\bullet,j} / 100$: effectif théorique sous hypothèse d'indépendance.

Exercice 3. Sexe & couleur des yeux : indépendantes ?

		couleur des yeux			total
		vert	bleu	marron	
sexe	garçon	10	11	28	49
	filles	7	18	26	51
	total	17	29	54	100

Tableau 2. effectifs conjoints observés

		couleur des yeux			total
		vert	bleu	marron	
sexe	garçon	8, 33	14, 21	26, 46	49
	filles	8, 67	14, 79	27, 54	51
	total	17	29	54	100

Tableau 3. effectifs conjoints théoriques

Chaque sujet d'un échantillon **aléatoire** de $n = 100$ français est un garçon ($i = 1$)/une fille ($i = 2$) et ses yeux sont verts ($j = 1$)/bleus ($j = 2$)/marron ($j = 3$).

$N_{i,j}$: nombre de garçons /filles aux yeux verts/bleus/marron.

$N_{i,\bullet}$: nbre de garçons/filles ; $N_{\bullet,j}$: nbre de sujets aux yeux verts/bleus/marron.

$N_{i,j}^* = N_{i,\bullet} \times N_{\bullet,j} / 100$: effectif théorique sous hypothèse d'indépendance.

- Hyp : $\mathcal{H}_0$: indépendance vs $\mathcal{H}_1$: non indépendance de sexe & couleur des yeux

Exercice 3. Sexe & couleur des yeux : indépendantes ?

		couleur des yeux			total
		vert	bleu	marron	
sexe	garçon	10	11	28	49
	filles	7	18	26	51
	total	17	29	54	100

Tableau 2. effectifs conjoints observés

		couleur des yeux			total
		vert	bleu	marron	
sexe	garçon	8, 33	14, 21	26, 46	49
	filles	8, 67	14, 79	27, 54	51
	total	17	29	54	100

Tableau 3. effectifs conjoints théoriques

Chaque sujet d'un échantillon **aléatoire** de $n = 100$ français est un garçon ($i = 1$)/une fille ($i = 2$) et ses yeux sont verts ($j = 1$)/bleus ($j = 2$)/marron ($j = 3$).

$N_{i,j}$: nombre de garçons /filles aux yeux verts/bleus/marron.

$N_{i,\bullet}$: nbre de garçons/filles ; $N_{\bullet,j}$: nbre de sujets aux yeux verts/bleus/marron.

$N_{i,j}^* = N_{i,\bullet} \times N_{\bullet,j} / 100$: effectif théorique sous hypothèse d'indépendance.

• Hyp : $\mathcal{H}_0$: indépendance vs $\mathcal{H}_1$: non indépendance de sexe & couleur des yeux

• SdT : $X = \sum_{i=1}^2 \sum_{j=1}^3 \left(N_{i,j} - N_{i,j}^* \right)^2 / N_{i,j}^* \sim \chi_2^2$ sous $\mathcal{H}_0$

Exercice 3. Sexe & couleur des yeux : indépendantes ?

		couleur des yeux			total
		vert	bleu	marron	
sexe	garçon	10	11	28	49
	filles	7	18	26	51
	total	17	29	54	100

Tableau 2. effectifs conjoints observés

		couleur des yeux			total
		vert	bleu	marron	
sexe	garçon	8, 33	14, 21	26, 46	49
	filles	8, 67	14, 79	27, 54	51
	total	17	29	54	100

Tableau 3. effectifs conjoints théoriques

Chaque sujet d'un échantillon **aléatoire** de $n = 100$ français est un garçon ($i = 1$)/une fille ($i = 2$) et ses yeux sont verts ($j = 1$)/bleus ($j = 2$)/marron ($j = 3$).

$N_{i,j}$: nombre de garçons /filles aux yeux verts/bleus/marron.

$N_{i,\bullet}$: nbre de garçons/filles ; $N_{\bullet,j}$: nbre de sujets aux yeux verts/bleus/marron.

$N_{i,j}^* = N_{i,\bullet} \times N_{\bullet,j} / 100$: effectif théorique sous hypothèse d'indépendance.

- Hyp : $\mathcal{H}_0$: indépendance vs $\mathcal{H}_1$: non indépendance de sexe & couleur des yeux
- SdT : $X = \sum_{i=1}^2 \sum_{j=1}^3 \left(N_{i,j} - N_{i,j}^* \right)^2 / N_{i,j}^* \sim \chi_{2,1-\alpha}^2$ sous $\mathcal{H}_0$
- ZdR : $\mathcal{H}_0$ est rejetée au seuil de 5% ($\alpha = 0,05$) ssi $X > \chi_{2,1-\alpha}^2$

Exercice 3. Sexe & couleur des yeux : indépendantes ?

		couleur des yeux			total
		vert	bleu	marron	
sexe	garçon	10	11	28	49
	filles	7	18	26	51
	total	17	29	54	100

Tableau 2. *effectifs conjoints observés*

		couleur des yeux			total
		vert	bleu	marron	
sexe	garçon	8, 33	14, 21	26, 46	49
	filles	8, 67	14, 79	27, 54	51
	total	17	29	54	100

Tableau 3. *effectifs conjoints théoriques*

Chaque sujet d'un échantillon **aléatoire** de $n = 100$ français est un garçon ($i = 1$)/une fille ($i = 2$) et ses yeux sont verts ($j = 1$)/bleus ($j = 2$)/marron ($j = 3$).

$N_{i,j}$: nombre de garçons /filles aux yeux verts/bleus/marron.

$N_{i,\bullet}$: nbre de garçons/filles ; $N_{\bullet,j}$: nbre de sujets aux yeux verts/bleus/marron.

$N_{i,j}^* = N_{i,\bullet} \times N_{\bullet,j} / 100$: effectif théorique sous hypothèse d'indépendance.

- Hyp : $\mathcal{H}_0$: indépendance vs $\mathcal{H}_1$: non indépendance de sexe & couleur des yeux
- SdT : $X = \sum_{i=1}^2 \sum_{j=1}^3 \left(N_{i,j} - N_{i,j}^* \right)^2 / N_{i,j}^* \underset{\text{ap.}}{\sim} \chi^2_2$ sous $\mathcal{H}_0$
- ZdR : $\mathcal{H}_0$ est rejetée au seuil de 5% ($\alpha = 0,05$) ssi $X > \chi^2_{2,1-\alpha}$
- Déc : $X_{\text{obs}} = (10 - 8,33)^2 / 8,33 + \dots + (26 - 27,54)^2 / 27,54 \approx 2,25$
 $X_{\text{obs}} \leq \chi^2_{2,0,95} \approx 6$; on ne rejette pas l'indépendance au seuil 5%

Exercice 3. Sexe & couleur des yeux : indépendantes ?

		couleur des yeux			total
		vert	bleu	marron	
sexe	garçon	10	11	28	49
	filles	7	18	26	51
	total	17	29	54	100

Tableau 2. effectifs conjoints observés

		couleur des yeux			total
		vert	bleu	marron	
sexe	garçon	8, 33	14, 21	26, 46	49
	filles	8, 67	14, 79	27, 54	51
	total	17	29	54	100

Tableau 3. effectifs conjoints théoriques

Chaque sujet d'un échantillon **aléatoire** de $n = 100$ français est un garçon ($i = 1$)/une fille ($i = 2$) et ses yeux sont verts ($j = 1$)/bleus ($j = 2$)/marron ($j = 3$).

$N_{i,j}$: nombre de garçons /filles aux yeux verts/bleus/marron.

$N_{i,\bullet}$: nbre de garçons/filles ; $N_{\bullet,j}$: nbre de sujets aux yeux verts/bleus/marron.

$N_{i,j}^* = N_{i,\bullet} \times N_{\bullet,j} / 100$: effectif théorique sous hypothèse d'indépendance.

• Hyp : $\mathcal{H}_0$: indépendance vs $\mathcal{H}_1$: non indépendance de sexe & couleur des yeux

• SdT : $X = \sum_{i=1}^2 \sum_{j=1}^3 \left(N_{i,j} - N_{i,j}^* \right)^2 / N_{i,j}^* \underset{\text{ap.}}{\sim} \chi_2^2$ sous $\mathcal{H}_0$

• ZdR : $\mathcal{H}_0$ est rejetée au seuil de 5% ($\alpha = 0,05$) ssi $X > \chi_{2,1-\alpha}^2$

• Déc : $X_{\text{obs}} = (10 - 8,33)^2 / 8,33 + \dots + (26 - 27,54)^2 / 27,54 \approx 2,25$
 $X_{\text{obs}} \leq \chi_{2,0,95}^2 \approx 6$; on ne rejette pas l'indépendance au seuil 5%

• p -valeur : $p\text{-val.} = \mathbb{P}(\chi_2^2 > X_{\text{obs}}) = 1 - \mathbb{P}(\chi_2^2 \leq 2,25) \approx 0,32$

Même exercice sous R

		<i>couleur des yeux</i>			total
		<i>vert</i>	<i>bleu</i>	<i>marron</i>	
sexe	garçon	10	11	28	49
	filles	7	18	26	51
total		17	29	54	100
<i>effectifs conjoints observés</i>					

Même exercice sous R

		<i>couleur des yeux</i>			
		<i>vert</i>	<i>bleu</i>	<i>marron</i>	<i>total</i>
<i>sexe</i>	<i>garçon</i>	10	11	28	49
	<i>filles</i>	7	18	26	51
<i>total</i>		17	29	54	100
<i>effectifs conjoints observés</i>					

`mydata = matrix(c(10,11,28,7,18,26),nrow=2,byrow=TRUE)` : les observations.

Même exercice sous R

		<i>couleur des yeux</i>			
		<i>vert</i>	<i>bleu</i>	<i>marron</i>	total
sexe	garçon	10	11	28	49
	filles	7	18	26	51
total		17	29	54	100
<i>effectifs conjoints observés</i>					

`mydata = matrix(c(10,11,28,7,18,26),nrow=2,byrow=TRUE)` : les observations.

`mytest = chisq.test(mydata)` : le test du χ^2 d'indépendance.

Même exercice sous R

		<i>couleur des yeux</i>			
		<i>vert</i>	<i>bleu</i>	<i>marron</i>	<i>total</i>
<i>sexe</i>	<i>garçon</i>	10	11	28	49
	<i>filles</i>	7	18	26	51
<i>total</i>		17	29	54	100
<i>effectifs conjoints observés</i>					

`mydata = matrix(c(10,11,28,7,18,26),nrow=2,byrow=TRUE)` : les observations.

`mytest = chisq.test(mydata)` : le test du χ^2 d'indépendance.

`mytest$expected` : le tableau des effectifs théoriques.

Même exercice sous R

		<i>couleur des yeux</i>			
		<i>vert</i>	<i>bleu</i>	<i>marron</i>	<i>total</i>
<i>sexe</i>	<i>garçon</i>	10	11	28	49
	<i>filles</i>	7	18	26	51
<i>total</i>		17	29	54	100

effectifs conjoints observés

`mydata = matrix(c(10,11,28,7,18,26),nrow=2,byrow=TRUE)` : les observations.

`mytest = chisq.test(mydata)` : le test du χ^2 d'indépendance.

`mytest$expected` : le tableau des effectifs théoriques.

`mytest$statistic` : la valeur observée de la statistique de test ($\approx 2,25$).

Même exercice sous R

		<i>couleur des yeux</i>			
		<i>vert</i>	<i>bleu</i>	<i>marron</i>	<i>total</i>
<i>sexe</i>	<i>garçon</i>	10	11	28	49
	<i>filles</i>	7	18	26	51
<i>total</i>		17	29	54	100

effectifs conjoints observés

`mydata = matrix(c(10,11,28,7,18,26),nrow=2,byrow=TRUE)` : les observations.

`mytest = chisq.test(mydata)` : le test du χ^2 d'indépendance.

`mytest$expected` : le tableau des effectifs théoriques.

`mytest$statistic` : la valeur observée de la statistique de test ($\approx 2,25$).

`mytest$p.value` : la p -valeur du test ($\approx 0,32$).

Même exercice sous R

		<i>couleur des yeux</i>			
		<i>vert</i>	<i>bleu</i>	<i>marron</i>	<i>total</i>
<i>sexe</i>	<i>garçon</i>	10	11	28	49
	<i>filles</i>	7	18	26	51
<i>total</i>		17	29	54	100

effectifs conjoints observés

`mydata = matrix(c(10,11,28,7,18,26),nrow=2,byrow=TRUE)` : les observations.

`mytest = chisq.test(mydata)` : le test du χ^2 d'indépendance.

`mytest$expected` : le tableau des effectifs théoriques.

`mytest$statistic` : la valeur observée de la statistique de test ($\approx 2,25$).

`mytest$p.value` : la p -valeur du test ($\approx 0,32$).

`qchisq(0.95,3)` : la valeur critique associée au seuil 5% ($\approx 6,0$).

Même exercice sous R

		<i>couleur des yeux</i>			
		<i>vert</i>	<i>bleu</i>	<i>marron</i>	<i>total</i>
<i>sexe</i>	<i>garçon</i>	10	11	28	49
	<i>filles</i>	7	18	26	51
<i>total</i>		17	29	54	100

effectifs conjoints observés

`mydata = matrix(c(10,11,28,7,18,26),nrow=2,byrow=TRUE)` : les observations.

`mytest = chisq.test(mydata)` : le test du χ^2 d'indépendance.

`mytest$expected` : le tableau des effectifs théoriques.

`mytest$statistic` : la valeur observée de la statistique de test ($\approx 2,25$).

`mytest$p.value` : la p -valeur du test ($\approx 0,32$).

`qchisq(0.95,3)` : la valeur critique associée au seuil 5% ($\approx 6,0$).

La valeur observée de la statistique de test (2,25) est inférieure à la valeur critique (6,0) ; de façon équivalente, la p -valeur (32%) est supérieure au seuil de significativité (5%) ; on ne rejette pas l'indépendance ($\mathcal{H}_0$) au seuil 5%.

Exercices de synthèse

Exercice 4

nbre de pannes	0	1	2	3	≥ 4
effectif observé	6	4	5	2	3

Nombre quotidien de pannes du tramway bordelais observé sur vingt jours

Doit-on rejeter au seuil 5% : *le nombre de pannes du tram bordelais en un jour est distribué selon $\mathcal{P}(1)$?*

Exercice 5

		classe			
		1	2	3	équipage
survie	oui	203	118	178	212
	non	122	167	528	673

Tableau de contingence classe/survie des passagers du Titanic

Doit-on rejeter au seuil 5% : *la survie d'un passager du Titanic est indépendante de sa classe ?*