Mathématiques pour les Sciences de la Vie
Probabilités – Statistique

Printemps 2022

S. Mousset

Université Claude Bernard Lyon I – France

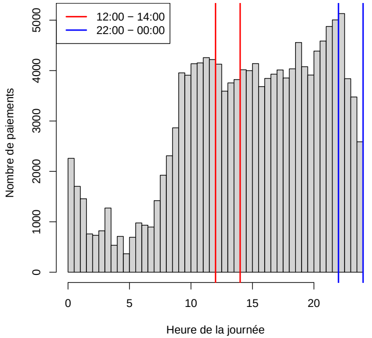
Un exemple : nombre d'événements par intervalle de temps

On dispose du fichier des paiements par carte de crédit sur deux jours ouvrables de septembre 2013 (données Kaggle.com).

- Heure du paiement (à la seconde près).
- Données sur l'achat.
- Caractère frauduleux ou non.
- Questions :
  - Sur un intervalle de temps, les achats se produisent-ils à vitesse constante (nombre d'achats / seconde) ?
  - Cela est-il valable en journée et le soir ?
- Vitesse constante λ événements par seconde $\Rightarrow \mathcal{P}(\lambda)$

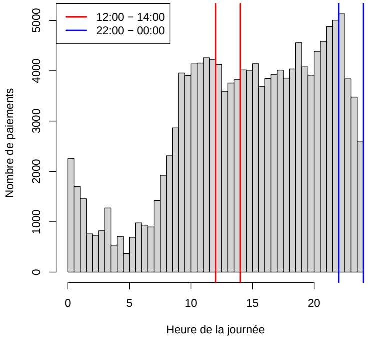
Test de conformité à la loi de Poisson

- Distribution des paiements (non frauduleux) sur une journée
- Deux intervalles : 12h00–14h00 et 22h00–24h00 (7200 secondes).



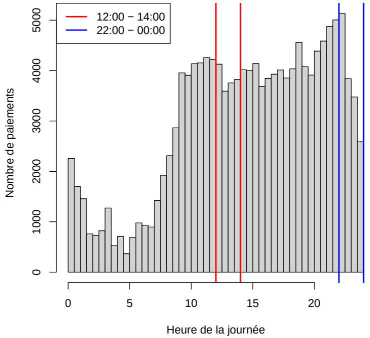
Test de conformité à la loi de Poisson

- Distribution des paiements (non frauduleux) sur une journée
- Deux intervalles : 12h00–14h00 et 22h00–24h00 (7200 secondes).
- On a l'impression que le taux est constant à midi et variable le soir.



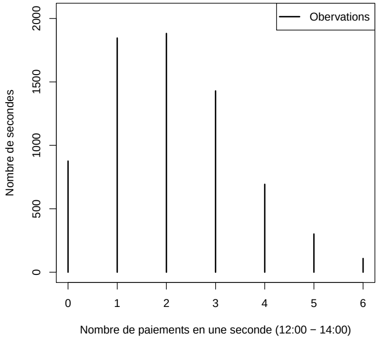
Test de conformité à la loi de Poisson

- Distribution des paiements (non frauduleux) sur une journée
- Deux intervalles : 12h00–14h00 et 22h00–24h00 (7200 secondes).
- On a l'impression que le taux est constant à midi et variable le soir.
- 12h00–14h00 : 15299 paiements 2.12 paiements/seconde.



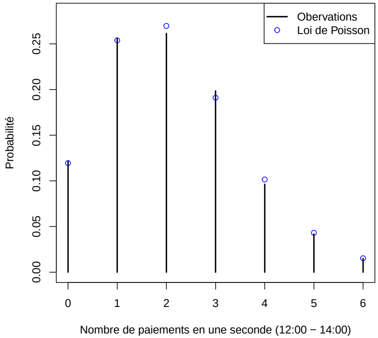
Test de conformité à la loi de Poisson

- Distribution des paiements (non frauduleux) sur une journée
- Deux intervalles : 12h00–14h00 et 22h00–24h00 (7200 secondes).
- On a l'impression que le taux est constant à midi et variable le soir.
- 12h00–14h00 : 15299 paiements 2.12 paiements/seconde.



Test de conformité à la loi de Poisson

- Distribution des paiements (non frauduleux) sur une journée
- Deux intervalles : 12h00–14h00 et 22h00–24h00 (7200 secondes).
- On a l'impression que le taux est constant à midi et variable le soir.
- 12h00–14h00 : 15299 paiements 2.12 paiements/seconde.
- Ajustement à $\mathcal{P}(\lambda = 2.12)$



Conformité à une distribution théorique à k classes

Classe	1	2	$\ldots$	i	$\ldots$	k	$\sum$
Effectif observé	N_1^{obs}	N_2^{obs}	$\ldots$	N_i^{obs}	$\ldots$	N_k^{obs}	N

Conformité à une distribution théorique à k classes

Classe Prob. selon H_0	1 p_1	2 p_2	$\ldots$ $\ldots$	i p_i	$\ldots$ $\ldots$	k p_k	$\sum$ 1
Effectif observé	N_1^{obs}	N_2^{obs}	$\ldots$	N_i^{obs}	$\ldots$	N_k^{obs}	N

Conformité à une distribution théorique à k classes

Classe	1	2	$\dots$	i	$\dots$	k	$\sum$
Prob. selon H_0	p_1	p_2	$\dots$	p_i	$\dots$	p_k	1
Effectif observé	N_1^{obs}	N_2^{obs}	$\dots$	N_i^{obs}	$\dots$	N_k^{obs}	N
Eff. selon H_0	$N_1^{H_0}$	$N_2^{H_0}$	$\dots$	$N_i^{H_0}$	$\dots$	$N_k^{H_0}$	N

Conformité à une distribution théorique à k classes

Classe Prob. selon H_0	1 p_1	2 p_2	$\ldots$ $\ldots$	i p_i	$\ldots$ $\ldots$	k p_k	$\sum$ 1
Effectif observé Eff. selon H_0	N_1^{obs} $N_1^{H_0}$	N_2^{obs} $N_2^{H_0}$	$\ldots$ $\ldots$	N_i^{obs} $N_i^{H_0}$	$\ldots$ $\ldots$	N_k^{obs} $N_k^{H_0}$	N N

Sous H_0, $N_i^{H_0} \sim \mathcal{B}(N, p_i)$

Conformité à une distribution théorique à k classes

Classe Prob. selon H_0	1 p_1	2 p_2	$\dots$ $\dots$	i p_i	$\dots$ $\dots$	k p_k	$\sum$ 1
Effectif observé Eff. selon H_0	N_1^{obs} $N_1^{H_0}$	N_2^{obs} $N_2^{H_0}$	$\dots$ $\dots$	N_i^{obs} $N_i^{H_0}$	$\dots$ $\dots$	N_k^{obs} $N_k^{H_0}$	N N

Sous H_0, $N_i^{H_0} \sim \mathcal{B}(N, p_i)$, si N est grand $\mathcal{B}(N, p_i) \rightarrow \mathcal{N}(N, Np_i(1-p_i))$.

Distribution d'échantillonnage sous H_0

$$N_i^{H_0} \sim \mathcal{B}(N, p_i) \to \mathcal{N}(N, Np_i(1-p_i)) \Rightarrow Z_i = \frac{N_i^{H_0} - Np_i}{\sqrt{Np_i(1-p_i)}} \sim \mathcal{N}(0,1)$$

Distribution d'échantillonnage sous H_0

$$N_i^{H_0} \sim \mathcal{B}(N, p_i) \to \mathcal{N}(N, Np_i(1-p_i)) \Rightarrow Z_i = \frac{N_i^{H_0} - Np_i}{\sqrt{Np_i(1-p_i)}} \sim \mathcal{N}(0,1)$$

- On s'intéresse au carré de l'écart à l'espérance $\mathbb{E}(N_i^{H_0}) = Np_i$

$$Z_i^2 = \frac{\left(N_i - \mathbb{E}(N_i^{H_0})\right)^2}{(1-p_i)\mathbb{E}(N_i^{H_0})} \sim \chi^2_{1\,\mathrm{ddl}}$$

Distribution d'échantillonnage sous H_0

$$N_i^{H_0} \sim \mathcal{B}(N, p_i) \to \mathcal{N}(N, Np_i(1-p_i)) \Rightarrow Z_i = \frac{N_i^{H_0} - Np_i}{\sqrt{Np_i(1-p_i)}} \sim \mathcal{N}(0,1)$$

- On s'intéresse au carré de l'écart à l'espérance $\mathbb{E}(N_i^{H_0}) = Np_i$

$$Z_i^2 = \frac{\left(N_i - \mathbb{E}(N_i^{H_0})\right)^2}{(1-p_i)\mathbb{E}(N_i^{H_0})} \sim \chi^2_{1\,\mathrm{ddl}}$$

- La loi $\chi^2_{k\,\mathrm{ddl}}$ est la loi suivie par la somme des carrés de k VA iid $\mathcal{N}(0,1)$.

Distribution d'échantillonnage sous H_0

- Le carré de l'écart relatif observé à l'espérance pour la classe i est

$$X_i^2 = (1 - p_i) Z_i^2 = \frac{(N_i - Np_i)^2}{Np_i}$$

Distribution d'échantillonnage sous H_0

- Le carré de l'écart relatif observé à l'espérance pour la classe i est

$$X_i^2 = (1 - p_i) Z_i^2 = \frac{(N_i - Np_i)^2}{Np_i}$$

- On s'intéresse à la somme des carrés des écarts relatifs sur toutes les classes

$$X^2 = \sum_{i=1}^{k} X_i^2 = \sum_{i=1}^{k} \frac{(N_i - Np_i)^2}{Np_i} \sim \chi^2_{k-1\,\mathrm{ddl}}$$

Distribution d'échantillonnage sous H_0

- Le carré de l'écart relatif observé à l'espérance pour la classe i est

$$X_i^2 = (1 - p_i) Z_i^2 = \frac{(N_i - Np_i)^2}{Np_i}$$

- On s'intéresse à la somme des carrés des écarts relatifs sur toutes les classes

$$X^2 = \sum_{i=1}^{k} X_i^2 = \sum_{i=1}^{k} \frac{(N_i - Np_i)^2}{Np_i} \sim \chi^2_{k-1\,\mathrm{ddl}}$$

- Un degré de liberté est perdu pour chaque paramètre déduit des observations (ici, $N = \sum N_i$ est fixé).

Distribution d'échantillonnage sous H_0

- Le carré de l'écart relatif observé à l'espérance pour la classe i est

$$X_i^2 = (1 - p_i) Z_i^2 = \frac{(N_i - N p_i)^2}{N p_i}$$

- On s'intéresse à la somme des carrés des écarts relatifs sur toutes les classes

$$X^2 = \sum_{i=1}^{k} X_i^2 = \sum_{i=1}^{k} \frac{(N_i - N p_i)^2}{N p_i} \sim \chi^2_{k-1\,\mathrm{ddl}}$$

- Un degré de liberté est perdu pour chaque paramètre déduit des observations (ici, $N = \sum N_i$ est fixé).
- Conditions d'application : convergence vers la loi normale : $N p_i$ assez grands ($\geqslant 5$ ou $\geqslant 10$).

Les paiements par carte de crédit (12h00 – 14h00)

Nombre de paiements	0	1	2	3	4	5	$\geqslant 6$	$\sum$
Nombres de secondes	876	1846	1882	1429	693	301	173	7200

Les paiements par carte de crédit (12h00 – 14h00)

Nombre de paiements	0	1	2	3	4	5	$\geqslant 6$	$\sum$
Distribution théorique	p_0	p_1	p_2	p_3	p_4	p_5	$1-\sum_{i=0}^{5} p_i$	1
Nombres de secondes	876	1846	1882	1429	693	301	173	7200

Les paiements par carte de crédit (12h00 – 14h00)

Nombre de paiements Distribution théorique	0 p_0	1 p_1	2 p_2	3 p_3	4 p_4	5 p_5	$\geqslant 6$ $1-\sum_{i=0}^{5} p_i$	$\sum$ 1
Nombres de secondes	876	1846	1882	1429	693	301	173	7200

- 15299 paiements en 7200 secondes $\lambda = 2.12$ paiements/seconde.

Les paiements par carte de crédit (12h00 – 14h00)

Nombre de paiements	0	1	2	3	4	5	$\geqslant 6$	$\sum$
Distribution théorique	p_0	p_1	p_2	p_3	p_4	p_5	$1-\sum_{i=0}^{5} p_i$	1
Nombres de secondes	876	1846	1882	1429	693	301	173	7200

- 15299 paiements en 7200 secondes $\lambda = 2.12$ paiements/seconde.
- $p_i = \frac{\lambda^i}{i!} e^{-\lambda}$

Les paiements par carte de crédit (12h00 – 14h00)

Nombre de paiements $\mathcal{P}(\lambda = 2.12)$	0 0.119	1 0.254	2 0.27	3 0.191	4 0.101	5 0.043	$\geqslant 6$ 0.022	$\sum$ 1
Nombres de secondes	876	1846	1882	1429	693	301	173	7200

- 15299 paiements en 7200 secondes $\lambda = 2.12$ paiements/seconde.
- $p_i = \frac{\lambda^i}{i!} e^{-\lambda}$

Les paiements par carte de crédit (12h00 – 14h00)

Nombre de paiements	0	1	2	3	4	5	$\geqslant 6$	$\sum$
$\mathcal{P}(\lambda = 2.12)$	0.119	0.254	0.27	0.191	0.101	0.043	0.022	1
Nombres de secondes	876	1846	1882	1429	693	301	173	7200
Attendu théorique Np_i	860	1827	1942	1375	731	310	155	7200

- 15299 paiements en 7200 secondes $\lambda = 2.12$ paiements/seconde.
- $p_i = \frac{\lambda^i}{i!} e^{-\lambda}$

Les paiements par carte de crédit (12h00 – 14h00)

Nombre de paiements $\mathcal{P}(\lambda = 2.12)$	0 0.119	1 0.254	2 0.27	3 0.191	4 0.101	5 0.043	$\geqslant 6$ 0.022	$\sum$ 1
Nombres de secondes Attendu théorique Np_i	876 860	1846 1827	1882 1942	1429 1375	693 731	301 310	173 155	7200 7200

$$X^2 = \frac{(876-860)^2}{860} +$$

Les paiements par carte de crédit (12h00 – 14h00)

Nombre de paiements	0	1	2	3	4	5	$\geqslant 6$	$\sum$
$\mathcal{P}(\lambda = 2.12)$	0.119	0.254	0.27	0.191	0.101	0.043	0.022	1
Nombres de secondes	876	1846	1882	1429	693	301	173	7200
Attendu théorique Np_i	860	1827	1942	1375	731	310	155	7200

$$X^2 = \frac{(876-860)^2}{860} + \frac{(1846-1827)^2}{1827} +$$

Les paiements par carte de crédit (12h00 – 14h00)

Nombre de paiements $\mathcal{P}(\lambda = 2.12)$	0 0.119	1 0.254	2 0.27	3 0.191	4 0.101	5 0.043	$\geqslant 6$ 0.022	$\sum$ 1
Nombres de secondes Attendu théorique Np_i	876 860	1846 1827	1882 1942	1429 1375	693 731	301 310	173 155	7200 7200

$$X^2 = \frac{(876 - 860)^2}{860} + \frac{(1846 - 1827)^2}{1827} + \ldots + \frac{(173 - 155)^2}{155}$$

Les paiements par carte de crédit (12h00 – 14h00)

Nombre de paiements	0	1	2	3	4	5	$\geqslant 6$	$\sum$
$\mathcal{P}(\lambda = 2.12)$	0.119	0.254	0.27	0.191	0.101	0.043	0.022	1
Nombres de secondes	876	1846	1882	1429	693	301	173	7200
Attendu théorique Np_i	860	1827	1942	1375	731	310	155	7200

$$X^2 = \frac{(876-860)^2}{860} + \frac{(1846-1827)^2}{1827} + \ldots + \frac{(173-155)^2}{155} = 8.77$$

Les paiements par carte de crédit (12h00 – 14h00)

Nombre de paiements $\mathcal{P}(\lambda = 2.12)$	0 0.119	1 0.254	2 0.27	3 0.191	4 0.101	5 0.043	$\geqslant 6$ 0.022	$\sum$ 1
Nombres de secondes Attendu théorique Np_i	876 860	1846 1827	1882 1942	1429 1375	693 731	301 310	173 155	7200 7200

$$X^2 = \frac{(876-860)^2}{860} + \frac{(1846-1827)^2}{1827} + \ldots + \frac{(173-155)^2}{155} = 8.77$$

7 classes, 2 ddl perdus (N et λ), on compare X^2 à $\chi^2_{\alpha=0.05,\,5\text{ ddl}} = 11.07$

Test χ^2 de conformité : déroulement

1. On reconnait un test χ^2 de conformité.

Test χ^2 de conformité : déroulement

1. On reconnait un test χ^2 de conformité.
2. Conditions d'application : effectif total supérieur à 50, tous les effectifs attendus théoriques supérieurs à 5 (ou 10)

Test χ^2 de conformité : déroulement

1. On reconnait un test χ^2 de conformité.
2. Conditions d'application : effectif total supérieur à 50, tous les effectifs attendus théoriques supérieurs à 5 (ou 10)
3. On énonce les hypothèses en concurrence :
   H_0 : le nombre de paiements en une seconde suit une loi de Poisson.
   H_1 : le nombre de paiements en une seconde ne suit pas une loi de Poisson.

Test χ^2 de conformité : déroulement

1. On reconnait un test χ^2 de conformité.
2. Conditions d'application : effectif total supérieur à 50, tous les effectifs attendus théoriques supérieurs à 5 (ou 10)
3. On énonce les hypothèses en concurrence :
   H_0 : le nombre de paiements en une seconde suit une loi de Poisson.
   H_1 : le nombre de paiements en une seconde ne suit pas une loi de Poisson.
4. On calcule les paramètres éventuels de la loi ($\lambda = 2.1248611$), les probabilités p_i de chaque classe, les effectifs attendus Np_i de chaque classe (éventuellement regrouper des classes pour satisfaire les C.A.), puis la statistique $X^2 = \sum_{i=1}^{k} \frac{(N_i - Np_i)^2}{Np_i} = 8.76579$

Test χ^2 de conformité : déroulement

1. On reconnait un test χ^2 de conformité.
2. Conditions d'application : effectif total supérieur à 50, tous les effectifs attendus théoriques supérieurs à 5 (ou 10)
3. On énonce les hypothèses en concurrence :
   H_0 : le nombre de paiements en une seconde suit une loi de Poisson.
   H_1 : le nombre de paiements en une seconde ne suit pas une loi de Poisson.
4. On calcule les paramètres éventuels de la loi ($\lambda = 2.1248611$), les probabilités p_i de chaque classe, les effectifs attendus Np_i de chaque classe (éventuellement regrouper des classes pour satisfaire les C.A.), puis la statistique $X^2 = \sum_{i=1}^{k} \frac{(N_i - Np_i)^2}{Np_i} = 8.76579$
5. Le nombre de ddl est $k - 1 - n_{\text{par}} = 5$ où $k = 7$ est le nombre de classes et $n_{\text{par}} = 1$ le nombre de paramètres estimés d'après l'observation. Comparer $X^2 = 8.76579$ à $\chi^2_{\alpha=0.05,\,5\text{ ddl}} = 11.07$ lu dans la table de χ^2.

Test χ^2 de conformité : déroulement

1. On reconnait un test χ^2 de conformité.
2. Conditions d'application : effectif total supérieur à 50, tous les effectifs attendus théoriques supérieurs à 5 (ou 10)
3. On énonce les hypothèses en concurrence :
   H_0 : le nombre de paiements en une seconde suit une loi de Poisson.
   H_1 : le nombre de paiements en une seconde ne suit pas une loi de Poisson.
4. On calcule les paramètres éventuels de la loi ($\lambda = 2.1248611$), les probabilités p_i de chaque classe, les effectifs attendus Np_i de chaque classe (éventuellement regrouper des classes pour satisfaire les C.A.), puis la statistique $X^2 = \sum_{i=1}^{k} \frac{(N_i - Np_i)^2}{Np_i} = 8.76579$
5. Le nombre de ddl est $k - 1 - n_{\text{par}} = 5$ où $k = 7$ est le nombre de classes et $n_{\text{par}} = 1$ le nombre de paramètres estimés d'après l'observation. Comparer $X^2 = 8.76579$ à $\chi^2_{\alpha=0.05,\,5\text{ ddl}} = 11.07$ lu dans la table de χ^2.
6. On a $X^2 < \chi^2_{\alpha=0.05,\,5\text{ ddl}} = 11.07$, donc on ne peut pas rejeter H_0 avec un risque $\alpha = 5\%$. On accepte H_0 avec un rique β inconnu : la distribution est Poissonienne, donc les paiements par carte de crédit se font à vitesse moyenne constante entre 12h00 et 14h00.

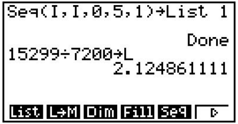
Avec vos calculatrices casio

Liste des valeurs des classes enregistrée dans la **Liste 1**

OPTN F1 F5 ALPHA (, ALPHA (, 0 , 5 , 1) → SHIFT 1 1 EXE

Enregistrer la valeur de lambda dans la variable **L**

1 5 2 9 9 ÷ 7 2 0 0 → ALPHA → EXE



Calcul des probabilités des classes qvec la loi de Poisson enregistrées dans la **Liste 2**

OPTN F5 F3 F6 F1 F1 SHIFT 1 1 , ALPHA →) → SHIFT 1 2 EXE

Calcul de la probabilité de la dernière classe enregistrée dans la variable **M**

1 − F1 F6 F6 F1 SHIFT 1 2 → ALPHA 7 EXE

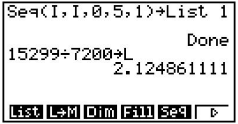
Avec vos calculatrices casio

Liste des valeurs des classes enregistrée dans la **Liste 1**

[OPTN] [F1] [F5] [ALPHA] [(] [,] [ALPHA] [(] [,] [0] [,]
[5] [,] [1] [)] [→] [SHIFT] [1] [1] [EXE]

Enregistrer la valeur de lambda dans la variable **L**

[1] [5] [2] [9] [9] [÷] [7] [2] [0] [0] [→]
[ALPHA] [→] [EXE]

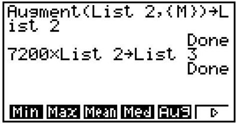


Ajout de cette probabilité en fin de **Liste 2**

[OPTN] [F1] [F6] [F5] [SHIFT] [1] [2] [,] [SHIFT] [×] [ALPHA]
[7] [SHIFT] [÷] [)] [→] [SHIFT] [1] [2] [EXE]

Calcul des attendus théoriques enregistrés dans la **Liste 3**

[7] [2] [0] [0] [×] [SHIFT] [1] [2] [→] [SHIFT] [1]
[3] [EXE]



Calcul des probabilités des classes qvec la loi de Poisson enregistrées dans la **Liste 2**

[OPTN] [F5] [F3] [F6] [F1] [F1] [SHIFT] [1] [1] [,] [ALPHA]
[→] [)] [→] [SHIFT] [1] [2] [EXE]

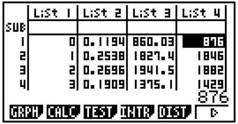
Calcul de la probabilité de la dernière classe enregistrée dans la variable **M**

[1] [−] [F1] [F6] [F6] [F1] [SHIFT] [1] [2] [→] [ALPHA]
[7] [EXE]

Entrer les effectifs observés dans la **Liste 4**

Effectuer le test de Chi2 de conformité

[MENU] [2] [F3] [F3] [F1] [F1] [4] [EXE] [▼] [F1] [3]
[EXE] [▼] [5] [EXE] [F1] [5] [EXE] [▼] [F2] [6] [EXE]



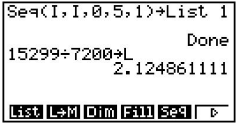
Avec vos calculatrices casio

Liste des valeurs des classes enregistrée dans la **Liste 1**

[OPTN] [F1] [F5] [ALPHA] [(] [,] [ALPHA] [(] [,] [0] [,] [5] [,] [1] [)] [→] [SHIFT] [1] [1] [EXE]

Enregistrer la valeur de lambda dans la variable **L**

[1] [5] [2] [9] [9] [÷] [7] [2] [0] [0] [→] [ALPHA] [→] [EXE]



Calcul des probabilités des classes qvec la loi de Poisson enregistrées dans la **Liste 2**

[OPTN] [F5] [F3] [F6] [F1] [F1] [SHIFT] [1] [1] [,] [ALPHA] [→] [)] [→] [SHIFT] [1] [2] [EXE]

Calcul de la probabilité de la dernière classe enregistrée dans la variable **M**

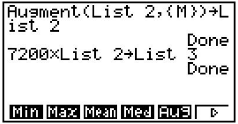
[1] [−] [F1] [F6] [F6] [F1] [SHIFT] [1] [2] [→] [ALPHA] [7] [EXE]

Ajout de cette probabilité en fin de **Liste 2**

[OPTN] [F1] [F6] [F5] [SHIFT] [1] [2] [,] [SHIFT] [×] [ALPHA] [7] [SHIFT] [÷] [)] [→] [SHIFT] [1] [2] [EXE]

Calcul des attendus théoriques enregistrés dans la **Liste 3**

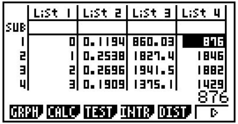
[7] [2] [0] [0] [×] [SHIFT] [1] [2] [→] [SHIFT] [1] [3] [EXE]



Entrer les effectifs observés dans la **Liste 4**

Effectuer le test de Chi2 de conformité

[MENU] [2] [F3] [F3] [F1] [F1] [4] [EXE] [▼] [F1] [3] [EXE] [▼] [5] [EXE] [F1] [5] [EXE] [▼] [F2] [6] [EXE]



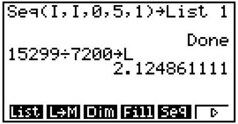
Avec vos calculatrices casio

Liste des valeurs des classes enregistrée dans la **Liste 1**

OPTN F1 F5 ALPHA (, ALPHA (, 0 , 5 , 1) → SHIFT 1 1 EXE

Enregistrer la valeur de lambda dans la variable **L**

1 5 2 9 9 ÷ 7 2 0 0 → ALPHA → EXE



Calcul des probabilités des classes qvec la loi de Poisson enregistrées dans la **Liste 2**

OPTN F5 F3 F6 F1 F1 SHIFT 1 1 , ALPHA →) → SHIFT 1 2 EXE

Calcul de la probabilité de la dernière classe enregistrée dans la variable **M**

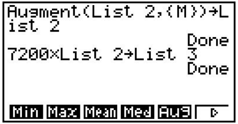
1 − F1 F6 F6 F1 SHIFT 1 2 → ALPHA 7 EXE

Ajout de cette probabilité en fin de **Liste 2**

OPTN F1 F6 F5 SHIFT 1 2 , SHIFT × ALPHA 7 SHIFT ÷) → SHIFT 1 2 EXE

Calcul des attendus théoriques enregistrés dans la **Liste 3**

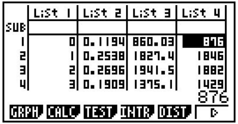
7 2 0 0 × SHIFT 1 2 → SHIFT 1 3 EXE

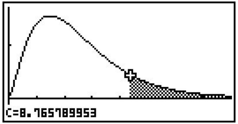


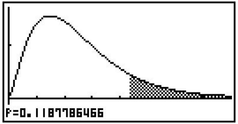
Entrer les effectifs observés dans la **Liste 4**

Effectuer le test de Chi2 de conformité

MENU 2 F3 F3 F1 F1 4 EXE ▼ F1 3 EXE ▼ 5 EXE F1 5 EXE ▼ F2 6 EXE

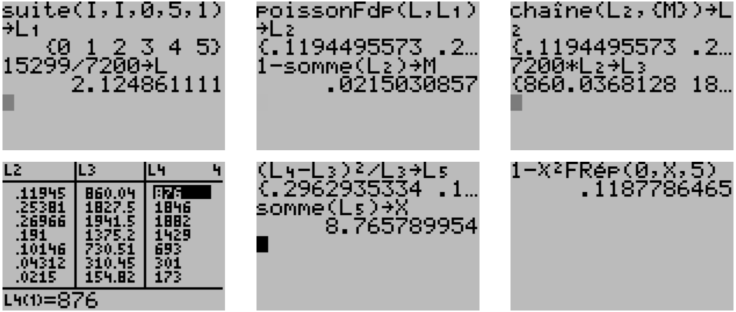






Avec vos calculatrices TI

- Pas de fonction dédiée
- Utilisation des listes
- Calcul “à la main”



Les paiements par carte de crédit (22h00 – 00h00)

Nombre de paiements	0	1	2	3	4	5	$\geqslant 6$	$\sum$
Nombres de secondes	1040	1906	1805	1203	681	331	234	7200

Les paiements par carte de crédit (22h00 – 00h00)

Nombre de paiements	0	1	2	3	4	5	$\geqslant 6$	$\sum$
Distribution théorique	p_0	p_1	p_2	p_3	p_4	p_5	$1 - \sum_{i=0}^{5} p_i$	1
Nombres de secondes	1040	1906	1805	1203	681	331	234	7200

Les paiements par carte de crédit (22h00 – 00h00)

Nombre de paiements	0	1	2	3	4	5	$\geqslant 6$	$\sum$
Distribution théorique	p_0	p_1	p_2	p_3	p_4	p_5	$1 - \sum_{i=0}^{5} p_i$	1
Nombres de secondes	1040	1906	1805	1203	681	331	234	7200

- 15039 paiements en 7200 secondes $\lambda = 2.09$ paiements/seconde.

Les paiements par carte de crédit (22h00 – 00h00)

Nombre de paiements Distribution théorique	0 p_0	1 p_1	2 p_2	3 p_3	4 p_4	5 p_5	$\geqslant 6$ $1-\sum_{i=0}^{5} p_i$	$\sum$ 1
Nombres de secondes	1040	1906	1805	1203	681	331	234	7200

- 15039 paiements en 7200 secondes $\lambda = 2.09$ paiements/seconde.
- $p_i = \frac{\lambda^i}{i!} e^{-\lambda}$

Les paiements par carte de crédit (22h00 – 00h00)

Nombre de paiements $\mathcal{P}(\lambda = 2.09)$	0 0.124	1 0.259	2 0.27	3 0.188	4 0.098	5 0.041	$\geqslant 6$ 0.02	$\sum$ 1
Nombres de secondes	1040	1906	1805	1203	681	331	234	7200

- 15039 paiements en 7200 secondes $\lambda = 2.09$ paiements/seconde.
- $p_i = \frac{\lambda^i}{i!} e^{-\lambda}$

Les paiements par carte de crédit (22h00 – 00h00)

Nombre de paiements $\mathcal{P}(\lambda = 2.09)$	0 0.124	1 0.259	2 0.27	3 0.188	4 0.098	5 0.041	$\geqslant 6$ 0.02	$\sum$ 1
Nombres de secondes Attendu théorique Np_i	1040 892	1906 1862	1805 1945	1203 1354	681 707	331 295	234 144	7200 7200

- 15039 paiements en 7200 secondes $\lambda = 2.09$ paiements/seconde.
- $p_i = \frac{\lambda^i}{i!} e^{-\lambda}$

Les paiements par carte de crédit (22h00 – 00h00)

Nombre de paiements	0	1	2	3	4	5	$\geqslant 6$	$\sum$
$\mathcal{P}(\lambda = 2.09)$	0.124	0.259	0.27	0.188	0.098	0.041	0.02	1
Nombres de secondes	1040	1906	1805	1203	681	331	234	7200
Attendu théorique Np_i	892	1862	1945	1354	707	295	144	7200

$$X^2 = \frac{(1040 - 892)^2}{892} +$$

Les paiements par carte de crédit (22h00 – 00h00)

Nombre de paiements	0	1	2	3	4	5	$\geqslant 6$	$\sum$
$\mathcal{P}(\lambda = 2.09)$	0.124	0.259	0.27	0.188	0.098	0.041	0.02	1
Nombres de secondes	1040	1906	1805	1203	681	331	234	7200
Attendu théorique Np_i	892	1862	1945	1354	707	295	144	7200

$$X^2 = \frac{(1040 - 892)^2}{892} + \frac{(1906 - 1862)^2}{1862} +$$

Les paiements par carte de crédit (22h00 – 00h00)

Nombre de paiements $\mathcal{P}(\lambda = 2.09)$	0 0.124	1 0.259	2 0.27	3 0.188	4 0.098	5 0.041	$\geqslant 6$ 0.02	$\sum$ 1
Nombres de secondes Attendu théorique Np_i	1040 892	1906 1862	1805 1945	1203 1354	681 707	331 295	234 144	7200 7200

$$X^2 = \frac{(1040 - 892)^2}{892} + \frac{(1906 - 1862)^2}{1862} + \ldots + \frac{(234 - 144)^2}{144}$$

Les paiements par carte de crédit (22h00 – 00h00)

Nombre de paiements	0	1	2	3	4	5	$\geqslant 6$	$\sum$
$\mathcal{P}(\lambda = 2.09)$	0.124	0.259	0.27	0.188	0.098	0.041	0.02	1
Nombres de secondes	1040	1906	1805	1203	681	331	234	7200
Attendu théorique Np_i	892	1862	1945	1354	707	295	144	7200

$$X^2 = \frac{(1040 - 892)^2}{892} + \frac{(1906 - 1862)^2}{1862} + \ldots + \frac{(234 - 144)^2}{144} = 114$$

Les paiements par carte de crédit (22h00 – 00h00)

Nombre de paiements $\mathcal{P}(\lambda = 2.09)$	0 0.124	1 0.259	2 0.27	3 0.188	4 0.098	5 0.041	$\geqslant 6$ 0.02	$\sum$ 1
Nombres de secondes Attendu théorique Np_i	1040 892	1906 1862	1805 1945	1203 1354	681 707	331 295	234 144	7200 7200

$$X^2 = \frac{(1040 - 892)^2}{892} + \frac{(1906 - 1862)^2}{1862} + \ldots + \frac{(234 - 144)^2}{144} = 114$$

7 classes, 2 ddl perdus (N et λ), on compare X^2 à $\chi^2_{\alpha=0.05,\, 5 \text{ ddl}} = 11.07$

Déroulement du test

1. On reconnait un test χ^2 de conformité.

Déroulement du test

1. On reconnait un test χ^2 de conformité.
2. Conditions d'application : effectif total supérieur à 50, tous les effectifs attendus théoriques supérieurs à 5 (ou 10)

Déroulement du test

1. On reconnait un test χ^2 de conformité.
2. Conditions d'application : effectif total supérieur à 50, tous les effectifs attendus théoriques supérieurs à 5 (ou 10)
3. On énonce les hypothèses en concurrence :
   H_0 : le nombre de paiements en une seconde suit une loi de Poisson.
   H_1 : le nombre de paiements en une seconde ne suit pas une loi de Poisson.

Déroulement du test

1. On reconnait un test χ^2 de conformité.
2. Conditions d'application : effectif total supérieur à 50, tous les effectifs attendus théoriques supérieurs à 5 (ou 10)
3. On énonce les hypothèses en concurrence :
   H_0 : le nombre de paiements en une seconde suit une loi de Poisson.
   H_1 : le nombre de paiements en une seconde ne suit pas une loi de Poisson.
4. On calcule les paramètres éventuels de la loi ($\lambda = 2.08875$), les probabilités p_i de chaque classe, les effectifs attendus Np_i de chaque classe (éventuellement regrouper des classes pour satisfaire les C.A.), puis la statistique $X^2 = \sum_{i=1}^{k} \frac{(N_i - Np_i)^2}{Np_i} = 114.3800314$

Déroulement du test

1. On reconnait un test χ^2 de conformité.
2. Conditions d'application : effectif total supérieur à 50, tous les effectifs attendus théoriques supérieurs à 5 (ou 10)
3. On énonce les hypothèses en concurrence :
   H_0 : le nombre de paiements en une seconde suit une loi de Poisson.
   H_1 : le nombre de paiements en une seconde ne suit pas une loi de Poisson.
4. On calcule les paramètres éventuels de la loi ($\lambda = 2.08875$), les probabilités p_i de chaque classe, les effectifs attendus Np_i de chaque classe (éventuellement regrouper des classes pour satisfaire les C.A.), puis la statistique $X^2 = \sum_{i=1}^{k} \frac{(N_i - Np_i)^2}{Np_i} = 114.3800314$
5. Le nombre de ddl est $k - 1 - n_{\text{par}} = 5$ où $k = 7$ est le nombre de classes et $n_{\text{par}} = 1$ le nombre de paramètres estimés d'après l'observation. Comparer $X^2 = 114.3800314$ à $\chi^2_{\alpha=0.05,\, 5 \text{ ddl}} = 11.07$ lu dans la table de χ^2.

Déroulement du test

1. On reconnait un test χ^2 de conformité.
2. Conditions d'application : effectif total supérieur à 50, tous les effectifs attendus théoriques supérieurs à 5 (ou 10)
3. On énonce les hypothèses en concurrence :
   H_0 : le nombre de paiements en une seconde suit une loi de Poisson.
   H_1 : le nombre de paiements en une seconde ne suit pas une loi de Poisson.
4. On calcule les paramètres éventuels de la loi ($\lambda = 2.08875$), les probabilités p_i de chaque classe, les effectifs attendus Np_i de chaque classe (éventuellement regrouper des classes pour satisfaire les C.A.), puis la statistique $X^2 = \sum_{i=1}^{k} \frac{(N_i - Np_i)^2}{Np_i} = 114.3800314$
5. Le nombre de ddl est $k - 1 - n_{\text{par}} = 5$ où $k = 7$ est le nombre de classes et $n_{\text{par}} = 1$ le nombre de paramètres estimés d'après l'observation. Comparer $X^2 = 114.3800314$ à $\chi^2_{\alpha=0.05,\, 5 \text{ ddl}} = 11.07$ lu dans la table de χ^2.
6. On a $X^2 > \chi^2_{\alpha=0.05,\, 5 \text{ ddl}} = 11.07$, donc on peut rejeter H_0 avec un risque $\alpha = 5\%$. La distribution n'est pas Poissonienne, donc les paiements par carte de crédit se font à des vitesse hétérogènes entre 22h00 et 00h00.

Un exemple de loi continue : la masse des grillons mâles

- Étude de 2015 de C.D. Kelly (université du Québec à Montréal)
- Sélection sexuelle : les femelles choisiraient les mâles ayant les meilleurs gènes.
- But : mettre en évidence une qualité (meilleure réponse immunitaire) chez les mâles préférés par les femelles.
- Données morphologiques sur 54 grillons, descendants élevés en laboratoire d'individus capturés en 2012 et 2013.
- Résultat de l'étude : pas de différence mise en évidence.

Un exemple de loi continue : la masse des grillons mâles

- Étude de 2015 de C.D. Kelly (université du Québec à Montréal)
- Sélection sexuelle : les femelles choisiraient les mâles ayant les meilleurs gènes.
- But : mettre en évidence une qualité (meilleure réponse immunitaire) chez les mâles préférés par les femelles.
- Données morphologiques sur 54 grillons, descendants élevés en laboratoire d'individus capturés en 2012 et 2013.
- Résultat de l'étude : pas de différence mise en évidence.

- **Masse** des grillons

Un exemple de loi continue : la masse des grillons mâles

- Étude de 2015 de C.D. Kelly (université du Québec à Montréal)
- Sélection sexuelle : les femelles choisiraient les mâles ayant les meilleurs gènes.
- But : mettre en évidence une qualité (meilleure réponse immunitaire) chez les mâles préférés par les femelles.
- Données morphologiques sur 54 grillons, descendants élevés en laboratoire d'individus capturés en 2012 et 2013.
- Résultat de l'étude : pas de différence mise en évidence.

- **Masse** des grillons
- **Longueur** du pronotum.

Un exemple de loi continue : la masse des grillons mâles

- Étude de 2015 de C.D. Kelly (université du Québec à Montréal)
- Sélection sexuelle : les femelles choisiraient les mâles ayant les meilleurs gènes.
- But : mettre en évidence une qualité (meilleure réponse immunitaire) chez les mâles préférés par les femelles.
- Données morphologiques sur 54 grillons, descendants élevés en laboratoire d'individus capturés en 2012 et 2013.
- Résultat de l'étude : pas de différence mise en évidence.

- **Masse** des grillons
- **Longueur** du pronotum.

Ces grandeurs suivent-elles une loi normale ?

Ajustement à une loi normale

- On dispose des 54 valeurs brutes

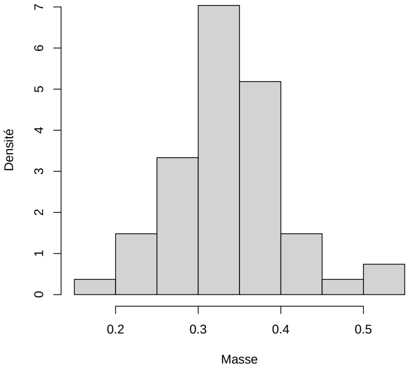
Valeurs observées (ordre croissant) :

0.1952, 0.2102, 0.2248, 0.2266, 0.2487, 0.2632,
0.2702, 0.2757, 0.2804, 0.2827, 0.2871, 0.2879,
0.2957, 0.2961, 0.3027, 0.3044, 0.305, 0.3091,
0.314, 0.3182, 0.3217, 0.3227, 0.3251, 0.3255,
0.3298, 0.3302, 0.3303, 0.3314, 0.3323, 0.3344,
0.3388, 0.3429, 0.3465, 0.3548, 0.3676, 0.3704,
0.3738, 0.3754, 0.3769, 0.3779, 0.3787, 0.3851,
0.3874, 0.3909, 0.3938, 0.3944, 0.3948, 0.4136,
0.4139, 0.4156, 0.4322, 0.4727, 0.5345, 0.5497

$$\sum x = 18.4 \quad \sum x^2 = 6.5$$

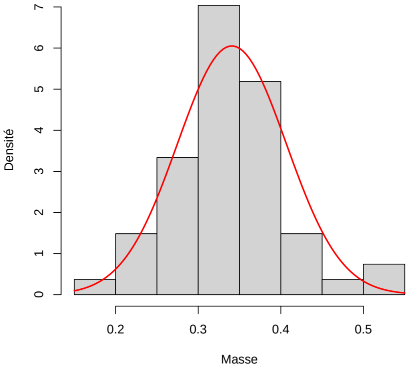
Ajustement à une loi normale

- On dispose des 54 valeurs brutes
- On veut appliquer un test χ^2 d'ajustement à une loi normale



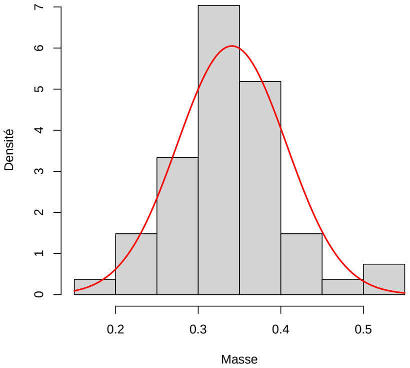
Ajustement à une loi normale

- On dispose des 54 valeurs brutes
- On veut appliquer un test χ^2 d'ajustement à une loi normale
- C.A. $n \geq 50$, $\mathbb{E}(N_i) \geqslant 10$



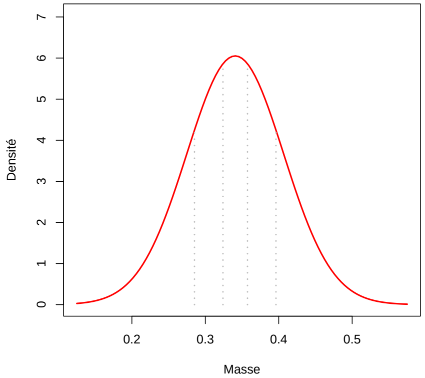
Ajustement à une loi normale

- On dispose des 54 valeurs brutes
- On veut appliquer un test χ^2 d'ajustement à une loi normale
- C.A. $n \geq 50$, $\mathbb{E}(N_i) \geqslant 10$
- On veut $\sim \frac{1}{5}$ des effectifs dans chaque classe.



Ajustement à une loi normale

- On dispose des 54 valeurs brutes
- On veut appliquer un test χ^2 d'ajustement à une loi normale
- C.A. $n \geq 50$, $\mathbb{E}(N_i) \geqslant 10$
- On veut $\sim \frac{1}{5}$ des effectifs dans chaque classe.
- On utilise les quantiles de la loi normale $\mathcal{N}(\bar{x}, \hat{\sigma}^2)$.

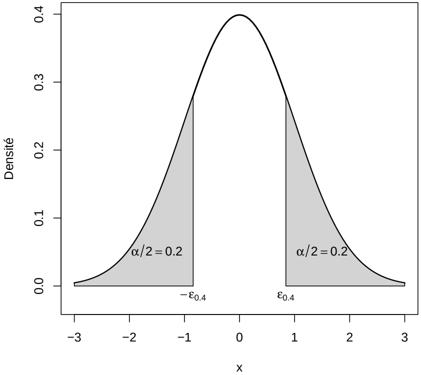


Quantiles de $\mathcal{N}(0,1)$

- On cherche les 20$^{\text{ème}}$, 40$^{\text{ème}}$, 60$^{\text{ème}}$ et 80$^{\text{ème}}$ percentiles de $\mathcal{N}(0,1)$.

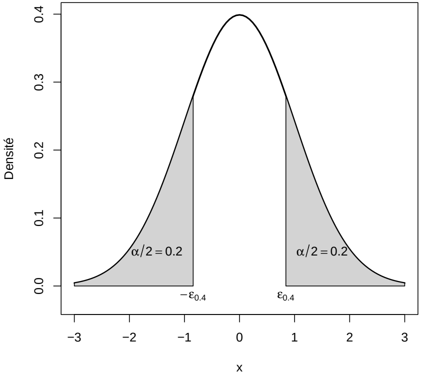
Quantiles de $\mathcal{N}(0, 1)$

- On cherche les $20^{\text{ème}}$, $40^{\text{ème}}$, $60^{\text{ème}}$ et $80^{\text{ème}}$ percentiles de $\mathcal{N}(0, 1)$.
- $q_{0.2} | \mathbb{P}(Z \leqslant q_{0.2}) = 0.20$



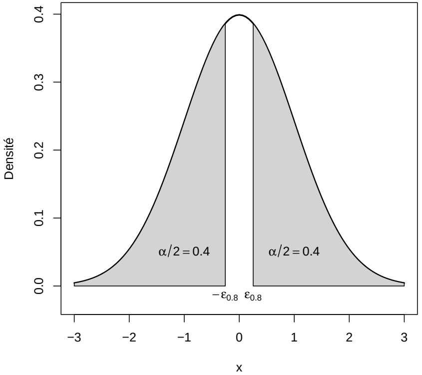
Quantiles de $\mathcal{N}(0,1)$

- On cherche les 20ème, 40ème, 60ème et 80ème percentiles de $\mathcal{N}(0,1)$.
- $q_{0.2} | \mathbb{P}(Z \leqslant q_{0.2}) = 0.20$
  $q_{0.2} = -\varepsilon_{\alpha=0.4}$, $q_{0.8} = \varepsilon_{\alpha=0.4}$



Quantiles de $\mathcal{N}(0,1)$

- On cherche les 20ème, 40ème, 60ème et 80ème percentiles de $\mathcal{N}(0,1)$.
- $q_{0.2} | \mathbb{P}(Z \leqslant q_{0.2}) = 0.20$
  $q_{0.2} = -\varepsilon_{\alpha=0.4}$, $q_{0.8} = \varepsilon_{\alpha=0.4}$
  $q_{0.4} = -\varepsilon_{\alpha=0.8}$, $q_{0.6} = \varepsilon_{\alpha=0.8}$



Quantiles de $\mathcal{N}(0,1)$

- On cherche les 20$^{\text{ème}}$, 40$^{\text{ème}}$, 60$^{\text{ème}}$ et 80$^{\text{ème}}$ percentiles de $\mathcal{N}(0,1)$.
- $q_{0.2} | \mathbb{P}(Z \leqslant q_{0.2}) = 0.20$
  $q_{0.2} = -\varepsilon_{\alpha=0.4}$, $q_{0.8} = \varepsilon_{\alpha=0.4}$
  $q_{0.4} = -\varepsilon_{\alpha=0.8}$, $q_{0.6} = \varepsilon_{\alpha=0.8}$

Quantile	0.2	0.4	0.6	0.8
$\mathcal{N}(0,1)$	-0.842	-0.253	0.253	0.842

Loi normale : table de l'écart-réduit.

La table donne, pour une probabilité α, la valeur ε telle que la probabilité que l'écart-réduit égale ou dépasse en valeur absolue ε vaut α.

α	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	∞	2.576	2.326	2.170	2.054	1.960	1.881	1.812	1.751	1.695
0.10	1.645	1.598	1.555	1.514	1.476	1.440	1.405	1.372	1.341	1.311
0.20	1.282	1.254	1.227	1.200	1.175	1.150	1.126	1.103	1.080	1.058
0.30	1.036	1.015	0.994	0.974	0.954	0.935	0.915	0.896	0.878	0.860
0.40	0.842	0.824	0.806	0.789	0.772	0.755	0.739	0.722	0.706	0.690
0.50	0.674	0.659	0.643	0.628	0.613	0.598	0.583	0.568	0.553	0.539
0.60	0.524	0.510	0.496	0.482	0.468	0.454	0.440	0.426	0.412	0.399
0.70	0.385	0.372	0.358	0.345	0.332	0.319	0.305	0.292	0.279	0.266
0.80	0.253	0.240	0.228	0.215	0.202	0.189	0.176	0.164	0.151	0.138
0.90	0.126	0.113	0.100	0.088	0.075	0.063	0.050	0.038	0.025	0.013

La probabilité α s'obtient par addition des nombres écrits en marge.
Exemple : pour $\alpha = 0.00 + 0.05 = 0.05$, $\varepsilon = 1.960$.

Quantiles de $\mathcal{N}(\bar{x}, \hat{\sigma}^2)$

n	$\sum x$	$\sum x^2$	$\bar{x}$	s^2	$\hat{\sigma}^2$	$\hat{\sigma}$
54	18.4	6.5	0.3407407	0.0042661	0.0043466	0.0659288

Quantiles de $\mathcal{N}(\bar{x}, \hat{\sigma}^2)$

n	$\sum x$	$\sum x^2$	$\bar{x}$	s^2	$\hat{\sigma}^2$	$\hat{\sigma}$
54	18.4	6.5	0.3407407	0.0042661	0.0043466	0.0659288

On déduit les quantiles de $\mathcal{N}(\bar{x}, \hat{\sigma}^2)$:

$$Z = \frac{X - \bar{x}}{\hat{\sigma}} \quad \Leftrightarrow \quad X = \bar{x} + Z\hat{\sigma}$$

Quantile	0.2	0.4	0.6	0.8
$\mathcal{N}(0, 1)$	-0.842	-0.253	0.253	0.842
$\mathcal{N}(\bar{x}, \hat{\sigma}^2)$	0.285	0.324	0.357	0.396

Quantiles de $\mathcal{N}(\bar{x}, \hat{\sigma}^2)$

n	$\sum x$	$\sum x^2$	$\bar{x}$	s^2	$\hat{\sigma}^2$	$\hat{\sigma}$
54	18.4	6.5	0.3407407	0.0042661	0.0043466	0.0659288

On déduit les quantiles de $\mathcal{N}(\bar{x}, \hat{\sigma}^2)$:

$$Z = \frac{X - \bar{x}}{\hat{\sigma}} \quad \Leftrightarrow \quad X = \bar{x} + Z\hat{\sigma}$$

Quantile	0.2	0.4	0.6	0.8
$\mathcal{N}(0,1)$	-0.842	-0.253	0.253	0.842
$\mathcal{N}(\bar{x}, \hat{\sigma}^2)$	0.285	0.324	0.357	0.396

Les effectifs observés et attendus théoriques sous H_0 sont

Intervalle	$\leqslant 0.285$	$\leqslant 0.324$	$\leqslant 0.357$	$\leqslant 0.396$	> 0.396
Effectifs observés	10	12	12	13	7
Effectifs attendus	10.8	10.8	10.8	10.8	10.8

Déroulement du test

1. On applique un test χ^2 de conformité à une loi normale.

Déroulement du test

1. On applique un test χ^2 de conformité à une loi normale.
2. C.A. : effectif total supérieur à 50 et tous les effectifs attendus théoriques supérieurs à 5 (ou 10)

Déroulement du test

1. On applique un test χ^2 de conformité à une loi normale.
2. C.A. : effectif total supérieur à 50 et tous les effectifs attendus théoriques supérieurs à 5 (ou 10)
3. Les hypothèses testées sont :
   H_0 : la masse des grillons mâles suit une loi normale.
   H_1 : la masse des grillons mâles ne suit pas une loi normale.

Déroulement du test

1. On applique un test χ^2 de conformité à une loi normale.
2. C.A. : effectif total supérieur à 50 et tous les effectifs attendus théoriques supérieurs à 5 (ou 10)
3. Les hypothèses testées sont :
   H_0 : la masse des grillons mâles suit une loi normale.
   H_1 : la masse des grillons mâles ne suit pas une loi normale.
4. Les paramètres de la loi normale sont $\bar{x} = 0.3407407$ et $\hat{\sigma}^2 = 0.0043466$.
   On a choisi les bornes des intervalles pour que les classes soient équiprobables. Les effectifs attendus sont $\frac{54}{5} = 10.8$ pour chaque classe.
   $X^2 = \sum_{i=1}^{k} \frac{(N_i - 10.8)^2}{10.8} = 2.111.$

Déroulement du test

1. On applique un test χ^2 de conformité à une loi normale.
2. C.A. : effectif total supérieur à 50 et tous les effectifs attendus théoriques supérieurs à 5 (ou 10)
3. Les hypothèses testées sont :
   H_0 : la masse des grillons mâles suit une loi normale.
   H_1 : la masse des grillons mâles ne suit pas une loi normale.
4. Les paramètres de la loi normale sont $\bar{x} = 0.3407407$ et $\hat{\sigma}^2 = 0.0043466$.
   On a choisi les bornes des intervalles pour que les classes soient équiprobables. Les effectifs attendus sont $\frac{54}{5} = 10.8$ pour chaque classe.
   $X^2 = \sum_{i=1}^{k} \frac{(N_i - 10.8)^2}{10.8} = 2.111$.
5. On a $k = 5$ classes et on a estimé $n_{\text{par}} = 2$ paramètres ($\bar{x}$ et $\hat{\sigma}^2$), on a donc $5 - 2 - 1 = 2$ ddl.
   La valeur seuil théorique est $\chi^2_{\alpha=0.05,\,2\text{ ddl}} = 5.991$ (voir la table de χ^2).

Déroulement du test

1. On applique un test χ^2 de conformité à une loi normale.
2. C.A. : effectif total supérieur à 50 et tous les effectifs attendus théoriques supérieurs à 5 (ou 10)
3. Les hypothèses testées sont :
   H_0 : la masse des grillons mâles suit une loi normale.
   H_1 : la masse des grillons mâles ne suit pas une loi normale.
4. Les paramètres de la loi normale sont $\bar{x} = 0.3407407$ et $\hat{\sigma}^2 = 0.0043466$.
   On a choisi les bornes des intervalles pour que les classes soient équiprobables. Les effectifs attendus sont $\frac{54}{5} = 10.8$ pour chaque classe.
   $X^2 = \sum_{i=1}^{k} \frac{(N_i - 10.8)^2}{10.8} = 2.111$.
5. On a $k = 5$ classes et on a estimé $n_{\text{par}} = 2$ paramètres ($\bar{x}$ et $\hat{\sigma}^2$), on a donc $5 - 2 - 1 = 2$ ddl.
   La valeur seuil théorique est $\chi^2_{\alpha=0.05,\,2\text{ ddl}} = 5.991$ (voir la table de χ^2).
6. $X^2 = 2.111 < \chi^2_{\alpha=0.05,\,2\text{ ddl}} = 5.991$: on ne peut pas rejeter H_0 avec un risque $\alpha = 5\%$.
   On accepte H_0 avec un risque β inconnu.
   La masse des grillons mâles suit donc une loi de distribution normale.

À la limite des conditions d'application...

- La convergence vers la loi χ^2 nécessite
  - Des grands échantillons
  - Des effectifs par classe assez grands

À la limite des conditions d'application. . .

- La convergence vers la loi χ^2 nécessite
  - Des grands échantillons
  - Des effectifs par classe assez grands
- Comment faire pour de petits échantillons ?

À la limite des conditions d'application...

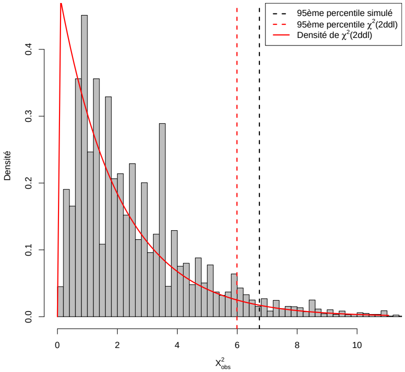
- La convergence vers la loi χ^2 nécessite
  - Des grands échantillons
  - Des effectifs par classe assez grands
- Comment faire pour de petits échantillons ? Obtenir la distribution sous H_0 avec des simulations.

À la limite des conditions d'application...

- La convergence vers la loi χ^2 nécessite
  - Des grands échantillons
  - Des effectifs par classe assez grands
- Comment faire pour de petits échantillons ? Obtenir la distribution sous H_0 avec des simulations.
- Exemple : ajustement à la loi normale $n = 54$, 5 classes

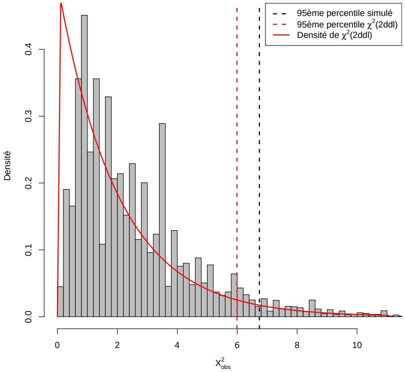
À la limite des conditions d'application. . .

- La convergence vers la loi χ^2 nécessite
  - Des grands échantillons
  - Des effectifs par classe assez grands
- Comment faire pour de petits échantillons ? Obtenir la distribution sous H_0 avec des simulations.
- Exemple : ajustement à la loi normale $n = 54$, 5 classes
  En fait, la valeur seuil simulée au risque $\alpha = 5\%$ est 6.74 (au lieu de 5.99).



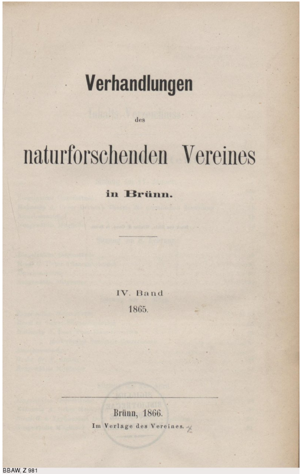
À la limite des conditions d'application...

- La convergence vers la loi χ^2 nécessite
  - Des grands échantillons
  - Des effectifs par classe assez grands
- Comment faire pour de petits échantillons ? Obtenir la distribution sous H_0 avec des simulations.
- Exemple : ajustement à la loi normale $n = 54$, 5 classes
  En fait, la valeur seuil simulée au risque $\alpha = 5\%$ est 6.74 (au lieu de 5.99).
  L'estimation du risque α associé à $\chi^2_{5\%,\,2\mathrm{ddl}} = 5.99$ est 0.075.



Conformité à une loi quelconque

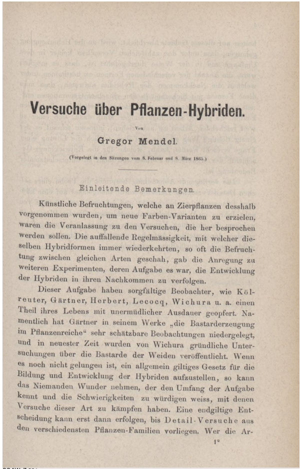
- Débats de l'association des naturalistes de Brünn (aujourd'hui Brno, Rép. Tchèque), 1866. Longtemps ignoré et “redécouvert” en 1900.



BBAW, Z 981

Conformité à une loi quelconque

- Débats de l'association des naturalistes de Brünn (aujourd'hui Brno, Rép. Tchèque), 1866. Longtemps ignoré et "redécouvert" en 1900.
- Sept caractères héritables du pois.



Versuche über Pflanzen-Hybriden.

Von

Gregor Mendel.

(Vorgelegt in den Sitzungen vom 8. Februar und 8. März 1865.)

Einleitende Bemerkungen.

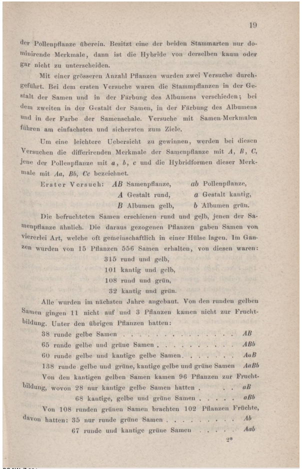
Künstliche Befruchtungen, welche an Zierpflanzen desshalb vorgenommen wurden, um neue Farben-Varianten zu erzielen, waren die Veranlassung zu den Versuchen, die her besprochen werden sollen. Die auffallende Regelmässigkeit, mit welcher dieselben Hybridformen immer wiederkehrten, so oft die Befruchtung zwischen gleichen Arten geschah, gab die Anregung zu weiteren Experimenten, deren Aufgabe es war, die Entwicklung der Hybriden in ihren Nachkommen zu verfolgen.

Dieser Aufgabe haben sorgfältige Beobachter, wie Kölreuter, Gärtner, Herbert, Lecocq, Wichura u. a. einen Theil ihres Lebens mit unermüdlicher Ausdauer geopfert. Namentlich hat Gärtner in seinem Werke „die Bastarderzeugung im Pflanzenreiche" sehr schätzbare Beobachtungen niedergelegt, und in neuester Zeit wurden von Wichura gründliche Untersuchungen über die Bastarde der Weiden veröffentlicht. Wenn es noch nicht gelungen ist, ein allgemein giltiges Gesetz für die Bildung und Entwicklung der Hybriden aufzustellen, so kann das Niemanden Wunder nehmen, der den Umfang der Aufgabe kennt und die Schwierigkeiten zu würdigen weiss, mit denen Versuche dieser Art zu kämpfen haben. Eine endgiltige Entscheidung kann erst dann erfolgen, bis Detail-Versuche aus den verschiedensten Pflanzen-Familien vorliegen. Wer die Ar-

1*

Conformité à une loi quelconque

- Débats de l'association des naturalistes de Brünn (aujourd'hui Brno, Rép. Tchèque), 1866. Longtemps ignoré et "redécouvert" en 1900.
- Sept caractères héritables du pois.
- Phénotypes de 556 individus de F2 $\frac{\mathrm{L}}{\mathrm{r}}\frac{\mathrm{J}}{\mathrm{v}} \times \frac{\mathrm{L}}{\mathrm{r}}\frac{\mathrm{J}}{\mathrm{v}}$.



19

der Pollenpflanze überein. Besitzt eine der beiden Stammarten nur dominirende Merkmale, dann ist die Hybride von derselben kaum oder gar nicht zu unterscheiden.

Mit einer grösseren Anzahl Pflanzen wurden zwei Versuche durchgeführt. Bei dem ersten Versuche waren die Stammpflanzen in der Gestalt der Samen und in der Färbung des Albumens verschieden; bei dem zweiten in der Gestalt der Samen, in der Färbung des Albumens und in der Farbe der Samenschale. Versuche mit Samen-Merkmalen führen am einfachsten und sichersten zum Ziele.

Um eine leichtere Uebersicht zu gewinnen, werden bei diesen Versuchen die differirenden Merkmale der Samenpflanze mit *A*, *B*, *C*, jene der Pollenpflanze mit *a*, *b*, *c* und die Hybridformen dieser Merkmale mit *Aa*, *Bb*, *Cc* bezeichnet.

Erster Versuch: *AB* Samenpflanze, *ab* Pollenpflanze,
A Gestalt rund, *a* Gestalt kantig,
B Albumen gelb, *b* Albumen grün.

Die befruchteten Samen erschienen rund und gelb, jenen der Samenpflanze ähnlich. Die daraus gezogenen Pflanzen gaben Samen von viererlei Art, welche oft gemeinschaftlich in einer Hülse lagen. Im Ganzen wurden von 15 Pflanzen 556 Samen erhalten, von diesen waren:

315 rund und gelb,
101 kantig und gelb,
108 rund und grün,
32 kantig und grün.

Alle wurden im nächsten Jahre angebaut. Von den runden gelben Samen gingen 11 nicht auf und 3 Pflanzen kamen nicht zur Fruchtbildung. Unter den übrigen Pflanzen hatten:

38 runde gelbe Samen . . . . *AB*
65 runde gelbe und grüne Samen . . . . *ABb*
60 runde gelbe und kantige gelbe Samen . . . . *AaB*
138 runde gelbe und grüne, kantige gelbe und grüne Samen *AaBb*

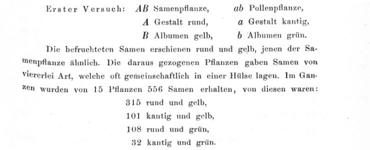
Von den kantigen gelben Samen kamen 96 Pflanzen zur Fruchtbildung, wovon 28 nur kantige gelbe Samen hatten . . . *aB*
68 kantige, gelbe und grüne Samen . . . . . *aBb*

Von 108 runden grünen Samen brachten 102 Pflanzen Früchte, davon hatten: 35 nur runde grüne Samen . . . . *Ab*
67 runde und kantige grüne Samen . . . . . *Aab*

2*

Conformité à une loi quelconque

- Débats de l'association des naturalistes de Brünn (aujourd'hui Brno, Rép. Tchèque), 1866. Longtemps ignoré et "redécouvert" en 1900.
- Sept caractères héritables du pois.
- Phénotypes de 556 individus de F2 $\frac{\mathrm{L}}{\mathrm{r}}\frac{\mathrm{J}}{\mathrm{v}} \times \frac{\mathrm{L}}{\mathrm{r}}\frac{\mathrm{J}}{\mathrm{v}}$.



Erster Versuch: *AB* Samenpflanze, *ab* Pollenpflanze,
A Gestalt rund, *a* Gestalt kantig,
B Albumen gelb, *b* Albumen grün.

Die befruchteten Samen erschienen rund und gelb, jenen der Samenpflanze ähnlich. Die daraus gezogenen Pflanzen gaben Samen von viererlei Art, welche oft gemeinschaftlich in einer Hülse lagen. Im Ganzen wurden von 15 Pflanzen 556 Samen erhalten, von diesen waren:

315 rund und gelb,
101 kantig und gelb,
108 rund und grün,
32 kantig und grün.

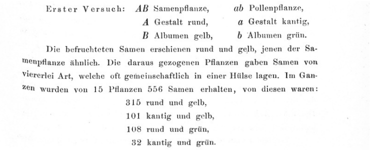
	[LJ]	[Lv]	[rJ]	[rv]
				
Observations	315	108	101	32

Conformité à une loi quelconque (de Mendel)

- Débats de l'association des naturalistes de Brünn (aujourd'hui Brno, Rép. Tchèque), 1866. Longtemps ignoré et "redécouvert" en 1900.
- Sept caractères héritables du pois.
- Phénotypes de 556 individus de F2 $\frac{\mathrm{L}}{\mathrm{r}}\frac{\mathrm{J}}{\mathrm{v}} \times \frac{\mathrm{L}}{\mathrm{r}}\frac{\mathrm{J}}{\mathrm{v}}$.

	[LJ]	[Lv]	[rJ]	[rv]

Observations	315	108	101	32

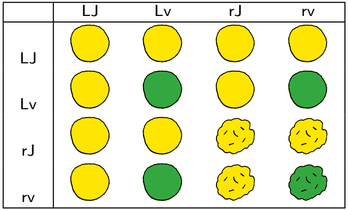


Erster Versuch: *AB* Samenpflanze, *ab* Pollenpflanze,
A Gestalt rund, *a* Gestalt kantig,
B Albumen gelb, *b* Albumen grün.

Die befruchteten Samen erschienen rund und gelb, jenen der Samenpflanze ähnlich. Die daraus gezogenen Pflanzen gaben Samen von viererlei Art, welche oft gemeinschaftlich in einer Hülse lagen. Im Ganzen wurden von 15 Pflanzen 556 Samen erhalten, von diesen waren:

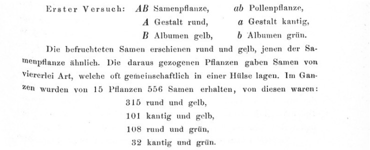
315 rund und gelb,
101 kantig und gelb,
108 rund und grün,
32 kantig und grün.

Ségrégation des allèles dans la F2



Conformité à une loi quelconque (de Mendel)

- Débats de l'association des naturalistes de Brünn (aujourd'hui Brno, Rép. Tchèque), 1866. Longtemps ignoré et “redécouvert” en 1900.
- Sept caractères héritables du pois.
- Phénotypes de 556 individus de F2 $\frac{\mathrm{L}}{\mathrm{r}}\frac{\mathrm{J}}{\mathrm{v}} \times \frac{\mathrm{L}}{\mathrm{r}}\frac{\mathrm{J}}{\mathrm{v}}$.



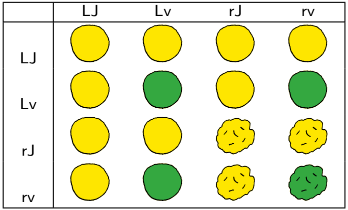
Erster Versuch: *AB* Samenpflanze, *ab* Pollenpflanze,
A Gestalt rund, *a* Gestalt kantig,
B Albumen gelb, *b* Albumen grün.

Die befruchteten Samen erschienen rund und gelb, jenen der Samenpflanze ähnlich. Die daraus gezogenen Pflanzen gaben Samen von viererlei Art, welche oft gemeinschaftlich in einer Hülse lagen. Im Ganzen wurden von 15 Pflanzen 556 Samen erhalten, von diesen waren:

315 rund und gelb,
101 kantig und gelb,
108 rund und grün,
32 kantig und grün.

	[LJ]	[Lv]	[rJ]	[rv]
Observations	315	108	101	32
Proba. th.	$9/16$	$3/16$	$3/16$	$1/16$

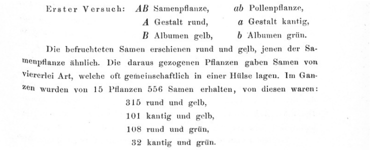
Ségrégation des allèles dans la F2



Conformité à une loi quelconque (de Mendel)

- Débats de l'association des naturalistes de Brünn (aujourd'hui Brno, Rép. Tchèque), 1866. Longtemps ignoré et "redécouvert" en 1900.
- Sept caractères héritables du pois.
- Phénotypes de 556 individus de F2 $\frac{\mathrm{L}}{\mathrm{r}}\frac{\mathrm{J}}{\mathrm{v}} \times \frac{\mathrm{L}}{\mathrm{r}}\frac{\mathrm{J}}{\mathrm{v}}$.

	[LJ]	[Lv]	[rJ]	[rv]
Observations	315	108	101	32
Proba. th.	$9/16$	$3/16$	$3/16$	$1/16$
Attendus th.	312.75	104.25	104.25	34.75

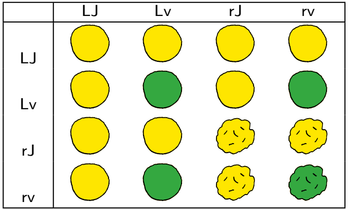


Erster Versuch: *AB* Samenpflanze, *ab* Pollenpflanze,
A Gestalt rund, *a* Gestalt kantig,
B Albumen gelb, *b* Albumen grün.

Die befruchteten Samen erschienen rund und gelb, jenen der Samenpflanze ähnlich. Die daraus gezogenen Pflanzen gaben Samen von viererlei Art, welche oft gemeinschaftlich in einer Hülse lagen. Im Ganzen wurden von 15 Pflanzen 556 Samen erhalten, von diesen waren:

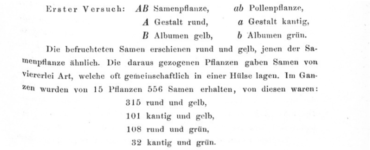
315 rund und gelb,
101 kantig und gelb,
108 rund und grün,
32 kantig und grün.

Ségrégation des allèles dans la F2



Conformité à une loi quelconque (de Mendel)

- Débats de l'association des naturalistes de Brünn (aujourd'hui Brno, Rép. Tchèque), 1866. Longtemps ignoré et "redécouvert" en 1900.
- Sept caractères héritables du pois.
- Phénotypes de 556 individus de F2 $\frac{\text{L}}{\text{r}}\frac{\text{J}}{\text{v}} \times \frac{\text{L}}{\text{r}}\frac{\text{J}}{\text{v}}$.



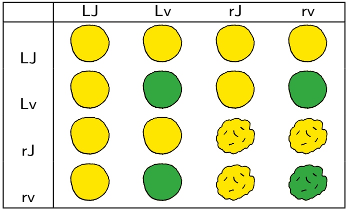
Erster Versuch: *AB* Samenpflanze, *ab* Pollenpflanze,
A Gestalt rund, *a* Gestalt kantig,
B Albumen gelb, *b* Albumen grün.

Die befruchteten Samen erschienen rund und gelb, jenen der Samenpflanze ähnlich. Die daraus gezogenen Pflanzen gaben Samen von viererlei Art, welche oft gemeinschaftlich in einer Hülse lagen. Im Ganzen wurden von 15 Pflanzen 556 Samen erhalten, von diesen waren:

315 rund und gelb,
101 kantig und gelb,
108 rund und grün,
32 kantig und grün.

	[LJ]	[Lv]	[rJ]	[rv]
Observations	315	108	101	32
Proba. th.	$9/16$	$3/16$	$3/16$	$1/16$
Attendus th.	312.75	104.25	104.25	34.75

Ségrégation des allèles dans la F2



$$X^2 = \sum \frac{(N_i - N_{\text{att}})^2}{N_{\text{att}}} = 0.470024$$

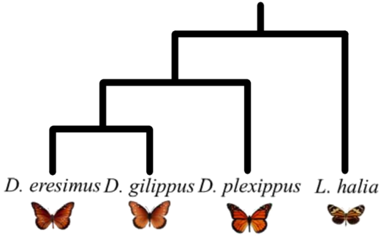
Déroulement du test

1. On souhaite appliquer un test χ^2 de conformité à une distribution théorique.
2. Les C.A. sont remplies ($N > 50$, tous les attendus théoriques supérieurs à 10).
3. Les hypothèses en concurrence sont :
   H_0 : Les fréquences théoriques des classes sont $\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16}$.
   H_1 : Au moins une fréquence est différente de la fréquence théorique proposée.
4. On a calculé les effectifs attendus (cf tableau). On trouve $X^2 = 0.470024$.
5. On a 4 classes et 3 degrés de liberté. Le seuil théorique sous H_0 est $\chi^2_{3\,\mathrm{ddl}} = 7.815$.
6. $X^2 < \chi^2_{3\,\mathrm{ddl}}$ donc on ne peut pas rejeter H_0 avec un risque α. On accepte H_0 avec un risque β inconnu.
   La distribution des phénotypes dans la F2 est conforme aux lois de de Mendel sur la ségrégation des allèles.

Conformité à une proportion théorique

Recherche d'hybridation passée : le test ABBA BABA

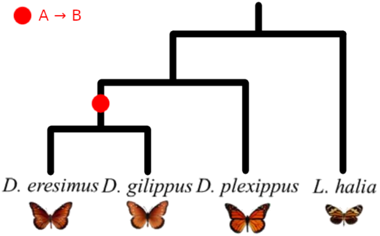
- Données : *Aardema et al.* (2016).
  3 espèces et 1 groupe externe *L. halia*.



Conformité à une proportion théorique

Recherche d'hybridation passée : le test ABBA BABA

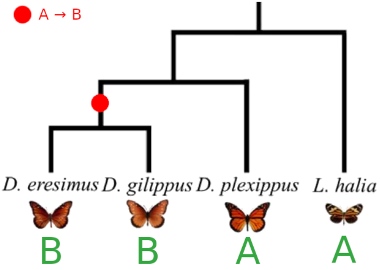
- Données : *Aardema et al.* (2016).
  3 espèces et 1 groupe externe *L. halia*.
- On compte les mutations. A est l'état ancestral et B l'état dérivé. On utilise les sites où deux espèces du triplet ont un état dérivé.



Conformité à une proportion théorique

Recherche d'hybridation passée : le test ABBA BABA

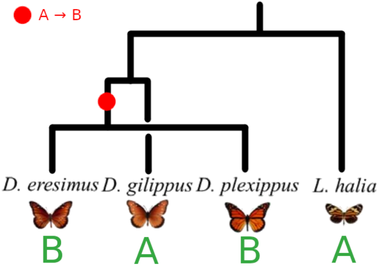
- Données : *Aardema et al.* (2016).
  3 espèces et 1 groupe externe *L. halia*.
- On compte les mutations. A est l'état ancestral et B l'état dérivé. On utilise les sites où deux espèces du triplet ont un état dérivé.
- Dans l'ordre *D. eresimus*, *D. plexipps*, *D. gilippus* et *L. halia*, un motif majoritaire : BBAA.



Conformité à une proportion théorique

Recherche d'hybridation passée : le test ABBA BABA

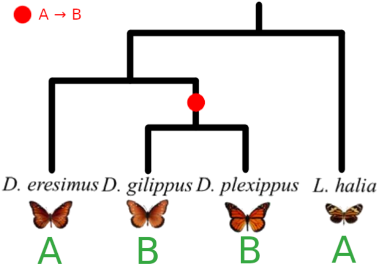
- Données : *Aardema et al.* (2016).
  3 espèces et 1 groupe externe *L. halia*.
- On compte les mutations. A est l'état ancestral et B l'état dérivé. On utilise les sites où deux espèces du triplet ont un état dérivé.
- Dans l'ordre *D. eresimus*, *D. plexipps*, *D. gilippus* et *L. halia*, un motif majoritaire : BBAA.
- Les motifs ABBA et BABA sont théoriquement plus rares et équiprobables.



Conformité à une proportion théorique

Recherche d'hybridation passée : le test ABBA BABA

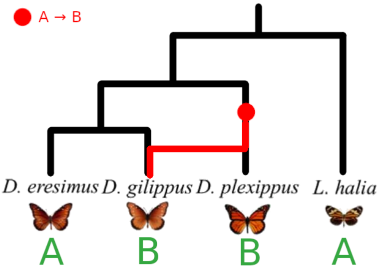
- Données : *Aardema et al.* (2016).
  3 espèces et 1 groupe externe *L. halia*.
- On compte les mutations. A est l'état ancestral et B l'état dérivé. On utilise les sites où deux espèces du triplet ont un état dérivé.
- Dans l'ordre *D. eresimus*, *D. plexipps*, *D. gilippus* et *L. halia*, un motif majoritaire : BBAA.
- Les motifs ABBA et BABA sont théoriquement plus rares et équiprobables.



Conformité à une proportion théorique

Recherche d'hybridation passée : le test ABBA BABA

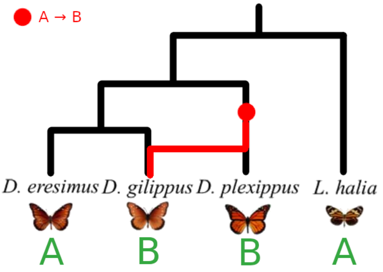
- Données : *Aardema et al.* (2016).
  3 espèces et 1 groupe externe *L. halia*.
- On compte les mutations. A est l'état ancestral et B l'état dérivé. On utilise les sites où deux espèces du triplet ont un état dérivé.
- Dans l'ordre *D. eresimus*, *D. plexipps*, *D. gilippus* et *L. halia*, un motif majoritaire : BBAA.
- Les motifs ABBA et BABA sont théoriquement plus rares et équiprobables.
- Une hybridation passée ⇒ excès de ABBA ou BABA.



Conformité à une proportion théorique

Recherche d'hybridation passée : le test ABBA BABA

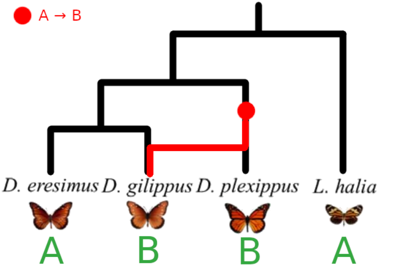
- Données : *Aardema et al.* (2016).
  3 espèces et 1 groupe externe *L. halia*.
- On compte les mutations. A est l'état ancestral et B l'état dérivé. On utilise les sites où deux espèces du triplet ont un état dérivé.
- Dans l'ordre *D. eresimus*, *D. plexipps*, *D. gilippus* et *L. halia*, un motif majoritaire : BBAA.
- Les motifs ABBA et BABA sont théoriquement plus rares et équiprobables.
- Une hybridation passée ⇒ excès de ABBA ou BABA.
- On teste l'équiprobabilité de ABBA et BABA.



Conformité à une proportion théorique

Recherche d'hybridation passée : le test ABBA BABA

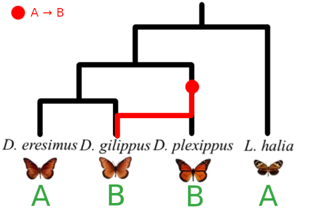
- Données : *Aardema et al.* (2016).
  3 espèces et 1 groupe externe *L. halia*.
- On compte les mutations. A est l'état ancestral et B l'état dérivé. On utilise les sites où deux espèces du triplet ont un état dérivé.
- Dans l'ordre *D. eresimus*, *D. plexipps*, *D. gilippus* et *L. halia*, un motif majoritaire : BBAA.
- Les motifs ABBA et BABA sont théoriquement plus rares et équiprobables.
- Une hybridation passée ⇒ excès de ABBA ou BABA.
- On teste l'équiprobabilité de ABBA et BABA.



	BBAA	ABBA	BABA
Observées	2818	1287	1630

Déroulement du test

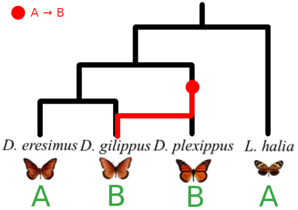
1. On applique un test χ^2 de conformité à la distribution équiprobable.



	BBAA	ABBA	BABA
Observées	2818	1287	1630

Déroulement du test

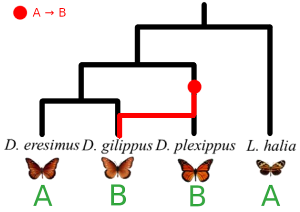
1. On applique un test χ^2 de conformité à la distribution équiprobable.
2. Les C.A. sont satisfaites (grands effectifs)



	BBAA	ABBA	BABA
Observées	2818	1287	1630

Déroulement du test

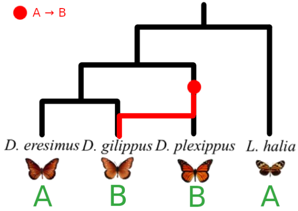
1. On applique un test χ^2 de conformité à la distribution équiprobable.
2. Les C.A. sont satisfaites (grands effectifs)
3. Les hypothèses testées sont :
   H_0 : les deux motifs ABBA et BABA sont équiprobables.
   H_1 : Les deux motifs ne sont pas équiprobables.



	BBAA	ABBA	BABA
Observées	2818	1287	1630
Attendues	2818	1458.5	1458.5

Déroulement du test

1. On applique un test χ^2 de conformité à la distribution équiprobable.
2. Les C.A. sont satisfaites (grands effectifs)
3. Les hypothèses testées sont :
   H_0 : les deux motifs ABBA et BABA sont équiprobables.
   H_1 : Les deux motifs ne sont pas équiprobables.
4. Sous H_0, on attend en moyenne $\frac{1}{2}(1287 + 1630) = 1458.5$ mutations donnant le motif ABBA et autant le motif BABA.
   $X^2 = 40.34$

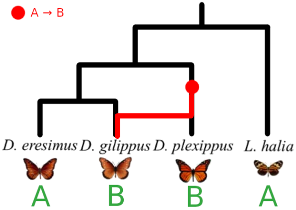


	BBAA	ABBA	BABA
Observées	2818	1287	1630
Attendues	2818	1458.5	1458.5

$$X^2 = \frac{(1287 - 1458.5)^2}{1458.5} + \frac{(1630 - 1458.5)^2}{1458.5}$$
$$= 40.34$$

Déroulement du test

1. On applique un test χ^2 de conformité à la distribution équiprobable.
2. Les C.A. sont satisfaites (grands effectifs)
3. Les hypothèses testées sont :
   H_0 : les deux motifs ABBA et BABA sont équiprobables.
   H_1 : Les deux motifs ne sont pas équiprobables.
4. Sous H_0, on attend en moyenne $\frac{1}{2}(1287 + 1630) = 1458.5$ mutations donnant le motif ABBA et autant le motif BABA.
   $X^2 = 40.34$
5. Il y a deux classes et 1 ddl. $\chi^2_{\alpha=0.05,\ 1\ \mathrm{ddl}} = 3.841$

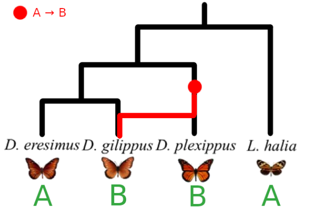


	BBAA	ABBA	BABA
Observées	2818	1287	1630
Attendues	2818	1458.5	1458.5

$$X^2 = \frac{(1287 - 1458.5)^2}{1458.5} + \frac{(1630 - 1458.5)^2}{1458.5}$$
$$= 40.34$$

Déroulement du test

1. On applique un test χ^2 de conformité à la distribution équiprobable.
2. Les C.A. sont satisfaites (grands effectifs)
3. Les hypothèses testées sont :
   H_0 : les deux motifs ABBA et BABA sont équiprobables.
   H_1 : Les deux motifs ne sont pas équiprobables.
4. Sous H_0, on attend en moyenne $\frac{1}{2}(1287+1630) = 1458.5$ mutations donnant le motif ABBA et autant le motif BABA.
   $X^2 = 40.34$
5. Il y a deux classes et 1 ddl. $\chi^2_{\alpha=0.05,\,1\ \mathrm{ddl}} = 3.841$
6. $X^2 > 3.841$ donc on rejette H_0 avec un risque $\alpha = 5\%$. Les motifs ABBA et BABA ne sont pas équiprobables.
   Il y a plus de BABA ce qui suggère des événements d'hybridation passée entre *D. plexipus* et *D. Eresimus*.



	BBAA	ABBA	BABA
Observées	2818	1287	1630
Attendues	2818	1458.5	1458.5

$$X^2 = \frac{(1287-1458.5)^2}{1458.5} + \frac{(1630-1458.5)^2}{1458.5}$$
$$= 40.34$$

Conclusions sur le test de conformité de χ^2

- Test de *conformité* dit aussi d'*ajustement* à une distribution théorique
  - Basé sur la convergence $\mathcal{B} \to \mathcal{N}$, d'où les conditions d'application.
  - Applicable à des distributions discrètes ou continues.
- Il existe aussi un test d'homogénéité des distributions (que vous verrez ultérieurement).
- Lorsque les C.A. ne sont pas remplies, possibilité (nécessité) de simuler sur ordinateur la distribution de X^2 sous H_0.

Conclusions sur les tests d'hypothèse

- Les tests d'hypothèse s'intéressent aux **valeurs théoriques** (μ) grâce à des observations sur un échantillon.
- Basés sur une variable aléatoire V sensible à un changement de la valeur théorique. V est appelée la **statistique** du test.
- H_0 et H_1 ne jouent pas le même rôle. H_0 est l'hypothèse qu'on souhaite invalider.
- Le statisticien connaît la distribution de V sous H_0 et choisit de considérer comme "inattendues" les valeurs les plus extrêmes de cette distribution.
- Le risque de première espèce est la probabilité de rejeter une H_0 alors qu'elle est vraie. Il est choisi par le statisticien.
- Le risque de deuxième espèce est le risque de se tromper en ne rejetant pas H_0. β est inconnu.
- Problème des tests multiples : parmi n tests indépendents, la probabilité qu'au moins un test invalide H_0 est $1 - (1 - \alpha)^n \to 1$.