LES TESTS STATISTIQUES
AVEC UNE CALCULATRICE OU GEOGEBRA
- Partie 3 -

Voici le dernier article *Tests statistiques avec la calculatrice ou GeoGebra*. Nous avons abordé les tests de conformité d'une moyenne et d'une proportion, et les tests de comparaison de deux proportions, de deux variances et de deux moyennes dans les deux précédents bulletins. Ici, nous présentons les tests du Khi-deux[1]. Comme dans les deux précédents articles, nous ne détaillerons pas le lien avec le travail mené habituellement à l'écrit.

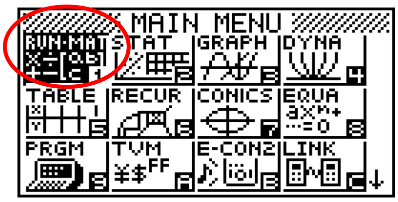
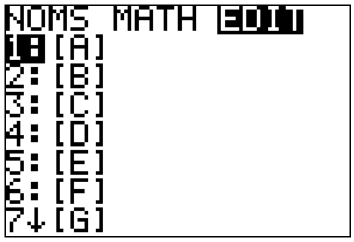
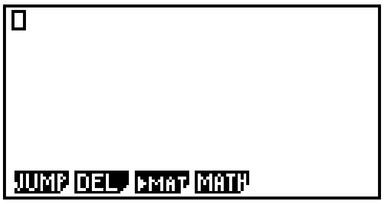
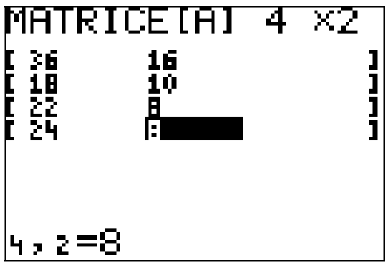
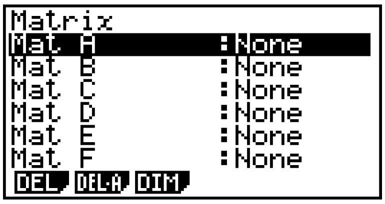
Certaines des fonctionnalités présentées ne sont pas présentes sur tous les modèles de calculatrices.

Saisie d'une matrice

Saisie de la matrice $\begin{pmatrix} 36 & 16 \\ 18 & 10 \\ 22 & 8 \\ 24 & 8 \end{pmatrix}$ qui sera utilisée dans un exemple traité.

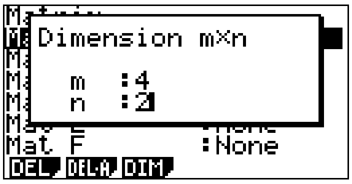
Avec une calculatrice

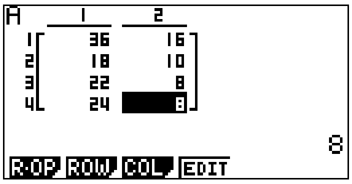
Saisie dans la matrice **A** (si elle ne contient pas de données que l'on souhaite conserver !)

sur CASIO 35+	**sur TI 83 PLUS .*fr***
touche $\boxed{\text{MENU}}$, option $\boxed{\text{RUN MAT}}$ 	**matrice** touches $\boxed{\text{2nde}}$, puis $\boxed{x^{-1}}$  onglet $\boxed{\text{EDIT}}$, ligne **1: [A]**
 touche $\boxed{\text{F3}}$ pour l'onglet $\boxed{\text{▶MAT}}$	On saisit la taille de la matrice et les valeurs des différents termes de la matrice.  Après validation du dernier terme, la matrice est prête.
 Touche $\boxed{\text{F3}}$ pour l'onglet $\boxed{\text{DIM}}$ où l'on entre la dimension de la matrice.	

[1] Des références théoriques sont consultables dans les articles du GRES en ligne sur R2math :

- exercices corrigés : http://r2math.enfa.fr/wp-content/uploads/2010/07/6-7-exercice.pdf
- test du Khi-2 d'indépendance : http://r2math.enfa.fr/wp-content/uploads/2010/07/9-10-independance.pdf
- Excel et le test du Khi-2 d'indépendance : http://r2math.enfa.fr/wp-content/uploads/2010/07/8-7-excel.pdf

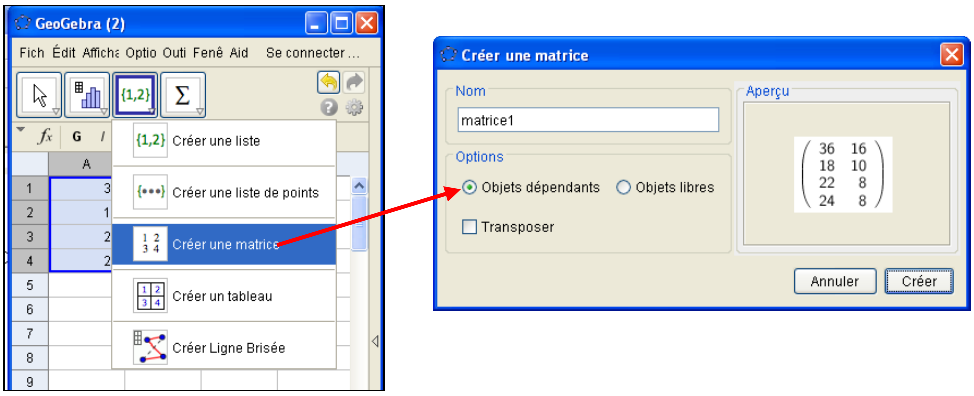
<table>
<tr><th>sur casio 35+</th></tr>
<tr><td>

On va saisir une matrice avec 4 lignes et 2 colonnes.</td></tr>
<tr><td>On saisit les différents termes de la matrice.


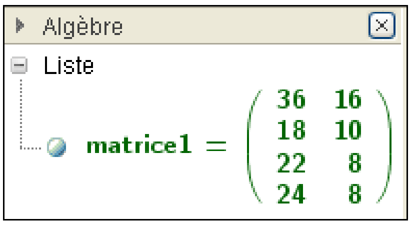
Après validation du dernier terme, la matrice est prête.</td></tr>
</table>

Avec GeoGebra

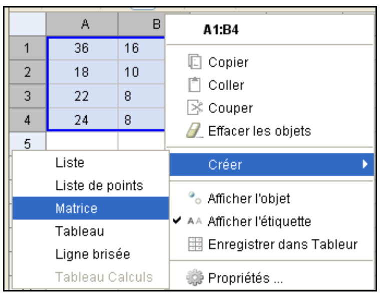
On saisit (ou on copie) les termes de la matrice dans le tableur de GeoGebra (menu ***Affichage*** ou touches **Ctrl** + **Maj** + **S**) en respectant la forme de la matrice. On sélectionne la plage de ces termes, on utilise le bouton , option ***Créer une matrice***.



L'option ***Objets dépendants*** lie la matrice au tableur (toute modification dans le tableur aura une incidence sur la matrice). Si on ne souhaite pas cette liaison, on utilise l'option ***Objets libres***.
On obtient la matrice correspondante : ***matrice1*** (dont on peut bien sûr changer le nom par défaut !), dans la fenêtre ***Algèbre*** :



Remarque :
Au lieu d'utiliser le bouton de la barre d'outils, on peut aussi faire un clic droit sur la zone sélectionnée et choisir ***Créer***, ***Matrice*** qui crée une matrice liée au tableur.



Test d'indépendance du Khi-deux

Exemple

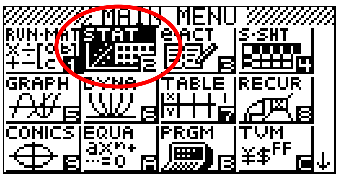
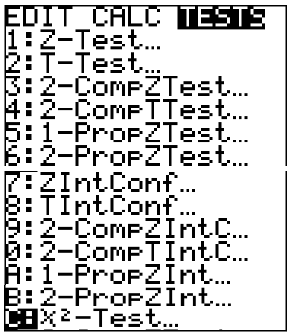
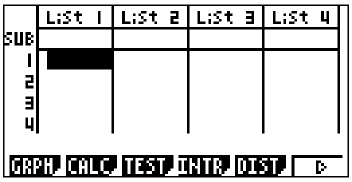
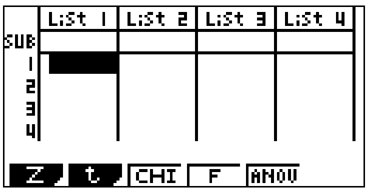
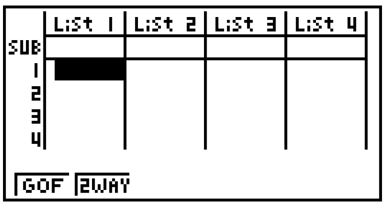
Dans un élevage bovin, comportant des bêtes de quatre races différentes, on a étudié 142 gestations et on en a déduit la distribution des effectifs donnée ci-dessous :

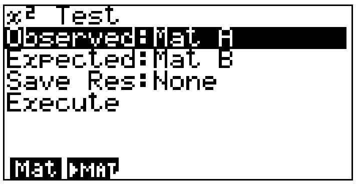
		Issue de la gestation	
		Vêlages	**Avortements**
Races	**1**	36	16
	2	18	10
	3	22	8
	4	24	8

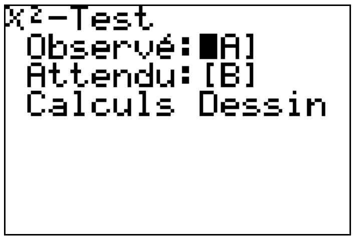
Doit-on en conclure qu'il existe une relation entre les races et l'issue de la gestation des vaches de l'élevage, au seuil de risque $\alpha = 0{,}05$?

Avec une calculatrice

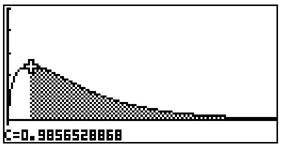
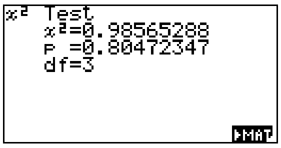
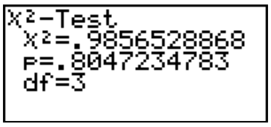
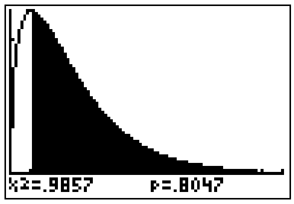
Le tableau des effectifs observés a été saisi, par exemple dans la matrice A selon les procédures décrites précédemment. Ainsi $A = \begin{pmatrix} 36 & 16 \\ 18 & 10 \\ 22 & 8 \\ 24 & 8 \end{pmatrix}$.

<table>
<tr><th>sur CASIO 35+</th><th>sur TI 83 PLUS .fr</th></tr>
<tr><td>touche MENU, option STAT
</td>
<td rowspan="4">touche stats, onglet TESTS

C:χ^2-Test...</td></tr>
<tr><td>
touche F3 pour l'onglet TEST</td></tr>
<tr><td>
touche F3 pour l'onglet CHI</td></tr>
<tr><td>
touche F2 pour l'onglet 2WAY</td></tr>
</table>

<table>
<tr><th colspan="2">sur CASIO 35+</th><th>sur TI 83 PLUS .fr</th></tr>
<tr>
<td>

Pour changer de matrice :
– touche F1 pour l'onglet Mat et choisir une matrice déjà saisie
ou
– touche F2, onglet ▶MAT pour construire la matrice des données. Dans ce cas, on saisit la matrice comme exposé au début de l'article...</td>
<td>On complète la fenêtre en fonction de :
- la matrice dans laquelle se trouvent les effectifs observés, ici A
- la matrice dans laquelle on obtiendra les effectifs théoriques, ici B</td>
<td>

Pour changer de matrice :
touches 2nde, puis matrice $\boxed{x^{-1}}$ et valider par entrer la matrice voulue.</td>
</tr>
</table>

DRAW sur CASIO	CALC sur CASIO	Calculs sur TI	Dessin sur TI
			

La variable de décision est $T = \sum_{i=1}^{4} \sum_{j=1}^{2} \frac{(Z_{i,j} - t_{i,j})^2}{t_{i,j}}$ où les $Z_{i,j}$ sont les variables aléatoires qui donnent le nombre d'observations de chaque couple de modalités (*Race* , *Issue*) et les $t_{i,j}$ sont les valeurs théoriques des effectifs de ces couples de modalités sous l'hypothèse d'indépendance des deux caractères étudiés. Sous l'hypothèse d'indépendance, la loi de T est approchée par la loi du Khi-deux à $(4 - 1)(2 - 1) = 3$ degrés de liberté lorsque les $t_{i,j}$ sont supérieurs ou égaux à 5.

χ^2 est la valeur observée de T. $\chi^2 \simeq 0{,}985\,652\,88$.

p est la probabilité que, sous l'hypothèse d'indépendance, la variable aléatoire T prenne une valeur supérieure à χ^2. Ici, **p** $= P(T > 0{,}985\,652\,88) \simeq 0{,}804\,723\,47$.

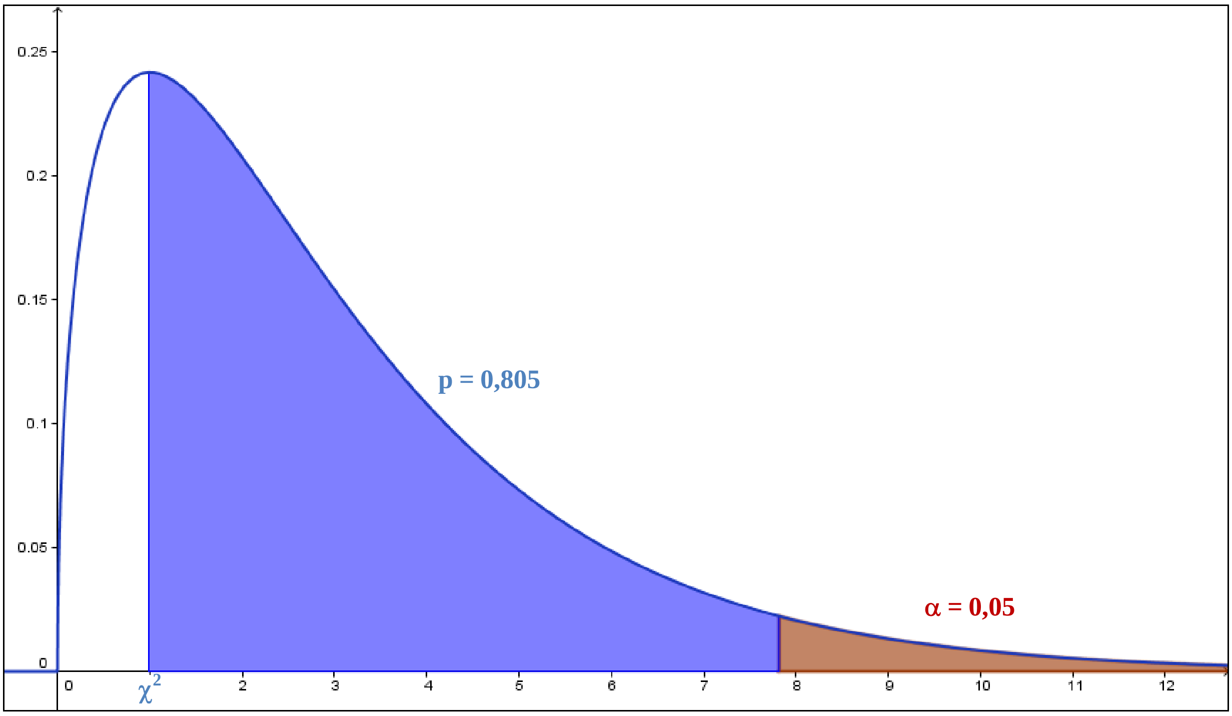
df est le nombre de degrés de liberté de T.

Exploitons les résultats donnés par la calculatrice pour donner la conclusion du test en comparant α avec la *p-value* **p**.

- Si $\alpha > 0{,}804\,723\,47$, on rejette l'hypothèse d'indépendance des deux caractères étudiés.
- Si $\alpha \leq 0{,}804\,723\,47$, on n'est pas en mesure de rejeter l'hypothèse d'indépendance des deux caractères étudiés.

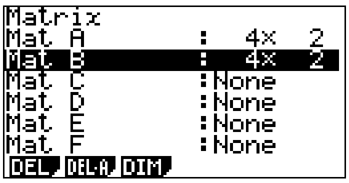
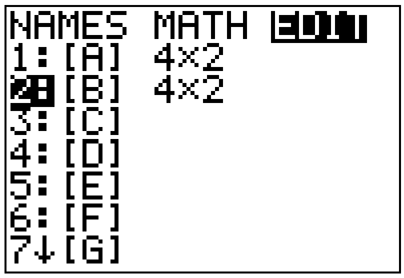
Revenons à notre exemple :

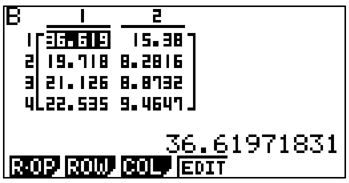
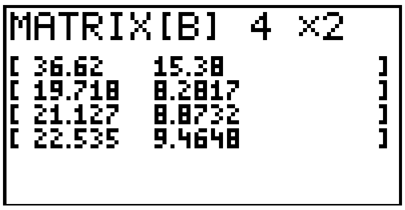
Comme α **= 0,05**, $\alpha <$ **p**, on n'est pas en mesure de rejeter l'hypothèse d'indépendance des deux caractères étudiés.



Remarque :

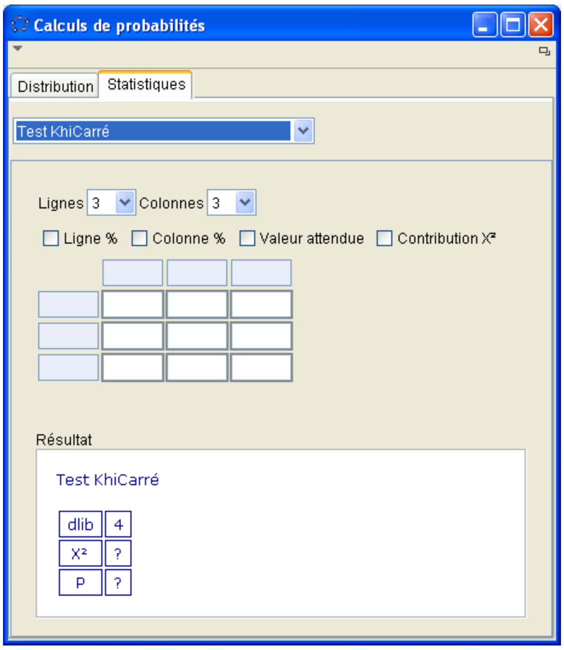
Pour afficher les valeurs théoriques attendues des effectifs de chaque couple de modalités, il suffit d'afficher la matrice **B** :

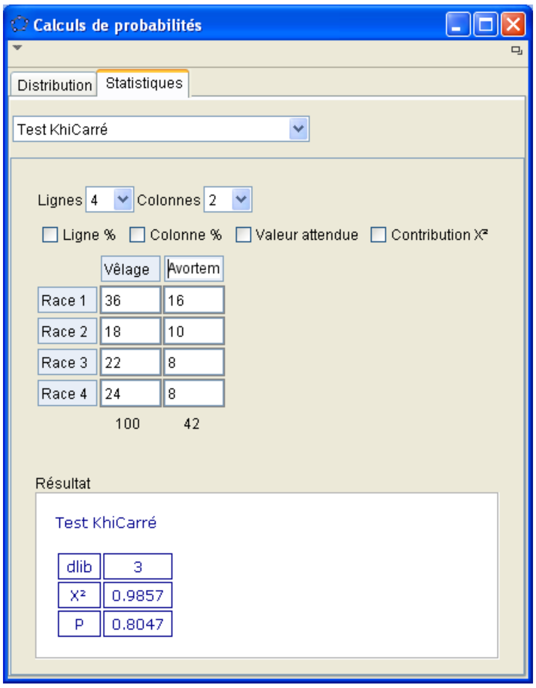
<table>
<tr><th>sur CASIO 35+</th><th>sur TI 83 PLUS .fr</th></tr>
<tr><td>touche [MENU], option [RUN MAT]
puis touche [F3] pour l'onglet [▶MAT]
On sélectionne la ligne de la matrice B et utilise la touche [EXE].

</td><td>touches [2nde], puis [matrice x^{-1}], onglet [EDIT]


et validation de la matrice voulue, ici B</td></tr>
<tr><td>
</td><td>
</td></tr>
</table>

Avec GeoGebra

On utilise l'onglet ***Statistiques*** proposé dans le module ***Calculs de probabilités*** (menu ***Affichage*** ou touches $\boxed{\textbf{Ctrl}}$ + $\boxed{\textbf{Maj}}$ + $\boxed{\textbf{P}}$). On choisit la commande ***TestKhiCarré***, on saisit la matrice des effectifs des couples de modalités observées dans la fenêtre qui s'ouvre :





On obtient :

dlib : le nombre de degrés de liberté de T.

χ^2 : la valeur observée de T. $\chi^2 \simeq 0{,}985\,7$.

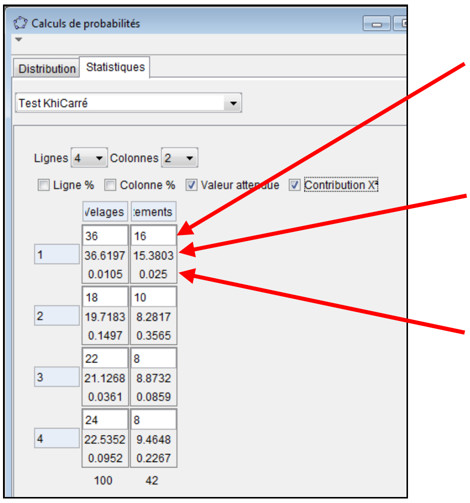
P est la probabilité que, sous l'hypothèse d'indépendance, la variable aléatoire T prenne une valeur supérieure à χ^2. Ici, **P** $= P(T > 0{,}985\,652\,88) \simeq 0{,}804\,7$.

Comme α **= 0,05**, $\alpha <$ **P**, on n'est pas en mesure de rejeter l'hypothèse d'indépendance des deux caractères étudiés.

Remarque :

Et à quoi servent les 4 cases à cocher  ?

Les options ***Ligne %*** et ***Colonne %*** permettent d'afficher les profils ligne ou colonne (comme dans les tableaux de contingence…).



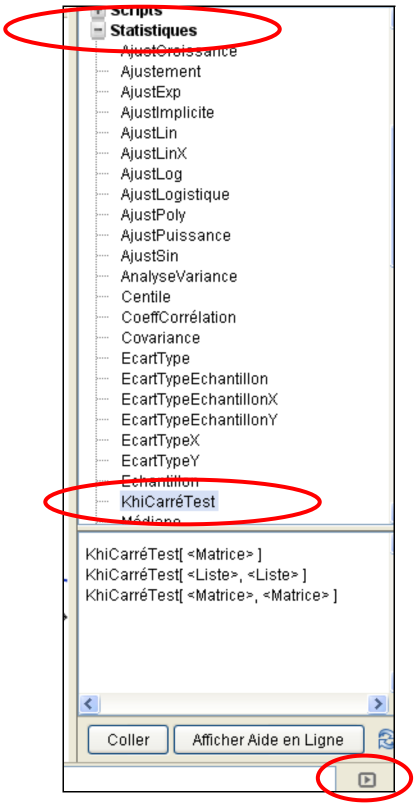
Les $n_{i,j}$ sont les nombres d'observations de chaque couple de modalités (*Race* , *Issue*)

L'option ***Valeur attendue*** permet d'afficher les effectifs théoriques $t_{i,j}$.

L'option ***Contribution*** χ^2 permet d'afficher les valeurs $\dfrac{(n_{i,j} - t_{i,j})^2}{t_{i,j}}$ dont la somme est égale à la valeur observée de T.

Avec GeoGebra (bis)

On peut aussi utiliser la commande ***KhiCarréTest*** accessible dans la liste des commandes de la rubrique ***Statistiques***. La liste des commandes s'obtient (et se ferme) avec la flèche à droite de la zone de saisie.



La syntaxe est : ***KhiCarréTest[Matrice]***

Matrice est la matrice des valeurs observées. La commande donne le résultat sous la forme d'une liste $\{p\text{-}value, T_{obs}\}$.

Si on a créé la matrice ***matrice1*** des effectifs des couples de modalités observées, la formule à utiliser est ***KhiCarréTest[matrice1]*** à écrire dans la ligne de saisie. On obtient {0.805,0.986}. Pour conclure, il reste à comparer α avec la p-value p comme précédemment.

Remarque :

Le nombre de chiffres des résultats est choisi par la commande ***Arrondi*** du menu ***Options***.

Remarque :

Si on dispose des valeurs théoriques attendues, une autre option est disponible pour la commande ***KhiCarréTest***.

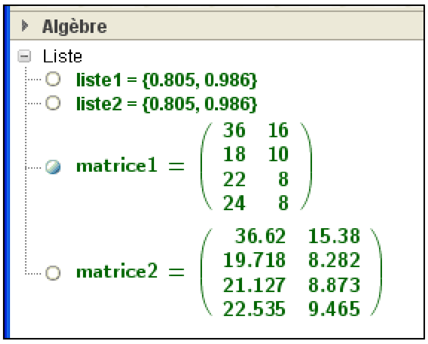
KhiCarréTest[Matrice des valeurs observées , Matrice des valeurs attendues]

réalise un test du Khi-deux à partir de la matrice des valeurs observées et de la matrice des valeurs attendues.

La commande donne le résultat sous la forme d'une liste $\{p\text{-}value, \chi^2\}$.

Reprenons notre exemple :

Si on dispose de la ***matrice2*** contenant les valeurs théoriques attendues :



Ici, la formule est ***KhiCarréTest[matrice1,matrice2]*** à écrire dans la ligne de saisie. On obtient {0.805,0.986}. On peut changer la précision de l'affichage des nombres dans le menu **Options**, rubrique **Arrondi**. Pour conclure, il reste à comparer α avec la p-value p comme précédemment. comme $\boldsymbol{\alpha = 0{,}05}$, $\alpha < \mathbf{P}$, on n'est pas en mesure de rejeter l'hypothèse d'indépendance des deux caractères étudiés.

Test du Khi-deux d'adéquation à une loi théorique

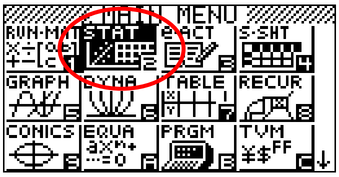
Exemple

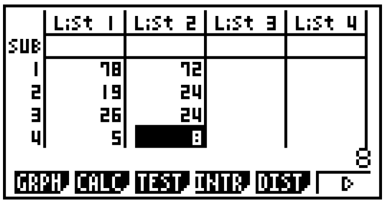
Un couple de cobayes à pelage gris et lisse a donné naissance à 128 descendants dont les pelages se répartissent de la manière suivante :

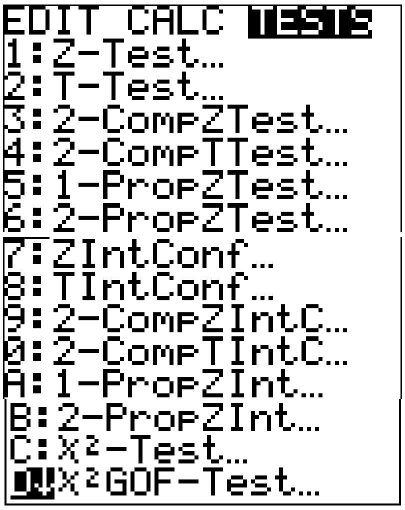
78 au pelage gris et lisse (*gl*) 19 au pelage blanc et rude (*br*)
26 au pelage blanc et lisse (*bl*) 5 au pelage gris et rude (*gr*)

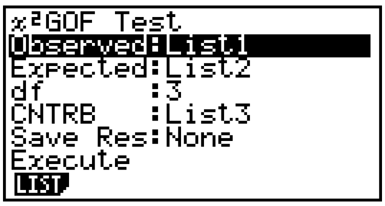
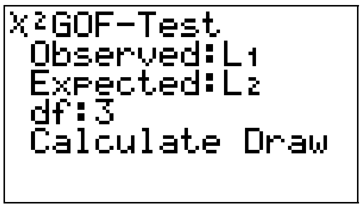
Si l'on admet que la transmission de ces deux caractères suit les lois de Mendel, les fréquences théoriques d'apparition f_{gl}, f_{bl}, f_{br}, f_{gr} des différentes catégories doivent être proportionnelles à 9, 3, 3 et 1.
Les résultats expérimentaux permettent-ils d'accepter ce mode de transmission au seuil de 5 % ?

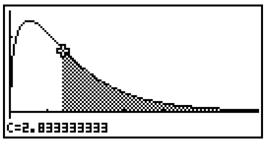
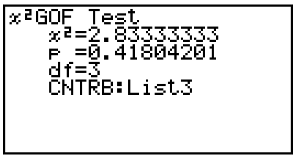
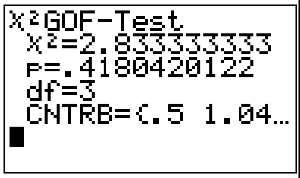
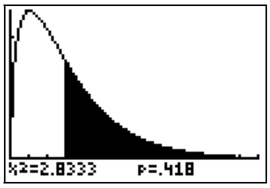
Avec une calculatrice

<table>
<tr><th>sur CASIO 35+</th><th>sur TI 83 PLUS .fr</th></tr>
<tr><td colspan="2">Les effectifs observés et les effectifs attendus sont saisis dans des listes (ici respectivement dans L1 = {78 , 19 , 26 , 5} et L2 = {72 , 24 , 24 , 8}).</td></tr>
<tr><td>touche [MENU], option [STAT]


On saisit les données observées et attendues dans les listes 1 et 2


touches [F3] pour l'onglet [TEST],
puis [F3] pour l'onglet [CHI]
puis [F1] pour l'onglet [GOF]</td>
<td>touche [stats], onglet [TESTS]


D:χ²GOF-Test
(GOF : Goodness Of Fit)</td></tr>
<tr><td>
</td>
<td>
</td></tr>
</table>

DRAW sur CASIO	CALC sur CASIO	Calculs sur TI	Dessin sur TI
			

La variable de décision est $T = \sum_{i=1}^{4} \frac{(Z_i - t_i)^2}{t_i}$ où les Z_i sont les variables aléatoires qui donnent le nombre d'observations de chaque modalité et les t_i sont leurs valeurs théoriques sous l'hypothèse d'adéquation de la répartition observée au modèle testé. Sous l'hypothèse d'adéquation, la loi de T est approchée par la loi du Khi-deux à $4 - 1 = 3$ degrés de libertés lorsque les t_i sont supérieurs ou égaux à 5.

χ^2 est la valeur observée de T. $\chi^2 \simeq 2{,}83\,333\,33$.

p est la probabilité que, sous l'hypothèse d'adéquation, la variable aléatoire T prenne une valeur supérieure à χ^2. Ici, **p** $= P(T > 2{,}83\,333\,33) \simeq 0{,}418\,042\,012\,2$.

df est le nombre de degrés de liberté de T.

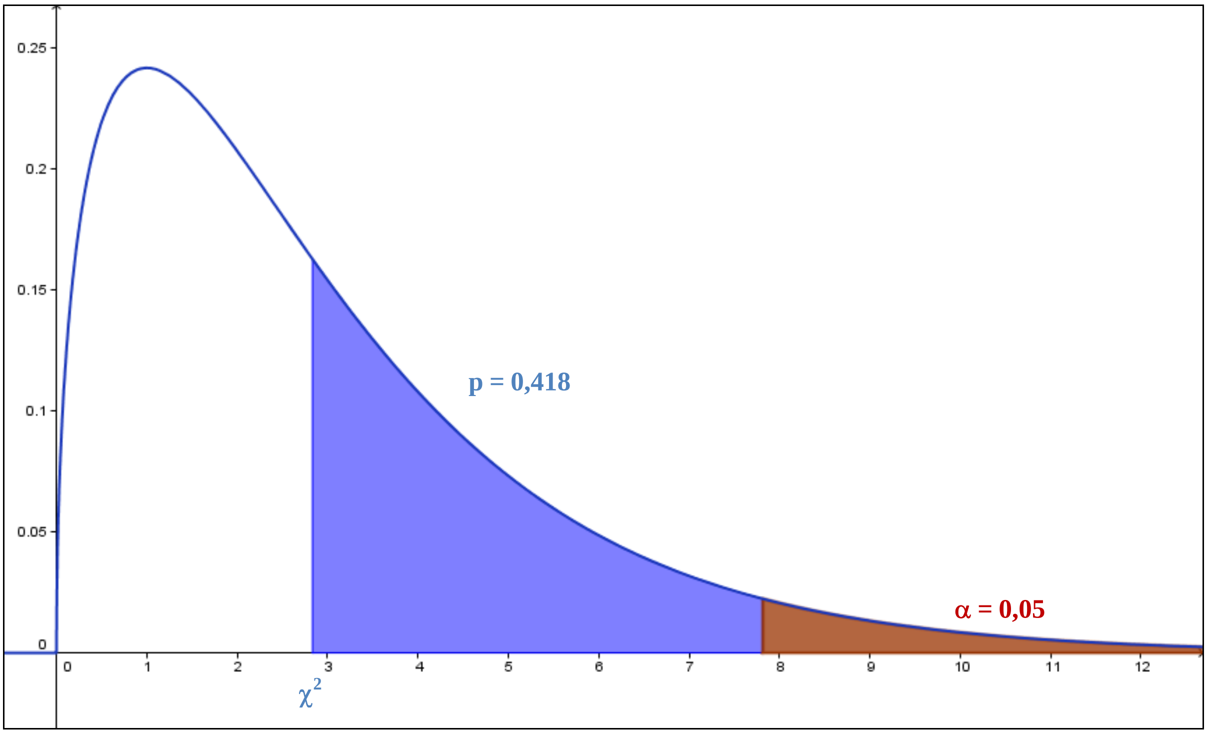
CNTRB est la liste des résultats des calculs $\frac{(Z_i - t_i)^2}{t_i}$ pour $1 \leq i \leq 4$.

Exploitons les résultats donnés par la calculatrice pour rédiger une règle de décision en comparant α avec la *p-value* **p**.

- Si $\alpha > 0{,}418\,042\,012\,2$, on rejette l'hypothèse d'adéquation.
- Si $\alpha \leq 0{,}418\,042\,012\,2$, on n'est pas en mesure de rejeter l'hypothèse d'adéquation.

Revenons à notre exemple :

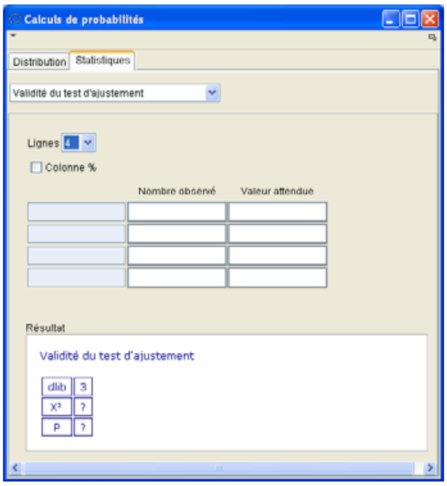
Comme **α = 0,05 :** $\alpha <$ **p**, on n'est pas en mesure de rejeter l'hypothèse d'adéquation à la distribution théorique des effectifs selon les lois de Mendel.

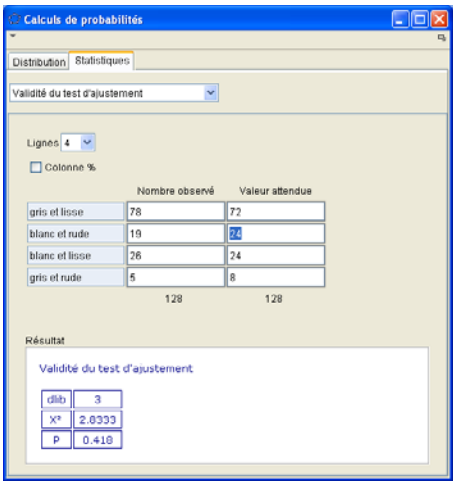


Avec GeoGebra

On utilise l'onglet ***Statistiques*** proposé dans le module ***Calculs de probabilités***. On choisit la commande ***Validité du test d'ajustement***.

Dans l'exemple, les valeurs théoriques sont $\frac{128}{(9+3+3+1)} \times 9 = 8 \times 9 = 72$, $8 \times 3 = 24$, $8 \times 3 = 24$ et $8 \times 1 = 8$.





La variable de décision est $T = \sum_{i=1}^{4} \frac{(N_i - t_i)^2}{t_i}$ où les N_i sont les variables aléatoires qui donnent le nombre d'observations de chaque modalité et les t_i sont les valeurs théoriques des effectifs de ces modalités. Sous l'hypothèse d'adéquation des données à la distribution théorique, la loi de T est approchée par la loi du Khi-deux à $(4 - 1) = 3$ degrés de libertés lorsque les t_i sont supérieurs ou égaux à 5.

dlib est le nombre de degrés de liberté de T.

χ^2 est la valeur observée de T. $\chi^2 \simeq 2{,}833\,3$.

P est la probabilité que, sous l'hypothèse d'adéquation des données à la distribution théorique, la variable aléatoire T prenne une valeur supérieure à χ^2. Ici, $\mathbf{P} = \mathrm{P}(T > 2{,}833\,3) \simeq 0{,}418$.

Comme $\boldsymbol{\alpha = 0{,}05}$ **:** $\alpha < \mathbf{p}$, on n'est pas en mesure de rejeter l'hypothèse d'adéquation à la distribution théorique des effectifs selon les lois de Mendel.

Avec GeoGebra (bis)

On peut aussi utiliser la commande ***KhiCarréTest***, disponible dans la liste des commandes de la rubrique **Statistiques**, dont la syntaxe est ***KhiCarréTest[liste1,liste2]***.

liste1 est la liste des effectifs observés et ***liste2*** est la liste des valeurs attendues sous l'hypothèse d'adéquation des données à la distribution testée. La commande donne le résultat sous la forme d'une liste $\{p\text{-}value, \chi^2\}$.

Reprenons notre exemple :

Ici, la formule est ***KhiCarréTest[{78,19,26,5},{72,24,24,8}]*** à écrire dans la ligne de saisie. On obtient $\{0.418, 2.833\}$. Pour conclure, il reste à comparer α avec la p-value p comme précédemment.

Remarque : On peut créer une liste directement dans la zone de saisie : les termes de la liste sont écrits entre accolades, séparés par des virgules (textes entre guillemets) ou à partir de cellules sélectionnées dans le tableur et commande *Créer*, *Liste* dans le menu contextuel obtenu par un clic droit.