

# Statistiques et probabilités


Les statistiques et les probabilités occupent une place importante dans l'enseignement de certaines classes préparatoires. Les principales fonctions nécessaires pour travailler dans ce domaine se trouvent dans les applications *Calculs* et *Tableur & listes*. L'application *Données & statistiques* permet d'effectuer des représentations graphiques de données statistiques, l'application *Graphiques & géométrie* ne permettant que la représentation de nuages de points. Des fonctions définies dans ce chapitre vous permettront d'étendre les possibilités de votre unité nomade.

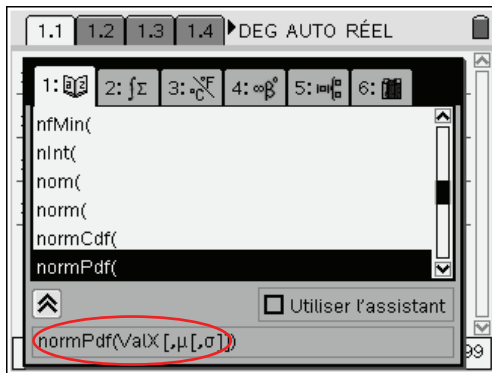
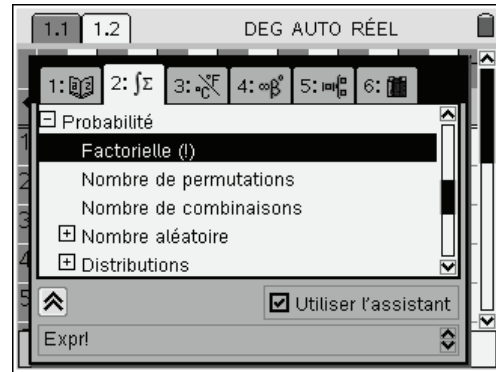
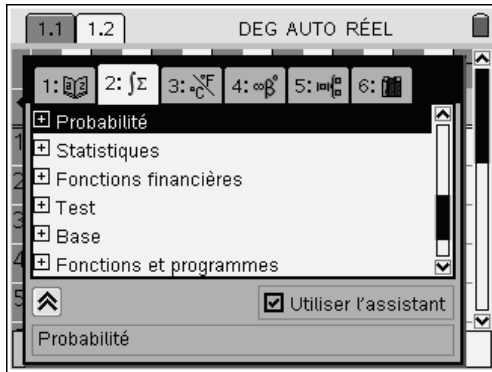
## Sommaire

1.	Les fonctions disponibles .....	2
1.1	Où trouver ces fonctions .....	2
1.2	Remarque importante sur les calculs de variance et d'écart type.....	3
2.	L'écriture de quelques fonctions utiles .....	5
2.1	Tableau de calcul, espérance, variance, écart type d'une variable aléatoire discrète.....	5
2.2	Texte des fonctions <b>tab</b> et <b>espvar</b> .....	6
3.	Utilisation en Statistiques : régression linéaire .....	7
4.	Lois discrètes usuelles.....	9
4.1	Les différentes fonctions présentes sur la TI-Nspire .....	9
4.2	Ajout de fonctions .....	11
4.3	Exemple de calcul utilisant les lois géométriques.....	12
5.	Lois continues usuelles.....	13
5.1	Utilisation directe de l'unité nomade TI-Nspire.....	13
5.2	Lois continues présentes sur la TI-Nspire.....	14
5.3	Quelques résultats classiques sur les lois normales.....	15
5.4	Approximations usuelles .....	16
5.5	Estimations et intervalles de confiance .....	20
5.6	Tests .....	24


# 1. Les fonctions disponibles

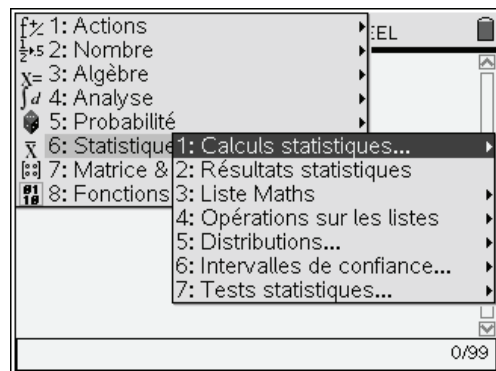
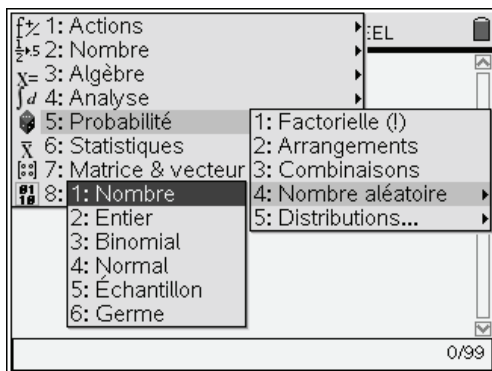
## 1.1 Où trouver ces fonctions



Vous aurez accès aux fonctions utilisables dans le catalogue () page 2, dans les rubriques : **Probabilité**, **Statistiques** mais aussi dans **Listes**. On peut y accéder aussi directement page 1 (taper la première lettre de la fonction cherchée).





☞ La syntaxe de la fonction sélectionnée se trouve affichée au bas gauche de l'écran.

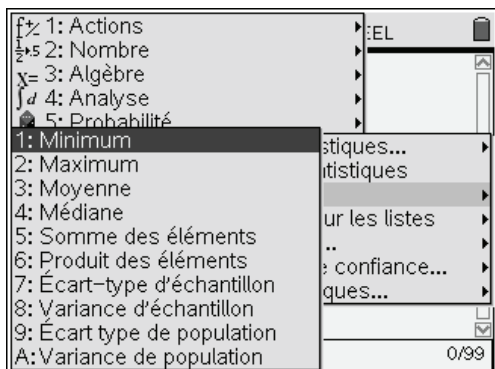
Dans l'application Calculs, on accède à ces fonctions dans les menus **Probabilité** et **Statistiques** (touches  **5**) ou **6**) :



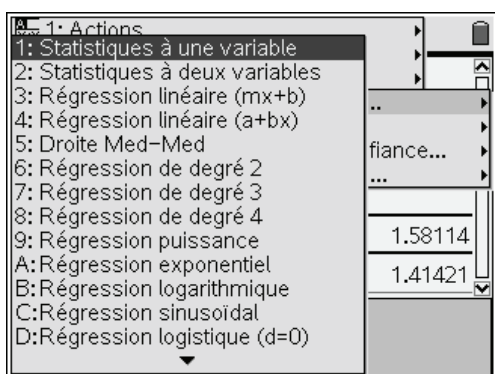
☞ Le symbole ! permettant le calcul des factorielles est disponible à l'aide du raccourci clavier  **1**) et également dans la palette de symboles () ligne 4.

Le nombre  $C_n^p$ , que l'on note  $\binom{n}{p}$ , se calcule par **nCr**( $n, p$ ),  $A_n^p$  par **nPr**( $n, p$ ) (resp.  **5**) **3**) et  **5**) **2**).

Vous trouverez d'autres fonctions utiles : calcul de la moyenne (**mean**), du maximum, du minimum, de la variance, ou encore de l'écart type des éléments d'une liste dans le menu **Liste Maths** accessible à partir du menu **Statistiques** (menu **6** **3**).



Certaines fonctions statistiques sont également utilisables à partir de l'application **Tableur & listes**, on accède à ces fonctions dans le menu **Statistiques/Calculs statistiques** (touches menu **4** **1**).



## 1.2 Remarque importante concernant le calcul de la variance et de l'écart type

**Attention**, il existe deux fonctions pour le calcul de l'écart type : **stdDevPop** et **stdDevSamp** et deux fonctions pour le calcul de la variance **varPop** et **varSamp**. L'une des deux ne donnant pas le résultat classiquement attendu en classes préparatoires.

La formule "usuelle" de calcul de la variance de la liste  $\{x_1, x_2, \dots, x_n\}$  est :

$$v = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Elle est donnée par la fonction **varPop**.

La fonction **varSamp** effectue le calcul à partir de la relation :

$$v = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

qui donne un estimateur sans biais de  $\sigma^2$ , variance de la population d'où est tiré l'échantillon  $\{x_1, x_2, \dots, x_n\}$  (voir page 20).

On utilise donc **varPop** et **stdDevPop** lorsque  $\{x_1, x_2, \dots, x_n\}$  représente la population entière, et **varSamp** et **stdDevSamp** lorsque  $\{x_1, x_2, \dots, x_n\}$  est un échantillon tiré d'une population dont on veut estimer les paramètres.

Prenons par exemple la liste  $\{1, 2, 3, 4, 5\}$ . La moyenne est 3. La variance est la moyenne des valeurs  $\{(-2)^2, (-1)^2, 0, 1^2, 2^2\}$ , c'est à dire  $10/5 = 2$ . L'écart type est égal à  $\sqrt{2}$ .

Expression	Valeur
$i := \text{seq}(i, i, 1, 5)$	$\{1, 2, 3, 4, 5\}$
$\text{varPop}(i)$	2
$\text{stDevPop}(i)$	$\sqrt{2}$
$\text{varSamp}(i)$	$\frac{5}{2}$
$\text{stDevSamp}(i)$	$\frac{\sqrt{10}}{2}$

La situation est identique dans l'application **Tableur & listes**. La valeur affichée **Sx** correspond à la valeur  $\sqrt{10}/2$ , et la valeur de  **$\sigma_x$**  correspond à la valeur  $\sqrt{2}$ . Ajoutez une page (**ctrl** **I**), choisissez l'application **Tableur & listes** (**menu** **3**). Entrez ensuite la liste dans la première colonne, puis ouvrez le menu **Statistiques à une variable** (**menu** **4** **1** **1**)

	A	B	C	D	E	F	G
1	1						
2	2						
3	3						
4	4						
5	5						
A5	5						

	A	B	C	D	E	F	G
1	1						
2	2						
3	3						
4	4						
5	5						
A5	5						

Validez en cliquant sur OK.

Statistiques à une variable
Liste des X1 : a[]
Liste des fréquences : 1
Liste des catégories :
Inclure catégories :
1ère colonne de résultat :

	A	B	C	D
			=OneVar(a[],	
3	3	$\sum x$		15.
4	4	$\sum x^2$		55.
5	5	$Sx := S_{n-1}x$	1.58114	1.58114
6		$\sigma_x := \sigma_n x$	1.41421	1.41421
D5		$= \frac{\sqrt{10}}{2}$		

Validez en cliquant sur OK, les résultats s'affichent.

Il est possible de redimensionner les colonnes afin de les rendre plus lisibles. Sélectionnez une cellule de la colonne à élargir, puis choisissez **Redimensionner, Largeur des colonnes** dans le menu contextuel accessible par  $\langle \text{ctrl} \rangle \langle \text{menu} \rangle$ . Déplacez ensuite la limite droite à l'aide du curseur validez par  $\langle \text{enter} \rangle$ . Les fonctions statistiques travaillent ici en mode approché, alors que les fonctions **varPop**, **stdDevPop**, **varSamp**, **stdDevSamp** font des calculs "exacts".

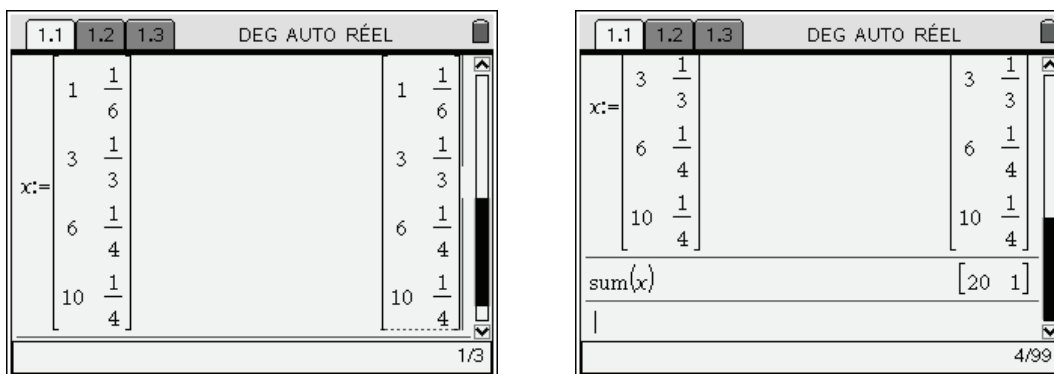
## 2. L'écriture de quelques fonctions utiles

Une bibliothèque de programmes **proba.tns** facilite certains calculs classiques, elle est téléchargeable sur le site [www.univers-ti-nspire.fr](http://www.univers-ti-nspire.fr). Lien direct : <http://www.univers-ti-nspire.fr/files/tns/proba.tns>.

### 2.1 Tableau de calcul, espérance, variance, écart type d'une variable aléatoire discrète

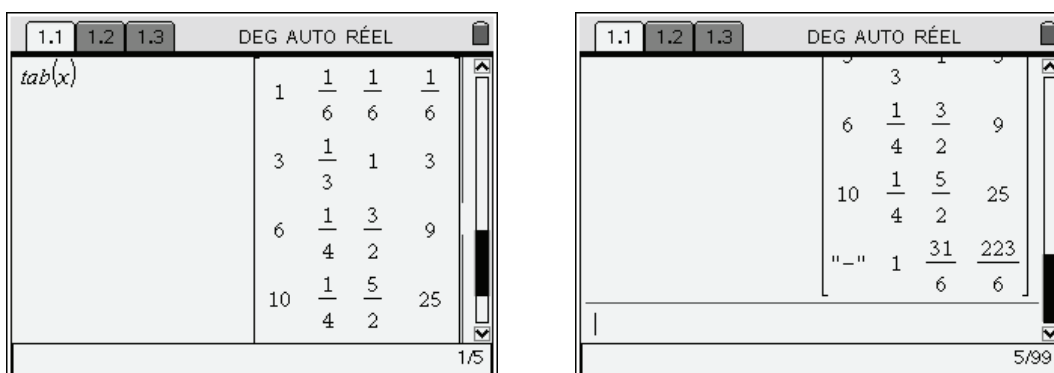
Considérons par exemple la variable aléatoire définie par  $X(\Omega) = \{1, 3, 6, 10\}$ ,  $p(X=1) = \frac{1}{6}$ ,  $p(X=3) = \frac{1}{3}$ ,  $p(X=6) = \frac{1}{4}$ ,  $p(X=10) = \frac{1}{4}$ .

Nous allons placer les éléments définissant cette variable aléatoire dans une matrice **x**. La première colonne contiendra les valeurs  $x_i$ , la seconde les probabilités  $p_i$ .



L'utilisation de **sum** ( $\langle \text{menu} \rangle \langle 6 \rangle \langle 3 \rangle \langle 5 \rangle$ ) faite dans le deuxième écran montre que la somme des probabilités est égale à 1. Jusque-là, tout va bien. Construisons à présent la matrice contenant également la colonne formée par les  $p_i x_i$  et celle formée par les  $p_i x_i^2$ .

Pour cela on peut utiliser une fonction **tab** dont vous trouverez la définition page 6 :



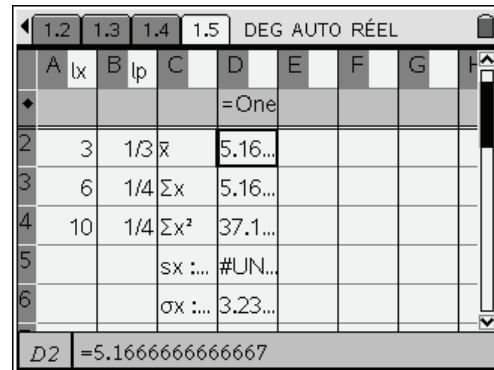
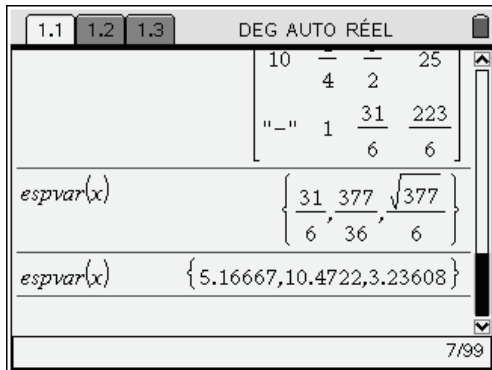
Une ligne contenant la somme des termes de chaque colonne a été ajoutée à la matrice. Cela peut faciliter la construction du tableau de calcul de l'espérance et de la variance.

On peut, par exemple, y lire que  $E(X) = \frac{31}{6}$  et  $E(X^2) = \frac{223}{6}$ .

Pour calculer ces deux dernières, il suffit d'utiliser la fonction **espvar**, que vous trouverez également au paragraphe suivant, qui utilise la fonction précédente pour retourner une liste formée par l'espérance, la variance, et l'écart type. Vous pouvez retrouver ces résultats à l'aide de l'application

Tableur & listes, entrez dans la colonne A la liste  $\{1,3,6,10\}$ , dans la colonne B, la liste  $\left\{\frac{1}{6}, \frac{1}{3}, \frac{1}{4}, \frac{1}{4}\right\}$ .

(menu) (4) (1) (1), nombre de listes : 1, Liste des  $x_i$  : a[ ], Liste des fréquences : b[ ], on valide (voir écran ci-dessous à gauche).



## 2.2 Texte des fonctions tab et espvar

L'écriture de ces deux fonctions utilise des fonctions de calcul matriciel, ce qui en fait l'intérêt. Si vous n'êtes pas familiarisé avec ces dernières, cela risque de vous paraître un peu mystérieux. Voici quelques explications permettant d'en suivre le fonctionnement. Ces explications sont beaucoup plus longues que la fonction **tab** !

☞ Ces calculs peuvent être également faits de façon très simple dans l'application **Tableur & listes**.

- $\mathbf{x}^*[i]$  forme le vecteur ligne obtenu à partir de la  $i$ -ième colonne de la matrice  $\mathbf{x}$  passée en argument. Cette matrice comporte les valeurs de  $x_i$  sur la première colonne, et celle de  $p_i$  dans la seconde.
- La fonction **mat▶list** permet de convertir ce vecteur ligne en liste. On obtient les listes **lx** et **lp** des valeurs et probabilités à utiliser pour les calculs suivants.
- À partir des quatre listes, **lx**, **lp**, **lx\*lp**, **lx^2\*lp**, il est possible de construire une matrice de quatre lignes contenant les valeurs de  $x_i$ , de  $p_i$ , de  $p_i x_i$  et de  $p_i x_i^2$ .
- En la transposant, on obtient la matrice **m** représentant le tableau dans sa présentation classique avec ses quatre colonnes.
- La fonction **sum** permet de faire la somme de chacune de ces colonnes et d'obtenir la matrice
 
$$\left[ \sum_{i=1}^n x_i \quad \sum_{i=1}^n p_i \quad \sum_{i=1}^n p_i x_i \quad \sum_{i=1}^n p_i x_i^2 \right].$$
- On "empile" ensuite cette matrice et la matrice **m** pour former le tableau dont la dernière ligne comporte les sommes de chaque colonne. Cela se fait en utilisant la fonction **colAugment**.
- Enfin, on modifie le terme situé sur la première colonne de la dernière ligne. Il correspond au cumul des valeurs de  $x_i$ , et n'est pas utile pour la suite. On utilise la fonction **rowDim** pour connaître le nombre de lignes de la matrice.

☞ L'opérateur de transposition  $^T$  s'obtient dans le menu (mat) (2) **Matrice**.

```

Define LibPub tab(mat)=Func
©mat:tableau de calcul proba/stat
Local lx,lp,u,v,m
lx:=mat▶list(matt[1])
lp:=mat▶list(matt [2])
m:={lx,lp,lx*lp,lx^2*lp}t
m:=colAugment(m,sum(m))
m[rowDim(m),1]:="--"
m
EndFunc

```

☞ *L'utilisation de la syntaxe*

Define LibPub... =

*permet de faire apparaître cette fonction dans le catalogue, voir [chapitre 17](#).*

La fonction **espvar** est beaucoup plus simple à comprendre.

On pourrait l'écrire directement, mais on peut aussi utiliser la fonction **tab** précédente. Il suffit de construire la matrice précédente et d'y aller chercher les informations utiles, c'est-à-dire les valeurs de

$\sum_{i=1}^n p_i x_i$  et de  $\sum_{i=1}^n p_i x_i^2$ , désignées par **ex** et **ex2** dans cette fonction.

```

Define LibPub espvar(x)=Func
©mat:calcul de esp et var
Local m,k,ex,ex2,v
m:= tab(x)
k:=rowDim(m)
ex:=m[k,3]
ex2:=m[k,4]
v:=ex2-ex^2
{ex,v,√(v)}
EndFunc

```

### 3. Utilisation en Statistique : régression linéaire

La statistique suivante donne l'évolution des stocks d'une entreprise :

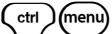
Année	2000	2001	2002	2003	2004	2005
Stock $Q_i$	6 400	7 200	8 700	10 400	12 600	15 000





Effectuer un ajustement à l'aide d'une régression linéaire, puis un ajustement à l'aide d'une fonction exponentielle. Que peut-on en conclure ? Donner une estimation du stock en 2009.

#### Solution

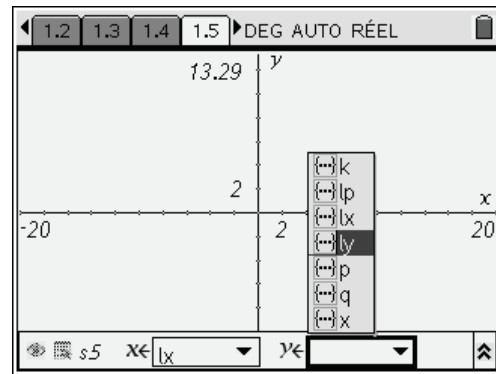
On entre dans l'éditeur la liste des numéros des années et la liste des  $Q_i$ .

$lx = \{1, 2, 3, 4, 5, 6\}$ ,  $ly = \{6400, 7200, 8700, 10400, 12600, 15000\}$ .

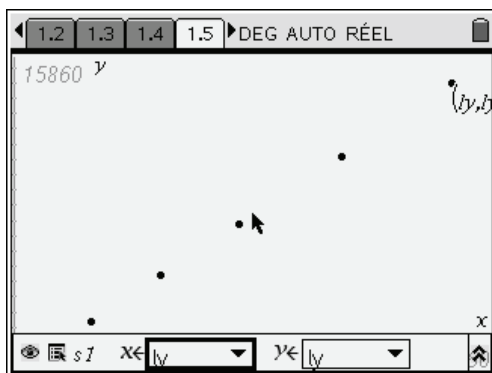
Pour représenter le nuage de points, on peut le faire sur la même page à l'aide de **Graphe rapide** que vous trouverez dans le menu contextuel accessible par .

On peut aussi ajouter une page Graphiques & géométrie en utilisant  , puis sélectionner **Nuage de points** dans le menu contextuel. Il suffit ensuite d'entrer les deux listes : appuyer sur , choisir la liste et valider,  pour passer à l'autre liste.

	A	B	C	D
	lx	ly		
	=seq(i,1,1,6)			
1	1	6400		
2	2	7200		
3	3	8700		
4	4	10400		
5	5	12600		



On valide, (esc), pour sortir de la ligne de saisie, (ctrl) (menu) (4) (9) pour choisir un zoom adapté aux données. Effectuons une régression linéaire, (ctrl) (left arrow) pour revenir à la page précédente, puis (menu) (4) (1) (3).



- 1: Actions
- 1: Statistiques à une variable
  - 2: Statistiques à deux variables
  - 3: Régression linéaire (mx+b)
  - 4: Régression linéaire (a+bx)
  - 5: Droite Med-Med
  - 6: Régression de degré 2
  - 7: Régression de degré 3
  - 8: Régression de degré 4
  - 9: Régression puissance
  - A: Régression exponentiel
  - B: Régression logarithmique
  - C: Régression sinusoidal
  - D: Régression logistique (d=0)

Régression linéaire (mx+b)

Liste des X : 'lx

Liste des Y : 'ly

Enregistrer RegEqn dans :

Liste des fréquences : 'y

Liste des catégories :

	B	C	D
	ly		
5)			=LinRegMx('lx','ly,
1	1	6400	Titre
2	2	7200	RegEqn
3	3	8700	m
4	4	10400	b
5	5	12600	r <sup>2</sup>
			.974042
D1	="Régression linéaire (mx+b)"		

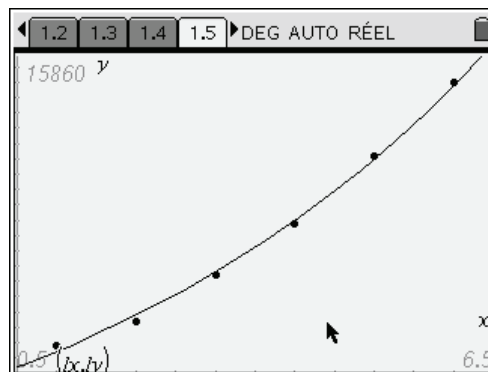
L'ajustement linéaire donne un coefficient de corrélation  $r \simeq 0,987$ .

☞ La troisième rubrique **Enregistrer RegEqn dans :** permet d'indiquer le nom de la variable dans laquelle sera mémorisée l'équation de la fonction servant à effectuer l'ajustement (f1 pour l'ajustement linéaire et f2 pour l'exponentiel).

(menu) (4) (1) (A). L'ajustement par une fonction de la forme  $Q = a * b^t$  donne un meilleur résultat comme l'indique la représentation graphique et le coefficient de corrélation  $r \simeq 0,998$ .



D	E	F
=LinRegMx('lx','ly,		=ExpReg('lx','ly,1):
m*x+b	Reg...	a*b^x
1740.	a	5213.58
3960.	b	1.19094
.974042	r <sup>2</sup>	.996284
.986936	r	.99814
F6 =.99814012936929		



Pour avoir l'estimation du stock en 2009, il suffit de calculer f2(10), ce qui donne approximativement 29 926, alors que l'ajustement linéaire aurait donné 21 360 (f1(10)).

## 4. Lois discrètes usuelles

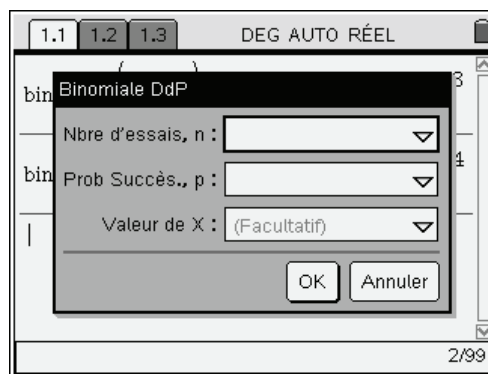
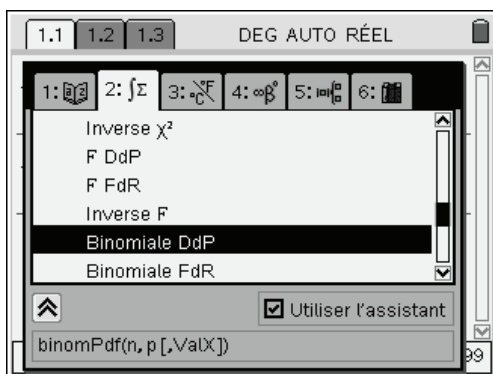
### 4.1 Les différentes fonctions présentes sur la TI-Nspire

Les lois discrètes usuelles : loi binomiale, loi géométrique, loi de Poisson sont directement intégrées. Il sera de plus très simple de définir d'autres lois lorsque vous en aurez l'utilité, voir par exemple la loi hypergéométrique.

Vous trouverez les fonctions correspondantes dans le menu **Distributions** du menu **Statistiques** (touches  $\text{menu}$  6 5) de l'application **Calculs**, et dans le menu **Distributions** du menu **Statistiques** (touches  $\text{menu}$  4 2) de l'application **Tableur & listes**.

Vous aurez aussi accès à ces fonctions dans le catalogue ( $\text{menu}$  2) page 2, dans les rubriques : **Probabilité**, et **Statistiques** sous-menu **Distributions**.

On peut soit utiliser l'assistant (voir écrans ci-dessous), soit entrer directement les paramètres.



- Loi binomiale de paramètres  $n$  et  $p$ ,  $n \in \mathbb{N}^*$ ,  $p \in ]0,1[$

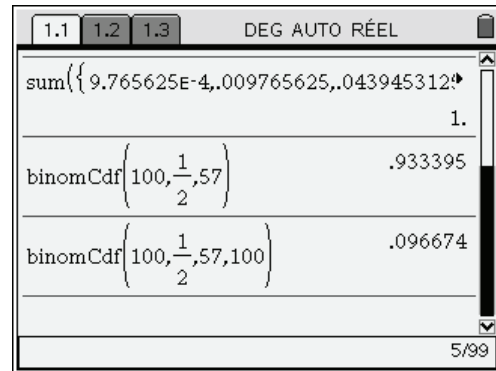
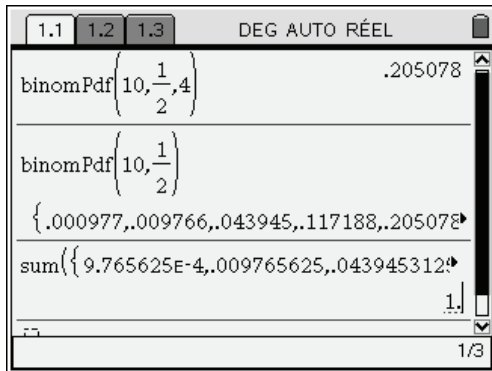
$$\text{On a } p(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \text{ pour tout } k \in \llbracket 0, n \rrbracket.$$

syntaxe : **binomPdf**( $n, p [, k]$ )<sup>1</sup> donne  $P(X = k)$

<sup>1</sup> Si  $k$  est omis, **binomPdf**( $n, p$ ) (resp. **binomCdf**( $n, p$ )) donne la liste des probabilités  $P(X = k)$  (resp.  $P(X \leq k)$ ) pour  $k \in \llbracket 0, n \rrbracket$ .

fonction de répartition : **binomCdf**( $n, p, [a, b]$ ) donne  $P(X \leq k)$  et **binomCdf**( $n, p, a, b$ ) donne  $P(a \leq X \leq b)$ .

Il est possible de calculer par exemple la probabilité d'avoir 4 "pile" lorsqu'on lance 10 fois une pièce de monnaie, et de vérifier que la somme des probabilités lorsque  $k$  varie de 0 à  $n$  est bien égale à 1. Dans le second écran, on calcule la probabilité d'obtenir au plus 57 "pile" au cours de 100 lancers ( $\approx 0,93$ ) et la probabilité d'obtenir au moins 57 "pile" au cours de 100 lancers ( $\approx 0,097$ ).



- Loi géométrique de paramètre  $p$  ( $p \in ]0, 1[$ )

Soit une expérience élémentaire dont l'issue est un succès ou un échec avec des probabilités respectives  $p$  et  $q = 1 - p$ . On renouvelle cette expérience jusqu'à l'obtention d'un succès. Cette loi caractérise le nombre d'expériences nécessaires pour obtenir le premier succès.  $P(X = k) = p(1 - p)^{k-1}$ ,  $k \in \mathbb{N}^*$

$$E(X) = \frac{1}{p} \text{ et } V(X) = \frac{q}{p^2}$$

syntaxe : **geomPdf**( $p, k$ )

fonction de répartition : **geomCdf**( $p, [a, b]$ ) donne  $P(a \leq X \leq b)$ , (par défaut  $a$  est égal à 1).

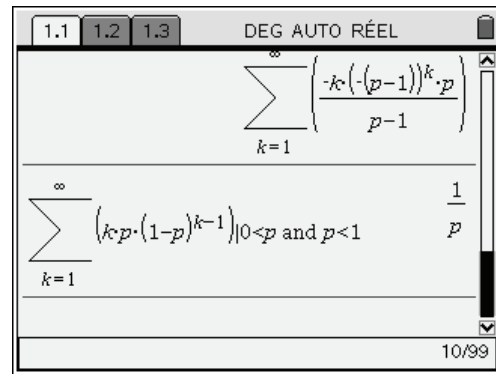
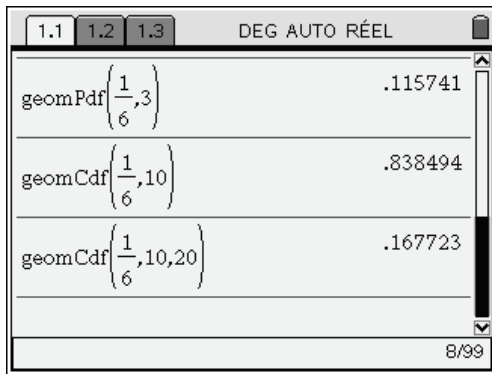
La fonction **geomPdf** permet par exemple de calculer la probabilité d'obtenir un "6" ou bout de 3 lancers d'un dé (non pipé). La fonction **geomCdf** permet dans l'exemple ci-dessous de calculer la probabilité d'obtenir un "6" en au plus 10 lancers, puis en au plus 20 lancers et au moins 10.

*Remarque* : pour calculer la probabilité  $P(X > k)$ , on peut utiliser la formule :

$$P(X > k) = 1 - P(X \leq k) = 1 - \mathbf{geomCdf}(p, k).$$

Le deuxième écran montre le calcul de l'espérance mathématique de la loi géométrique de paramètre  $p$ , l'utilisation de **geomPdf** ne permet pas ici d'obtenir le résultat, les valeurs données par cette fonction sont des valeurs approchées, il faut donc entrer la 0 ci-dessus valeur formelle de la probabilité et de plus préciser que  $0 < p < 1$  à l'aide de l'opérateur "sachant que"  $\textcircled{1}$ . On peut calculer de même  $E(X^2)$  et en déduire la variance :

$$E(X) = \frac{1}{p} \text{ et } V(X) = \frac{1-p}{p^2}.$$

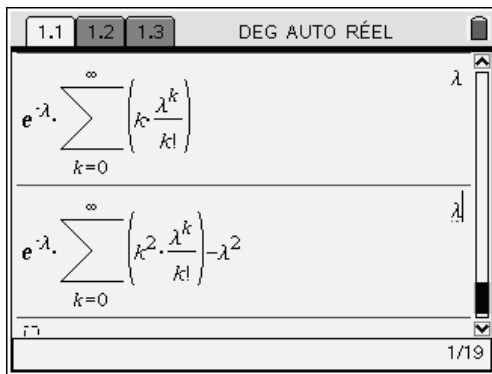


- Loi de Poisson de paramètre  $\lambda \in \mathbb{R}_+^*$

syntaxe : **poissPdf**( $\lambda, k$ ) donne  $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$ , pour  $k \in \mathbb{N}$

fonction de répartition : **poissCdf**( $\lambda, [a, b]$ ) donne  $P(a \leq X \leq b)$ . Par défaut  $a$  est égal à 0.

On peut vérifier que l'espérance mathématique et la variance sont égales au paramètre  $\lambda$ .



## 4.2 Ajout de fonctions

Comme on l'a vu dans les exemples ci-dessus les fonctions considérées donnent des résultats en valeur approchée. Si vous avez besoin de résultats formels sur une TI-Nspire CAS, comme on le voit dans les deux derniers exemples, vous pouvez réécrire ces fonctions comme ci-dessous pour la loi hypergéométrique. Vous retrouverez ces fonctions dans la bibliothèque **proba**, téléchargeable sur le site [www.univers-ti-nspire.fr](http://www.univers-ti-nspire.fr). Lien direct : <http://www.univers-ti-nspire.fr/files/tns/proba.tns>.

- Loi hypergéométrique  $N, n$  et  $p$  :

Define LibPub lhyp( $m, n, p, k$ ) =  $nCr(m^*p, k) * nCr(m^*(1-p), n-k) / nCr(m, n)$

Define LibPub fhyp( $m, n, p, k$ ) =  $\sum (nCr(m^*p, i) * nCr(m^*(1-p), n-i), i, 0, k) / nCr(m, n)$

☞ On a utilisé **m** car **N** et **n** sont interprétés de la même façon par l'unité nomade TI-Nspire. De plus, dans la définition des fonctions de répartition, on a "manuellement" fait la mise en facteurs nécessaire pour diminuer le temps nécessaire au calcul.

Voici les définitions des fonctions donnant la loi de probabilité (l) et la fonction de répartition (f) des trois lois usuelles précédentes : binomiale, géométrique et Poisson :

Define LibPub lbinom( $n, p, k$ ) =  $nCr(n, k) * p^k * (1-p)^{(n-k)}$

Define LibPub fbinom( $n, p, k$ ) =  $\sum (lbinom(n, p, i), i, 0, k)$

Define LibPub lgeom( $p, k$ ) =  $p * (1-p)^{(k-1)}$

Define LibPub fgeom( $p, k$ ) =  $\sum (lgeom(p, i), i, 1, k)$

Define LibPub IPoisson( $\lambda, k$ )= $e^{-(\lambda)} \cdot \lambda^k / (k!)$   
 Define LibPub fPoisson( $\lambda, k$ )= $\sum (\lambda^i / (i!), i, 0, k) \cdot e^{-(\lambda)}$

### 4.3 Exemple de calcul utilisant les lois géométriques

On considère une variable aléatoire  $X$  suivant une loi géométrique de paramètre  $p = 1/10$ .

Nous allons calculer les probabilités  $P(X = 2)$  et  $P(X > 3)$ , puis calculer l'espérance et la variance. Cela ne pose aucun problème en utilisant les fonctions décrites dans ce chapitre.

1. Calcul de  $P(X = 2)$  et de  $P(X > 3) = 1 - P(X \leq 3)$  (avec vérification).

TI-Nspire CAS screenshot showing the calculation of  $P(X = 2)$  and  $P(X > 3)$  using built-in functions. The screen displays:

$\text{geomPdf}\left(\frac{1}{10}, 2\right)$	.09
$1 - \text{geomCdf}\left(\frac{1}{10}, 1, 3\right)$	.729

The status bar at the bottom indicates page 2/99.

TI-Nspire CAS screenshot showing the calculation of  $P(X > 3)$  using the  $1 - \text{geomCdf}$  function and a summation formula. The screen displays:

$1 - \text{geomCdf}\left(\frac{1}{10}, 1, 3\right)$	.729
$1 - \sum_{k=1}^3 \left( \frac{1}{10} \cdot \left(\frac{9}{10}\right)^{k-1} \right)$	.729

The status bar at the bottom indicates page 3/99.

TI-Nspire CAS screenshot showing the calculation of  $P(X > 3)$  using a summation formula and its simplified form. The screen displays:

$1 - \sum_{k=1}^3 \left( \frac{1}{10} \cdot \left(\frac{9}{10}\right)^{k-1} \right)$	.729
$\left(\frac{9}{10}\right)^3$	.729

The status bar at the bottom indicates page 4/99.

☞ (On a également fait un calcul direct de cette probabilité :  $X > 3$ , ou encore  $X \geq 4$  signifie que les trois premiers essais ont été des échecs.)

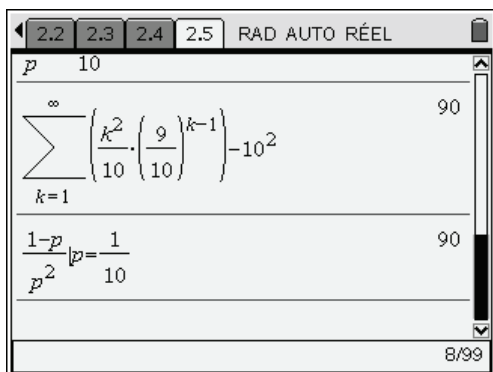
3. Calcul de l'espérance.

TI-Nspire CAS screenshot showing the calculation of the expected value of  $X$ . The screen displays:

$\sum_{k=1}^{\infty} \left( \frac{k}{10} \cdot \left(\frac{9}{10}\right)^{k-1} \right)$	10
$\frac{1}{p} = \frac{1}{1/10}$	10

The status bar at the bottom indicates page 6/99.

4. Calcul de la variance et vérification.



## 5. Lois continues usuelles

### 5.1 Utilisation directe de l'unité nomade TI-Nspire

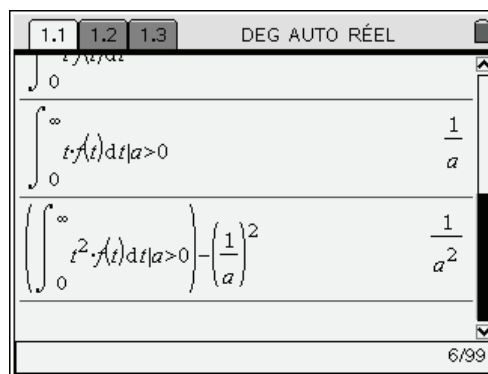
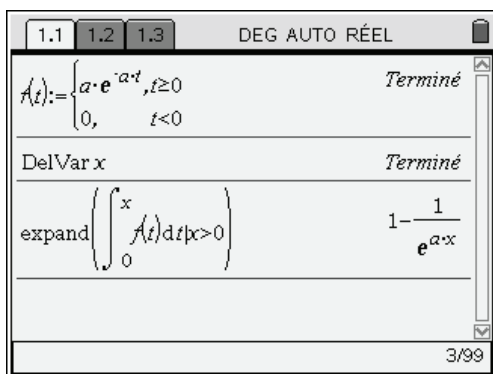
Rappelons pour commencer la définition de deux lois continues classiques, non existantes dans les menus, mais facilement programmables.

1. La loi uniforme sur  $[a, b]$  :  $d(t) = \begin{cases} \frac{1}{b-a} & t \in [a, b] \\ 0 & t \notin [a, b] \end{cases}$ .
2. La loi exponentielle de paramètre  $a$  :  $d(t) = \begin{cases} a e^{-at} & t \geq 0 \\ 0 & t < 0 \end{cases}$ .

La détermination des fonctions de répartition des deux premières ne pose pas de problème :

1. La loi uniforme sur  $[a, b]$  :  $F(t) = \begin{cases} 0 & t \in ]-\infty, a[ \\ \frac{t-a}{b-a} & t \in [a, b] \\ 1 & t \in [b, +\infty[ \end{cases}$ .
2. La loi exponentielle de paramètre  $a > 0$  :  $F(t) = \begin{cases} 0 & t \in ]-\infty, 0] \\ 1 - e^{-at} & t \in [0, +\infty[ \end{cases}$ .

On peut vérifier ce dernier résultat avec la calculatrice. On peut de la même façon calculer l'espérance et la variance de cette loi, à condition de bien préciser le signe de  $a$  (écran de droite).



Des calculs de ce type pourront être faits de la même façon avec d'autres lois que vous rencontrerez dans des exercices.

Vous retrouverez les fonctions concernant la loi uniforme et la loi exponentielle dans la bibliothèque **proba**, téléchargeable sur le site [www.univers-ti-nspire.fr](http://www.univers-ti-nspire.fr).

Lien direct : <http://www.univers-ti-nspire.fr/files/tns/proba.tns>.

## 5.2 Lois continues présentes sur la TI-Nspire

- Loi normale de moyenne  $\mu$  et d'écart type  $\sigma$

**normPdf** (densité de probabilité de la loi normale)

calcule la valeur de la densité de probabilité de la loi normale de moyenne  $\mu$  et d'écart type  $\sigma$ , en un réel  $x$  spécifié. Les valeurs par défaut sont  $\mu = 0$  et  $\sigma = 1$ . La densité de probabilité est définie par :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, (\sigma > 0)$$

syntaxe : **normPdf**( $x$  [,  $\mu$ ,  $\sigma$ ])

fonction de répartition : **normCdf**( $a$ ,  $b$ , [,  $\mu$ ,  $\sigma$ ]) donne la probabilité  $P(a \leq X \leq b)$ .

**InvNorm** permet de calculer la valeur  $x$  telle que  $P(X \leq x) = \alpha$ ,  $\alpha \in [0,1]$ , lorsque  $X$  suit la loi normale de moyenne  $\mu$  et d'écart type  $\sigma$ .

Syntaxe : **invNorm**( $\alpha$  [,  $\mu$ ,  $\sigma$ ]).

- Loi de Student à  $df$  degrés de liberté ( $df \in \mathbb{N}^*$ ).

**tPdf** (densité d'une loi de Student)

calcule la densité de probabilité de la loi de Student à  $df$  degrés de liberté en un réel  $x$  spécifié. La densité de probabilité est définie par :

$$f(x) = \frac{\Gamma((df+1)/2) (1+x^2/df)^{-(df+1)/2}}{\Gamma(df/2) \sqrt{\pi df}}$$

$$\text{où } \Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$$

$$E(X) = 0 \text{ et } V(X) = \frac{df}{df-2} \text{ si } df > 2$$

syntaxe : **tPdf**( $x$ ,  $df$ )

fonction de répartition : **tCdf**( $a$ ,  $b$ ,  $df$ ) donne la probabilité  $P(a \leq X \leq b)$ .

- Loi de Fisher à  $df_1$  et  $df_2$  degrés de liberté ( $(df_1, df_2) \in \mathbb{N}^{*2}$ )

**FPdf** (densité d'une loi de Fisher)

calcule la densité de probabilité de la distribution de Fisher en un réel  $x \in \mathbb{R}_+^*$  spécifié. La densité de probabilité s'exprime sous la forme :

$$f(x) = \frac{\Gamma((n+d)/2)}{\Gamma(n/2)\Gamma(d/2)} \left(\frac{n}{d}\right)^{n/2} x^{n/2-1} (1+nx/d)^{-(n+d)/2}$$

avec :  $n = df_1$  et  $d = df_2$ .

$$E(X) = \frac{df_2}{df_2 - 2} \text{ si } df_2 > 2 \text{ et}$$

$$V(X) = \left( \frac{df_2}{df_2 - 2} \right)^2 \frac{2(df_1 + df_2 - 2)}{df_1(df_2 - 4)} \text{ si } df_2 > 4.$$

syntaxe : **FPdf**( $x, df_1, df_2$ )

fonction de répartition : **FCdf**( $a, b, df_1, df_2$ ) donne la probabilité  $P(a \leq X \leq b)$ .

- Loi du Khi 2 ( $\chi^2$ ) (ou de Pearson) à  $df$  degrés de liberté ( $df \in \mathbb{N}^*$ )

**chi2Pdf** (densité de probabilité d'une loi du Khi 2)

calcule la densité de probabilité de la loi du Khi 2, en une valeur spécifiée  $x \in \mathbb{R}_+^*$ . La densité de probabilité est définie par :

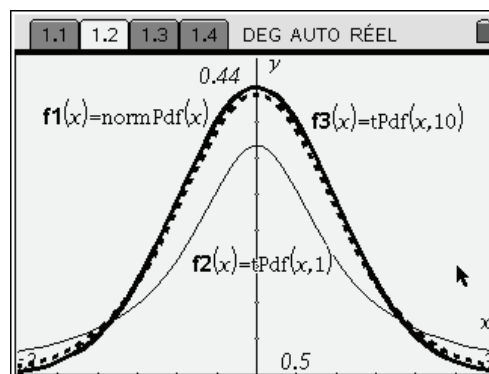
$$f(x) = \frac{1}{\Gamma(df/2)} (1/2)^{df/2} x^{df/2-1} e^{-x/2}, x > 0$$

$$E(X) = df \text{ et } V(X) = 2df$$

syntaxe : **chi2Pdf**( $x, df$ )

fonction de répartition : **chi2Cdf**( $a, b, df$ ) donne la probabilité  $P(a \leq X \leq b)$ .

**Exemple** : représentation de la loi normale centrée réduite (trait épais) et des lois de Student de, respectivement, 1 (trait fin) et 10 (en pointillés gras) degrés de liberté.



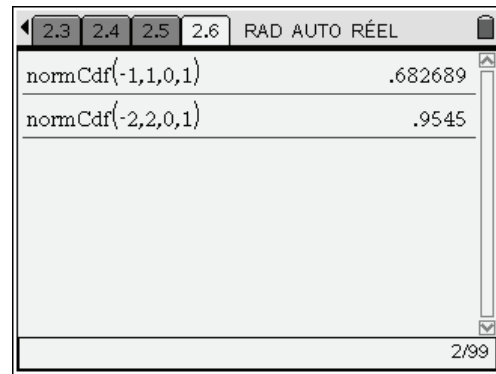
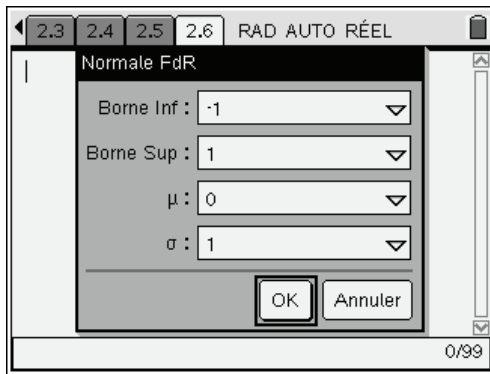
### 5.3 Quelques résultats classiques sur les lois normales

On considère une loi normale de paramètres  $m$  et  $\sigma$ .

1. On demande de calculer  $P(m - \sigma \leq X \leq m + \sigma)$  et  $P(m - 2\sigma \leq X \leq m + 2\sigma)$ .
2. On demande également de déterminer, en fonction de  $\sigma$ , la valeur de  $a$  telle que  $P(m - a \leq X \leq m + a) = 85/100$ .

Pour la question 1, il suffit d'utiliser les fonctions **normCdf**, avec les bornes  $-1, 1$  puis  $-2, 2$  et les paramètres  $m = 0$  et  $\sigma = 1$ , car :

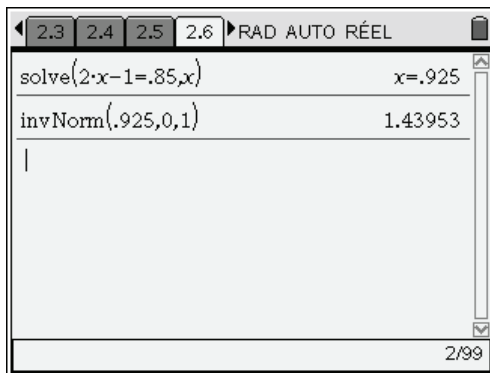
$$P(m - a.\sigma \leq X \leq m + a.\sigma) = P\left(-a \leq \frac{X - m}{\sigma} \leq a\right)$$



Deuxième question.

$$P(m-a \leq X \leq m+a) = 85/100 \text{ équivaut à } P(|X-m| < a) = P\left(\left|\frac{X-m}{\sigma}\right| < \frac{a}{\sigma}\right) \simeq 2\Phi\left(\frac{a}{\sigma}\right) - 1 = 0.85$$

où  $\Phi$  est la fonction de répartition de la loi normale centrée réduite. La fonction **invNorm** permet d'obtenir le résultat :  $a \simeq 1.44\sigma$ .



## 5.4 Approximations usuelles

On peut visualiser à l'aide de la calculatrice les approximations usuelles des diverses lois.

Si  $\begin{cases} n \geq 30 \\ np \geq 15 \\ npq > 5 \end{cases}$  la loi binomiale  $\mathcal{B}(n, p)$  peut être approximée par la loi normale  $\mathcal{N}(np, \sqrt{npq})$ .

Si  $\begin{cases} p \leq 0,1 \\ n \geq 30 \\ np < 15 \end{cases}$  la loi binomiale  $\mathcal{B}(n, p)$  peut être approximée par la loi de Poisson  $\mathcal{P}(np)$ .

Si  $\lambda \geq 15$  la loi de Poisson  $\mathcal{P}(\lambda)$  peut être approximée par la loi normale  $\mathcal{N}(\lambda, \sqrt{\lambda})$ .

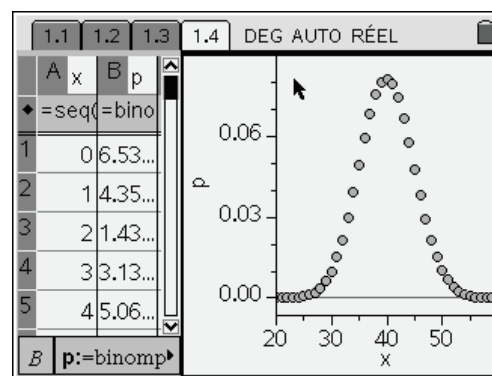
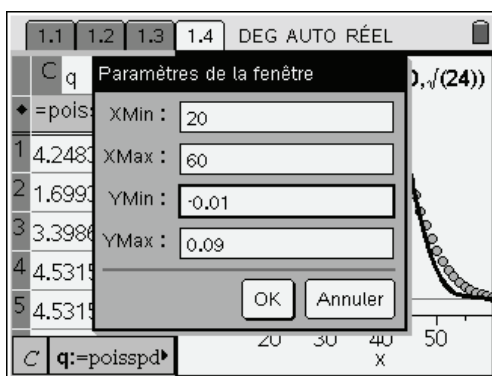
Si  $df \geq 30$  la loi de Student à  $df$  degrés de liberté peut être approximée par la loi normale centrée réduite (voir représentation graphique page 15).

Les graphiques suivants illustrent ces problèmes d'approximation. On travaille avec l'application **Tableur & listes**, dans la colonne A on place les entiers de 0 à 100 (= **seq**( $i, i, 0, 100$ )), on la nomme x. Dans la colonne B, on place les probabilités de la loi binomiale de paramètres  $n=100$  et  $p=0,4$ ,  $\mathcal{B}(100; 0,4)$  (= **binomPdf**(100,0.4)), on la nomme p, et dans la colonne C les probabilités de la loi de Poisson de paramètre 40,  $\mathcal{P}(40)$  (= **poissPdf**(40,a[])), on la nomme q.

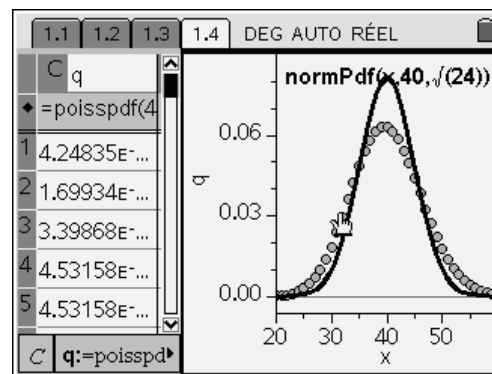
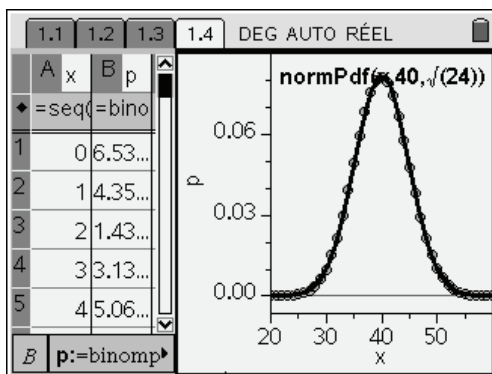


1.1	1.2	1.3	1.4
A x	B p	C q	
♦=seq(=binompdf(100,.4)=poisspdf(40,a[ ]))			
1	0	6.53319E-23	4.24835E-18
2	1	4.35546E-21	1.69934E-16
3	2	1.4373E-19	3.39868E-15
4	3	3.13012E-18	4.53158E-14
5	4	5.06036E-17	4.53158E-13
C q:=poisspdf(40,a[ ])			

On sélectionne les deux premières colonnes. **(ctrl) (menu) 8 (Graphe rapide)** permet de représenter la première loi, **(ctrl) 5 1** permet de modifier le partage d'écran et **(ctrl) (menu) 1 1** de régler la fenêtre.



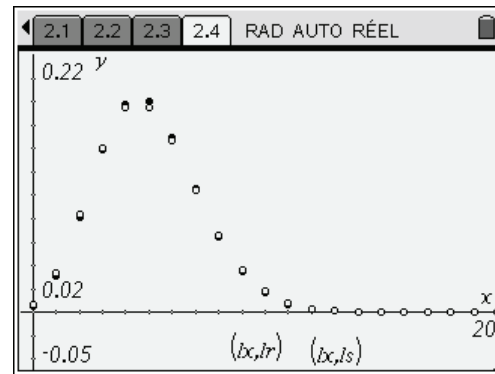
**(menu) 4 9** permet de tracer la courbe représentative de la densité de la loi normale de paramètres  $np = 40$  et  $\sqrt{npq} = \sqrt{24}$ . Pour obtenir l'écran de droite, on clique sur la lettre p représentant l'axe des ordonnées, et on choisit la variable q.



On voit bien que, pour ces valeurs des paramètres, l'approximation par la loi normale  $\mathcal{N}(40, \sqrt{24})$  est bonne, alors qu'elle ne l'est pas par la loi de Poisson.

On peut également étudier un cas où la loi de Poisson donne une bonne approximation de la loi binomiale. On colle dans la première colonne la liste lx définie par **seq(x,x,0,100)** (**seq** s'obtient dans le catalogue, ou si l'on connaît la syntaxe peut être tapé directement), puis dans lr : **binomPdf(100, 0.05)** et dans ls : **poissPdf(5, lx)**. Plutôt que d'utiliser comme ci-dessus **Graphe rapide**, on insère une nouvelle page avec l'application **Graphiques & géométrie**, ce qui permet d'avoir des représentations plus lisibles et on définit les deux nuages de points (lx, lr), (lx, ls).

A	lx	B	lr	C	ls
			=binompdf(100,.05)		=poisspdf(5,lx)
1	0		.005921		.006738
2	1		.031161		.03369
3	2		.081182		.084224
4	3		.139576		.140374
5	4		.178143		.175467



Ils se superposent presque parfaitement ( $p = 0,05 \leq 0,1$ ,  $n = 100 \geq 30$  et  $np = 5 < 15$ ).

### Exemple 1

On lance une paire de dés non pipés,  $X_i$  représente la somme des deux nombres obtenus au  $i$ -ième lancer. Déterminer le nombre de lancers pour que la moyenne des résultats  $X_i$  diffère de 7 de moins de 0,1 avec une probabilité supérieure à 0,95.

Les  $X_i$  constituent une suite de variables indépendantes, de même loi de moyenne 7 et d'écart type  $\sqrt{\frac{35}{6}}$ .

Le théorème central limite permet de dire que la variable représentant la moyenne  $F_n = \frac{X_1 + X_2 + \dots + X_n}{n}$  centrée, réduite,  $Y_n = \frac{F_n - 7}{\sqrt{\frac{35}{6n}}}$  converge en loi vers la variable normale

centrée réduite, quand  $n$  tend vers  $+\infty$ .

En utilisant cette approximation, nous avons donc :

$$P(|F_n - 7| < 0,1) = P(|Y_n| < 0,1 \cdot \sqrt{\frac{6n}{35}}) \simeq 2\Phi\left(0,1 \cdot \sqrt{\frac{6n}{35}}\right) - 1$$

où  $\Phi$  est la fonction de répartition de la loi normale centrée réduite. Nous voulons donc trouver  $n$  tel que :

$$2\Phi\left(0,1 \cdot \sqrt{\frac{6n}{35}}\right) - 1 \geq 0,95, \text{ soit } \Phi\left(0,1 \cdot \sqrt{\frac{6n}{35}}\right) \geq 0,975$$

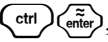
1.1	1.2	1.3	1.4	DEG AUTO RÉEL
				invNorm(.975)
				1.95996
				solve(.1 * sqrt(6*n/35) = 1.96, n)
				n=2240.93

On doit avoir  $0,1 \cdot \sqrt{\frac{6n}{35}} \geq 1,96$ , il suffit donc de prendre  $n \geq 2241$ .

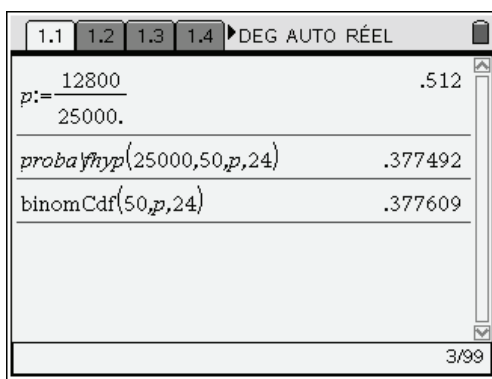
### Exemple 2 : Approximation de la loi hyper-géométrique par une loi binomiale, puis par une loi normale

Un jour d'élections, dans une ville comptant 25 000 votants, 12 800 votent pour la municipalité sortante. On dépouille 50 bulletins tirés au hasard parmi les 25 000. Quelle est la probabilité que, sur cet échantillon, la majorité sortante soit minoritaire ? (Au maximum 24 bulletins favorables.)

On se trouve ici en présence d'une variable aléatoire suivant une loi hypergéométrique de paramètres  $N = 25\,000$ ,  $p = 12\,800 / 25\,000$  et  $n = 50$ .

Attention, si vous demandez un calcul exact, vous risquez d'obtenir un résultat totalement inutilisable et un temps de calcul important. Il est préférable soit de valider par , soit de mettre un point à la fin d'un des nombres (comme le montre l'écran précédent) pour que le calcul se fasse en mode approché.

On peut comparer avec le résultat que l'on obtiendrait en prenant une loi binomiale de paramètres  $n = 50$  et  $p = 12\,800 / 25\,000$  :

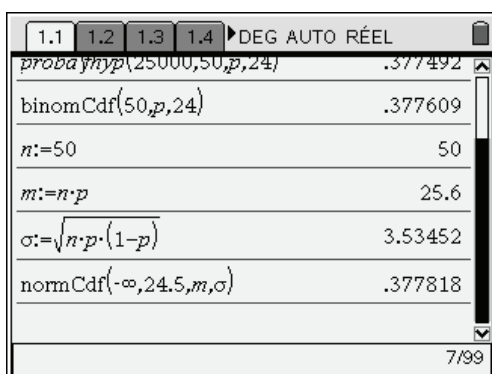


Expression	Résultat
$p := \frac{12800}{25000}$	.512
$\text{proba}^{\text{hyp}}(25000, 50, p, 24)$	.377492
$\text{binomCdf}(50, p, 24)$	.377609

Le résultat obtenu est très proche du précédent, ce qui est conforme à la théorie sur l'approximation par une loi binomiale dans le cas où  $N > 10n$ .

On peut enfin effectuer une nouvelle approximation, en utilisant à présent une loi normale de paramètres  $m = np$  et  $\sigma = \sqrt{np(1-p)}$ , et en utilisant une correction de continuité.

On calcule  $P(X \leq 24,5)$  :



Expression	Résultat
$\text{proba}^{\text{hyp}}(25000, 50, p, 24)$	.377492
$\text{binomCdf}(50, p, 24)$	.377609
$n := 50$	50
$m := n \cdot p$	25.6
$\sigma := \sqrt{n \cdot p \cdot (1-p)}$	3.53452
$\text{normCdf}(-\infty, 24.5, m, \sigma)$	.377818

On peut voir que cette approximation est également satisfaisante, ce qui était prévisible puisque les trois conditions que l'on impose en général pour ce type d'approximation :  $n \geq 30$ ,  $np \geq 15$ ,  $np(1-p) > 5$  étaient bien toutes vérifiées.

Vous remarquerez que dans les conditions de l'exercice, un sondage sur un aussi petit nombre de personnes risque de donner un résultat faux plus d'une fois sur trois.

Reprenons rapidement le calcul avec un échantillon de 1 000 personnes :

1.1 1.2 1.3 1.4 ▶ DEG AUTO RÉEL	
$\sigma := \sqrt{n \cdot p \cdot (1-p)}$	31.55452
$\text{normCdf}(-\infty, 24.5, m, \sigma)$	.377818
$n := 1000$	1000
$m := n \cdot p$	512.
$\sigma := \sqrt{n \cdot p \cdot (1-p)}$	15.8068
$\text{normCdf}(-\infty, 499.5, m, \sigma)$	.214531
	11/99

Cette fois, il n'y a plus qu'environ une "chance" sur cinq de se tromper.

(Ce qui n'est pas totalement négligeable...)

## 5.5 Estimations et intervalles de confiance

### • Estimateur sans biais

Soient  $\mu$  (moyenne) et  $\sigma$  (écart type) les paramètres d'un caractère  $X$  dans une population  $\mathcal{P}$  de taille  $N$ . On prélève tous les échantillons de taille  $n$ , on note  $(n, \bar{x}_i, s_i)$  les paramètres de  $X$ .

Si  $(n, \bar{x}, s)$  sont les paramètres d'un échantillon pris au hasard,  $\bar{x}$  est un estimateur sans biais de  $\mu$ , par contre  $s$  n'est pas un estimateur sans biais de  $\sigma$ , il faut prendre pour cela  $S$  défini par  $S^2 = \frac{n}{n-1} s^2$ ,

soit  $S = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})^2}$  (les  $x_j$  étant les valeurs de  $X$  dans l'échantillon).

### • Distribution d'échantillonnage des moyennes

On note  $\bar{X}$  la variable qui suit la loi de probabilité déterminée par la distribution d'échantillonnage des moyennes  $\{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k\}$  ( $k$  est le nombre d'échantillons,  $N = nk$ ).

$$E(\bar{X}) = E(X) = \mu$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \text{ dans le cas d'un tirage non exhaustif (avec remise)}$$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \text{ dans le cas d'un tirage exhaustif (sans remise)}$$

☞ Si  $N$  est grand  $\sqrt{\frac{N-n}{N-1}} \simeq 1$ , on peut supposer le tirage non exhaustif,

si  $\sigma$  est inconnu on le remplace dans le calcul par son estimateur  $S$ .

### • Distribution d'échantillonnage des fréquences (ou proportions)

On se place dans le cas où le caractère  $X$  ne peut prendre que les valeurs 0 et 1 avec respectivement les probabilités  $p$  et  $q = 1 - p$ . Soit  $F$  la variable prenant pour valeur la fréquence  $f_i$  d'apparitions de "l'événement favorable" dans le  $i$ -ième échantillon.

$$E(F) = p$$

$$\sigma_F = \sqrt{\frac{pq}{n}} \text{ dans le cas d'un tirage non exhaustif (avec remise)}$$

$\sigma_F = \sqrt{\frac{pq}{n}} \sqrt{\frac{N-n}{N-1}}$  dans le cas d'un tirage exhaustif (sans remise)

☞ Si  $N$  est grand  $\sigma_F \simeq \sqrt{\frac{pq}{n}}$ , on peut supposer le tirage non exhaustif. Si  $p$  est inconnue on la remplace par son estimation  $f$ , fréquence observée sur un échantillon, et  $n$  par  $n-1$ , si  $n$  petit.

• **Intervalle de confiance d'une moyenne**

Si  $n \geq 30$ , que la population mère soit normalement distribuée ou non, on peut supposer que la variable centrée réduite  $T = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$  suit approximativement une loi normale centrée réduite.

$$\text{Si } P(|T| \leq t_{\alpha/2}) = P\left(\mu - t_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + t_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

ce qui revient à dire que la probabilité qu'une valeur  $\bar{x}$  de  $\bar{X}$  appartienne à l'intervalle  $\left[\mu - t_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \mu + t_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right]$  est égale à  $1 - \alpha$ , ou encore que  $\mu \in \left[\bar{x} - t_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + t_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right]$  avec la probabilité  $1 - \alpha$  ou le risque d'erreur  $\alpha$ .

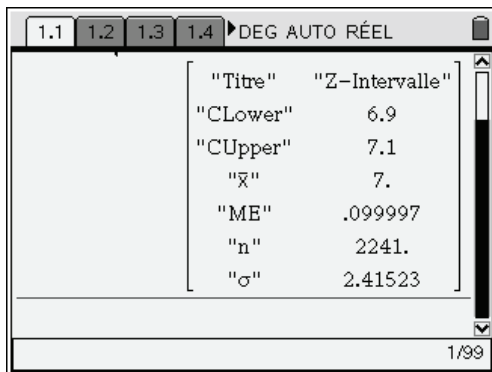
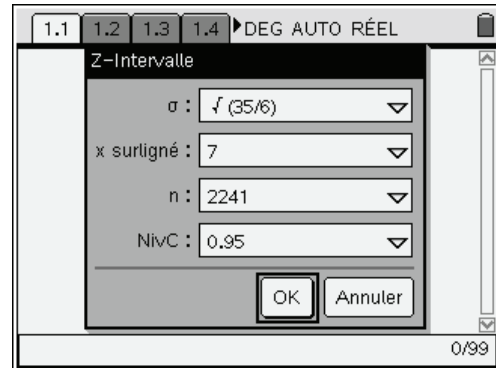
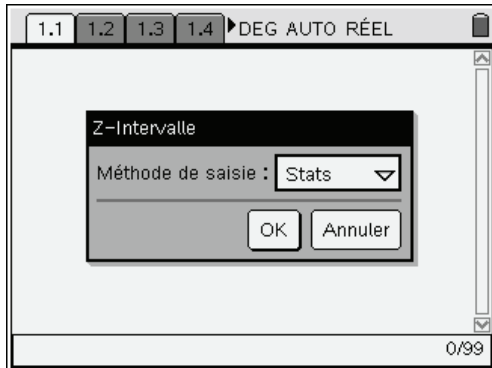
L'intervalle  $\left[\bar{x} - t_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + t_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right]$  est appelé intervalle de confiance de l'estimateur de  $\mu$  au niveau de confiance  $1 - \alpha$  ou au risque d'erreur  $\alpha$ .

☞ Si le tirage est exhaustif on remplace  $\frac{\sigma}{\sqrt{n}}$  par  $\frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$   
si  $\sigma$  est inconnu on le remplace par son estimateur  $S$ .

Si  $n < 30$ , si la population mère est normalement distribuée et si l'échantillon est non exhaustif, on peut supposer que la variable centrée réduite  $T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$  ( $S$  estimateur de  $\sigma$  voir page 20) suit une loi

de Student à  $n-1$  degrés de liberté, les calculs sont les mêmes que ci-dessus. Seuls changent les  $t_{\alpha/2}$  qui sont donnés ici par une loi de Student au lieu de la loi normale. Dans la pratique on utilise la fonction **zInterval** (menu) 6) 6) 1) dans le premier cas :  $\sigma$  connu ; **tInterval** (menu) 6) 6) 2) dans le second cas :  $\sigma$  inconnu.

**Exemple.** On peut contrôler le résultat de l'exemple du paragraphe précédent. On utilise **zIntervall** avec l'option **Stats** et les paramètres  $\sigma = \sqrt{\frac{35}{6}}$ ,  $\bar{x} = 7$ ,  $n = 2241$  et  $C.Level = 0,95$ . On retrouve bien l'intervalle  $[6,9; 7,1]$ .



- **Intervalle de confiance d'une fréquence (ou proportion)**

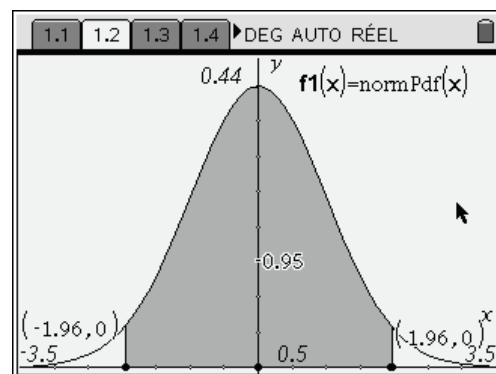
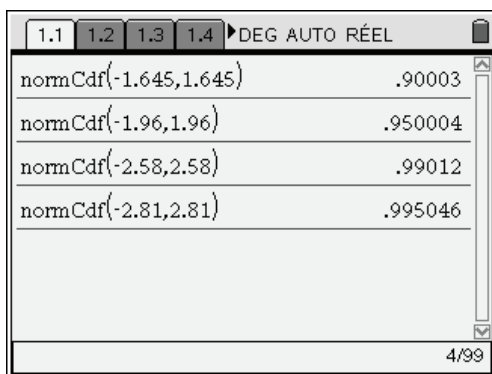
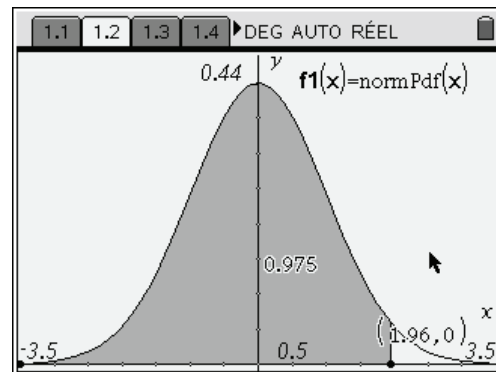
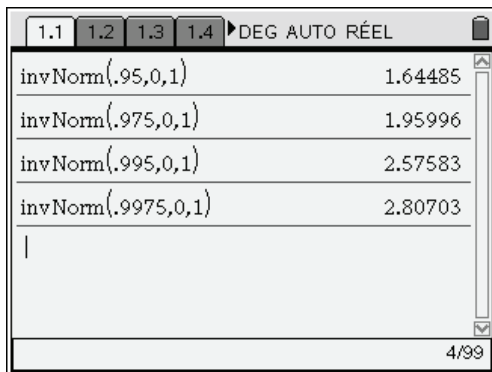
Avec les notations vues page 20, on appelle intervalle de confiance de l'estimateur de  $p$  au niveau de confiance  $1-\alpha$  ou au risque d'erreur  $\alpha$  l'intervalle  $\left[ f - t_{\alpha/2} \sqrt{\frac{pq}{n}}, f + t_{\alpha/2} \sqrt{\frac{pq}{n}} \right]$ . Utiliser **zIntervall\_1Prop** (menu  $\left[ \text{menu} \right] \left[ 6 \right] \left[ 6 \right] \left[ 5 \right]$ ).

☞ Si le tirage est exhaustif on remplace  $\sqrt{\frac{pq}{n}}$  par  $\sqrt{\frac{pq}{n}} \sqrt{\frac{N-n}{N-1}}$   
 si  $p$  est inconnue on remplace  $\sqrt{\frac{pq}{n}}$  par son estimateur  $\sqrt{\frac{f(1-f)}{n-1}}$ .

- **Calcul des  $t_{\alpha/2}$**

Si  $T$  suit la loi normale centrée réduite  $\mathcal{N}(0,1)$ , notons  $t_{\alpha/2}$  le réel positif tel que  $P(|T| \leq t_{\alpha/2}) = 1 - \alpha$ .

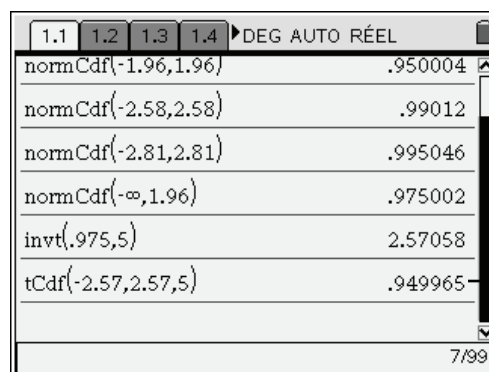
$t_{\alpha/2}$  peut être donné par la fonction **InvNorm** (menu  $\leftarrow$  5  $\leftarrow$  5  $\leftarrow$  3) en prenant pour valeur de Area  $1 - \alpha/2$ , voir le graphique ci-dessous.



On peut ainsi dresser le tableau suivant :

Risque $\alpha$	0,5 %	1 %	5 %	10 %
Seuil de confiance $1 - \alpha$	99,5 %	99 %	95 %	90 %
$t_{\alpha/2}$	2,81	2,58	1,96	1,645

Pour un effectif faible ( $n < 30$ ) on peut procéder de la même façon pour calculer  $t_{\alpha/2}$  en utilisant la loi de Student. Par exemple au risque d'erreur de 5 % et pour 5 degrés de liberté  $t_{\alpha/2} \simeq 2,57$  (voir calcul ci-dessous).



vérification

☞ La valeur est supérieure à celle donnée par la loi normale, ceci vient du fait que la distribution de Student est plus aplatie.

Dans le menu Intervalles de confiance il existe d'autres fonctions, par exemple :

**zInterval\_2Samp** et **tInterval\_2Samp** ( (menu) 6 6 3 et 4 ) qui calculent un intervalle de confiance pour la différence de deux moyennes lorsque respectivement les écarts types sont connus, ou ne le sont pas ; **zInterval\_2Prop** ( (menu) 6 6 6 ) calcule un intervalle de confiance pour la différence de deux proportions.

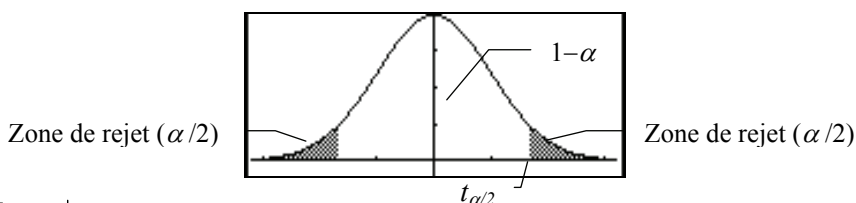
## 5.6 Tests

### • Tests de conformité d'un échantillon à une population

Comparaison d'une moyenne (resp. fréquence) observée à une moyenne (resp. fréquence) théorique.

Sous l'hypothèse nulle  $H_0 : \mu = \mu_0$ , la variable  $\frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}}$  peut être assimilée à une variable normale centrée réduite ou à une variable suivant une loi de Student, suivant la taille de l'échantillon (voir page 21).

**Test bilatéral** : on teste l'hypothèse nulle  $H_0 : \mu = \mu_0$  et l'alternative  $\mu \neq \mu_0$ . Au seuil de signification  $\alpha$  la zone de rejet est la partie grisée.

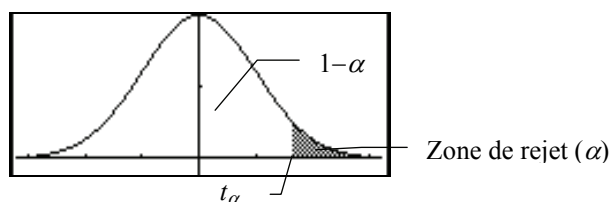


Si la valeur de  $\frac{|\bar{X} - \mu_0|}{\sigma_{\bar{X}}}$  est inférieure à  $t_{\alpha/2}$  on accepte l'hypothèse  $H_0$  ;

sinon on rejette l'hypothèse  $H_0$ .

$t_{\alpha/2}$  est donné par la loi normale centrée réduite quand  $n \geq 30$  et par la loi de Student à  $n-1$  degrés de liberté dans le cas contraire.

**Test unilatéral à droite** : on teste l'hypothèse nulle  $H_0 : \mu = \mu_0$  et l'alternative  $\mu > \mu_0$ . Au seuil de signification  $\alpha$  la zone de rejet est la partie grisée :



Si la valeur de  $\frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}}$  est inférieure à  $t_\alpha$  on accepte l'hypothèse  $H_0$  ;<sup>2</sup>

sinon, on rejette l'hypothèse  $H_0$  (avec les mêmes remarques que ci-dessus).

**Test unilatéral à gauche** : on teste l'hypothèse nulle  $H_0 : \mu = \mu_0$  et l'alternative  $\mu < \mu_0$ . La zone de rejet est à gauche, symétrique de celle à droite et la condition d'acceptation de  $H_0$  :

la valeur de  $\frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}}$  est supérieure à  $-t_\alpha$ .

<sup>2</sup> $t_\alpha$  est directement donné par **invNorm** ( (menu) 6 5 3 ) ou **invT** (Student (menu) 6 5 6 ).

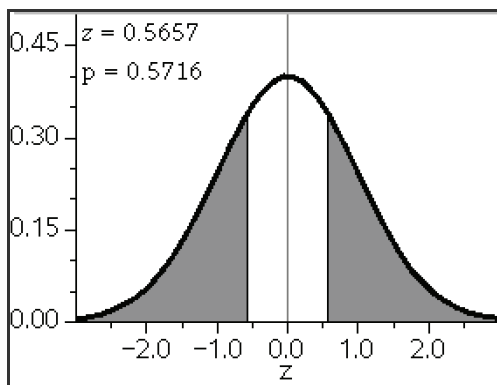
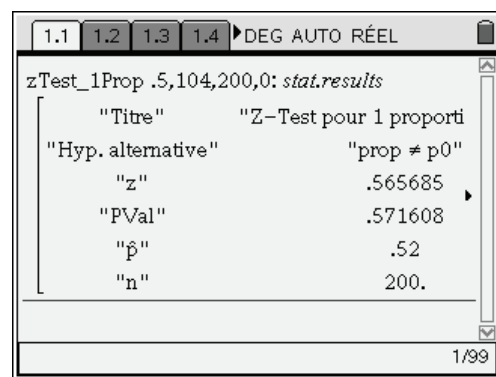
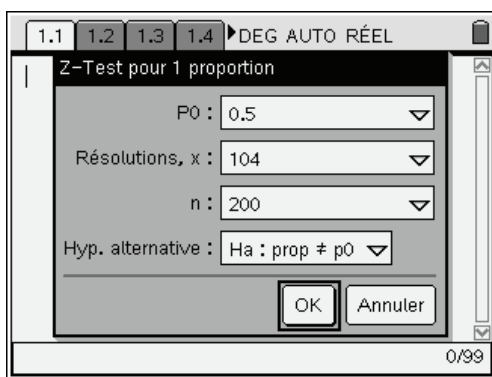


Pour effectuer ce genre de test, on dispose de la fonction **zTest** (menu 6 7 1 ou menu 4 4 1) lorsque l'écart type  $\sigma$  est connu, et de la fonction **tTest** (menu 6 7 2 ou menu 4 4 2) dans le cas contraire. L'échantillon peut être représenté soit sous forme de liste (Data), soit par sa moyenne ( $\bar{x}$ ) et sa taille (n), (Stats).

Pour une proportion on utilisera, de façon analogue à ce que l'on a vu pour une moyenne, la variable  $\frac{F - p_0}{\sigma_F}$ , voir page 20.

**zTest\_1Prop** (menu 6 7 5 ou menu 4 4 5) effectue le test d'une proportion de réussites inconnue (prop). Elle utilise comme données d'entrée le nombre de réussites dans l'échantillon (Successes, x) et la taille (n) de l'échantillon. L'hypothèse nulle  $H_0 : prop = p_0$  est testée contre l'une des hypothèses alternatives suivantes :  $prop \neq p_0$ ,  $prop < p_0$ ,  $prop > p_0$ .

Prenons par exemple l'exercice 1 de la page 30 :  $p_0 = 0,5$ , l'hypothèse nulle  $H_0 : prop = 0,5$ , on effectue un test bilatéral, Successes,  $x = 104$ ,  $n = 200$ .



La valeur  $z$  étant inférieure à 1,96, au seuil de confiance de 95%, l'hypothèse  $H_0$  ne peut être rejetée.

### • Tests d'homogénéité de deux échantillons

Il s'agit de comparer deux moyennes, ou deux proportions, observées. Le premier échantillon caractérisé par  $(n_1, \bar{x}_1, s_1)$  est prélevé dans une population  $\mathcal{P}_1$  de paramètres  $(\mu_1, \sigma_1)$ , le second caractérisé par  $(n_2, \bar{x}_2, s_2)$  est prélevé dans une population  $\mathcal{P}_2$  de paramètres  $(\mu_2, \sigma_2)$ .

Considérons le cas des **moyennes** ; le principe est le même que pour les tests de conformité en utilisant la variable  $\frac{D}{\sigma_D}$  où  $D = \bar{X}_1 - \bar{X}_2$ , ( $\bar{X}_1$  et  $\bar{X}_2$  sont indépendantes).

Si  $n_1 \geq 30$  et  $n_2 \geq 30$ ,  $\sigma_D = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ , les  $\sigma_i$  éventuellement remplacés par leur estimateur  $S_i$ , sous l'hypothèse nulle  $H_0: \mu_1 = \mu_2$ , la variable  $\frac{D}{\sigma_D}$  peut être assimilée à une variable normale centrée réduite.

Sinon, en supposant que  $\sigma_1 = \sigma_2 = \sigma$ ,  $\sigma_D = \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ ,  $\sigma$  pouvant être remplacé par son estimateur  $S = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}}$ , la variable sous  $H_0$  suit, dans ce cas, une loi de Student à  $n_1 + n_2 - 2$  degrés de liberté.

Deux fonctions permettent d'effectuer des tests de comparaison des moyennes de deux échantillons indépendants : **zTest\_2Prop** ((menu) 6) 7) 6) ou ((menu) 4) 4) 6)) lorsque les écarts types  $\sigma_1$  et  $\sigma_2$  sont connus ; **zTest\_2Samp** ((menu) 6) 7) 3) ou ((menu) 4) 4) 3)) dans le cas contraire.

Pour les **proportions** : l'hypothèse nulle  $H_0$  est  $p_1 = p_2 (= p_0)$ ,  $D = F_1 - F_2$ , et pour des échantillons de grande taille  $\sigma_D \simeq \sqrt{p_0(1-p_0)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ ,  $p_0$  est estimé par  $\frac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$ , les  $f_i$  étant les fréquences observées.

La fonction **zTest\_2Prop** ((menu) 6) 7) 6) ou ((menu) 4) 4) 6)) permet d'effectuer un tel test, elle utilise comme données d'entrée le nombre de réussites ( $x_1$  et  $x_2$ ) et la taille des échantillons ( $n_1$  et  $n_2$ ).

Citons de plus un test de comparaison de deux écarts types (test de Fisher) ; **FTest\_2Samp** ((menu) 6) 7) 9)) teste l'hypothèse nulle  $H_0: \sigma_1 = \sigma_2$ . La moyenne des populations et les écarts types sont tous inconnus.

**Exemple 1.** Une machine remplit des sachets dont le poids théorique moyen est de 170 mg. On prélève un échantillon de 20 sachets de la production de cette machine, on obtient les résultats suivants :

178, 170, 173, 173, 172, 172, 165, 173, 165, 169,  
175, 170, 173, 168, 175, 171, 165, 174, 168, 180 ;

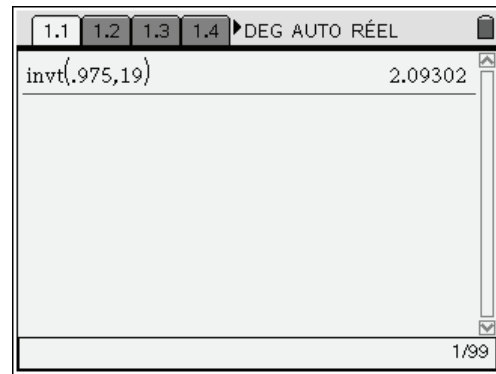
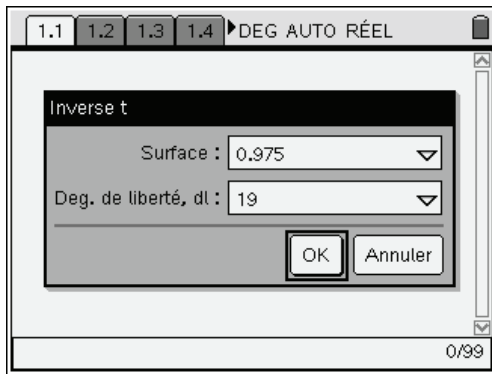
Cette fabrication est-elle conforme aux exigences au seuil de signification de 5% ?

Un sachet est refusé si son poids diffère du poids moyen de plus de 5 mg, la machine est considérée comme étant bien réglée si la proportion de sachets refusés est inférieure à 10%.

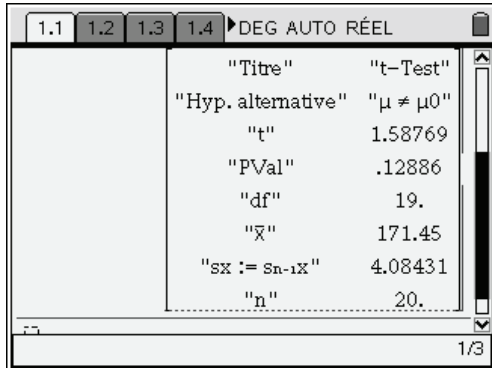
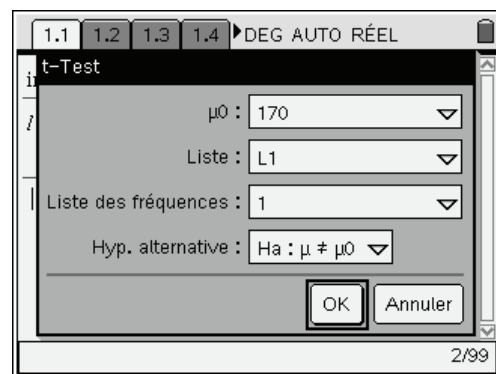
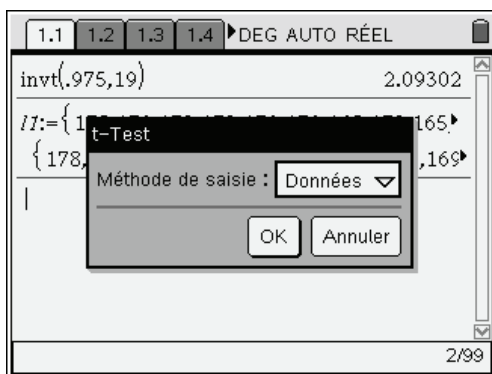
La machine est-elle bien réglée ?

Vu les remarques faites précédemment on effectue un test bilatéral à l'aide de **tTest** ((menu) 6) 7) 2)), on choisit pour Méthode de saisie : Données, la liste des valeurs est contenue dans L1, à l'aide de la fonction **invT** on calcule  $t_\alpha$  correspondant au seuil de signification de 5%.

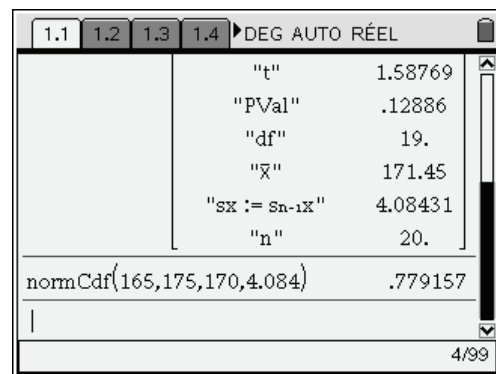
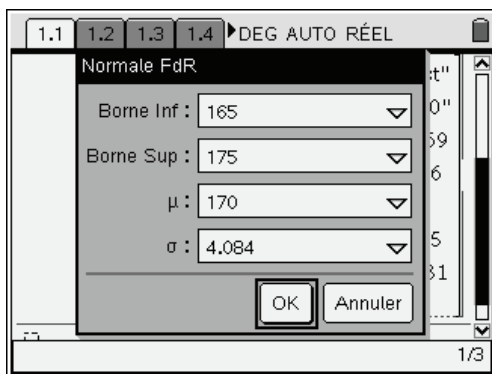
$t_\alpha \simeq 2,093$  comme l'indique le calcul ci-dessous.



On entre la liste dans la variable L1. **tTest** (menu 6) (7) (2) permet d'obtenir  $S_x$  qui donne une estimation de l'écart type de la production. De plus, on a bien  $PVal > 0,05$ , donc au seuil de signification de 5% on accepte l'hypothèse  $H_0$  que la moyenne des sachets est bien 170 mg.



On suppose donc que le poids suit la loi normale  $\mathcal{N}(170,4.084)$ . La fonction **normCdf** permet de calculer la probabilité que le poids soit compris entre 165 et 175 mg.

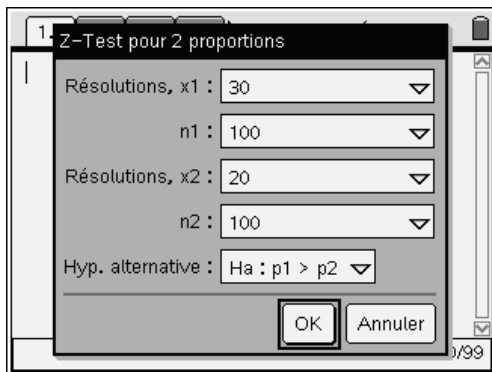


La probabilité que le paquet soit rejeté est donc de 22%. **Conclusion** : la machine est mal réglée.

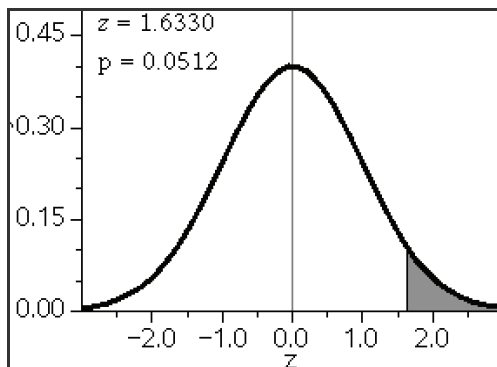
**Exemple 2.** 20% des piles provenant d'une certaine fabrication peuvent fonctionner plus de 100 heures. Un traitement spécial appliqué à un échantillon de 100 piles a permis d'obtenir un fonctionnement de plus de 100 heures pour 30 d'entre elles. L'amélioration apparente est-elle significative au seuil de 5%, de 10% ?

Il s'agit de comparer deux proportions, on suppose que les valeurs sont normalement distribuées,  $n=100 > 30$ , on utilise la loi normale, donc la fonction **zTest\_2Prop** (menu 6) 7) 6) ou menu 4) 4) 6).

L'hypothèse  $H_0$  : le nouveau traitement n'a pas amélioré la durée de vie des piles ( $p_1 = p_2$ ), est testée contre l'hypothèse alternative  $p_1 > p_2$  (test unilatéral à droite).



"Z-Test pour 2 proportions"	
"Titre"	"Z-Test pour 2 proportions"
"alternative"	"p1 > p2"
"z"	1.63299
"PVal"	.051235
"p1"	.3
"p2"	.2
"p"	.25
"n1"	100.
"n2"	100.



Au seuil de 5%,  $z$  est légèrement inférieur à  $t_{0,05} \simeq 1,645$  (ce qui équivaut à dire que  $p > 0,05$ ), on ne peut donc pas rejeter l'hypothèse  $H_0$ , par contre au seuil de signification de 10% on peut la rejeter et donc admettre que le nouveau traitement a amélioré la vie des piles.

### • Test du Khi 2

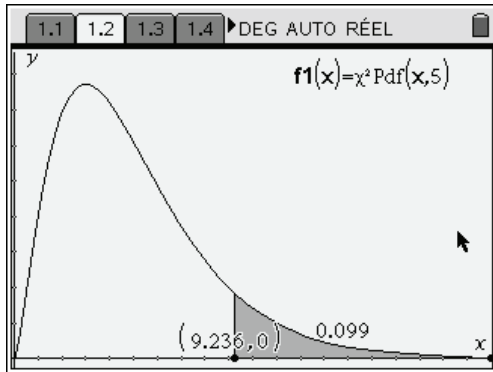
Il s'agit de comparer une distribution d'un caractère observé sur un échantillon donné et une distribution théorique basée sur un modèle probabiliste. L'hypothèse nulle  $H_0$  consiste à supposer que l'on a concordance des deux distributions.

On calcule  $\chi_c^2 = \sum_{k=1}^r \frac{(n_k - np_k)^2}{np_k}$ , où  $n_k$  est l'effectif observé possédant le caractère  $k$ ,  $p_k$  la

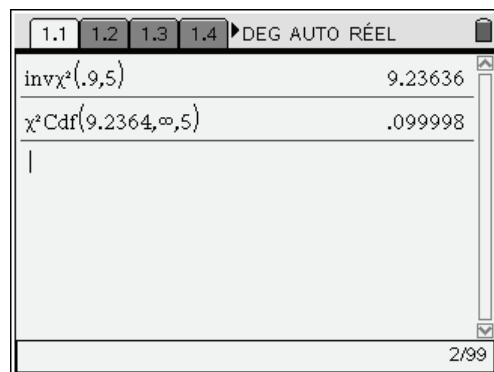
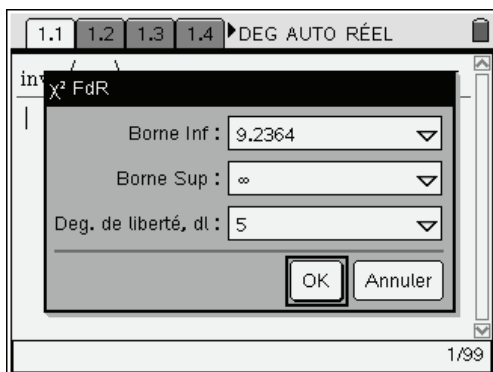
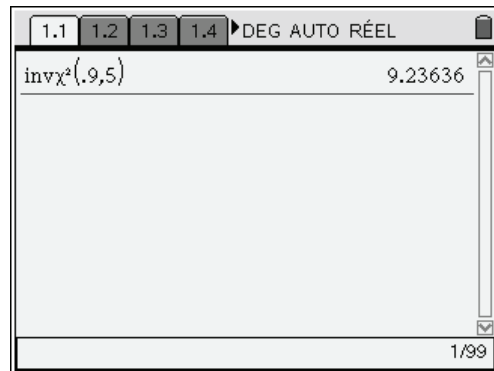
probabilité théorique d'obtenir ce caractère,  $np_k$  étant alors l'effectif théorique. La valeur de  $\chi_c^2$  sera d'autant plus grande que les deux distributions diffèrent. Le nombre de degrés de liberté est égal à  $r - 1$ .

Un seuil de signification  $\alpha$  étant fixé les tables usuelles du Khi 2 donne le nombre  $\chi_\alpha^2$  tel que  $P(\chi^2 \geq \chi_\alpha^2) = \alpha$ . Par exemple pour 5 degrés de liberté si  $\chi_\alpha^2 = 9,236$ ,  $P(\chi^2 \geq \chi_\alpha^2) \simeq 0,1$  (voir ci-dessous).

- Si  $\chi_c^2 \leq \chi_\alpha^2$ , l'hypothèse  $H_0$  est acceptable.
- Si  $\chi_c^2 > \chi_\alpha^2$ , l'hypothèse  $H_0$  est à rejeter.



Pour obtenir  $\chi_\alpha^2$  directement à partir de  $\alpha$  à l'aide de la calculatrice, on utilise la fonction **Inv  $\chi^2$**  ((menu) 6 5 9), mais attention il faut entrer comme premier paramètre  $1 - \alpha$  et non  $\alpha$ , car Area correspond à la probabilité  $P(\chi^2 \leq \chi_\alpha^2)$ .



## Exercices

### 1 Sondage et loi normale

Lors d'une élection, à la sortie des urnes on interroge 200 électeurs choisis au hasard. 52% d'entre eux ont voté pour un candidat A. Peut-on en conclure que A va être élu, en admettant un risque d'erreur de 5% ? Même question avec un risque d'erreur de 1%. Quelle devrait être la taille de l'échantillon pour pouvoir affirmer que A sera élu avec un niveau de confiance de 0,95 ?

### 2 Suivi d'une production

Une machine automatique produit des pièces dont le diamètre théorique est de 100 mm. On suppose que les diamètres des pièces produites sont distribués suivant une loi normale. On prélève au hasard un échantillon de 20 pièces. On obtient les dimensions suivantes :

92, 95, 106, 96, 100, 91, 96, 89, 104, 91,  
92, 92, 94, 98, 96, 103, 105, 95, 107, 94.

Au seuil de confiance de 95% peut-on affirmer que le diamètre moyen de la production est de 100 mm ?

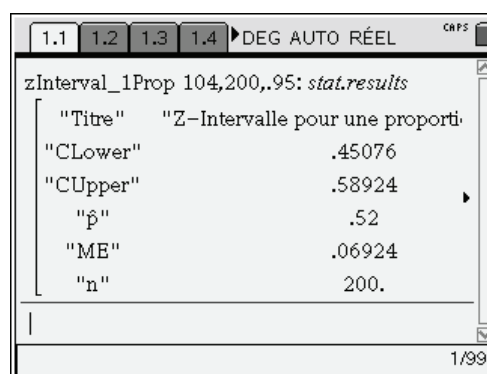
### 3 Recherche d'une ville test

Un référendum, à l'échelon national, a donné 55% de oui, 40% de non, 5% de blancs ou nuls. On recherche une ville test pour les élections futures. On en trouve une qui, sur 10 000 votants, a donné 56,5% de oui, 39% de non, 4,5% de blancs ou nuls. Peut-on dire que cette ville reflète la physionomie de la nation au risque de 1% ? et au risque de 1‰ ?

## Solutions des exercices

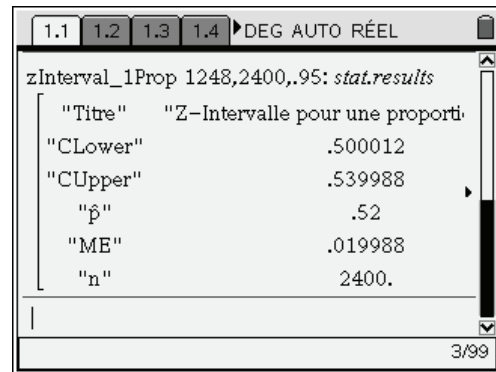
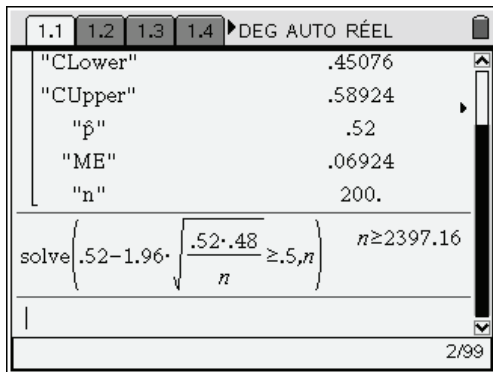
### 1 Sondage et loi normale

La population est supposée normalement répartie, nous sommes en présence de "grands" échantillons, on cherche à estimer une proportion. On utilise ici la loi normale, la fonction est : `zInterval_1Prop` (menu  $\left(\begin{smallmatrix} \text{menu} \\ \text{6} \end{smallmatrix}\right) \left(\begin{smallmatrix} \text{6} \\ \text{5} \end{smallmatrix}\right)$ ).



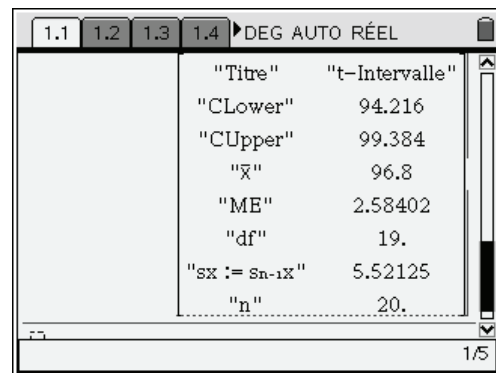
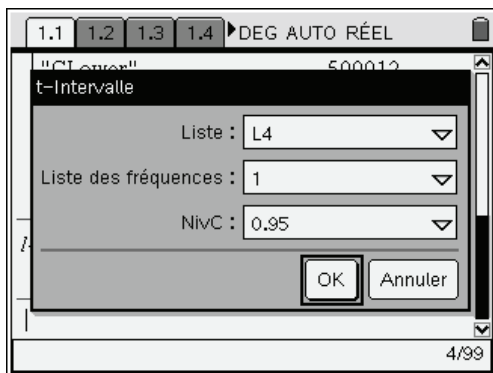
On obtient l'intervalle  $[0.45, 0.59]$ , donc au risque d'erreur de 5%, (ou au niveau de confiance de 0.95), il n'est pas possible de dire que A sera élu, il en sera évidemment de même au risque de 1%, car plus on diminue le risque plus l'intervalle est important, ici :  $[0.429, 0.611]$ .

Pour être sûr que la borne inférieure de l'intervalle soit supérieure à 50% il suffit de prendre  $n$  tel que  $0.52 - t_{\alpha/2} \sqrt{\frac{0.52 \times 0.48}{n}} \geq 0.5$ ,  $t_{\alpha/2} = 1.96$  pour  $\alpha = 5\%$ . Ce qui donne une taille d'échantillon de 2 400 personnes...



## 2 Suivi d'une production

La distribution est normale et l'on a un échantillon de petite taille ; on utilise donc ici la loi de Student. La fonction **tInterval** permet d'obtenir un intervalle de confiance dans ce cas. On saisit la liste des valeurs dans l'éditeur, (menu) 6) 6) 2), on choisit l'option Données, on remplit la boîte de dialogue :



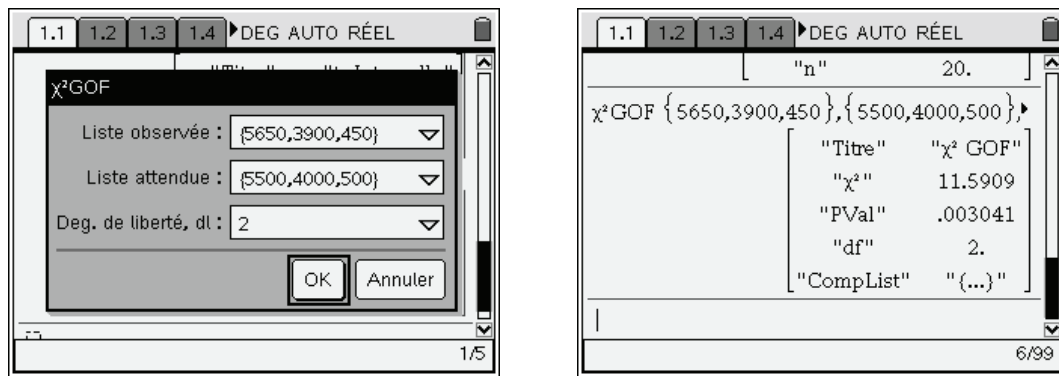
La réponse à la question est non car  $100 \notin [94.22, 99.38]$ .

## 3 Recherche d'une ville test

Hypothèse  $H_0$  : la ville reflète la physionomie de la nation.

On effectue un test du  $\chi^2$ . **Chi2 GOF...**

Liste Observée = {5650, 3900, 450}, Liste attendue = {5500, 4000, 500} (liste théorique), Deg de liberté = 2. On valide.



Le calcul du Khi 2 donne 11.59, le nombre de degrés de liberté est égal à 2.

La probabilité  $P\text{Value} = P(\chi^2 \geq 11.59) = 0.003$ , est inférieure à 0.01, donc au seuil de 0.01 l'hypothèse  $H_0$  doit être rejetée, par contre au seuil de 0.001 elle peut être acceptée ( $\chi_{0,01}^2 = 9.21 < \chi_c^2 < \chi_{0,001}^2 = 13.82$ ).