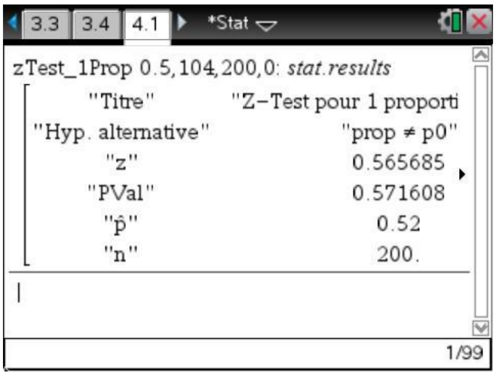
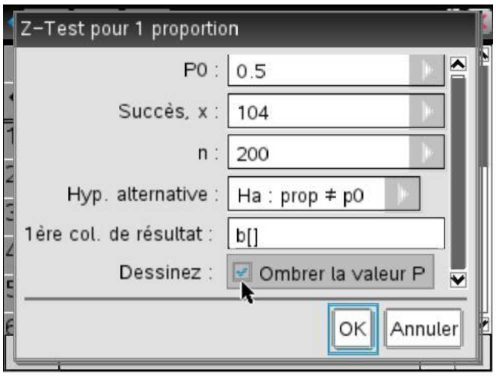
Pour effectuer ce genre de test, on dispose de la fonction **zTest** ([menu][6][7][1] application Calculs ou [menu][4][4][1] application Tableur & listes) lorsque l'écart type σ est connu, et de la fonction **tTest** ([menu][6][7][2] ou [menu][4][4][2]) dans le cas contraire. L'échantillon peut être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

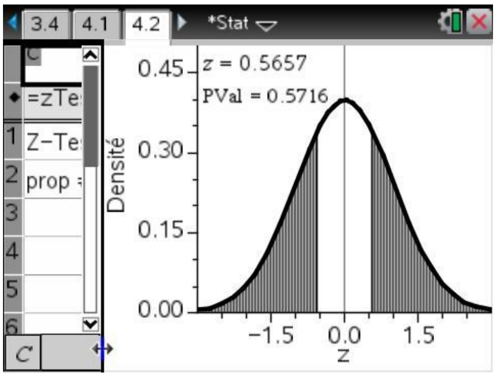
Pour une proportion on utilisera, de façon analogue à ce que l'on a vu pour une moyenne, la variable $\dfrac{F - p_0}{\sigma_F}$, voir page 20.

zTest_1Prop ([menu][6][7][5] application Calculs ou [menu][4][4][5] application Tableur & listes) 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 $(n_1, \overline{x}_1, s_1)$ est prélevé dans une population $\mathcal{P}_1$ de paramètres (μ_1, σ_1), le second caractérisé par $(n_2, \overline{x}_2, s_2)$ est prélevé dans une population $\mathcal{P}_2$ de paramètres (μ_2, σ_2).

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

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

Sinon, en supposant que $\sigma_1 = \sigma_2 = \sigma$, $\sigma_D = \sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$, σ pouvant être remplacé par son estimateur $S = \sqrt{\dfrac{(n_1 - 1){s_1}^2 + (n_2 - 1){s_2}^2}{n_1 + n_2 - 2}}$, 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 application Calculs ou menu 4 4 6 application Tableur & listes) lorsque les écarts types σ_1 et σ_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 $\sigma_D \simeq \sqrt{p_0(1 - p_0)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}$, p_0 est estimé par $\dfrac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$, 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 (x1 et x2) et la taille des échantillons (n1 et n2).

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 :

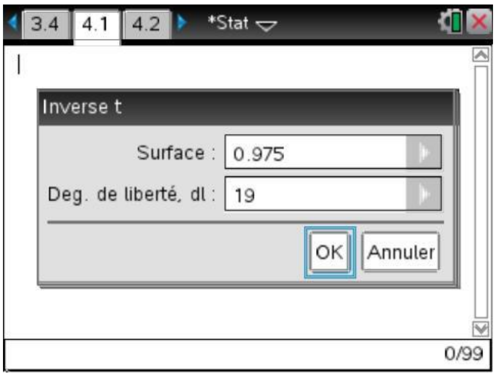
178, 170, 173, 173, 172, 172, 165, 173, 165, 169,
175, 170, 173, 168, 175, 171, 165, 174, 168, 180 ;

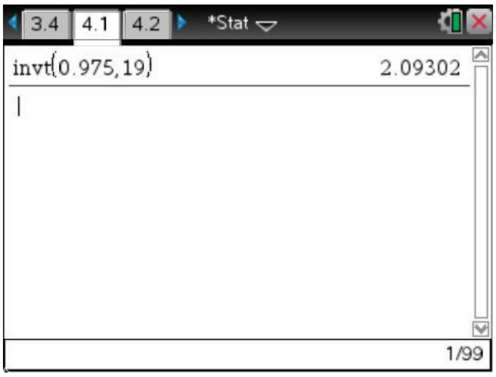
Cette fabrication est-elle conforme aux exigences au seuil de signification de 5% ?

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 ?

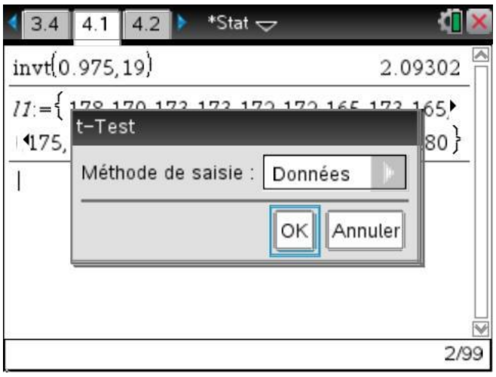
Vu les remarques faites précédemment on effectue un test bilatéral à l'aide de **tTest** (menu 6 7 2), on choisit pour Méthode de saisie : Données, la liste des valeurs est contenue dans L1, à l'aide de la fonction **invt** on calcule t_α correspondant au seuil de signification de 5%.

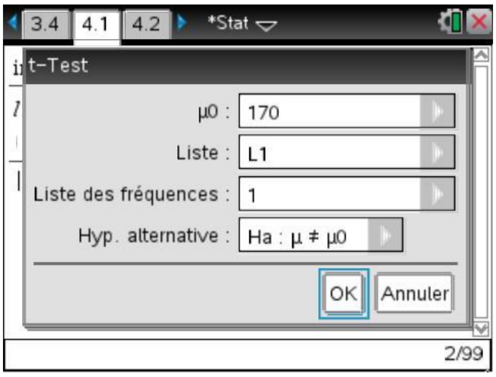
$t_\alpha \simeq 2,093$ comme l'indique le calcul ci-dessous.

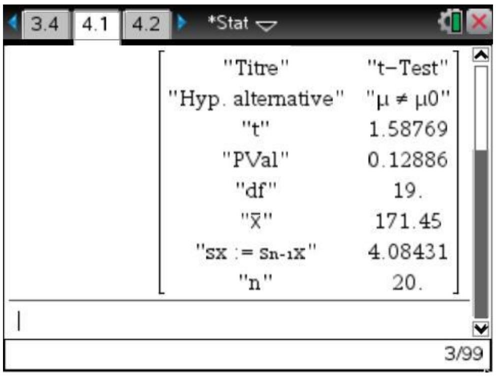




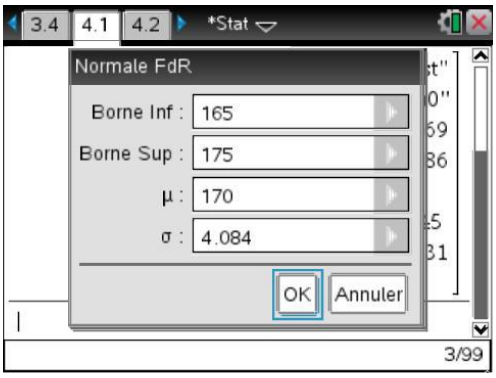
On entre la liste dans la variable L1. **tTest** (menu 6 7 2) permet d'obtenir Sx qui donne une estimation de l'écart type de la production. De plus, on a bien $PVal > 0,05$, donc au seuil de signification de 5% on accepte l'hypothèse H_0 que la moyenne des sachets est bien 170 mg.

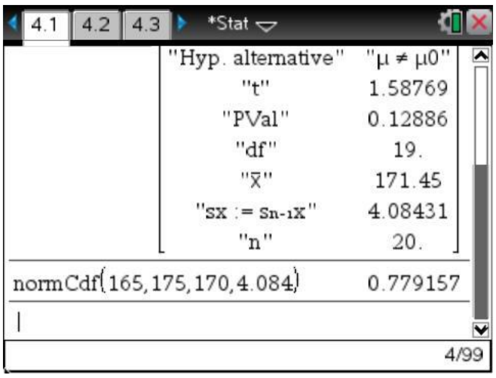






On suppose donc que le poids suit la loi normale $\mathcal{N}(170, 4.084)$. La fonction **normCdf** permet de calculer la probabilité que le poids soit compris entre 165 et 175 mg.





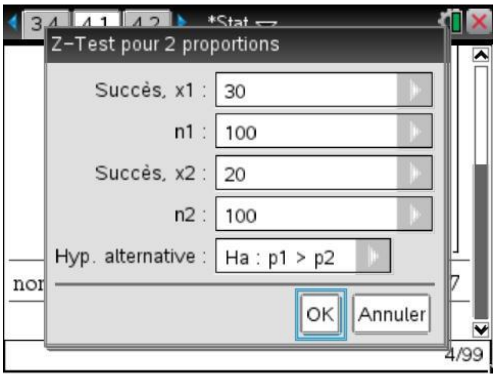
La probabilité que le paquet soit rejeté est donc de 22%. **Conclusion** : la machine est mal réglée.

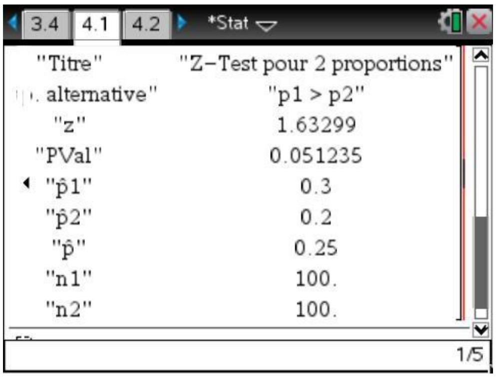
Exemple 2. 20% des piles provenant d'une certaine fabrication peuvent fonctionner plus de 100 heures. Un traitement spécial appliqué à un échantillon de 100 piles a permis d'obtenir un fonctionnement de plus de 100 heures pour 30 d'entre elles. L'amélioration apparente est-elle significative au seuil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

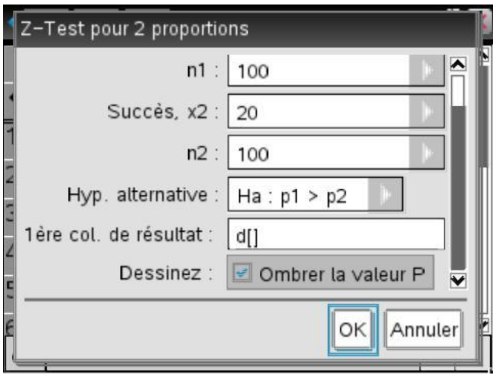
L'hypothèse H_0 : le nouveau traitement n'a pas amélioré la durée de vie des piles ($p_1 = p_2$), est testée contre l'hypothèse alternative $p_1 > p_2$ (test unilatéral à droite).

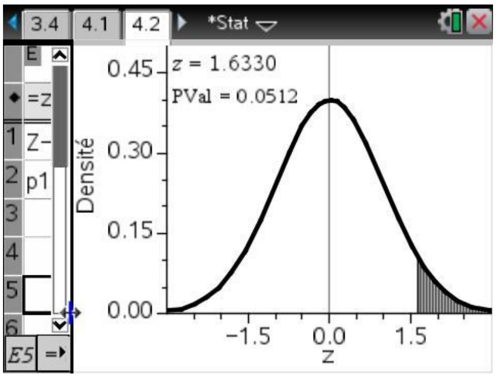
[menu] [6] [7] [6] dans l'application Calculs :





[menu] [4] [4] [6] dans l'application Tableur & listes (permet une représentation graphique) :





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.

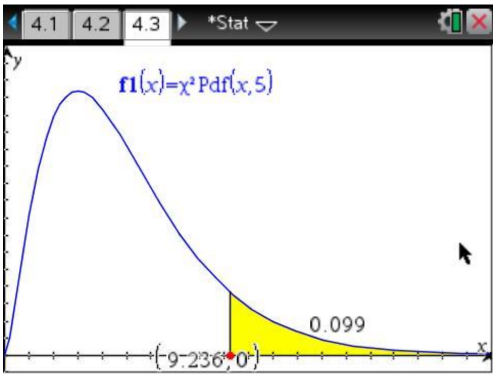
- **Test du Khi 2**

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.

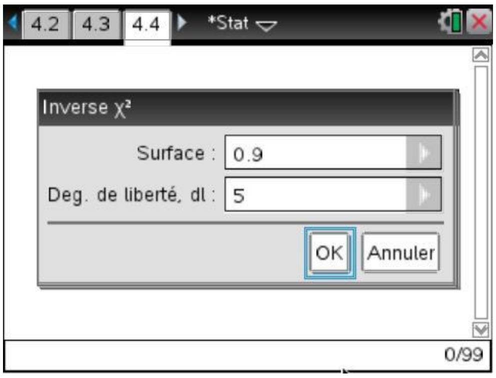
On calcule $\chi_c^2 = \sum_{k=1}^{r} \frac{(n_k - np_k)^2}{np_k}$, où n_k est l'effectif observé possédant le caractère k, p_k la probabilité théorique d'obtenir ce caractère, np_k étant alors l'effectif théorique. La valeur de χ_c^2 sera d'autant plus grande que les deux distributions diffèrent. Le nombre de degrés de liberté est égal à $r - 1$.

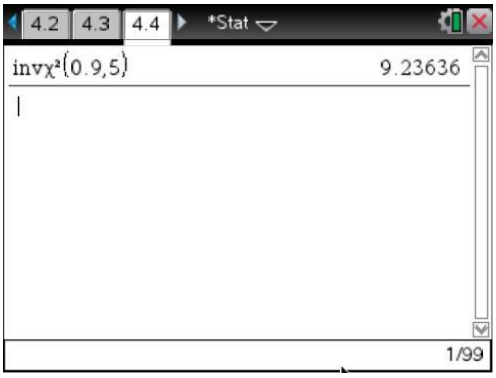
Un seuil de signification α étant fixé les tables usuelles du Khi 2 donne le nombre χ_α^2 tel que $P(\chi^2 \geq \chi_\alpha^2) = \alpha$. Par exemple pour 5 degrés de liberté si $\chi_\alpha^2 = 9{,}236$, $P(\chi^2 \geq \chi_\alpha^2) \simeq 0{,}1$ (voir ci-dessous).

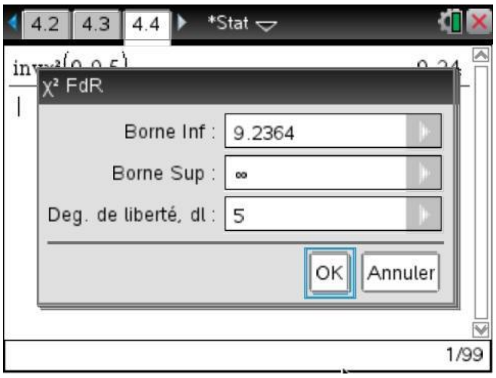
- Si ${\chi_c}^2 \leq \chi_\alpha^2$, l'hypothèse H_0 est acceptable.
- Si ${\chi_c}^2 > \chi_\alpha^2$, l'hypothèse H_0 est à rejeter.

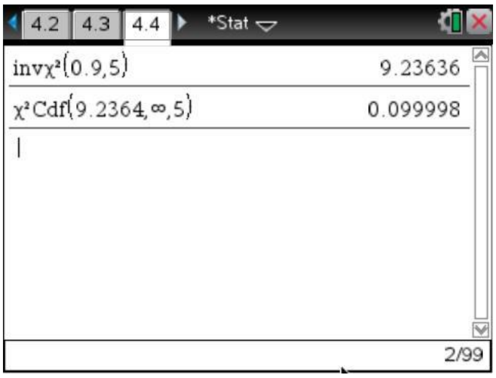


Pour obtenir χ_α^2 directement à partir de α à l'aide de la calculatrice, on utilise la fonction **Inv** χ^2 (menu 6 5 9), mais attention il faut entrer comme premier paramètre $1-\alpha$ et non α, car Area correspond à la probabilité $P(\chi^2 \leq \chi_\alpha^2)$.









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 ?

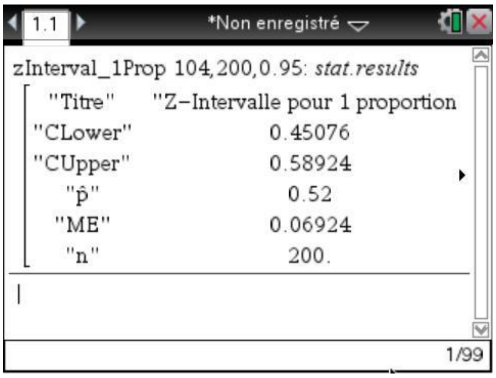
3 Recherche d'une ville test

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

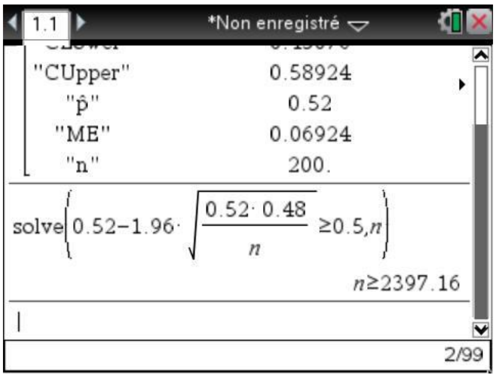
1 Sondage et loi normale

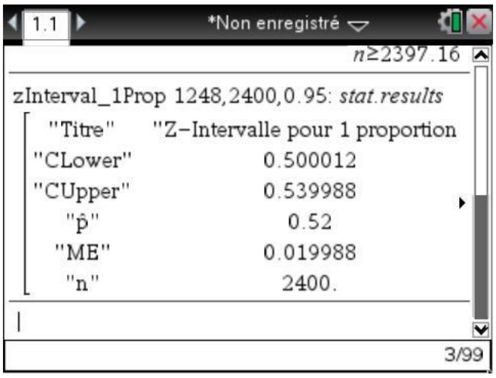
La population est supposée normalement répartie, nous sommes en présence de ''grands'' échantillons, on cherche à estimer une proportion. On utilise ici la loi normale, la fonction est : **zInterval_1Prop** (menu 6 6 5).



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 : $[0.429, 0.611]$.

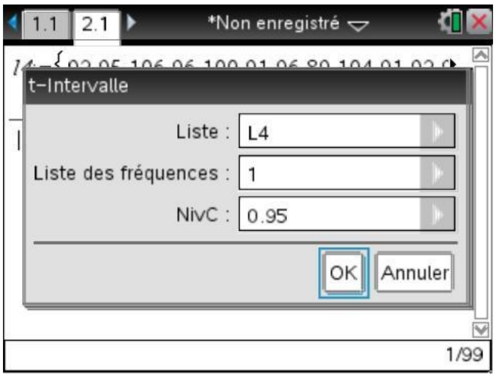
Pour être sûr que la borne inférieure de l'intervalle soit supérieure à 50% il suffit de prendre n tel que $0.52 - t_{\alpha/2}\sqrt{\dfrac{0.52\times 0.48}{n}} \geq 0.5$, $t_{\alpha/2} = 1.96$ pour $\alpha = 5\%$. Ce qui donne une taille d'échantillon de 2 400 personnes...

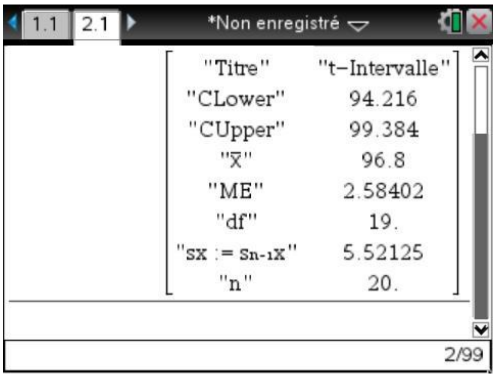




2 Suivi d'une production

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 :





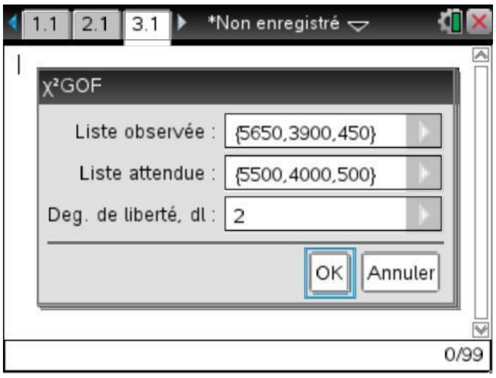
La réponse à la question est non car $100 \notin [94.22, 99.38]$.

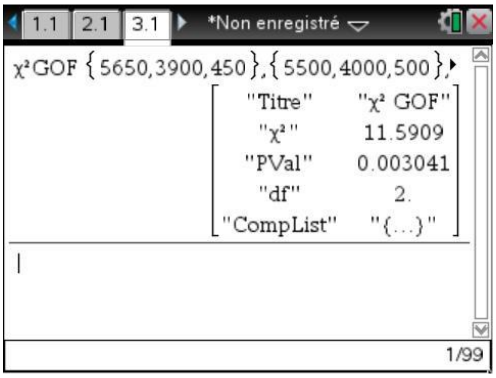
3 Recherche d'une ville test

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

On effectue un test du χ^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 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 ($\chi^2_{0,01} = 9.21 < \chi_c{}^2 < \chi^2_{0.001} = 13.82$).