Le génie pour l'industrie

École de technologie supérieure
Service des enseignements généraux
Local B-2500 514-396-8938
Site internet : http://www.seg.etsmtl.ca

MAT350

Probabilités et statistiques

Utilisation de la TI-NSPIRE

dans le cadre du cours MAT350

Par Sylvie gervais

Rédigé en juin 2012

Table des matières

Partie 1

PREMIÈRE PARTIE

1.1 Statistiques descriptives

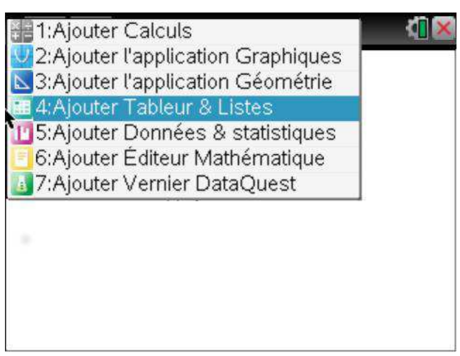
1.1.1 Données présentées en série

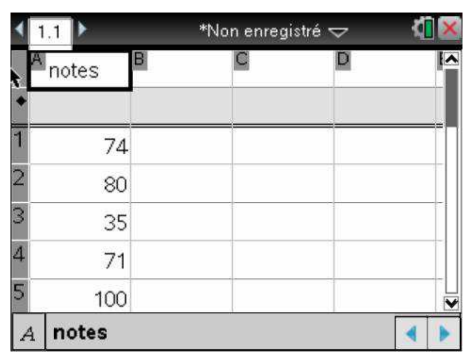
Voyons à partir d'un exemple comment obtenir les statistiques descriptives lorsque les données sont présentées en série.

Exemple 1.1

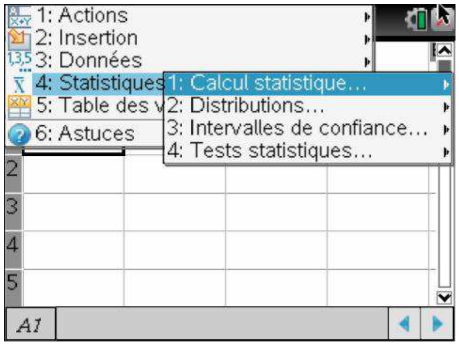
Étude des résultats obtenus par un groupe d'étudiants : 74, 80, 35, 71, 100, 75, 68, 81, 77 et 70.

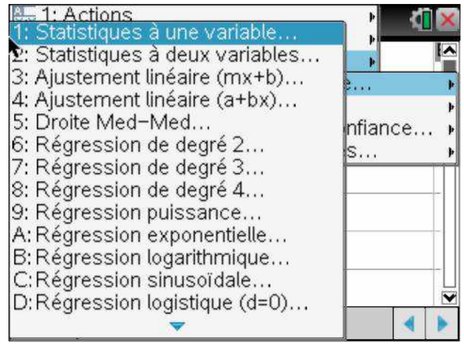
1. Ouvrir un *Tableur & listes* : [CTRL] [doc] [4]
2. Entrer les résultats dans une colonne et nommer cette colonne "notes" tel qu'illustré ci-dessous :



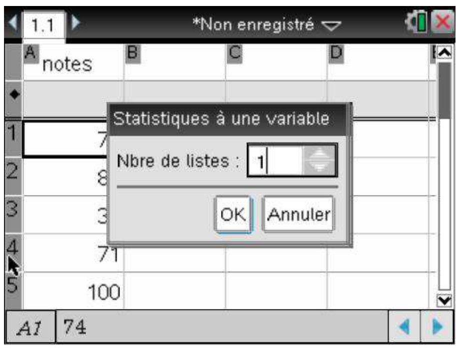


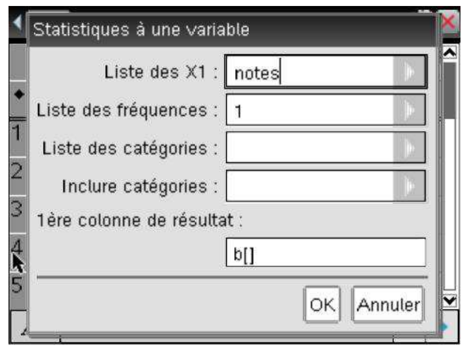
3. Les fonctions de cette sections se trouvent dans le menu *Statistiques* / *Calcul statistique* / *Statistiques à une variable* : [menu] [4] [1] [1]





4. Remplir les informations demandées tel qu'illustré ci-dessous :



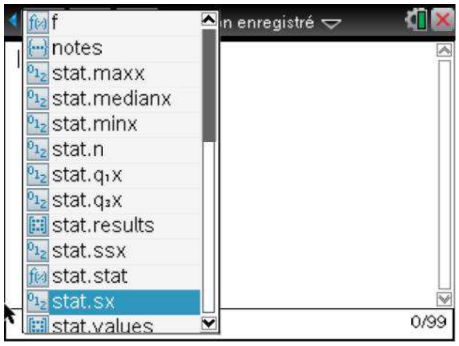


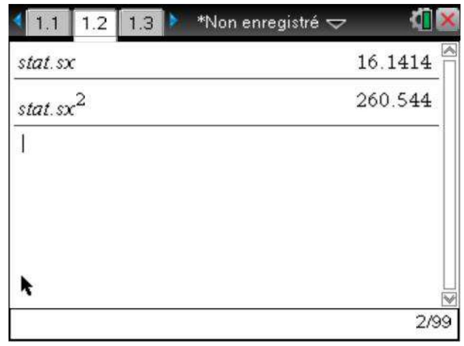
5. On obtient alors les mesures échantillonnales suivantes :

B	C
	=OneVar(notes,1): CopyVar S
Titre	Statistiques à une variable
$\bar{x}$	73.1
Σx	731.
Σx^2	55781.
$sx := s_{n-1}x$	16.1414
$\sigma x := \sigma_n x$	15.3131
n	10.
MinX	35.
Q_1X	70.
MedianX	74.5
Q_3X	80.
MaxX	100.
$SSX := \Sigma(x-\bar{x})^2$	2344.9

B	C	D	E
	=OneVar(notes,1): CopyVar S		
Titre	Statistiques à une variable		
$\bar{x}$	73.1		moyenne
Σx	731.		
Σx^2	55781.		
$sx := s_{n-1}x$	16.1414		écart-type
$\sigma x := \sigma_n x$	15.3131		
n	10.		
MinX	35.		minimum
Q_1X	70.		q1
MedianX	74.5		médiane
Q_3X	80.		q3
MaxX	100.		maximum
$SSX := \Sigma(x-\bar{x})^2$	2344.9		

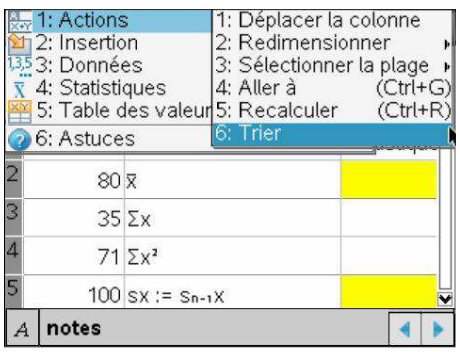
6. Dans une fenêtre de calcul, on peut obtenir chacune des quantités obtenues par l'opération précédente. Ouvrir une fenêtre *Calculs* : [CTRL] [doc] [1] et en appuyant sur la touche [var], on retrouve la terminologie utilisée par défaut pour ces quantités. Par exemple, si on veut obtenir la variance des notes, qui est l'écart-type au carré, on procède de la façon suivante :

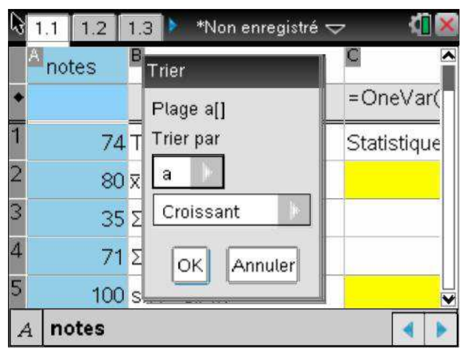


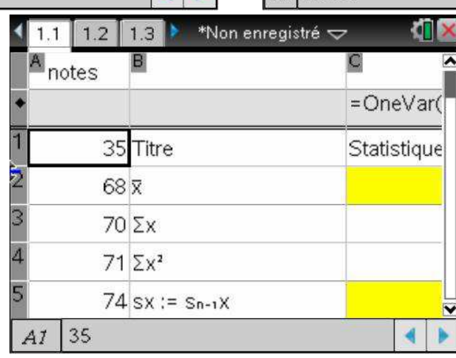


Remarques :

- On aurait pu aussi obtenir les différentes statistiques descriptives précédentes à partir d'une feuille de calculs. Une fois les données entrées dans une liste, à partir d'une feuille de calculs, sélectionner [menu] [6] [1] [1].
- Il sera commode éventuellement dans le cours de trier une colonne de données. Pour ce faire, il suffit de placer le curseur dans la cellule titre de la colonne et de sélectionner [menu] [1] [6] tel qu'illustré ci-dessous.

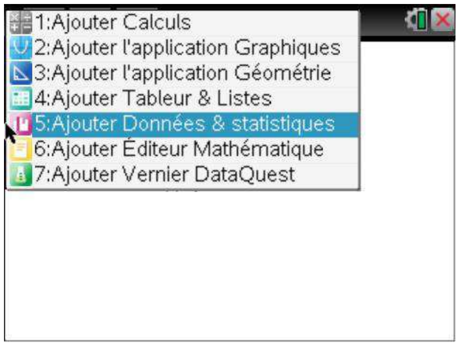


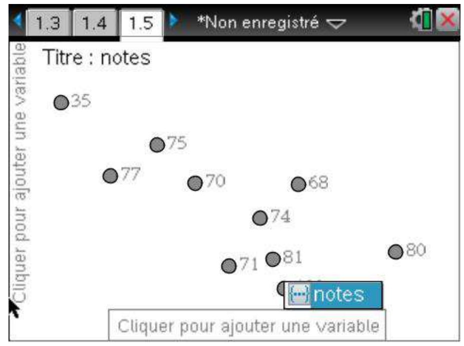




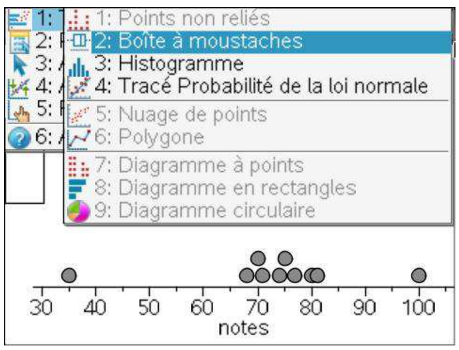
Voyons maintenant comment obtenir un graphique pour illustrer ces données.

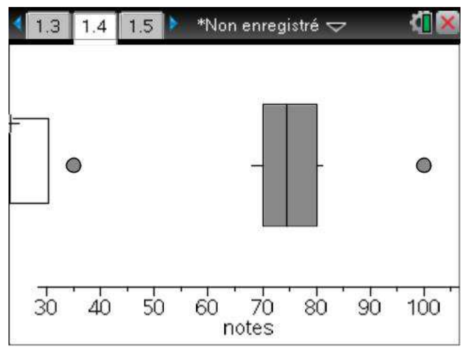
1. Ouvrir un *Données & statistiques* : [CTRL] [doc] [5]
2. Placer le curseur dans le rectangle "Cliquer pour ajouter une variable". Sélectionner la variable "notes".





3. Pour faire une boîte à moustaches, sélectionner *Menu/Type de tracé/Boîte à moustaches* : [menu] [1] [2]





4. Pour rajouter un titre : *Menu/Actions/Insérer du texte* : [menu] [3] [3]
5. Pour modifier l'échelle du graphique : *Menu/Fenêtre & Zoom/Réglage de la fenêtre* : [menu] [5] [1]

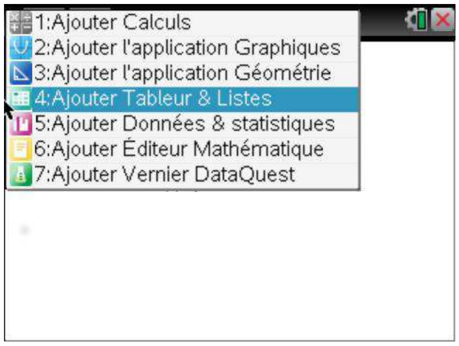
1.1.2 Données groupées par valeurs ou en classes

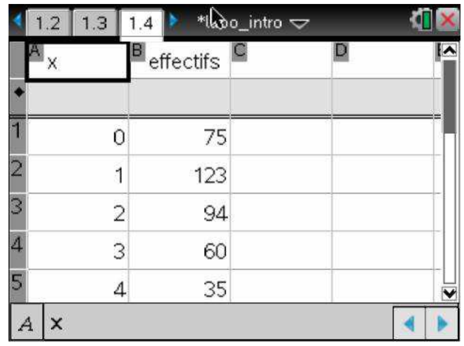
Exemple 1.2

On s'intéresse au nombre d'erreurs d'assemblage d'un échantillon de 396 appareils. On a observé les résultats suivants :

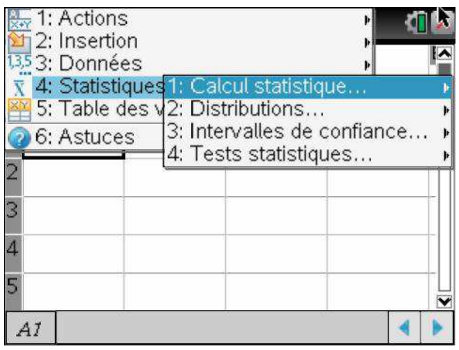
Nombre d'erreurs	effectifs
0	75
1	123
2	94
3	60
4	35
5	9
Total	396

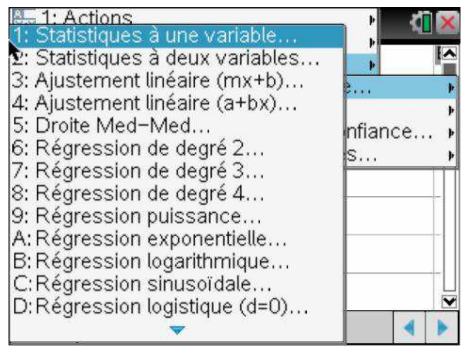
1. Ouvrir un *Tableur & listes* : [CTRL] [doc] [4]
2. Entrer les résultats dans deux colonnes et les nommer "x" et "effectifs" tel qu'illustré ci-dessous :



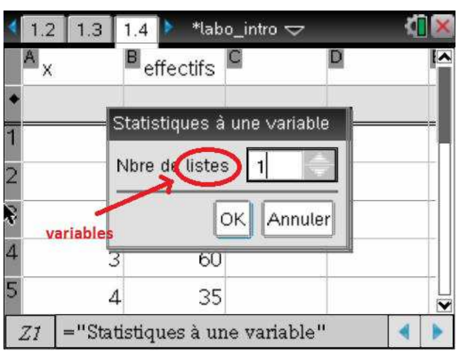


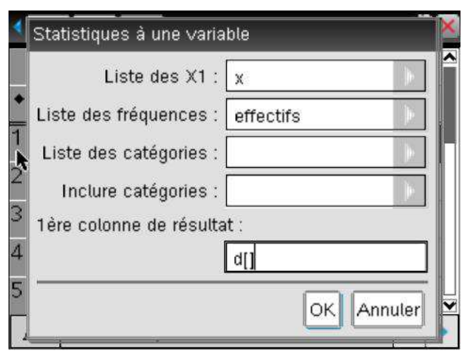
3. Les fonctions de cette sections se trouvent dans le menu *Statistiques / Calcul statistique / Statistiques à une variable* : [menu] [4] [1] [1]



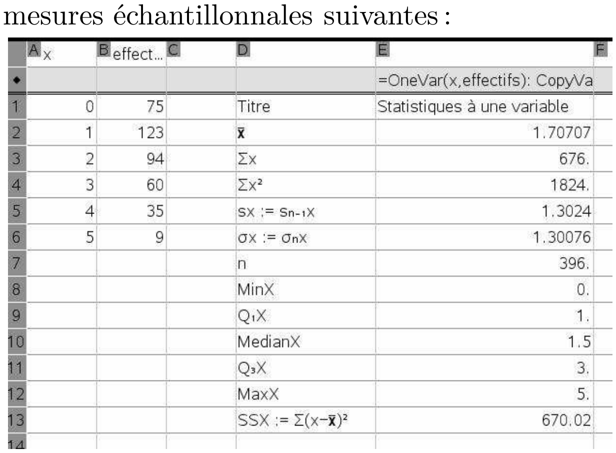


4. Remplir les informations demandées tel qu'illustré ci-dessous :



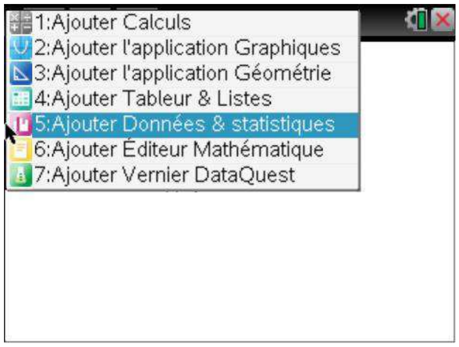


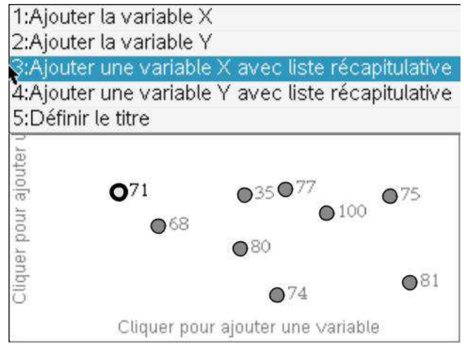
5. On obtient alors les mesures échantillonnales suivantes :



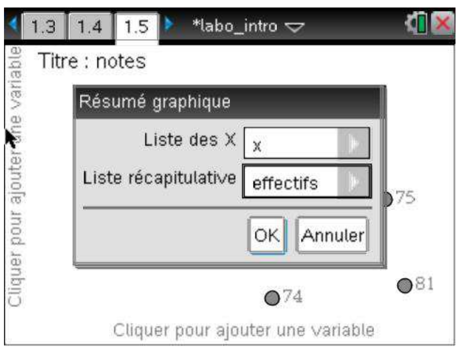
	A x	B effect…	C	D	E
◆					=OneVar(x,effectifs): CopyVa
1	0	75		Titre	Statistiques à une variable
2	1	123		x̄	1.70707
3	2	94		Σx	676.
4	3	60		Σx²	1824.
5	4	35		sx := sn-1x	1.3024
6	5	9		σx := σnx	1.30076
7				n	396.
8				MinX	0.
9				Q1X	1.
10				MedianX	1.5
11				Q3X	3.
12				MaxX	5.
13				SSX := Σ(x−x̄)²	670.02

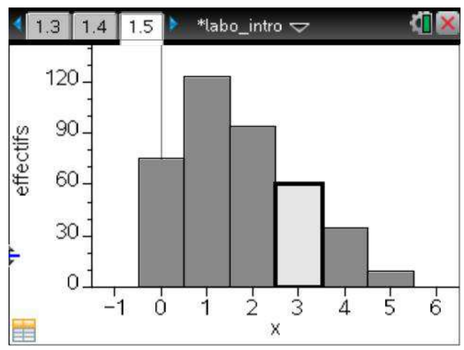
6. Pour faire un graphique illustrant ces données, ouvrir un *Données & statistiques* : [CTRL] [doc] [5]
7. Pour tenir compte qu'il ne s'agit que d'une variable avec ses effectifs associés, sélectionner [CTRL] [menu] et choisir "*Ajouter une variable X avec liste récapitulative*".



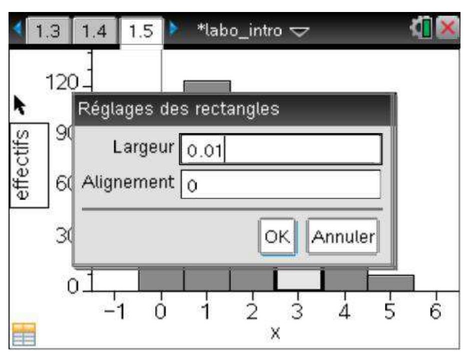


8. Remplir les informations demandées tel qu'illustré ci-dessous. On obtient alors par défaut un histogramme, graphique qu'il faudra modifier pour obtenir un diagramme à bâtons.

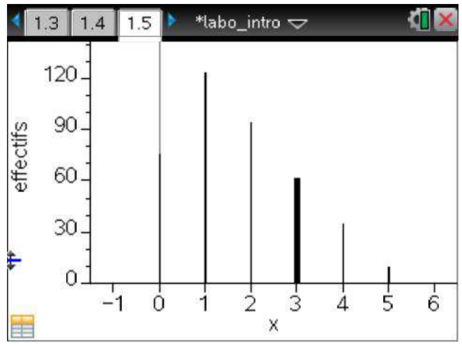




9. Pour obtenir un diagramme à bâtons, il faut ajuster la largeur des rectangles en sélectionnant *Menu* /*Propriétés du tracé* / *Propriétés de l'histogramme* / *Réglage des rectangles* : [menu] [2] [2] [2]



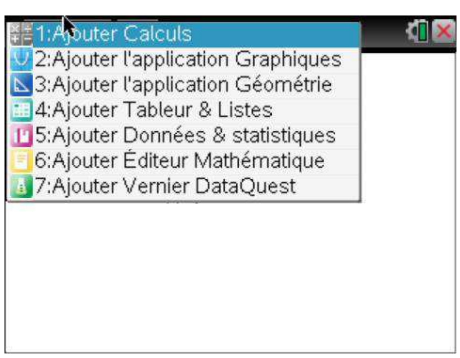
10. Et on obtient alors le diagramme à bâtons illustrant ces données.

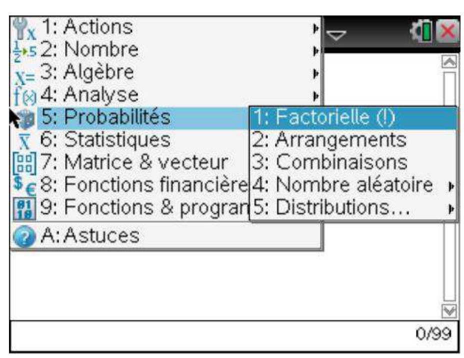


11. Pour transférer des données à partir d'Excel sur la Nspire, brancher la calculatrice sur un poste sur lequel le logiciel est installé. Par la suite, copier le fichier dans le répertoire de la calculatrice via l'onglet "Contenu".

1.2 Probabilités

1. Ouvrir une feuille *Calculs* : [CTRL] [doc] [1]
2. Les fonctions de cette sections se trouvent dans le menu *probabilités* : [menu] [5]





Formules de dénombrement

1. Une **permutation** de n objets distincts représente le nombre de façons différentes de disposer ces n objets et se calcule de la façon suivante :

$$n! = n(n-1)(n-2)\cdots 2\cdot 1$$

Menu de la TI	Directement dans la feuille de calculs
[menu] [5] [1]	*n !*

2. Le nombre de **combinaisons** de k objets parmi n, noté C_k^n, représente le nombre de façons de choisir k objets parmi n objets distincts **en ne tenant pas compte de l'ordre**. On calcule C_k^n de la façon suivante :

$$C_k^n = \frac{n!}{k!\cdot(n-k)!}$$

Menu de la TI	Directement dans la feuille de calculs
[menu] [5] [3]	*nCr(n,x)*

3. Le nombre d'**arrangements** de k objets parmi n, noté A_k^n, représente le nombre de façons de choisir k objets parmi n objets distincts **en tenant compte de l'ordre**. On calcule A_k^n de la façon suivante :

$$A_k^n = \frac{n!}{(n-k)!}$$

Menu de la TI	Directement dans la feuille de calculs
[menu] [5] [2]	*nPr(n,x)*

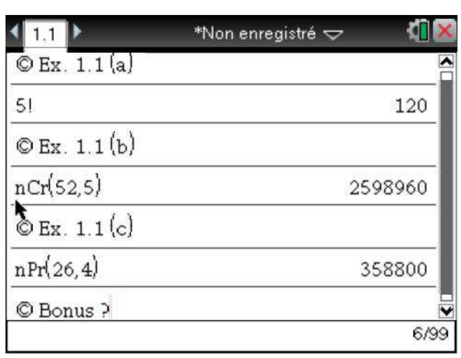
Exemple 1.3

a) Considérons un jeu de cartes de 52 cartes (sans les jokers). On choisit 5 cartes au hasard. De combien de façons différentes peut-on disposer les 5 cartes choisies dans nos mains ?

b) Combien de mains de 5 cartes différentes est-il possible d'obtenir ?

c) On veut étiqueter les pièces produites par un robot à l'aide d'un code comprenant 4 lettres différentes (de A à Z). De combien de codes différents dispose-t-on ?

Solution :



BONUS Si on revient au problème de cartes et que le tirage se fait avec remise, combien de mains de 5 cartes est-il maintenant possible d'obtenir ?

1.3 Variables aléatoires

1.3.1 Variables aléatoires générales

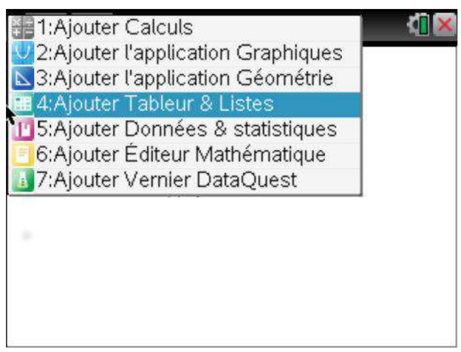
On peut utiliser la TI pour calculer l'espérance et la variance d'une variable aléatoire générale discrète. Pour illustrer la procédure, considérons l'exemple suivant.

Exemple 1.4

Considérons un dé truqué de façon telle que la probabilité d'obtenir un 6 est deux fois plus élevée que celle d'obtenir chacune des autres faces. Calculer l'espérance et la variance du résultat du lancé de ce dé.

Solution :

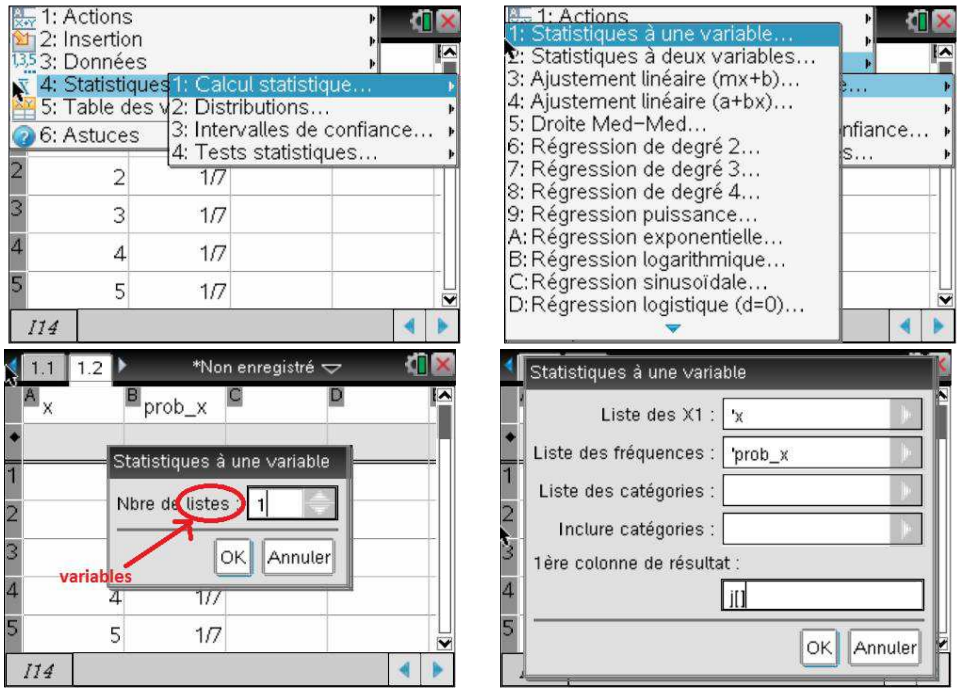
1. Ouvrir un *Tableur & listes* : [CTRL] [doc] [4]



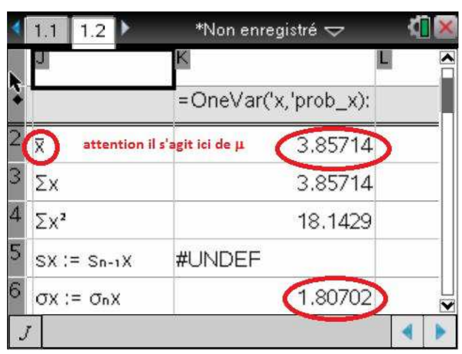
2. Entrer le support de X dans une colonne et la fonction de masse dans une autre colonne.

	A x	B prob_x	C	D
◆				
1	1	1/7		
2	2	1/7		
3	3	1/7		
4	4	1/7		
5	5	1/7		
6	6	2/7		
7				

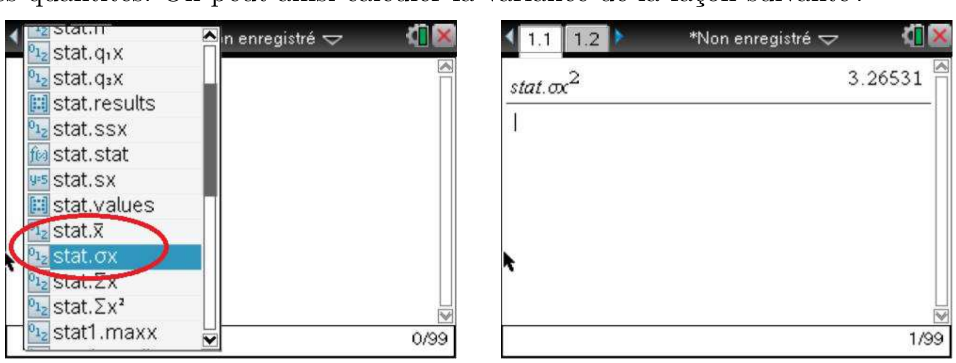
3. Calculer l'espérance et l'écart-type de la variable aléatoire X à partir de l'utilitaire de calculs statistiques : [menu] [4] [1] [1]



4. Dans la colonne qui contient les résultats (identifiée à la fenêtre précédente), on obtient l'espérance (μ) et l'écart-type (σ) de la variable aléatoire X aux endroits identifiés ci-dessous :

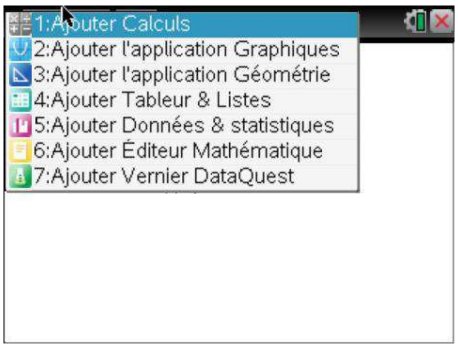


5. Dans la fenêtre de calcul, on peut obtenir chacune des quantités obtenues par l'opération précédente. En appuyant sur la touche [var], on retrouve la terminologie utilisée par défaut pour ces quantités. On peut ainsi calculer la variance de la façon suivante :



1.3.2 Quelques modèles discrètes

1. Ouvrir une feuille *Calculs* : [CTRL] [doc] [1]
2. Les fonctions de cette sections se trouvent dans le menu *probabilités* : [menu] [5] [5]



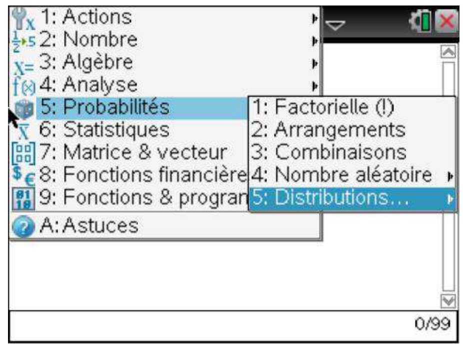
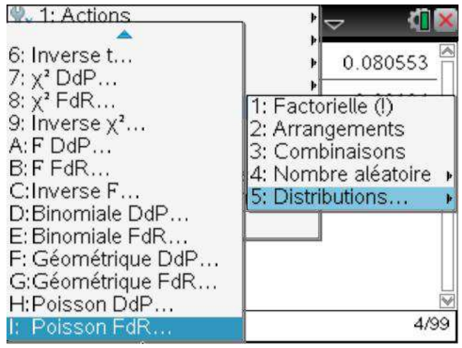


Tableau 1.1 Quelques modèles discrets

Lois	Probabilité recherchée	Menu TI	Appel direct de la fonction
Binomiale	$P(X = c)$	[menu] [5] [5] [D]	*binomPdf(n,p,c)*
$X \sim B(n, p)$	$P(a \leq X \leq b)$	[menu] [5] [5] [E]	*binomCdf(n,p,a,b)*
Poisson	$P(X = c)$	[menu] [5] [5] [H]	*poissPdf(λ,c)*
$X \sim P(\lambda)$	$P(a \leq X \leq b)$	[menu] [5] [5] [I]	*poissCdf(λ,a,b)*
Géométrique	$P(X = c)$	[menu] [5] [5] [F]	*geomPdf(p,c)*
$X \sim Geom(p)$	$P(a \leq X \leq b)$	[menu] [5] [5] [G]	*geomCdf(p,a,b)*

Remarques :

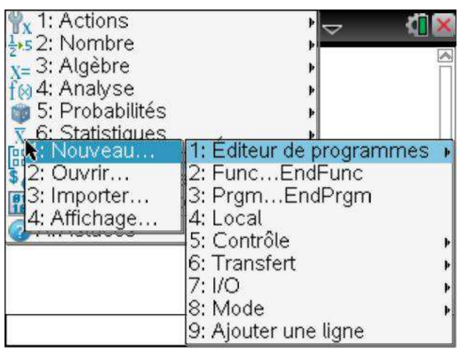
1. Si la TI est configurée en français, on trouvera les noms suivants dans les menus (DdP au lieu de Pdf et FdR au lieu de Cdf). Cependant, lorsqu'on appelle directement la fonction, on doit utiliser le nom anglais, même si la calculatrice est configurée en français.

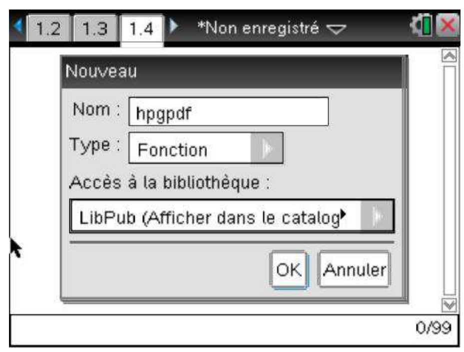


2. Pour les lois discrètes, la TI ne prend pas ∞ dans les paramètres. Ainsi, pour calculer $P(X \geq x)$, il suffit d'utiliser le fait que $P(X \geq x) = 1 - P(X < x)$.
3. La loi hypergéométrique n'est pas configurée dans la TI. On peut toujours faire les calculs en utilisant la fonction de masse ou encore programmer les deux fonctions suivantes.

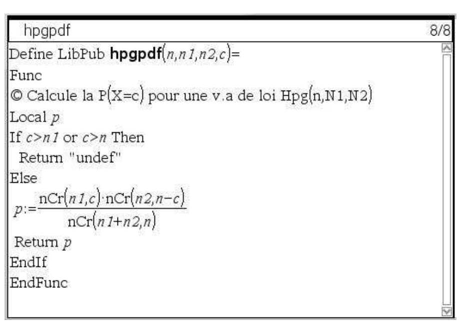
   3.1 Ouvrir l'éditeur de programmes et fonctions : [menu] [9] [1] [1]

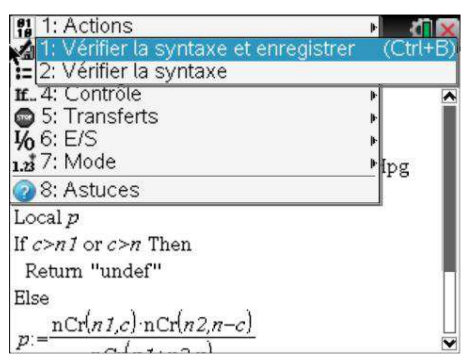
   3.2 Créer la fonction **hpgPdf**



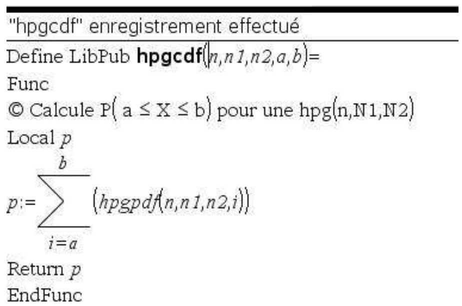


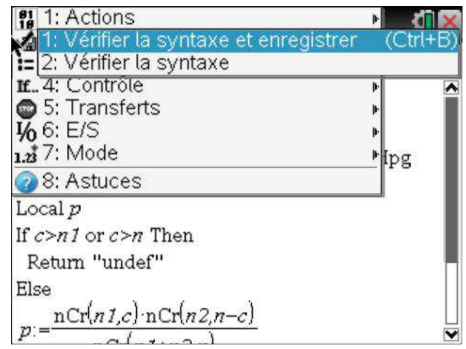
   3.3 Faire [CTRL] [6] pour ne plus avoir l'écran divisé en deux. Construire la fonction tel qu'illustré ci-dessous. Ne pas oublier de vérifier la syntaxe et d'enregistrer la fonction [menu] [2] [1]



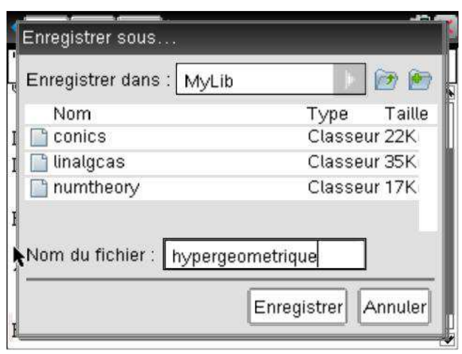


3.4 Construire la deuxième fonction **hpgCdf** fonction tel qu'illustré ci-dessous. Nouveau programme [menu] [9] [1] [1], [CTRL] [6] pour ne plus avoir l'écran divisé en deux et [menu] [2] [1] pour vérifier la syntaxe et enregistrer la fonction.

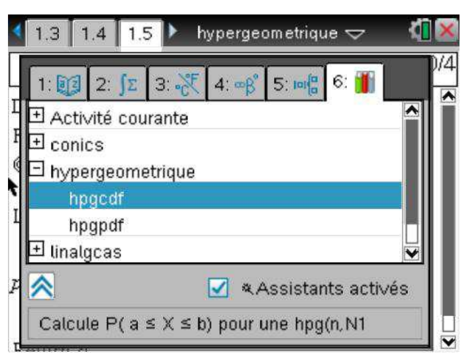




3.5 Il faut ensuite enregistrer le classeur dans Mylib [CTRL] [save] et rafraîchir les bibliothèques pour avoir accès à la fonction dans le catalogue [doc] [6].



3.6 Les deux fonctions seront maintenant disponibles dans le catalogue en tout temps [catalogue] [6].



Exemple 1.5

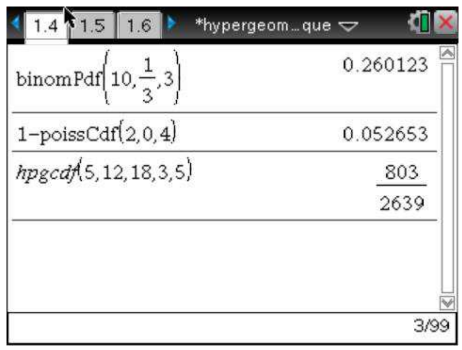
Calculer les probabilités suivantes :

a) $P(X = 3)$ si $X \sim B(10, 1/3)$,

b) $P(X \geq 5)$ si $X \sim P(2)$,

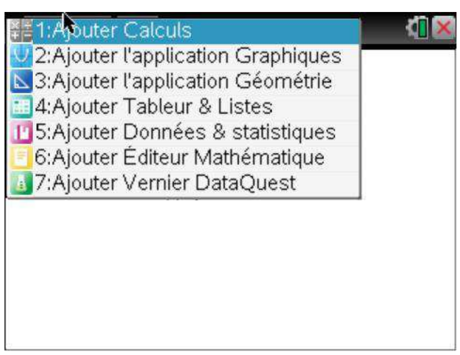
c) $P(3 \leq X \leq 5)$ si $X \sim Hpg(5, 12, 18)$.

Solution :



1.3.3 Quelques modèles continus

1. Ouvrir une feuille *Calculs* : [CTRL] [doc] [1]
2. Les fonctions de cette sections se trouvent dans le menu *probabilités* : [menu] [5] [5]



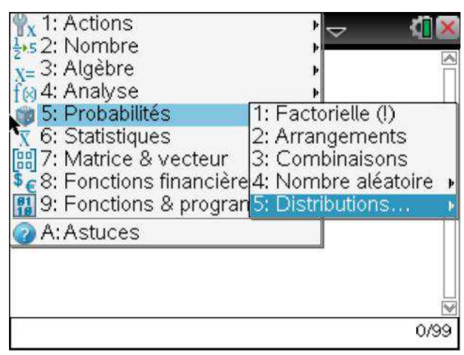


Tableau 1.2 Quelques modèles continus

Lois	TI	$P(a \leq X \leq b)$	Loi inverse	utilité
Normale $X \sim N(\mu, \sigma^2)$	Menu TI	[menu] [5] [5] [2]	[menu] [5] [5] [3]	Estimation tests d'hypothèses
	Fonction	$normCdf(a, b, \mu, \sigma)$	$invnorm(\alpha, \mu, \sigma)$	
Student $X \sim t_\nu$	Menu TI	[menu] [5] [5] [5]	[menu] [5] [5] [6]	Estimation tests d'hypothèses
	Fonction	$tCdf(a, b, \nu)$	$invt(\alpha, \nu)$	
Khi-deux $X \sim \chi^2_\nu$	Menu TI	[menu] [5] [5] [8]	[menu] [5] [5] [9]	Tests d'ajustement
	Fonction	$\chi^2 Cdf(a, b, \nu)$	$inv\chi^2(\alpha, \nu)$	
Fisher $X \sim F_{\nu_1, \nu_2}$	Menu TI	[menu] [5] [5] [B]	[menu] [5] [5] [C]	Test d'égalité des variances Régression linéaire ANOVA
	Fonction	$FCdf(a, b, \nu_1, \nu_2)$	$invF(\alpha, \nu_1, \nu_2)$	

Remarques :

1. La loi exponentielle n'est pas définie dans la TI. On peut construire une fonction **expCdf** (voir la procédure présentée pour créer les fonctions **hpgPdf** et **hpgCdf** dans la section des modèles discrets). On peut toutefois tout simplement utiliser le fait que si $X \sim Exp(\theta)$, on a alors
$$P(X \leq x) = 1 - e^{\frac{-x}{\theta}}.$$
2. Pour les lois inverses, la quantité α représente la surface **à gauche** du point recherché. Autrement dit, on cherche la valeur c telle que $P(X \leq c) = \alpha$.

Exemple 1.6

Supposons que la taille des étudiants de l'ÉTS est distribuée selon une loi normale de moyenne 168 cm et d'écart-type 10 cm. On veut former une équipe de basketball constituée des 10% des étudiants les plus grands. À partir de quelle taille peut-on prétendre faire partie de l'équipe ?

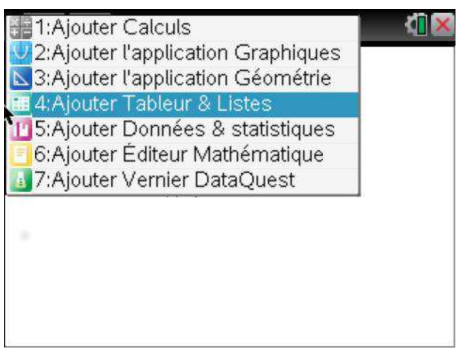
Solution :

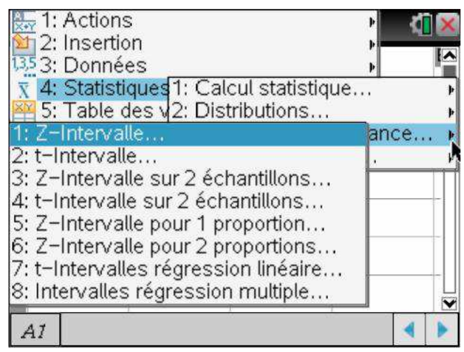
Partie 2

DEUXIÈME PARTIE

2.1 Estimation

1. Ouvrir un *Tableur & listes* : [CTRL] [doc] [4]
2. Les outils utilisés pour l'estimation se trouvent dans le menu *Statistiques/Intervalles de confiance* : [menu] [4] [3]





où chacun de ces choix représentent les intervalles de confiance dans les contextes suivants :

Utilitaire d'intervalles de confiance	
1 : Z-Intervalle	I.C pour une moyenne μ dans le cas où σ est connu
2 : t-Intervalle	I.C pour une moyenne μ dans le cas où σ est inconnu
3 : Z-intervalle sur 2 échantillons	I.C pour la différence de deux moyennes $\mu_1 - \mu_2$ dans le cas où les écarts-types σ_1 et σ_2 sont connus
4 : t-intervalle sur 2 échantillons	I.C pour la différence de deux moyennes $\mu_1 - \mu_2$ dans le cas où les écarts-types σ_1 et σ_2 sont inconnus et supposés égaux (groupé = oui) ou non supposés égaux (groupé = non)
5 : Z-intervalle pour une proportion	I.C pour une proportion p
6 : Z-intervalle pour 2 proportions	I.C pour la différence de deux proportions $p_1 - p_2$ (ce sujet n'est pas couvert dans le cours)
7 : t-intervalles régression linéaire	I.C pour estimer la pente de la droite de régression linéaire simple ou pour estimer une prévision toujours dans le cadre de la régression linéaire simple (voir les détails dans le chapitre sur la régression)
8 : Intervalles régression multiple	I.C dans le cadre de la régression multiple (voir les détails dans le chapitre sur la régression)

Dans le cadre du cours, on utilise ces outils dans les différentes situations suivantes.

Utilitaire d'intervalles de confiance dans le cadre du cours		
Paramètres	Cas	Menu TI
μ	σ connu	[menu] [4] [3] [1]
	σ inconnu	[menu] [4] [3] [2]
$\mu_1 - \mu_2$	Les variances σ_1^2 et σ_2^2 sont connues	[menu] [4] [3] [3]
	Les variances σ_1^2 et σ_2^2 sont inconnues mais supposées égales	[menu] [4] [3] [4] Groupé = OUI
	Les variances σ_1^2 et σ_2^2 sont inconnues mais non supposées égales	[menu] [4] [3] [4] Groupé = NON
p	Toujours	[menu] [4] [3] [5]
β_1 ou une prévision y_0	Dans le cadre de la régression (voir la section sur la régression)	[menu] [4] [3] [7]

Illustrons comment utiliser ces fonctions dans le cadre d'un exemple.

Exemple 2.1

Des essais sur la durée de vie en heures de 16 ampoules ont donné les résultats suivants :
434, 405, 451, 423, 431 , 463, 418, 425, 423, 438, 422, 407, 394, 444, 419, 433.

a) Estimer par un intervalle de confiance de niveau 95% la durée de vie moyenne de ce type d'ampoules.

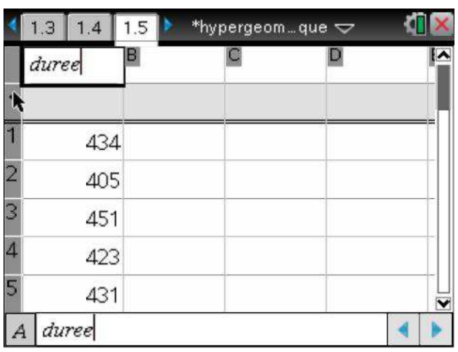
b) Quelle est la marge d'erreur ?

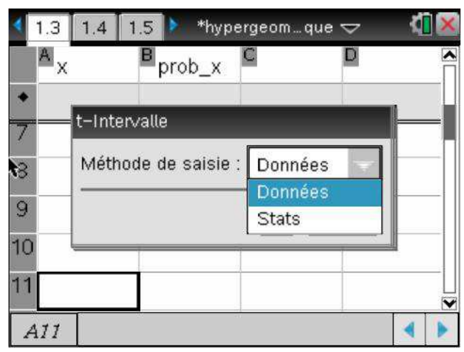
c) Quelle aurait dû être la taille de l'échantillon pour que l'erreur d'estimation ne dépasse pas 6 heures dans 19 cas sur 20 ?

Solution :

Dans ce problème, on veut estimer une *moyenne* dans le cas où l'écart-type dans la population est *inconnu* ([menu] [4] [3] [2]). Pour répondre à ces questions, il faut d'abord exécuter l'utilitaire d'intervalles de confiance de la façon suivante.

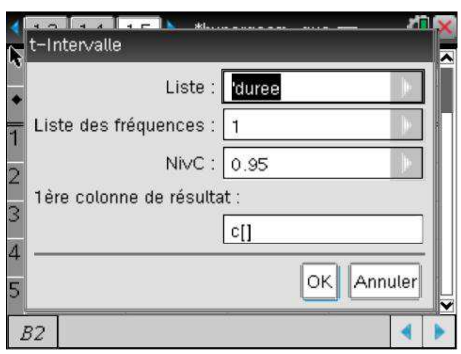
1. Entrer les données dans une liste
2. Sélectionner : [menu] [4] [3] [2] et choisir l'option *Données*



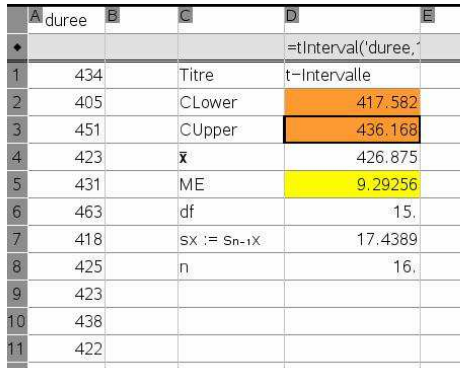


Lorsqu'on ne dispose que des résultats échantillonnaux ($\bar{x}$ et s), on choisit l'option *Stats*.

3. Remplir les informations demandées tel qu'illustré ci-dessous.



4. On obtient alors les résultats suivants



	A duree	B	C	D	E
◆				=tInterval('duree,'	
1	434		Titre	t−Intervalle	
2	405		CLower	417.582	
3	451		CUpper	436.168	
4	423		x̄	426.875	
5	431		ME	9.29256	
6	463		df	15.	
7	418		sx := sn-1x	17.4389	
8	425		n	16.	
9	423				
10	438				
11	422				

On peut maintenant répondre aux questions demandées.

a) L'intervalle de confiance est donné par $[417.582; 436.168]$.

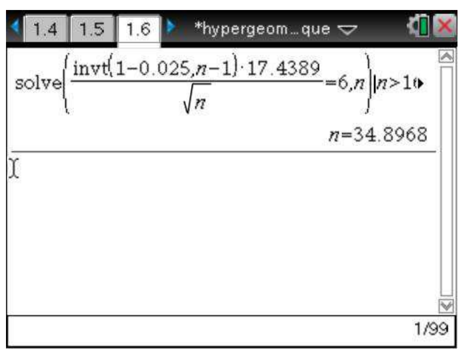
b) La marge d'erreur est

$$ME = t_{n-1;\alpha/2}\frac{s}{\sqrt{n}} = 9.29256.$$

c) Avec ces 16 ampoules, on a obtenu une marge d'erreur de 9.29 heures. On cherche donc n tel que $ME \leq 6$ avec $1-\alpha = 0.95$. Autrement dit, on veut résoudre l'équation suivante :

$$ME = t_{n-1;\alpha/2}\frac{s}{\sqrt{n}} \leq 6.$$

Avec la fonction *solve* de la TI, on obtient



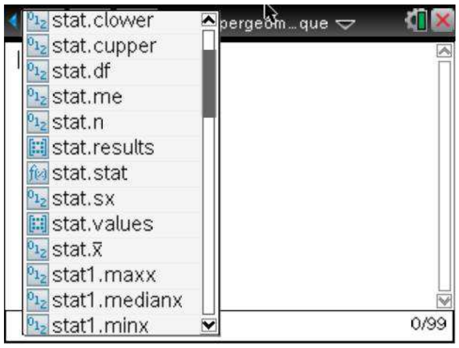
Remarques :

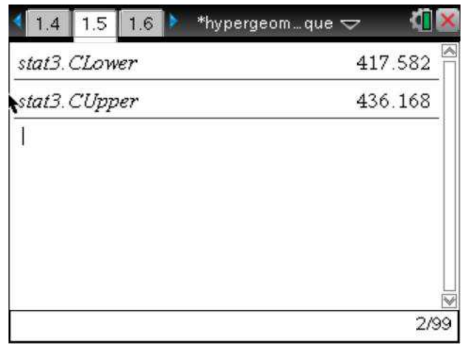
1. Cet utilitaire ne tient pas compte du facteur de correction. Il faut donc être en mesure de construire les intervalles de confiance sans utiliser cet outil. Par exemple, si nous avions eu un facteur de correction dans l'exemple précédent, l'intervalle à calculer aurait été :

$$\bar{x} \pm t_{n-1;\alpha/2}\frac{s}{\sqrt{n}}\sqrt{1-\frac{n}{N}}.$$

Toutes ces quantités s'obtiennent facilement avec la TI : $\bar{x}$ et s peuvent être obtenus à partir de l'utilitaire de *calculs statistiques* [menu] [4] [1] [1] (voir la section des statistiques descriptives) et $t_{n-1;\alpha/2}$ est tout simplement la fonction $invt(1-\alpha/2, n-1)$.

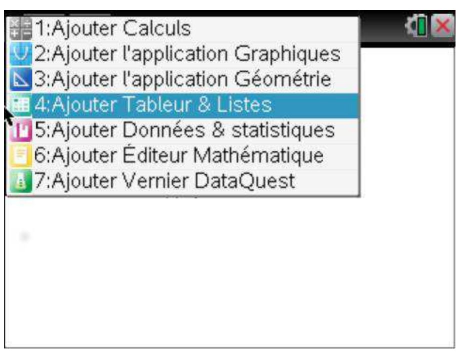
2. Dans la fenêtre de calculs, on peut obtenir chacune des quantités obtenues par l'utilitaire d'*intervalles de confance*. En appuyant sur la touche [var], on retrouve la terminologie utilisée par défaut pour ces quantités. On peut ainsi, par exemple, obtenir les bornes de l'intervalle de confiance directement dans une feuille de calculs de la façon suivante.

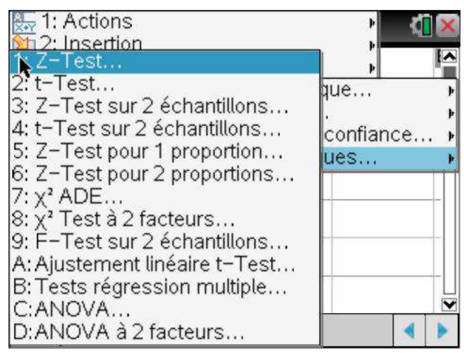




2.2 Tests d'hypothèses

1. Ouvrir un *Tableur & listes* : [CTRL] [doc] [4]
2. Les outils utilisés pour les tests d'hypothèses se trouvent dans le menu *Statistiques/Tests statistiques* : [menu] [4] [4]





où chacun de ces choix représentent les tests d'hypothèses dans les contextes suivants :

Utilitaire de tests statistiques	
1 : Z-Test	Test sur une moyenne μ dans le cas où σ est connu
2 : t-Test	Test sur une moyenne μ dans le cas où σ est inconnu
3 : Z-Test sur 2 échantillons	Test sur l'égalité de deux moyennes, $H_0 : \mu_1 = \mu_2$ dans le cas où les écarts-types σ_1 et σ_2 sont connus
4 : t-Test sur 2 échantillons	Test sur l'égalité de deux moyennes, $H_0 : \mu_1 = \mu_2$ dans le cas où les écarts-types σ_1 et σ_2 sont inconnus et supposés égaux (groupé = oui) ou non supposés égaux (groupé = non)
5 : Z-Test pour une proportion	Test sur une proportion p
6 : Z-Test pour 2 proportions	Test sur l'égalité de deux proportions, $H_0 : p_1 = p_2$ (ce sujet n'est pas couvert dans le cours)
7 : χ^2 ADE	Test du khi-carré. Test d'ajustement ou test d'indépendance
8 : χ^2 Test à 2 facteurs	Tests du khi-carré à deux facteurs (ce sujet n'est pas couvert dans le cours)
9 : F-Test sur 2 échantillons	Test de l'égalité des variances
A : Ajustement linéaire t-Test	Permet de performer une analyse de régression linéaire simple (voir les détails dans le chapitre sur la régression)
B : Tests régression multiple	Permet de performer une analyse de régression multiple (voir les détails dans le chapitre sur la régression)
C : ANOVA	Analyse de variance à un facteur (ce sujet n'est pas couvert dans le cours)
D : ANOVA à 2 facteurs	Analyse de variance à deux facteurs (ce sujet n'est pas couvert dans le cours)

Dans le cadre du cours, on utilise ces outils dans les différentes situations suivantes.

Utilitaire de tests statistiques dans le cadre du cours		
Test	Cas	Menu TI
$H_0 : \mu = \mu_0$	σ connu	[menu] [4] [4] [1]
	σ inconnu	[menu] [4] [4] [2]
$H_0 : \mu_1 = \mu_2$	Les variances σ_1^2 et σ_2^2 sont connues	[menu] [4] [4] [3]
	Les variances σ_1^2 et σ_2^2 sont inconnues mais supposées égales	[menu] [4] [4] [4] Groupé = OUI
	Les variances σ_1^2 et σ_2^2 sont inconnues mais non supposées égales	[menu] [4] [4] [4] Groupé = NON
$H_0 : p = p_0$	Toujours	[menu] [4] [4] [5]
$H_0 : \sigma_1^2 = \sigma_2^2$	Toujours	[menu] [4] [4] [9]
$H_0 : \beta_1 = 0$	Dans le cadre de la régression (voir la section sur la régression)	[menu] [4] [4] [A]
$H_0 : \beta_1 = \beta_2 = \ldots = \beta_p = 0$	Dans le cadre de la régression multiple (voir la section sur la régression)	[menu] [4] [4] [B]

Illustrons comment utiliser ces outils dans le cadre de quelques exemples.

Exemple 2.2

On étudie la proportion de circuits intégrés défectueux d'une production. Sur un échantillon aléatoire de 500 circuits intégrés, on a trouvé 32 circuits défectueux. Peut-on conclure que la proportion de circuits défectueux excède 4% avec un risque d'erreur $\alpha = 5\%$?

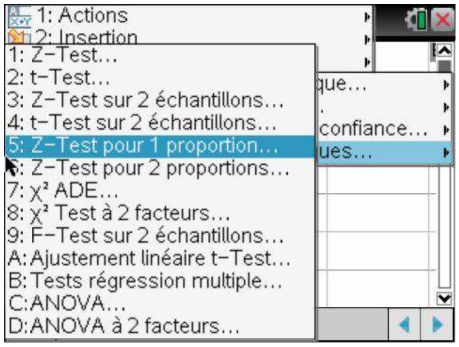
Solution :

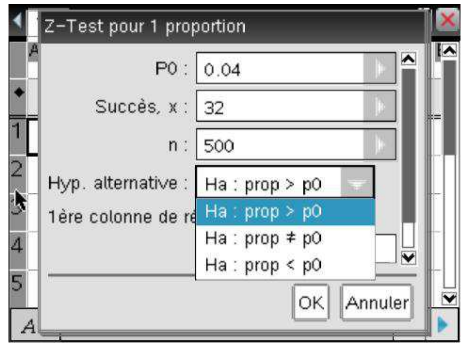
Il s'agit d'un test sur une proportion où on veut confronter les hypothèses :

$$H_0 \; : \; p = 0.04 \quad \text{vs} \quad H_1 \; : \; p > 0.04$$

avec un risque d'erreur $\alpha = 5\%$.

1. À partir d'une feuille *Tableur & listes*, sélectionner : [menu] [4] [4] [5] et remplir les informations demandées tel qu'illustré ci-dessous.



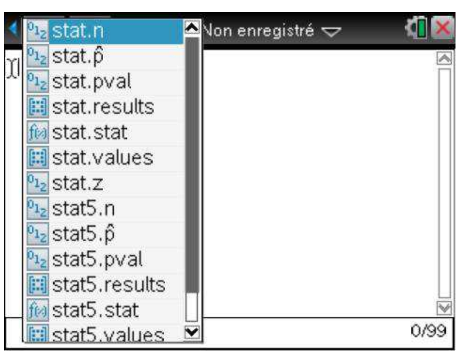


2. On obtient alors les résultats suivants

B	C
	=zTest_1Prop(0.04,32,500,1):
Titre	Z-Test pour 1 proportion un...
Hyp. alternative...	prop > p0
z	2.73861
PVal	0.003085
p̂	0.064
n	500.

3. Ainsi, puisque $\alpha_p = 0.003085 < 0.05 = \alpha$, on rejette H_0 au niveau α. On peut donc croire que la proportion de circuits défectueux excède 4% de façon significative.

4. Dans une fenêtre de calculs, on peut obtenir chacune des quantités obtenues par cette analyse. En appuyant sur la touche [var], on retrouve la terminologie utilisée par défaut pour ces quantités. On peut ainsi, par exemple, obtenir la valeur-p du test de la façon suivante



Exemple 2.3

Un chercheur veut comparer les scores à un test d'attention d'un échantillon aléatoire de 15 sujets après leur avoir fait consommer de la caféine et d'un autre échantillon aléatoire de 15 sujets à qui il fait consommer un placebo (café décaféiné). Il obtient les résultats suivant :

Résultats échantillonnaux	Caféine	Placebo
Score moyen	49	44
Écart-type	2.14	2.94
Nombre de sujets	15	15

Peut-on croire que la caféine augmente significativement le score moyen des individus à ce tests d'aptitude au niveau 5% ?

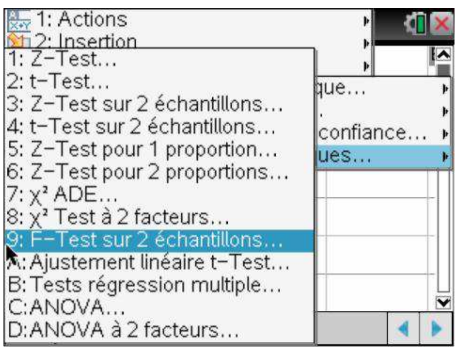
Solution :

Dans ce problème, on veut comparer deux moyennes en confrontant les hypothèses suivantes :

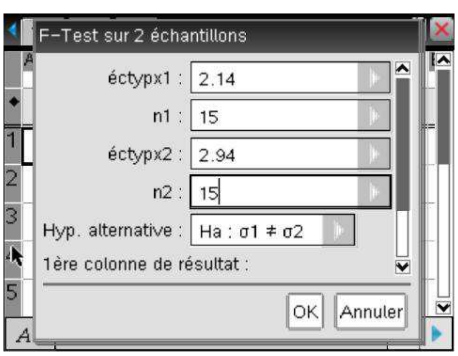
$$H_0 \ : \ \mu_1 = \mu_2 \ \text{ vs } \ H_1 \ : \ \mu_1 > \mu_2$$

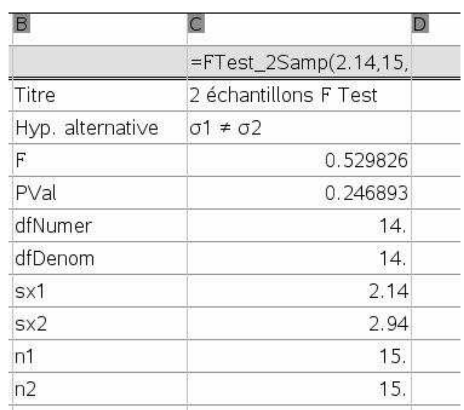
avec un risque d'erreur $\alpha = 5\%$. Puisque les variances σ_1^2 et σ_2^2 sont inconnues, on doit d'abord procéder à un test d'égalité des variances pour déterminer si on pourra considérer que les variances sont égales ou non.

1. À partir d'une feuille *Tableur & listes*, sélectionner : [menu] [4] [4] [9] et choisir l'option *Stats*.



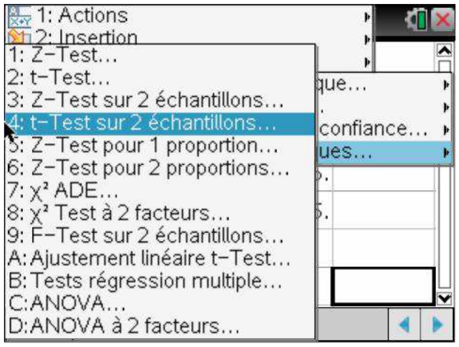
2. Remplir les informations demandées tel qu'illustré ci-dessous pour obtenir les résultats suivants.

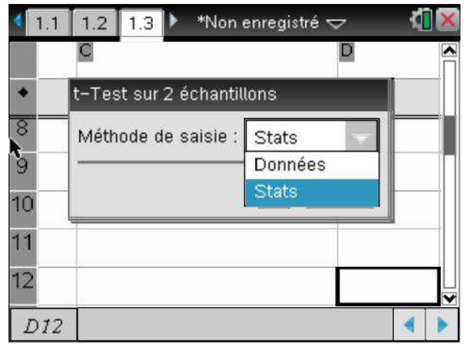




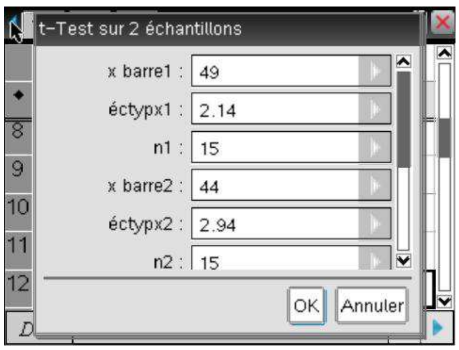
3. Puisque $\alpha_p = 0.246893 > 0.20$, on ne rejette pas l'hypothèse que les variances soient égales. Rappelons que lorsque ce n'est pas mentionné, on utilise un $\alpha = 20\%$ dans le cadre de ce test.

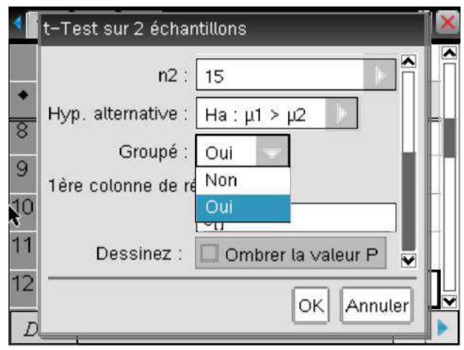
4. À partir d'une feuille *Tableur & listes*, sélectionner : [menu] [4] [4] [4] et choisir l'option *Stats*.





5. Remplir les informations demandées tel qu'illustré ci-dessous en sélectionnant l'option *OUI* pour le champ *Groupé* puisqu'on a établi que les variances pouvaient être considérées égales.





6. On obtient les résultats suivants

E	F
	=tTest_2Samp(49,2.14,15,
Titre	t-Test sur 2 échantillons...
Hyp. alternative	μ1 > μ2
t	5.32534
PVal	0.000006
df	28.
x̄1	49.
x̄2	44.
sx1	2.14
sx2	2.94
sp	2.5713
n1	15.
n2	15.

7. Puisque $\alpha_p = 0.000006 < 0.05 = \alpha$, on rejette H_0 au niveau α. On peut donc croire que la caféine augmente significativement le score moyen des individus à ce tests d'aptitude avec un risque d'erreur $\alpha = 5\%$.

Exemple 2.4

Une manufacture produit des segments pour piston allant dans des moteurs automobiles. On a choisi au hasard 15 segments pour lesquels on a calculé un diamètre moyen de 74.039 mm et un écart-type de 0.01 mm. On veut tester l'hypothèse que la moyenne des segments est de 74.035 mm. Quelle devrait être la taille échantillonnale minimale pour pouvoir détecter avec une probabilité de 95% que le diamètre des segments est en réalité de 74.030, tout en conservant une probabilité de 1% de rejeter à tort l'hypothèse nulle ?

Solution :

Il s'agit d'un test sur une moyenne où on veut confronter les hypothèses :

$$H_0 \ : \ \mu = 74.035 \quad \text{vs} \quad H_1 \ : \ \mu \neq 74.035$$

avec un risque d'erreur $\alpha = 1\%$ et où l'écart-type σ est inconnu. On cherche n tel que $\alpha = 0.01$, $\beta = 0.05$, $\mu_0 = 74.035$, $\mu_1 = 74.030$, $s = 0.01$ et le test est bilatéral. En utilisant la formule correspondante dans le tableau suivant :

Tableau 2.1 Calcul de n pour contrôler les risques α et β dans un test sur une moyenne

Calcul de n pour contrôler les risques α et β			
paramètre	Cas	Type de test	Calcul de n
μ	σ connu	Test unilatéral	$n = \frac{(z_\alpha + z_\beta)^2 \cdot \sigma^2}{(\mu_1 - \mu_0)^2}$
		Test bilatéral	$n = \frac{(z_{\alpha/2} + z_\beta)^2 \cdot \sigma^2}{(\mu_1 - \mu_0)^2}$
	σ inconnu	Test unilatéral	$n = \frac{(t_{n-1;\alpha} + t_{n-1;\beta})^2 \cdot s^2}{(\mu_1 - \mu_0)^2}$
		Test bilatéral	$n = \frac{(t_{n-1;\alpha/2} + t_{n-1;\beta})^2 \cdot s^2}{(\mu_1 - \mu_0)^2}$

On peut utiliser la fonction *solve* de la TI pour obtenir le résultat suivant

$$\text{solve}\left(\frac{(\text{invt}(1-0.005,n-1)+\text{invt}(1-0.05,n-1))^2 \cdot (0.01)^2}{(74.03-74.035)^2}=n,n\right)\Big|n>15 \qquad n=74.3203$$

Remarques :

1. Dans la fenêtre de calcul, on peut obtenir chacune des quantités obtenues par l'utilitaire de *tests d'hypothèses*. En appuyant sur la touche [var], on retrouve la terminologie utilisée par défaut pour ces quantités. Voir par exemple l'étape 4. de la solution de l'exemple 2.2.

2. Cet utilitaire ne tient pas compte du facteur de correction. Il faut donc être en mesure de faire des tests d'hypothèses sans utiliser cet outil. Les tableaux suivants présentent comment calculer la valeur-p et la puissance des tests dans les différents contextes vus au cours.

TABLEAU 2.2 Règle de décision et valeur-p d'un test sur une moyenne

Tests d'hypothèses sur μ au seuil de signification α				
Cas	Calcul de $\sigma_{\bar{X}}$ et loi de $\bar{X}$	$H_0\ :\ \mu = \mu_0$ vs	Règle de décision et zone de rejet	Calcul de α_p On rejette H_0 si $\alpha_p < \alpha$
σ connu	**Sous H_0, on a $\bar{X} \sim N(\mu_0\ ,\ \sigma_{\bar{x}}^2)$** $\sigma_{\bar{X}} = \begin{cases} \frac{\sigma}{\sqrt{n}} & \text{Si population infinie ou tirage avec remise} \\ \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} & \text{Si population finie et tirage sans remise} \end{cases}$	$H_1\ :\ \mu > \mu_0$	Rejeter H_0 si $\bar{X} \geq C$ où $C = \mu_0 + z_\alpha \cdot \sigma_{\bar{x}}$	$\alpha_p = P(\bar{X} \geq \bar{x}_{obs} \mid \mu = \mu_0)$ $\overset{TI}{=}$ *normcdf* $(\bar{x}_{obs},\ \infty,\ \mu_0,\ \sigma_{\bar{x}})$
		$H_1\ :\ \mu < \mu_0$	Rejeter H_0 si $\bar{X} \leq C$ où $C = \mu_0 - z_\alpha \cdot \sigma_{\bar{x}}$	$\alpha_p = P(\bar{X} \leq \bar{x}_{obs} \mid \mu = \mu_0)$ $\overset{TI}{=}$ *normcdf* $(-\infty,\ \bar{x}_{obs},\ \mu_0,\ \sigma_{\bar{x}})$
		$H_1\ :\ \mu \neq \mu_0$	Rejeter H_0 si $\bar{X} \leq C_1$ ou $\bar{X} \geq C_2$ où $C_2 = \mu_0 + z_{\alpha/2} \cdot \sigma_{\bar{x}}$ $C_1 = \mu_0 - z_{\alpha/2} \cdot \sigma_{\bar{x}}$	$\alpha_p = 2 \cdot min\{a, b\}$ $a = P(\bar{X} \geq \bar{x}_{obs} \mid \mu = \mu_0)$ $b = P(\bar{X} \leq \bar{x}_{obs} \mid \mu = \mu_0)$
σ inconnu	**Sous H_0, on a $T = \frac{\bar{X}-\mu_0}{\sigma_{\bar{x}}} \sim t_{n-1}$** $\sigma_{\bar{X}} \approx \begin{cases} \frac{s}{\sqrt{n}} & \text{Si population infinie ou tirage avec remise} \\ \frac{s}{\sqrt{n}} \cdot \sqrt{1 - \frac{n}{N}} & \text{Si population finie et tirage sans remise} \end{cases}$	$H_1\ :\ \mu > \mu_0$	Rejeter H_0 si $\bar{X} \geq C$ où $C = \mu_0 + t_{n-1;\alpha} \cdot \sigma_{\bar{x}}$	$\alpha_p = P(T \geq \frac{\bar{x}_{obs}-\mu_0}{\sigma_{\bar{x}}} \mid \mu = \mu_0)$ $\overset{TI}{=}$ *tcdf* $\left(\frac{\bar{x}_{obs}-\mu_0}{\sigma_{\bar{x}}}\ ,\ \infty\ ,\ n-1\right)$
		$H_1\ :\ \mu < \mu_0$	Rejeter H_0 si $\bar{X} \leq C$ où $C = \mu_0 - t_{n-1;\alpha} \cdot \sigma_{\bar{x}}$	$\alpha_p = P(T \leq \frac{\bar{x}_{obs}-\mu_0}{\sigma_{\bar{x}}} \mid \mu = \mu_0)$ $\overset{TI}{=}$ *tcdf* $\left(-\infty\ ,\ \frac{\bar{x}_{obs}-\mu_0}{\sigma_{\bar{x}}}\ ,\ n-1\right)$
		$H_1\ :\ \mu \neq \mu_0$	Rejeter H_0 si $\bar{X} \leq C_1$ ou $\bar{X} \geq C_2$ où $C_2 = \mu_0 + t_{n-1;\alpha/2} \cdot \sigma_{\bar{x}}$ $C_1 = \mu_0 - t_{n-1;\alpha/2} \cdot \sigma_{\bar{x}}$	$\alpha_p = 2 \cdot P(T \geq \left\lvert\frac{\bar{x}_{obs}-\mu_0}{\sigma_{\bar{x}}}\right\rvert \mid \mu = \mu_0)$ $\overset{TI}{=}$ $2 \cdot$ *tcdf* $\left(\left\lvert\frac{\bar{x}_{obs}-\mu_0}{\sigma_{\bar{x}}}\right\rvert\ ,\ \infty\ ,\ n-1\right)$

Puissance d'un test sur μ au seuil de signification α où en réalité $\mu = \mu_1$

Cas	Calcul de $\sigma_{\bar{X}}$ et loi de $\bar{X}$	$H_0 \ : \ \mu = \mu_0$ vs	Règle de décision et zone de rejet (calculée sous H_0)	Calcul de la puissance $1-\beta$ (calculée sous H_1)
σ connu	**Sous H_1, on a $\bar{X} \sim N(\mu_1 \ , \ \sigma_{\bar{x}}^2)$** $\sigma_{\bar{X}} = \begin{cases} \frac{\sigma}{\sqrt{n}} & \text{Si population infinie ou tirage avec remise} \\ \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} & \text{Si population finie et tirage sans remise} \end{cases}$	$H_1 \ : \ \mu > \mu_0$	Rejeter H_0 si $\bar{X} \geq C$ où $C = \mu_0 + z_\alpha \cdot \sigma_{\bar{x}}$	$1-\beta = P(\bar{X} \geq C \mid \mu = \mu_1)$ $\overset{TI}{=} \ normalcdf\,(C, \ \infty, \ \mu_1, \ \sigma_{\bar{x}})$
		$H_1 \ : \ \mu < \mu_0$	Rejeter H_0 si $\bar{X} \leq C$ où $C = \mu_0 - z_\alpha \cdot \sigma_{\bar{x}}$	$1-\beta = P(\bar{X} \leq C \mid \mu = \mu_1)$ $\overset{TI}{=} \ normalcdf\,(-\infty, \ C, \ \mu_1, \ \sigma_{\bar{x}})$
		$H_1 \ : \ \mu \neq \mu_0$	Rejeter H_0 si $\bar{X} \leq C_1$ ou $\bar{X} \geq C_2$ où $C_2 = \mu_0 + z_{\alpha/2} \cdot \sigma_{\bar{x}}$ $C_1 = \mu_0 - z_{\alpha/2} \cdot \sigma_{\bar{x}}$	$1-\beta =$ $1 \ - \ P(C_1 \leq \bar{X} \leq C_2 \mid \mu = \mu_1)$ $\overset{TI}{=} 1 - normalcdf\,(C_1, \ C_2, \ \mu_1, \ \sigma_{\bar{x}})$
σ inconnu	**Sous H_1, on a $T = \frac{\bar{X}-\mu_1}{\sigma_{\bar{x}}} \sim t_{n-1}$** $\sigma_{\bar{X}} \approx \begin{cases} \frac{s}{\sqrt{n}} & \text{Si population infinie ou tirage avec remise} \\ \frac{s}{\sqrt{n}} \cdot \sqrt{1 - \frac{n}{N}} & \text{Si population finie et tirage sans remise} \end{cases}$	$H_1 \ : \ \mu > \mu_0$	Rejeter H_0 si $\bar{X} \geq C$ où $C = \mu_0 + t_{n-1;\alpha} \cdot \sigma_{\bar{x}}$	$1-\beta = P(T \geq \frac{C-\mu_1}{\sigma_{\bar{x}}} \mid \mu = \mu_1)$ $\overset{TI}{=} \ tcdf\left(\frac{C-\mu_1}{\sigma_{\bar{x}}}, \ \infty, \ n-1\right)$
		$H_1 \ : \ \mu < \mu_0$	Rejeter H_0 si $\bar{X} \leq C$ où $C = \mu_0 - t_{n-1;\alpha} \cdot \sigma_{\bar{x}}$	$1-\beta = P(T \leq \frac{C-\mu_1}{\sigma_{\bar{x}}} \mid \mu = \mu_1)$ $\overset{TI}{=} \ tcdf\left(-\infty, \ \frac{C-\mu_1}{\sigma_{\bar{x}}}, \ n-1\right)$
		$H_1 \ : \ \mu \neq \mu_0$	Rejeter H_0 si $\bar{X} \leq C_1$ ou $\bar{X} \geq C_2$ où $C_2 = \mu_0 + t_{n-1;\alpha/2} \cdot \sigma_{\bar{x}}$ $C_1 = \mu_0 - t_{n-1;\alpha/2} \cdot \sigma_{\bar{x}}$	$1-\beta =$ $1 - P(\frac{C_1-\mu_1}{\sigma_{\bar{x}}} \leq T \leq \frac{C_2-\mu_1}{\sigma_{\bar{x}}} \mid \mu = \mu_1)$ $\overset{TI}{=} 1 - \ tcdf\left(\frac{C_1-\mu_1}{\sigma_{\bar{x}}}, \ \frac{C_2-\mu_1}{\sigma_{\bar{x}}}, \ n-1\right)$

TABLEAU 2.4 Règle de décision et valeur-p d'un test sur une proportion

<table>
<tr><th colspan="4">Tests d'hypothèses sur p au seuil de signification α</th></tr>
<tr><th>Calcul de $\sigma_{\hat{P}}$ et loi de $\hat{P}$</th><th>$H_0 \ : \ p = p_0$
vs</th><th>Règle de décision
et zone de rejet</th><th>Calcul de α_p
On rejette H_0 si $\alpha_p < \alpha$</th></tr>
<tr><td rowspan="3">Sous H_0, on a $\hat{P} \approx N(p_0 \ , \ \sigma^2_{\hat{P}_{(0)}})$

$\sigma_{\hat{P}_{(0)}} = \begin{cases} \sqrt{\frac{p_0(1-p_0)}{n}} & \text{Si population infinie ou tirage avec remise} \\ \sqrt{\frac{p_0(1-p_0)}{n}} \cdot \sqrt{\frac{N-n}{N-1}} & \text{Si population finie et tirage sans remise} \end{cases}$

Conditions d'application
$n \geq 30$
$np_0 \geq 5$
$n(1-p_0) \geq 5$</td><td>$H_1 \ : \ p > p_0$</td><td>Rejeter H_0 si $\hat{P} \geq C$
où
$C = p_0 + z_\alpha \cdot \sigma_{\hat{p}_{(0)}}$</td><td>$\alpha_p = P(\hat{P} \geq \hat{p}_{obs} \mid p = p_0)$
$\overset{TI}{=} \ normcdf\left(\hat{p}_{obs}, \ \infty, \ p_0, \ \sigma_{\hat{p}_{(0)}}\right)$</td></tr>
<tr><td>$H_1 \ : \ p < p_0$</td><td>Rejeter H_0 si $\hat{P} \leq C$
où
$C = p_0 - z_\alpha \cdot \sigma_{\hat{p}_{(0)}}$</td><td>$\alpha_p = P(\hat{P} \leq \hat{p}_{obs} \mid p = p_0)$
$\overset{TI}{=} \ normcdf\left(-\infty, \ \hat{p}_{obs}, \ p_0, \ \sigma_{\hat{p}_{(0)}}\right)$</td></tr>
<tr><td>$H_1 \ : \ p \neq p_0$</td><td>Rejeter H_0 si $\hat{P} \leq C_1$ ou $\hat{P} \geq C_2$
où
$C_2 = p_0 + z_{\alpha/2} \cdot \sigma_{\hat{p}_{(0)}}$
$C_1 = p_0 - z_{\alpha/2} \cdot \sigma_{\hat{p}_{(0)}}$</td><td>$\alpha_p = 2 \cdot min\{a, b\}$

$a = P(\hat{P} \geq \hat{p}_{obs} \mid p = p_0)$
$b = P(\hat{P} \leq \hat{p}_{obs} \mid p = p_0)$</td></tr>
</table>

TABLEAU 2.5 Calcul de la puissance pour un test sur une proportion

Puissance d'un test sur p au seuil de signification α où en réalité $p = p_1$			
Calcul de $\sigma_{\hat{P}}$ et loi de $\hat{P}$	$H_0 \ : \ p = p_0$ vs	Règle de décision et zone de rejet (calculée sous H_0)	Calcul de la puissance $1-\beta$ (calculée sous H_1)
Sous H_1, on a $\hat{P} \approx N(p_1 \ , \ \sigma^2_{\hat{P}_{(1)}})$ $\sigma_{\hat{P}_{(1)}} = \begin{cases} \sqrt{\frac{p_1(1-p_1)}{n}} & \text{Si population infinie ou tirage avec remise} \\ \sqrt{\frac{p_1(1-p_1)}{n}} \cdot \sqrt{\frac{N-n}{N-1}} & \text{Si population finie et tirage sans remise} \end{cases}$ **Conditions d'application** $n \geq 30$ $np_1 \geq 5$ $n(1-p_1) \geq 5$	$H_1 \ : \ p > p_0$	Rejeter H_0 si $\hat{P} \geq C$ où $C = p_0 + z_\alpha \cdot \sigma_{\hat{p}_{(0)}}$	$1 - \beta = P(\hat{P} \geq C \mid p = p_1)$ $\overset{TI}{=} \ normcdf\left(C, \ \infty, \ p_1, \ \sigma_{\hat{p}_{(1)}}\right)$
	$H_1 \ : \ p < p_0$	Rejeter H_0 si $\hat{P} \leq C$ où $C = p_0 - z_\alpha \cdot \sigma_{\hat{p}_{(0)}}$	$1 - \beta = P(\hat{P} \leq C \mid p = p_1)$ $\overset{TI}{=} \ normcdf\left(-\infty, \ C, \ p_1, \ \sigma_{\hat{p}_{(1)}}\right)$
	$H_1 \ : \ p \neq p_0$	Rejeter H_0 si $\hat{P} \leq C_1$ ou $\hat{P} \geq C_2$ où $C_2 = p_0 + z_{\alpha/2} \cdot \sigma_{\hat{p}_{(0)}}$ $C_1 = p_0 - z_{\alpha/2} \cdot \sigma_{\hat{p}_{(0)}}$	$1 - \beta = 1 - P(C_1 \leq \hat{P} \leq C_2 \mid p = p_1)$ $\overset{TI}{=} 1 - normcdf\left(C_1, \ C_2, \ p_1, \ \sigma_{\hat{p}_{(1)}}\right)$

Tableau 2.6 Règles de décision et valeur-p d'un test sur deux moyennes

Tests d'hypothèses sur deux moyennes au seuil de signification α				
Contexte	Statistique du test et distribution	$H_0 \ : \ \mu_1 = \mu_2$ vs	Règle de décision et zone de rejet	Calcul de α_p On rejette H_0 si $\alpha_p < \alpha$
Contexte 1 σ_1^2 et σ_2^2 **connues**	Statistique du test $Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ Sous H_0, on a $Z \sim N(0,\ 1)$	$H_1 \ : \ \mu_1 > \mu_2$	Rejeter H_0 si $Z \geq z_\alpha$	$\alpha_p = P(Z \geq Z_{obs} \mid \mu_1 = \mu_2)$ $\overset{TI}{=}$ *normcdf* $(Z_{obs},\ \infty,\ 0,\ 1)$
		$H_1 \ : \ \mu_1 < \mu_2$	Rejeter H_0 si $Z \leq -z_\alpha$	$\alpha_p = P(Z \leq Z_{obs} \mid \mu_1 = \mu_2)$ $\overset{TI}{=}$ *normcdf* $(-\infty,\ Z_{obs},\ 0,\ 1)$
		$H_1 \ : \ \mu_1 \neq \mu_2$	Rejeter H_0 si $Z \leq -z_{\alpha/2}$ ou $Z \geq z_{\alpha/2}$	$\alpha_p = 2 \cdot P(Z \geq \lvert Z_{obs} \rvert \mid \mu_1 = \mu_2)$ $\overset{TI}{=}$ $2 \cdot$ *normcdf* $(\lvert Z_{obs} \rvert,\ \infty,\ 0,\ 1)$
Contexte 2 σ_1^2 et σ_2^2 **inconnues** mais supposées égales $\sigma_1^2 = \sigma_2^2$	Statistique du test $T = \frac{\bar{X}_1 - \bar{X}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ où $s_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$ Sous H_0, on a $T \sim t_{n_1+n_2-2}$	$H_1 \ : \ \mu_1 > \mu_2$	Rejeter H_0 si $T \geq t_{n_1+n_2-2;\alpha}$	$\alpha_p = P(T \geq T_{obs} \mid \mu_1 = \mu_2)$ $\overset{TI}{=}$ *tcdf* $(T_{obs},\ \infty,\ n_1 + n_2 - 2)$
		$H_1 \ : \ \mu_1 < \mu_2$	Rejeter H_0 si $T \leq -t_{n_1+n_2-2;\alpha}$	$\alpha_p = P(T \leq T_{obs} \mid \mu_1 = \mu_2)$ $\overset{TI}{=}$ *tcdf* $(-\infty,\ T_{obs},\ n_1 + n_2 - 2)$
		$H_1 \ : \ \mu_1 \neq \mu_2$	Rejeter H_0 si $T \leq -t_{n_1+n_2-2;\alpha/2}$ ou $T \geq t_{n_1+n_2-2;\alpha/2}$	$\alpha_p = 2 \cdot P(T \geq \lvert T_{obs} \rvert \mid \mu_1 = \mu_2)$ $\overset{TI}{=}$ $2 \cdot$ *tcdf* $(\lvert T_{obs} \rvert,\ \infty,\ n_1 + n_2 - 2)$

Tableau 2.7 Règles de décision et valeur-p d'un test sur deux moyennes (suite)

Tests d'hypothèses sur deux moyennes au seuil de signification α (suite)

Contexte	Statistique du test et distribution	$H_0 \ : \ \mu_1 = \mu_2$ vs	Règle de décision et zone de rejet	Calcul de α_p On rejette H_0 si $\alpha_p < \alpha$
Contexte 3 σ_1^2 et σ_2^2 **inconnues** et ne sont pas supposées égales $\sigma_1^2 \neq \sigma_2^2$	Statistique du test $T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$ Sous H_0, on a $T \sim t_\nu$ où $\nu = \frac{\left(S_1^2/n_1 + S_2^2/n_2\right)^2}{(S_1^2/n_1)^2/(n_1-1) + (S_2^2/n_2)^2/(n_2-1)}.$	$H_1 \ : \ \mu_1 > \mu_2$	Rejeter H_0 si $T \geq t_{\nu;\alpha}$	$\alpha_p = P(T \geq T_{obs} \mid \mu_1 = \mu_2)$ $\overset{TI}{=} \ tcdf\,(T_{obs}, \ \infty, \ \nu)$
		$H_1 \ : \ \mu_1 < \mu_2$	Rejeter H_0 si $T \leq -t_{\nu;\alpha}$	$\alpha_p = P(T \leq T_{obs} \mid \mu_1 = \mu_2)$ $\overset{TI}{=} \ tcdf\,(-\infty, \ T_{obs}, \ \nu)$
		$H_1 \ : \ \mu_1 \neq \mu_2$	Rejeter H_0 si $T \leq -t_{\nu;\alpha/2}$ ou $T \geq t_{\nu;\alpha/2}$	$\alpha_p = 2 \cdot P(T \geq \lvert T_{obs}\rvert \mid \mu_1 = \mu_2)$ $\overset{TI}{=} \ 2 \cdot tcdf\,(\lvert T_{obs}\rvert, \ \infty, \ \nu)$

Choix entre le contexte 2 ou 3 : test de l'égalité des variances

On veut confronter les hypothèses : $H_0 \ : \ \sigma_1^2 = \sigma_2^2$ vs $H_1 \ : \ \sigma_1^2 \neq \sigma_2^2$

On calcule la statistique : $F = \frac{S_1^2}{S_2^2}$ et la valeur-p associée à ce résultat

$$
\begin{aligned}
\text{si } F_{obs} \geq 1 : \quad & \alpha_p = 2 \cdot P(F \geq F_{obs} \mid H_0) \\
& \overset{TI}{=} 2 \cdot Fcdf\,(F_{obs}, \ \infty, \ n_1 - 1, \ n_2 - 1) \\
\text{si } F_{obs} < 1 : \quad & \alpha_p = 2 \cdot P(F \leq F_{obs} \mid H_0) \\
& \overset{TI}{=} 2 \cdot Fcdf\,(-\infty, \ F_{obs}, \ n_1 - 1, \ n_2 - 1)
\end{aligned}
$$

2.3 Régression linéaire

2.3.1 Régression linéaire simple

Une analyse de régression linéaire simple se fait à partir des tableaux suivants :

TABLEAU 2.8 Tableau d'analyse de la variance (ANOVA) pour la régression linéaire simple

Tableau d'analyse de la variance (ANOVA)					
Source de variation	Somme des carrés	Degrés de liberté	Moyenne des carrés	Statistique F	valeur-p
Régression	SCR	1	$MCR = \frac{SCR}{1}$	$F = \frac{MCR}{MCE}$	$\alpha_p = P(F_{1;n-2} \geq F_{obs})$
Erreur	SCE	$n-2$	$MCE = \frac{SCE}{n-2}$		
Totale	SCT	$n-1$			

TABLEAU 2.9 Tableau des estimateurs des paramètres pour la régression linéaire simple

Estimation des paramètres				
Paramètres	Estimateurs	Écart-type de l'estimateur	Statistique T sous $H_0 : \beta_1 = 0$	valeur-p
β_0	$\hat{\beta}_0$	$s(\hat{\beta}_0) = s\sqrt{\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]}$	$T = \frac{\hat{\beta}_0}{s(\hat{\beta}_0)} \sim t_{n-2}$	$\alpha_p = 2 \cdot P(T_{n-2} \geq \lvert T_{obs}\rvert)$
β_1	$\hat{\beta}_1$	$s(\hat{\beta}_1) = \frac{s}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$	$T = \frac{\hat{\beta}_1}{s(\hat{\beta}_1)} \sim t_{n-2}$	$\alpha_p = 2 \cdot P(T_{n-2} \geq \lvert T_{obs}\rvert)$

où

$$
\begin{aligned}
SCT = &\quad \textstyle\sum_{i=1}^{n}(y_i-\bar{y})^2 \\
SCR = &\quad \textstyle\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2 = \hat{\beta}_1^2\cdot\sum_{i=1}^{n}(x_i-\bar{x})^2 = \hat{\beta}_1^2\cdot\left(\sum_{i=1}^{n}x_i^2-n\bar{x}^2\right) = R^2\cdot SCT \\
SCE = &\quad \textstyle\sum_{i=1}^{n}e_i^2 = \sum_{i=1}^{n}(y_i-\hat{y}_i)^2 = (n-2)s^2 \\
\hat{\beta}_0 = &\quad \bar{y}-\hat{\beta}_1\bar{x} \\
\hat{\beta}_1 = &\quad \frac{\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}}{\sum_{i=1}^{n}x_i^2-n\bar{x}^2} = r\frac{s_y}{s_x}.
\end{aligned}
$$

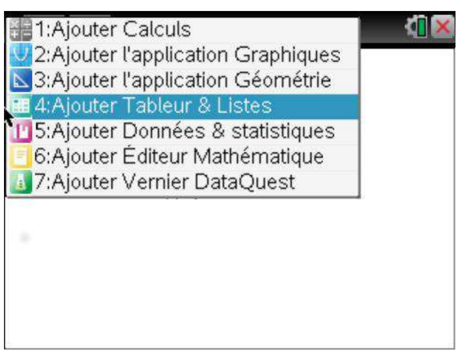
Illustrons comment utiliser la TI pour obtenir ces différentes quantités à l'aide d'un exemple.

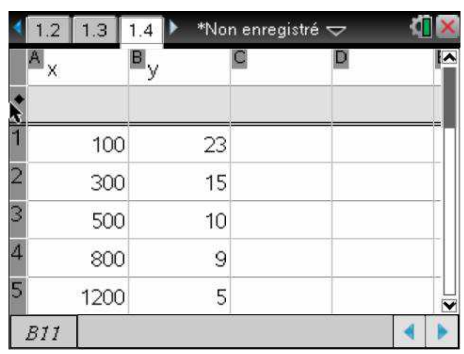
Exemple 2.5

Une entreprise située à Montréal se demande s'il existe un lien linéaire entre la durée (en min) des appels interurbains faits par la compagnie et la distance (en km) de Montréal de la ville où a été logé l'appel. On a observé les résultats suivants :

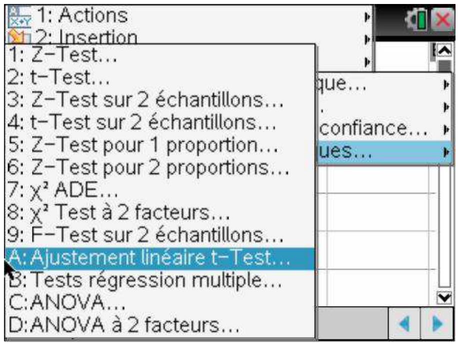
X (en km)	100	300	500	800	1200	50	700	1800	600	1500
Y (en min)	23	15	10	9	5	30	10	7	12	8

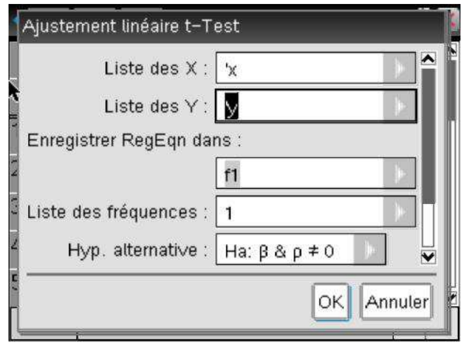
1. Ouvrir un *Tableur & listes* : [CTRL] [doc] [4]
2. Entrer les données dans deux listes X et Y





3. Sélectionner : [menu] [4] [4] [A] et remplir les informations demandées tel qu'illustré ci-dessous.





4. On obtient alors les résultats suivants :

Tableau 2.10 Résultats de l'utilitaire de régression ([menu] [4] [4] [A])

C	D
	=LinRegtTest('x,'y,1,0): C
Titre	Régression linéaire t–Te…
Hyp. alternative	β & ρ ≠ 0
RegEqn	a+b*x
t	-3.51085
PVal	0.007952
df	8.
a	20.7873
b	-0.010447
s	5.2155
SESlope	0.002976
r²	0.606417
r	-0.778728
Resid	{3.2573846529417,-2.65…

On voudrait répondre aux questions suivantes :

a) Calculer le coefficient de corrélation linéaire entre X et Y.

b) Peut-on affirmer que la durée d'un appel interurbain et la distance de Montréal de la ville où a été logé l'appel sont des variables linéairement dépendantes au niveau 5% ?

c) Obtenir la droite des moindres carrés.

d) Calculer les résidus.

e) Que vaut la variance résiduelle ?

f) Calculer les trois sommes de carrés : SCT, SCR et SCE.

g) Calculer le coefficient de détermination et interpréter sa valeur.

h) Remplir toutes les cases du tableau d'analyse de la variance et du tableau des coefficients en utilisant la TI.

ANALYSE DE VARIANCE

	Degré de liberté	Somme des carrés	Moyenne des carrés	F	Valeur critique de F
Régression					
Résidus					
Total					

	Coefficients	Erreur-type	Statistique t	Probabilité
Constante				
Distance				

i) Estimer la durée d'un appel interurbain pour un appel fait dans une ville située à 150 km de Montréal par un intervalle de confiance de niveau 90%.

j) Estimer la durée *moyenne* d'un appel interurbain pour des appels faits à 150 km de Montréal par un intervalle de confiance de niveau 90%.

k) Obtenir le nuage de points et tracer la droite des moindres carrés à même ce graphique.

l) Obtenir le graphique des résidus.

Solution :

a) Directement à partir des résultats obtenus à l'étape 4, on a $r = 0.778728$.

b) On veut confronter les hypothèses

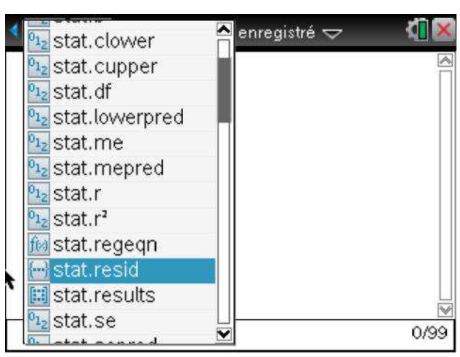
$$H_0 \; : \; \rho = 0 \;\; \text{vs} \;\; H_1 \; : \; \rho \neq 0$$

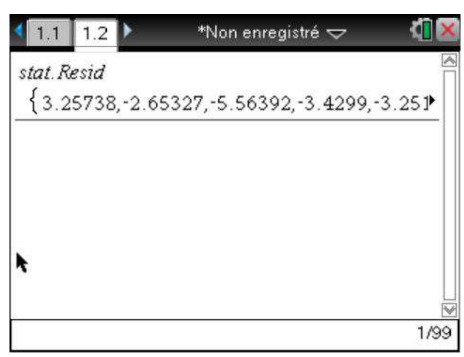
où ρ est le coefficient de corrélation linéaire théorique entre X et Y. La statistique du test $T = r\sqrt{\frac{n-2}{1-r^2}} \sim t_{n-2}$ si H_0 est vraie. Encore une fois, à partir des résultats obtenus à l'étape 4, on obtient $T_{obs} = -3.51085$ et $\alpha_p = 2 \cdot P(T \geq |T_{obs}|) = 0.007952$. Donc, puisque $\alpha_p < \alpha = 5\%$, on rejette H_0 au niveau 5% et ainsi on peut croire que le lien linéaire est significatif.

c) $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$. On obtient :

$$\hat{Y} = 20.7873 - 0.010447 \cdot X$$

d) Dans une feuille de calculs, on peut voir tous les résidus via le bouton [var] qui sont sauvegardés dans la variable *stat.resid*.



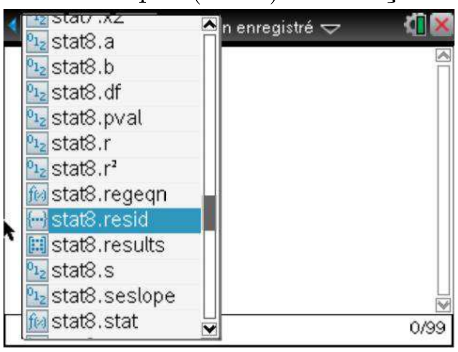


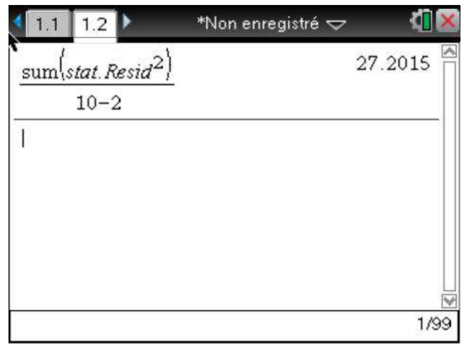
e) La variance résiduelle se calcule de la façon suivante :

$$s^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-2} = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-2}.$$

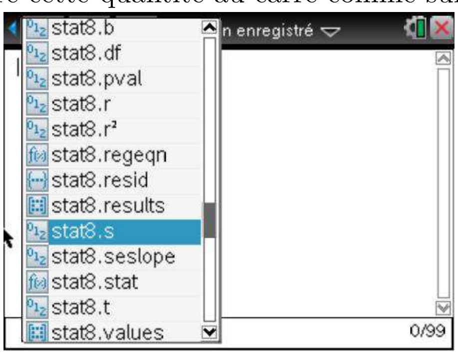
On peut obtenir la variance résiduelle très rapidement avec la Nspire des deux façons suivantes :

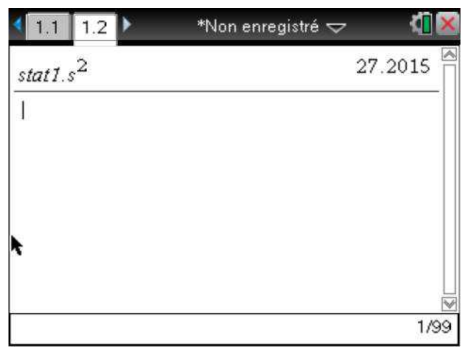
- On peut obtenir le résultat en additionnant directement les carrés des résidus et en divisant le tout par $(n-2)$ de la façon suivante :





- On peut aussi tout simplement prendre l'écart-type résiduel, s, obtenu à l'étape 4 et mettre cette quantité au carré comme suit :

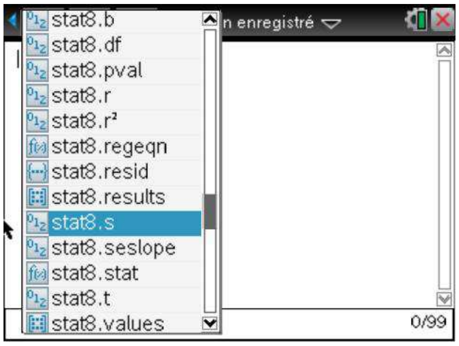




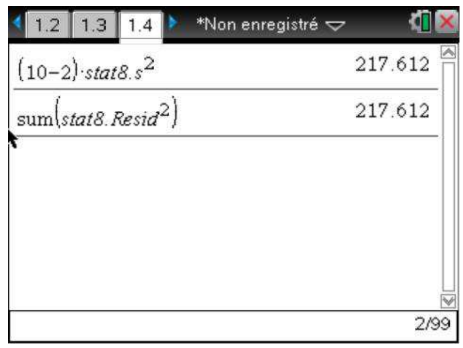
f) Plusieurs approches peuvent être utilisées pour obtenir ces quantités, en voici quelques-unes.

SCE D'abord pour la somme des carrés des résidus, SCE, on peut l'obtenir de 2 façons très simples :

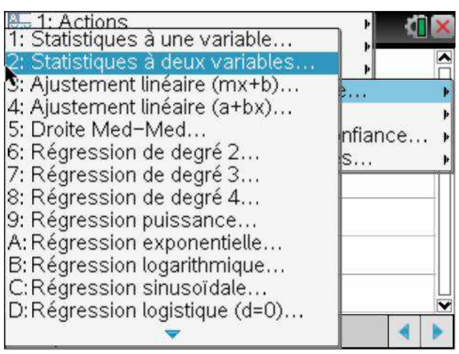
- On sait que $SCE = \sum_{i=1}^{n} e_i^2 = (n-2)s^2$. On peut tout simplement prendre l'écart-type résiduel, s, obtenu à l'étape 4 et mettre cette quantité au carré et multiplier ce résultat par $(n-2)$ comme suit :

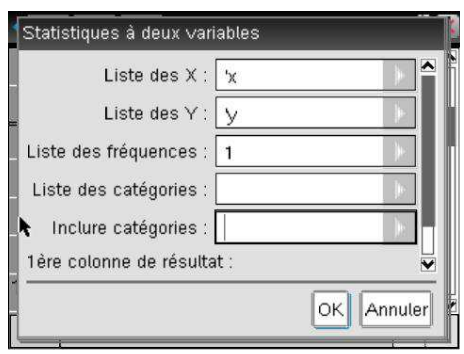


- On peut aussi obtenir le même résultat en additionnant directement les carrés des résidus de la façon suivante :



SCT Pour obtenir SCT, on peut utiliser l'utilitaire de calculs statistiques à deux variables en effectuant les opérations suivantes : [menu] [4] [1] [2] et ensuite remplir les informations demandées tel qu'illustré ci-dessous.





On observe ensuite directement

$$SCT = 552.9$$

SCR Plusieurs approches permettent d'obtenir cette somme de carrés. Par exemple :

- $SCR = SCT - SCE = 552.9 - 217.612 = 335.288$
- On peut aussi utiliser l'équivalence suivante :

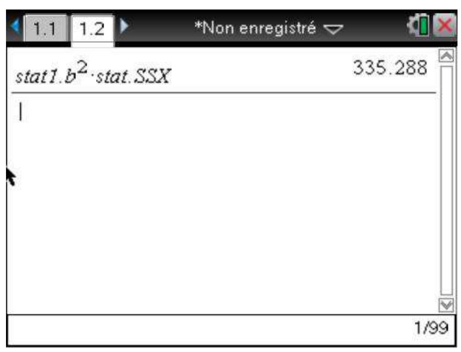
$$SCR = \hat{\beta}_1^2 \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$$

TABLEAU 2.11 Résultats des calculs statistiques pour deux variables ([menu] [4] [1] [2])

Titre	Statistiques à deux...
x̄	755.
Σx	7550.
Σx²	8.7725E6
sx := sn-1x	584.261
σx := σnx	554.279
n	10.
ȳ	12.9
Σy	129.
Σy²	2217.
sy := sn-1y	7.83794
σy := σny	7.43572

Σxy	65300.
r	-0.778728
MinX	50.
Q₁X	300.
MedianX	650.
Q₃X	1200.
MaxX	1800.
MinY	5.
Q₁Y	8.
MedianY	10.
Q₃Y	15.
MaxY	30.
SSX := Σ(x−x̄)²	3.07225E6
SSY := Σ(y−ȳ)²	552.9

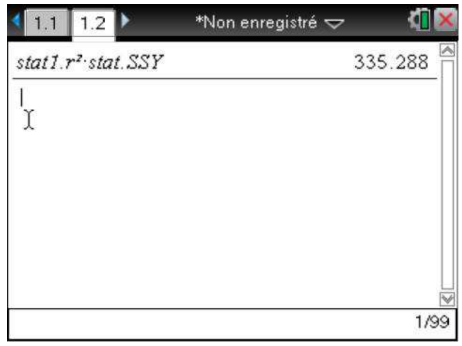
où $\hat{\beta}_1$ s'obtient directement des résultats du tableau 2.10 ($b = -0.010447$) et $\sum_{i=1}^{n}(x_i - \bar{x})^2$ s'obtient suite à l'analyse présentée au tableau 2.11 ($SSX = 3.07225 \times 10^6$). On peut obtenir directement ces quantités via le bouton [var] tel qu'illustré ci-dessous.



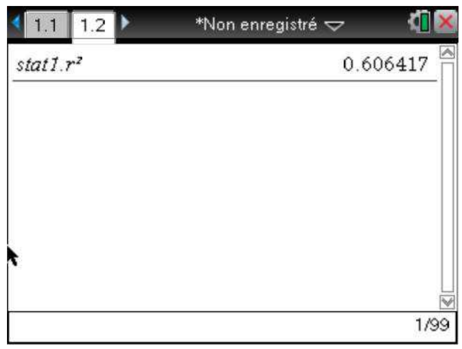
- Enfin, on peut utiliser le fait que :

$$SCR = R^2 \cdot SCT$$

où R^2 s'obtient directement des résultats du tableau 2.10 ($r^2 = 0.606417$) et SCT s'obtient suite à l'analyse présentée au tableau 2.11 ($SSY = 552.9$). On peut obtenir directement ces quantités via le bouton [var] tel qu'illustré ci-dessous.



g) On obtient directement le coefficient de détermination directement des résultats du tableau 2.10 ($r^2 = 0.606417$) ou en utilisant le bouton [var] tel qu'illustré ci-dessous :



h) Comparez vos résultats aux tableaux obtenus suite à une analyse de régression sur un logiciel statistique (ici, *Statgraphics* a été utilisé).

Régression simple - Y en fonction de X
Variable à expliquer: Y
Variable explicative: X
Modèle linéaire: Y = a + b*X

Coefficients

	Estimation des moindres carrés	*Erreur type*	*t*	*Probabilité*
Ordonnée	20,7873	2,78695	7,45879	0,0001
Pente	-0,0104467	0,00297556	-3,51085	0,0080

Analyse de variance

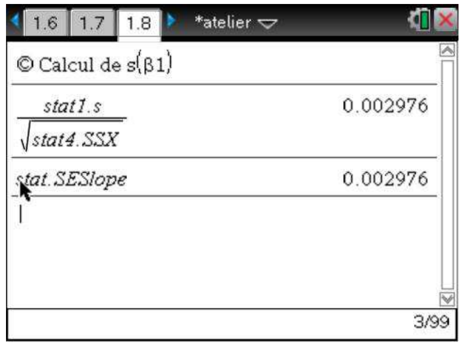
Source	*Somme des carrés*	*Ddl*	*Carré moyen*	*F*	*Probabilité*
Modèle	335,288	1	335,288	12,33	0,0080
Résidu	217,612	8	27,2015		
Total (Corr.)	552,9	9			

Coefficient de corrélation = -0,778728
R-carré = 60,6417 %

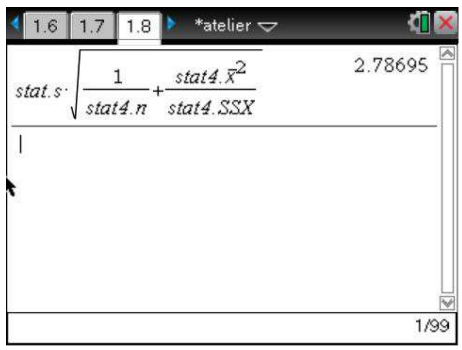
Remarques :

Pour le calcul de l'écart-type des estimateurs (colonne erreur type ci-dessous), on procède de la façon suivante :

1. $s(\hat{\beta}_1) = \frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}} = \frac{s}{\sqrt{SSX}} = SESlope$. On retrouve ces quantités directement dans les tableaux 2.10 de la page 37 et 2.11 de la page 41.



2. $s(\hat{\beta}_0) = s\sqrt{\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right]} = s\sqrt{\left[\frac{1}{n} + \frac{\bar{x}^2}{SSX}\right]}$.



i) Le tableau suivant présente les formules pour obtenir les intervalles de confiance d'une prévision et de la prévision moyenne :

Intervalles de confiance pour une prévision dans le contexte de la régression linéaire simple

L'intervalle de confiance de niveau $1 - \alpha$ pour estimer la valeur y_0 lorsque $x = x_0$ est donné par

$$\hat{y}_0 \ \pm \ t_{n-2;\alpha/2} \cdot s \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

Et l'intervalle de confiance de niveau $1 - \alpha$ pour estimer la **valeur moyenne** d'observations de y_0 lorsque $x = x_0$ est donné par

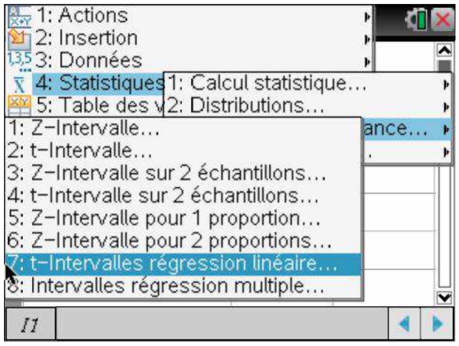
$$\hat{y}_0 \ \pm \ t_{n-2;\alpha/2} \cdot s \cdot \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

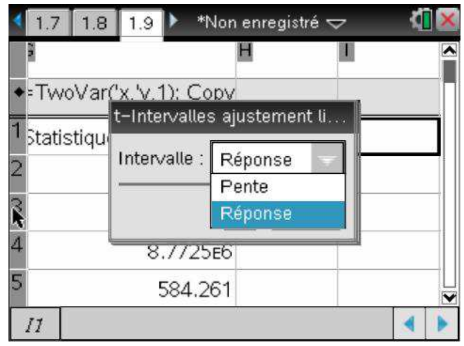
Dans notre exemple, l'équation de la droite des moindres carrés est :

$$\hat{y} = 20.7873 - 0.010447x.$$

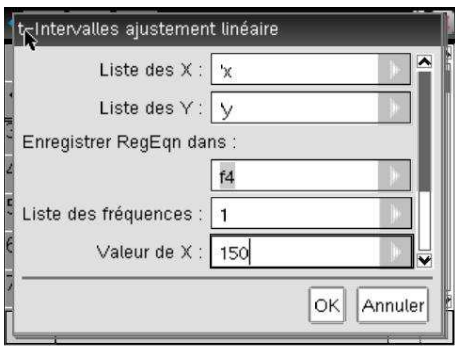
On veut estimer la durée d'un appel interurbain pour un appel fait à 150 km de Montréal par un intervalle de confiance de niveau 90%.

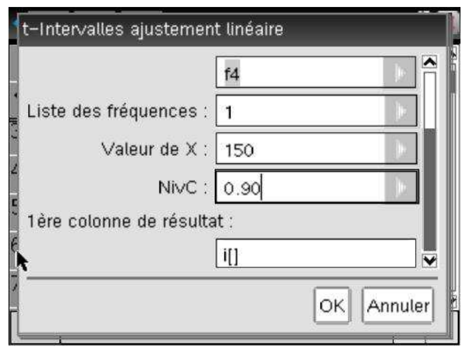
1. Retourner au classeur où se trouvent les données X et Y et sélectionner : [menu] [4] [3] [7]
2. Sélectionner l'option *Réponse* puisqu'on veut un intervalle de confiance pour une prévision. Si on veut estimer la pente de la droite, on choisit alors l'option *Pente*.





3. Remplir les informations demandées tel qu'illustré ci-dessous.





4. On obtient les résultats suivants :

K	L
	=LinRegtIntervals('x,'y,1,1,15(
Titre	t−Intervalle avec régression...
RegEqn	a+b*x
ŷ	19.2203
df	8.
CLower	14.6802
CUpper	23.7604
ME	4.54008
SE	2.4415
LowerPred	8.51174
UpperPred	29.9288
MEPred	10.7085
SEPred	5.75868
a	20.7873
b	-0.010447
r²	0.606417
r	-0.778728
Resid	{3.2573846529417,-2.65326...
xval	150.

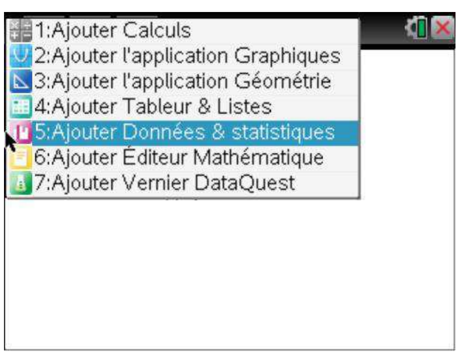
Ainsi, l'intervalle de confiance de niveau 90% pour estimer la durée d'un appel fait à 150 km de Montréal est $[8.5117; 29.9288]$. Autrement dit, les bornes de l'intervalle recherché sont les valeurs correspondant à *LowerPred* et *UpperPred* du tableau précédent.

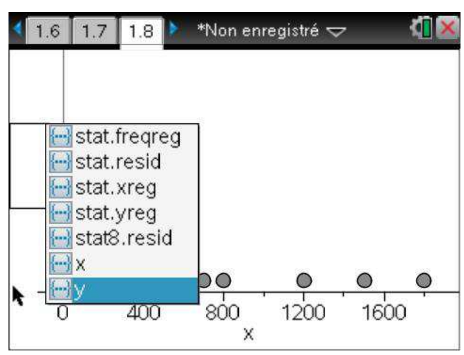
j) On veut maintenant estimer la durée *moyenne* d'un appel interurbain pour des appels faits à 150 km de Montréal par un intervalle de confiance de niveau 90%. Les bornes de l'intervalle recherché seront maintenant les valeurs correspondant à *CLower* et *CUpper* du tableau obtenu à l'étape précédente. Autrement dit, on obtient l'intervalle $[14.6802; 23.7604]$.

K	L
	=LinRegtIntervals('x,'y,1,1,15(
Titre	t-Intervalle avec régression...
RegEqn	a+b*x
ŷ	19.2203
df	8.
CLower	14.6802
CUpper	23.7604
ME	4.54008
SE	2.4415
LowerPred	8.51174
UpperPred	29.9288
MEPred	10.7085
SEPred	5.75868
a	20.7873
b	-0.010447
r²	0.606417
r	-0.778728
Resid	{3.2573846529417,-2.65326...
xval	150.

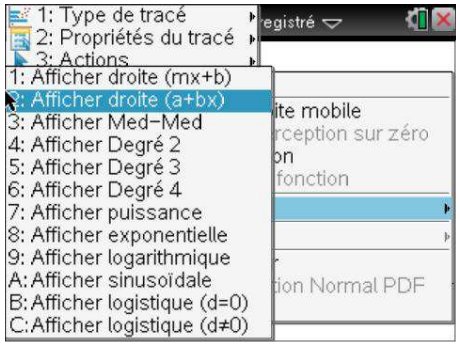
k) On peut tracer le nuage de points et la droite des moindres carrés en suivant les étapes suivantes :

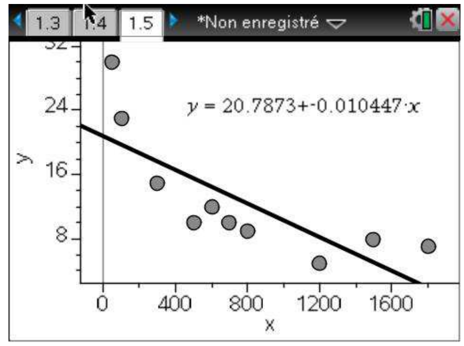
1. Ouvrir une feuille *Données & Statistiques* : [CTRL] [doc] [5]
2. Sélectionner la variable explicative X sur l'axe des x et la variable expliquée Y sur l'axe des y



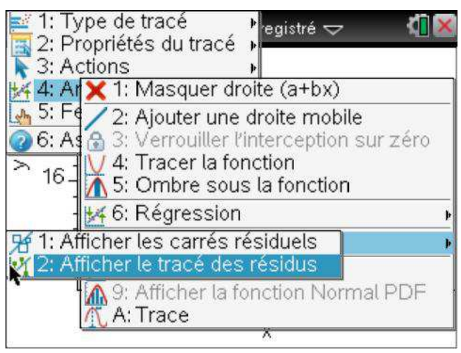


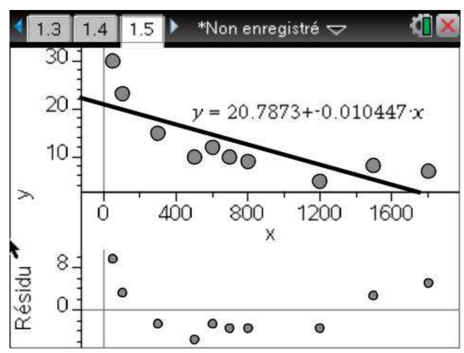
3. Ajouter la droite des moindres carrés : [menu] [4] [6] [2]



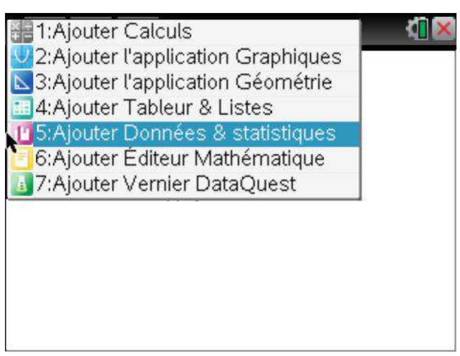


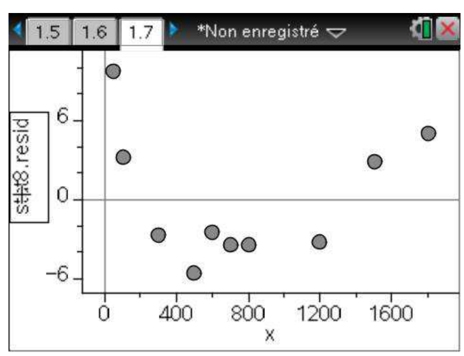
1) On obtient le graphique des résidus en suivant les étapes suivantes :
   1. Ouvrir une autre feuille *Données & Statistiques* : [CTRL] [doc] [5]
   2. Sélectionner la variable explicative X sur l'axe des x et les résidus *stat.resid* sur l'axe des y
   3. On peut obtenir le graphique des résidus directement à partir du nuage de points auquel on a ajouté la droite des moindres carrés.
      1. Retourner à la page où se trouve le nuage de points et la droite des moindres carrés et sélectionner : [menu] [4] [7] [2]
      2. On obtient alors sur une même page le nuage de points, l'équation de la droite des moindres carrés ainsi que le graphique des résidus.





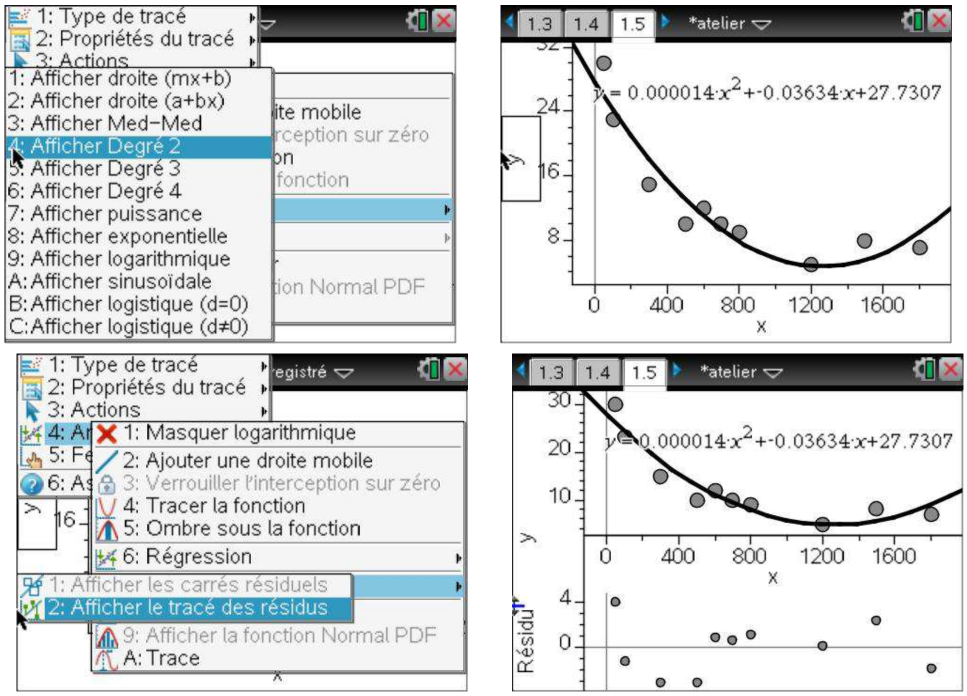
On peut aussi obtenir le graphique des résidus dans une autre page graphique de la façon suivante :



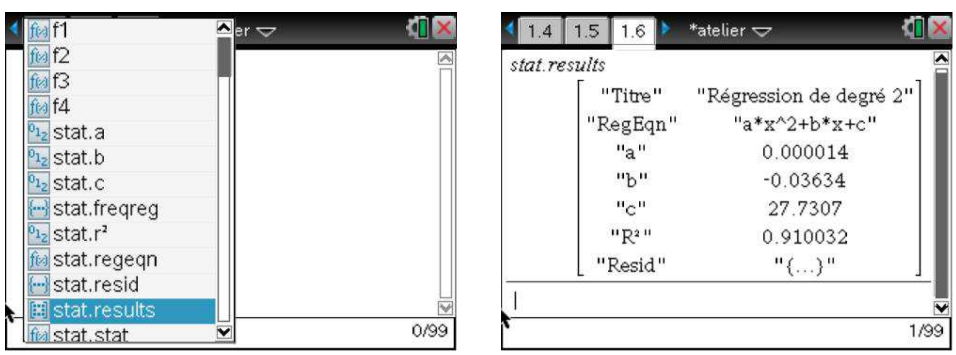


Modèles non linéaires

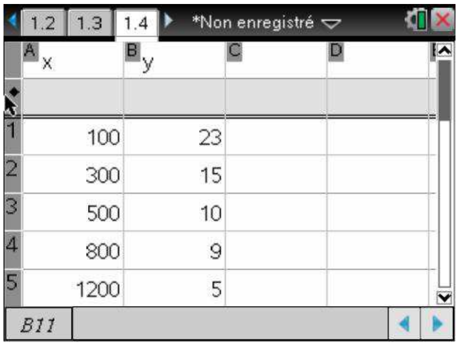
En observant le graphique des résidus obtenus, on pourrait pousser plus loin l'analyse de ces données en étudiant par exemple la pertinence d'un modèle non linéaire. Par exemple, si on essaie d'ajuster les données à une équation du second degré de la façon suivante :

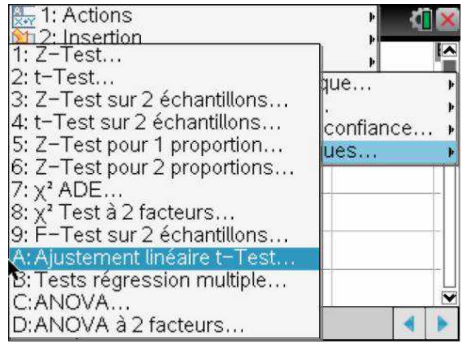


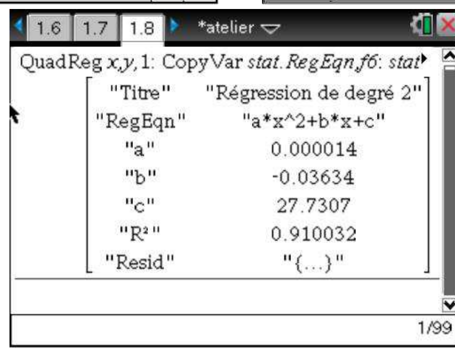
On peut retrouver les résultats de l'analyse de régression logarithmique en utilisant le bouton [var] tel qu'illustré ci-dessous.



Pour faire une analyse de régression non linéaire sans passer par l'interface graphique, ouvrir une feuille de calculs (ou directement dans le classeur contenant les données) et sélectionner [menu] [6] [1] [6] :


1.2 1.3 1.4 *Non enregistré
A x
B y
C
D
1 100 23
2 300 15
3 500 10
4 800 9
5 1200 5
B11


1: Actions
2: Insertion
1: Z-Test…
2: t-Test…
3: Z-Test sur 2 échantillons…
4: t-Test sur 2 échantillons…
5: Z-Test pour 1 proportion…
6: Z-Test pour 2 proportions…
7: χ² ADE…
8: χ² Test à 2 facteurs…
9: F-Test sur 2 échantillons…
A: Ajustement linéaire t-Test…
B: Tests régression multiple…
C:ANOVA…
D:ANOVA à 2 facteurs…
que…
confiance…
ues…


1.6 1.7 1.8 *atelier
QuadReg x,y,1: CopyVar stat.RegEqn,f6: stat
"Titre" "Régression de degré 2"
"RegEqn" "a*x^2+b*x+c"
"a" 0.000014
"b" -0.03634
"c" 27.7307
"R²" 0.910032
"Resid" "{...}"
1/99

2.3.2 Régression linéaire multiple

Une analyse de régression linéaire multiple se fait à partir des tableaux suivants :

Tableau 2.12 Tableau d'analyse de la variance (ANOVA) pour la régression linéaire multiple

Tableau d'analyse de la variance (ANOVA)					
Source de variation	Somme des carrés	Degrés de liberté	Moyenne des carrés	Statistique F sous H_0 : $\beta_1 = \cdots = \beta_p = 0$	valeur-p
Régression	SCR	p	$MCR = \frac{SCR}{p}$	$F = \frac{MCR}{MCE}$	$\alpha_p = P(F_{p;n-p-1} \geq F_{obs})$
Erreur	SCE	$n-p-1$	$MCE = \frac{SCE}{n-p-1}$		
Totale	SCT	$n-1$			

Tableau 2.13 Tableau des estimateurs des paramètres pour la régression linéaire multiple

Estimation des paramètres				
Paramètres	Estimateurs	Écart-type de l'estimateur	Statistique T sous $H_0 : \beta_i = 0$	valeur-p
β_0	$\hat{\beta}_0$	$s(\hat{\beta}_0)$	$T = \frac{\hat{\beta}_0}{s(\hat{\beta}_0)} \sim t_{n-p-1}$	$\alpha_p = 2 \cdot P(T_{n-p-1} \geq \lvert T_{obs} \rvert)$
β_1	$\hat{\beta}_1$	$s(\hat{\beta}_1)$	$T = \frac{\hat{\beta}_1}{s(\hat{\beta}_1)} \sim t_{n-p-1}$	$\alpha_p = 2 \cdot P(T_{n-p-1} \geq \lvert T_{obs} \rvert)$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
β_p	$\hat{\beta}_p$	$s(\hat{\beta}_p)$	$T = \frac{\hat{\beta}_p}{s(\hat{\beta}_p)} \sim t_{n-p-1}$	$\alpha_p = 2 \cdot P(T_{n-p-1} \geq \lvert T_{obs} \rvert)$

Intervalle de confiance pour une prévision dans le contexte de la régression linéaire multiple

L'intervalle de confiance de niveau $1-\alpha$ pour estimer la valeur y_0 lorsque $\boldsymbol{x} = \boldsymbol{x_0}$ est donné par

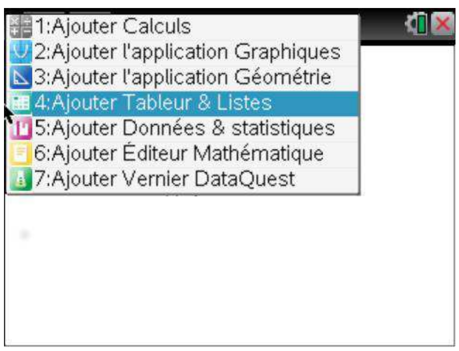
$$\left[\hat{y}_0 - t_{n-p-1;\alpha/2}\sqrt{s^2\left(1+\boldsymbol{x_0'}(\boldsymbol{X'X})^{-1}\boldsymbol{x_0}\right)}\ ;\ \hat{y}_0 + t_{n-p-1;\alpha/2}\sqrt{s^2\left(1+\boldsymbol{x_0'}(\boldsymbol{X'X})^{-1}\boldsymbol{x_0}\right)}\right]$$

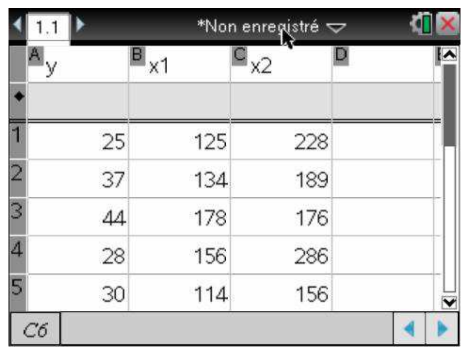
où $s^2 = \hat{\sigma}^2 = MCE = \frac{SCE}{n-p-1}$ est la variance résiduelle.

On procède donc de façon similaire à la régression linéaire simple. Illustrons-le à l'aide d'un exemple.

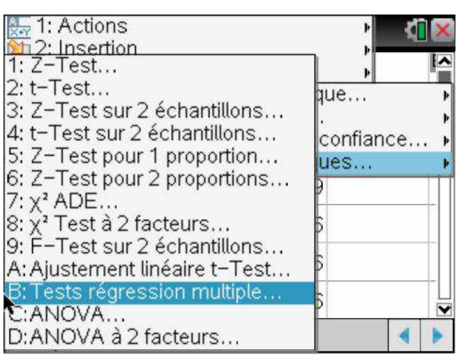
Exemple 2.6

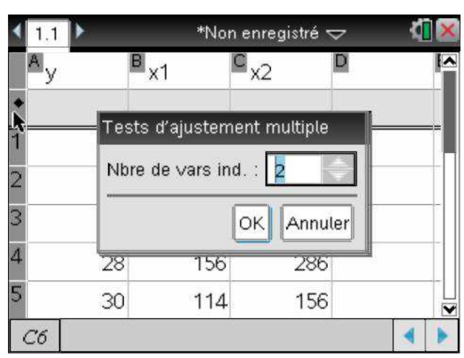
1. Ouvrir un *Tableur & listes* : [CTRL] [doc] [4]
2. Entrer les données des trois listes suivantes.





3. Sélectionner : [menu] [4] [4] [B]
4. Remplir les informations demandées tel qu'illustré ci-dessous (on a 2 variables indépendantes).





5. On obtient les résultats suivants :

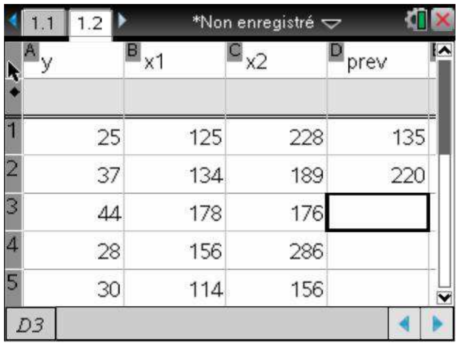
D	E
	=MultRegTests('y,'x1,'x2): Co
Titre	Test de régressions multiple...
RegEqn	b0+b1*x1+b2*x2+...
F	9.38635
PVal	0.09628
R²	0.90372
adjR²	0.80744
s	3.36204
DW	2.72403
dfReg	2.
SSReg	212.193
MSReg	106.097
dfError	2.
SSError	22.6066
MSError	11.3033
bList	{20.19074397837,0.2438914...
tList	{1.9156738462926,3.619731...
PList	{0.1954788573054,0.068564...
SEList	{10.539760730901,0.067378...
ŷList	{26.580775408857,32.89755...
Resid	{-1.5807754088568,4.10244...
sResid	{-0.59887592652921,1.4016...

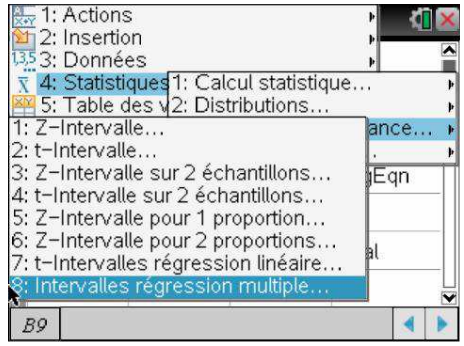
6. Toutes ces quantités sont maintenant accessibles via la touche [var] à partir d'une feuille de calculs.

7. Pour obtenir des intervalles de confiance pour une prévision, on procède de la façon suivante :

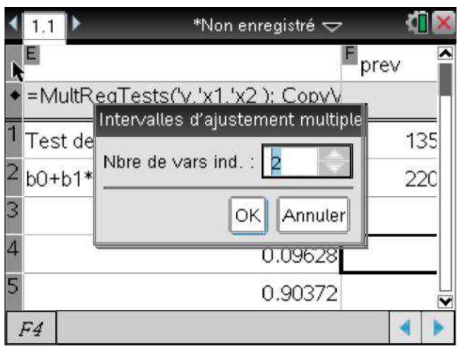
   7.1 Entrer dans une liste les valeurs de x_1 et x_2 pour lesquelles on veut obtenir une prévision. Supposons ici qu'on veuille estimer y par un intervalle de confiance de niveau 95% pour $x_1 = 135$ et $x_2 = 220$. Appellons cette liste *prev*.

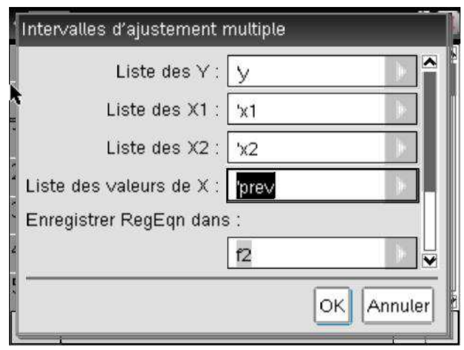
   7.2 Sélectionner : [menu] [4] [3] [8]

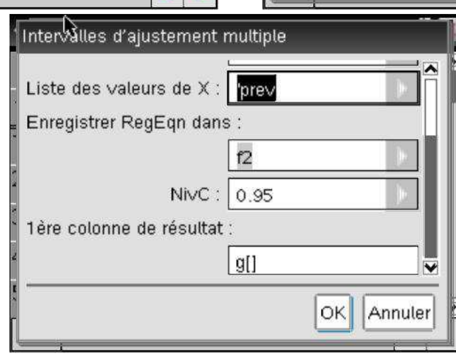




7.3 Remplir les informations demandées tel qu'illustré ci-dessous (2 variables indépendantes).







7.4 On obtient les résultats suivants :

G	H
	=MultRegIntervals('y,'x1,'x
Titre	Intervalles de régression...
RegEqn	b0+b1*x1+b2*x2+...
ŷ	29.8652
dfError	2.
CLower	22.7645
CUpper	36.9658
ME	7.10066
SE	1.6503
LowerPred	13.7507
UpperPred	45.9796
MEPred	16.1144
SEPred	3.74523
bList	{20.19074397837,0.2438...
Resid	{-1.5807754088568,4.10...

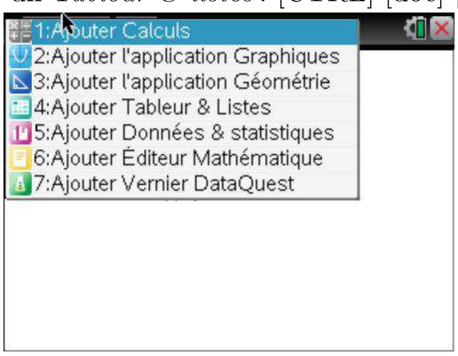
7.5 Comme dans le cas de la régression linéaire simple, les bornes de l'intervalle de confiance pour la valeur de y_0 correspondent aux valeurs *LowerPred* et *UpperPred*, alors que les bornes de l'intervalle de confiance pour la valeur *moyenne* de y_0 sont données par *CLower* et *CUpper*.

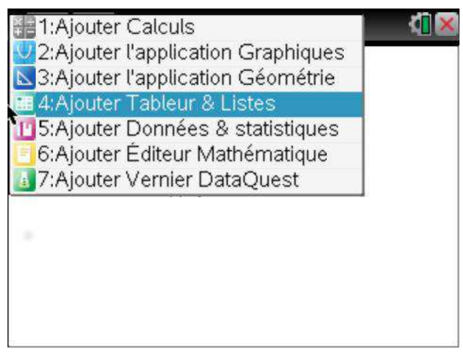
Annexes

A.1 Classeurs, activités et pages

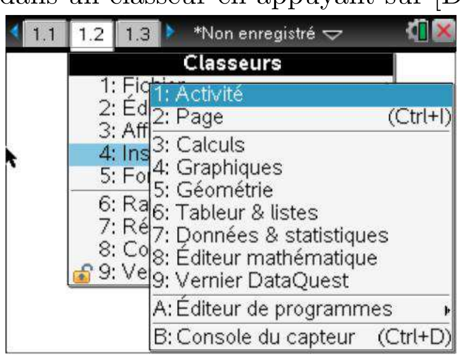
Commençons par créer un classeur qui contiendra une page de type *Calculs* et une page de type *Tableur & listes*.

1. Ouvrir une feuille *Calculs* : [CTRL] [doc] [1]
2. Ouvrir un *Tableur & listes* : [CTRL] [doc] [4]





3. On peut naviguer d'une page à l'autre en appuyant sur [CTRL] et sur les flèches ◃ ou ▹ du pavé tactile.
4. Pour ajouter une activité dans un classeur en appuyant sur [DOC] [4] [1]



A.2 Quelques commandes de base

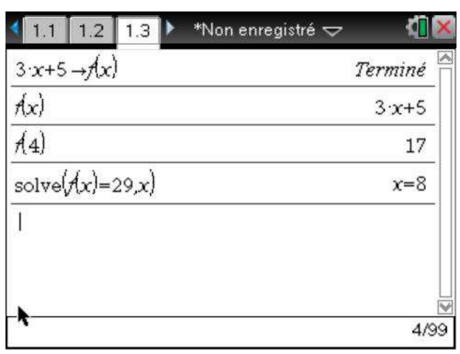
Voyons quelques commandes qui seront très utiles dans le cadre du cours.

1. La touche [esc] : pour revenir à l'étape précédente en tout temps
2. La touche [menu] : qui présente un menu contextuel des fonctions disponibles selon où l'on se trouve
3. La touche on : pour revenir à l'écran d'accueil
4. La touche [enter] : pour confirmer l'exécution d'une commande
5. La touche : pour voir les différents modèles d'écriture mathématique
6. La touche : pour voir le catalogue des fonctions disponibles
7. La combinaison [CTRL] : pour ouvrir la palette de symboles
8. La touche [sto →] (que l'on accède en appyant sur [CTRL] [var]) : pour définir des variables ou des fonctions
9. La fonction *solve* qui permet de résoudre une équation

Exemple 1.7

1. Définir la fonction $f(x) = 3x + 5$
2. Évaluer la fonction pour $x = 4$
3. Résoudre l'équation $3x + 5 = 29$

Solution :



Rédigé par Sylvie Gervais
Service des enseignements généraux,
École de technologie supérieure.

Version révisée en août 2012.