Utilisation des calculatrices en probabilités et statistique

Sylvain Mousset
sylvain.mousset@univ-lyon1.fr

9 février 2022

Ce document a été établi pour vous permettre d'utiliser au mieux vos calculatrices pour réaliser les calculs statistiques des UE "Mathématiques pour les sciences de la vie" et "Biostatistique-Bioinformatique". Le but n'est pas d'illustrer ou d'expliquer des concepts (ils sont expliqués en cours) mais de décrire l'ensemble des fonctions et des tests implémentés sur vos calculatrices afin de vous permettre d'utiliser ces outils. En ce sens, ce document dépasse le strict programme des deux UE mentionnées plus haut mais les étudiants de chaque UE retrouveront les tests qui y sont enseignés.
Il est possible que des erreurs se soient glissées dans ce document. Si vous pensez en avoir trouvé une, faites en part à l'auteur par mail.

Table des matières

1 Distributions

La table 1 indique les équivalences entre les fonctions de distributions vues en cours et les commandes utilisables sur le logiciel R et sur deux modèles de calculatrices fréquemment utilisé par les étudiants (Texas Instrument TI-82 Stats.fr et Casio Graph 35+).

TABLE 1 – Distributions et commandes associées (R, TI-82 stats.fr, Casio Graph35+)

Notation du cours	Définition	R	TEXAS INSTRUMENTS	CASIO
Loi normale, $X \sim \mathcal{N}(\mu, \sigma^2)$				
$f_{\mathcal{N}(\mu,\sigma^2)}(x)$	$\lim_{\varepsilon\to 0}\left(\frac{1}{\varepsilon}\mathbb{P}(x \leqslant X \leqslant x+\varepsilon)\right)$	`dnorm(`x`,mean=`μ`,sd=`σ`)`	`normalFdp(`x, μ, σ`)`	`NormPD(`μ, σ, x`)`
$F_{\mathcal{N}(\mu,\sigma^2)}(x)$	$\mathbb{P}(X \leqslant x)$	`pnorm(`x`,mean=`μ`,sd=`σ`)`	–	–
$F_{\mathcal{N}(\mu,\sigma^2)}(b) - F_{\mathcal{N}(\mu,\sigma^2)}(a)$	$\mathbb{P}(a \leqslant X \leqslant b)$	–	`normalFRép(`a, b, μ, σ`)`	`NormCD(`a, b, σ, μ`)`
$F^{-1}_{\mathcal{N}(\mu,\sigma^2)}(p)$	$x \,\vert\, \mathbb{P}(X \leqslant x) = p$	`qnorm(`p`,mean=`μ`,sd=`σ`)`	`FracNormale(`p, μ, σ`)`	`InvNormCD(`p, σ, μ`)`
Loi de Student, $X \sim \mathcal{T}(k \text{ ddl})$				
$f_{\mathcal{T}(k \text{ ddl})}(x)$	$\lim_{\varepsilon\to 0}\left(\frac{1}{\varepsilon}\mathbb{P}(x \leqslant X \leqslant x+\varepsilon)\right)$	`dt(`x`,df=`k`)`	`studentFdp(`x, k`)`	`tPD(`x, k`)`
$F_{\mathcal{T}(k \text{ ddl})}(x)$	$\mathbb{P}(X \leqslant x)$	`pt(`x`,df=`k`)`	–	–
$F_{\mathcal{T}(k \text{ ddl})}(b) - F_{\mathcal{T}(k \text{ ddl})}(a)$	$\mathbb{P}(a \leqslant X \leqslant b)$	–	`studentFRép(`a, b, k`)`	`tCD(`a, b, k`)`
$F^{-1}_{\mathcal{T}(k \text{ ddl})}(p)$	$x \,\vert\, \mathbb{P}(X \leqslant x) = p$	`qt(`p`,df=`k`)`	–	`InvTCD(`p, k`)`
Loi de χ^2, $X \sim \chi^2(k \text{ ddl})$				
$f_{\chi^2(k \text{ ddl})}(x)$	$\lim_{\varepsilon\to 0}\left(\frac{1}{\varepsilon}\mathbb{P}(x \leqslant X \leqslant x+\varepsilon)\right)$	`dchisq(`x`,df=`k`)`	`X2Fdp(`x, k`)`	`ChiPD(`x, k`)`
$F_{\chi^2(k \text{ ddl})}(x)$	$\mathbb{P}(X \leqslant x)$	`pchisq(`x`,df=`k`)`	–	–
$F_{\chi^2(k \text{ ddl})}(b) - F_{\chi^2(k \text{ ddl})}(a)$	$\mathbb{P}(a \leqslant X \leqslant b)$	–	`X2FRép(`a, b, k`)`	`ChiCD(`a, b, k`)`
$F^{-1}_{\chi^2(k \text{ ddl})}(p)$	$x \,\vert\, \mathbb{P}(X \leqslant x) = p$	`qchisq(`p`,df=`k`)`	–	`InvChiCD(`$1-p, k$`)`
Loi de Fisher-Snedecor, $X \sim \mathcal{F}(k_1,\ k_2 \text{ ddl})$				
$f_{\mathcal{F}(k_1,\ k_2 \text{ ddl})}(x)$	$\lim_{\varepsilon\to 0}\left(\frac{1}{\varepsilon}\mathbb{P}(x \leqslant X \leqslant x+\varepsilon)\right)$	`df(`x`,df1=`k_1`,df2=`k_2`)`	`FFdp(`x, k_1, k_2`)`	`FPD(`x, k_1, k_2`)`
$F_{\mathcal{F}(k_1,\ k_2 \text{ ddl})}(x)$	$\mathbb{P}(X \leqslant x)$	`pf(`x`,df1=`k_1`,df2=`k_2`)`	–	–
$F_{\mathcal{F}(k_1,\ k_2 \text{ ddl})}(b) - F_{\mathcal{F}(k_1,\ k_2 \text{ ddl})}(a)$	$\mathbb{P}(a \leqslant X \leqslant b)$	–	`FFRép(`a, b, k_1, k_2`)`	`FCD(`a, b, k_1, k_2`)`
$F^{-1}_{\mathcal{F}(k_1,\ k_2 \text{ ddl})}(p)$	$x \,\vert\, \mathbb{P}(X \leqslant x) = p$	`qf(`p`,df1=`k_1`,df2=`k_2`)`	–	`InvFCD(`$1-p, k_1, k_2$`)`
Loi binomiale, $X \sim \mathcal{B}(n, p)$				
$f_{\mathcal{B}(n,p)}(x)$	$\mathbb{P}(X = x)$	`dbinom(`x`,size=`n`,prob=`p`)`	`binomFdp(`n, p, x`)`	`BinominalPD(`x, n, p`)`
$F_{\mathcal{B}(n,p)}(x)$	$\mathbb{P}(X \leqslant x)$	`pbinom(`x`,size=`n`,prob=`p`)`	`binomFRép(`n, p, x`)`	`BinominalCD(`x, n, p`)`
$F^{-1}_{\mathcal{B}(n,p)}(p_x)$	$\sup\left(\left\{k \in \mathbb{N} \,\vert\, p \leqslant F_{\mathcal{B}(n,p)}(k)\right\}\right)$	`qbinom(`p_x`,size=`n`,prob=`p`)`	–	`InvBinominalCD(`p_x, n, p`)`
Loi de Poisson, $X \sim \mathcal{P}(\lambda)$				
$f_{\mathcal{P}(\lambda)}(x)$	$\mathbb{P}(X = x)$	`dpois(`x`,lambda=`λ`)`	`poissonFdp(`λ, x`)`	`PoissonPD(`x, λ`)`
$F_{\mathcal{P}(\lambda)}(x)$	$\mathbb{P}(X \leqslant x)$	`ppois(`x`,lambda=`λ`)`	`poissonFRép(`x, λ`)`	`PoissonCD(`x, λ`)`
$F^{-1}_{\mathcal{P}(\lambda)}(p)$	$\sup\left(\left\{k \in \mathbb{N} \,\vert\, p \leqslant F_{\mathcal{P}(\lambda)}(k)\right\}\right)$	`qpois(`p`,lambda=`λ`)`	–	`InvPoissonCD(`p, λ`)`

2 Estimation ponctuelle de la moyenne et de la variance

2.1 Les trois types de présentation des données

Vos données peuvent vous être présentées de l'une des trois façons suivantes :

1. **Données brutes** : $X_1, X_2, \ldots, X_n$ constituent les n mesures de l'échantillon.
   Dans ce cas, on enregistrera une liste pour les observations. Vos calculatrices disposent d'éditeurs de listes mais vous pouvez aussi les enregistrer directement entre accolades séparées par des virgules.

2. **Données regroupées** : $\begin{matrix} X_1 & X_2 & \ldots & X_k \\ n_1 & n_2 & \ldots & n_k \end{matrix}$ constituent les $n = \sum_{i=1}^{k} n_i$ données.
   Dans ce cas, on enregistrera deux listes pour les observations. Une liste contiendra les valeurs X_i et une atre contiendra les effectifs n_i.

3. **Données résumées** : l'échantillon contient n mesures dont la somme est $\sum X = \sum_{i=1}^{n} X_i$ et la somme des carrés est $\sum X^2 = \sum_{i=1}^{n} X_i^2$.
   Dans ce cas vous devrez calculer la moyenne de l'échantillon $\bar{X} = \frac{\sum X}{n}$ et l'estimation de l'écart-type de la population $\hat{\sigma} = \sqrt{\frac{n}{n-1}\left(\frac{1}{n}\sum X^2 - \bar{X}^2\right)}$. Il est préférable de stocker ces valeurs dans les mémoires de votre calculatrices pour conserver tous les chiffres avec la précision de votre machine.

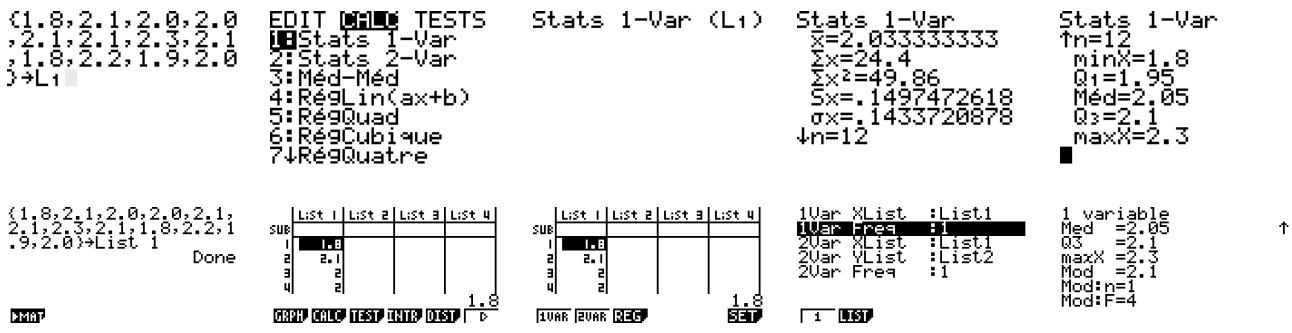
2.2 Estimation à partir de donées brutes

On veut estimer la moyenne et la variance de la distribution d'une variable aléatoire dans la population à partir de données d'un échantillon. Pour illustrer ce calcul on utilisera un échantillon constitué de 12 observations indépendantes. Les données brutes correspondantes sont présentées dans le tableau 2.

TABLE 2 – Données brutes

X_i	1.8	2.1	2.0	2.0	2.1	2.1	2.3	2.1	1.8	2.2	1.9	2.0

FIGURE 1 – Estimations ponctuelles de la moyenne et de l'écart-type sur des données brutes



À gauche sur calculatrices TI à droite sur Casio en bas

1. Les données sont entrées dans une première liste (`L1` ou `List 1`).
2. Sur TI, on choisit un calcul statistique sur une variable. On donne en argument entre parenthèses la liste où sont stockées les données brutes.
3. Sur Casio, on choisit le menu `CALC` puis on configure le calcul avec SET. Là on indique que les valeurs sont dans `List 1` et on indique que les effectifs valent 1 pour toutes les valeurs observées. On revient au menu précédent en appuyant sur la touche `EXIT` et on choisit `1VAR`
4. On lit les résultats du calcul. La moyenne de l'échantillon est $\bar{x} = 2.0333$ l'estimation de l'écart-type (notée $\hat{\sigma}$ dans le cours) est notée `Sx` et la valeur de l'écart-type dans l'échantillon (notée s dans le cours) est notée σx. On a $\hat{\mu} = \bar{x} = 2.0333$, $s = 0.1434$ et $\hat{\sigma} = 0.1498$ (résultats arrondis à la $4^{\text{ème}}$ décimale).

2.3 Estimation ponctuelle à partir de données regroupées

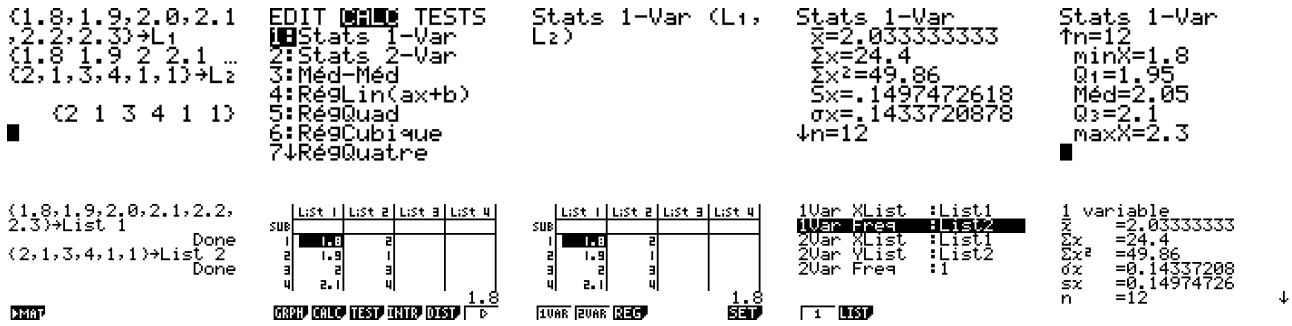
Les données de la table 2 sont regroupées sous la forme de la table 3. Le calcul de l'estimation de la moyenne et de l'écart-type est présenté sur les captures d'écran de la figure 2.

n	$\sum X$	$\sum X^2$
12	24.4	49.86

TABLE 3 – Données groupées

X_i	1.8	1.9	2.0	2.1	2.2	2.3
n_i	2	1	3	4	1	1

FIGURE 2 – Estimations ponctuelles de la moyenne et de l'écart-type sur des données groupées



De gauche à droite

1. Les valeurs de la variable aléatoire sont entrées dans une première liste (ici `L1`) ou `List 1`). Les effectifs correspondants sont saisis dans une seconde liste (ici `L2` ou `List 2`) de même dimension.
2. Sur TI, on choisit un calcul statistique sur une variable puis on donne en arguments entre parenthèses et séparées par une virgule la liste des valeurs `L1` puis la liste des effectifs `L2`.
3. Sur Casio, on choisit le menu `CALC` puis on configure le calcul avec SET. Là on indique que les valeurs sont dans `List 1` et les effectifs dans `list 2`. On revient au menu précédent en appuyant sur la touche `EXIT` et on choisit `1VAR`
4. On lit les résultats du calcul. Ils sont évidemment identiques au cas de la figure 1 sur les données non regroupées. On a $\hat{\mu} = \bar{x} = 2.0333$, $s = 0.1434$ et $\hat{\sigma} = 0.1498$ (résultats arrondis à la $4^{\text{ème}}$ décimale).

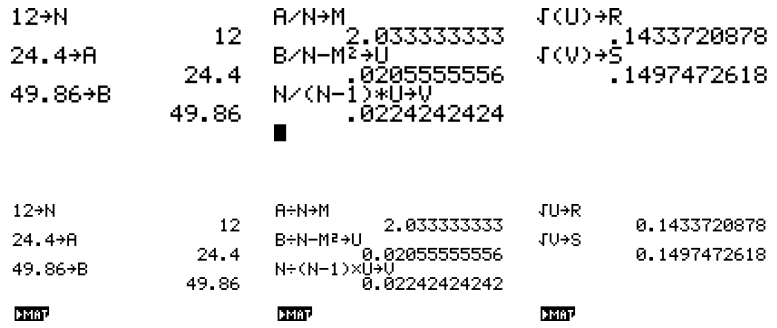
2.4 Estimation ponctuelle à partir de données résumées

À présent on ne dispose plus que d'un résumé des valeurs de la table 2, par exemple sous la forme suivante du tableau 2.4 : vous devez alors appliquer les formules classiques pour calculer la moyenne $\bar{x}$, la variance observée dans l'échantillon s^2 et l'estimation de la variance dans la population $\hat{\sigma^2}$.

$$\bar{x} = \frac{\sum X}{n}, \quad s^2 = \frac{\sum X^2}{n} - \bar{x}^2, \quad \hat{\sigma}^2 = \frac{n}{n-1} s^2 \tag{1}$$

Pour faire le moins possible d'erreurs d'arrondis, il est préférable de stocker les résultats des calculs dans les mémoires de votre calculatrice comme illustré sur la figure 3.

FIGURE 3 – Estimations ponctuelles de la moyenne et de l'écart-type sur des données résumées



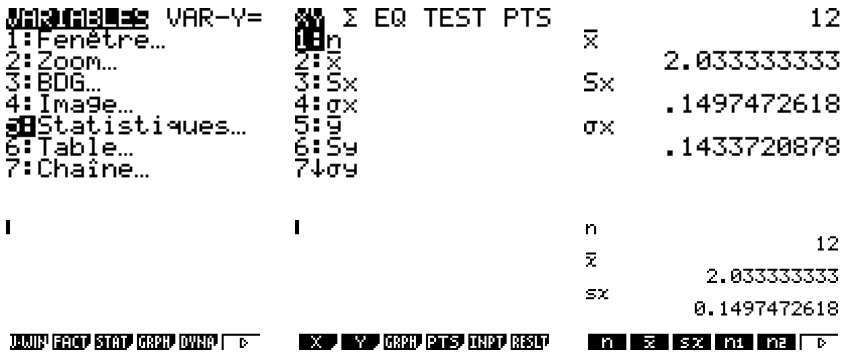
De gauche à droite

1. On entre l'effectif dans la variable `N`, la somme des valeurs dans `A` et la somme des carrés dans `B`.
2. On stocke la moyenne dans `M`, la variance observée dans l'échantillon dans `U` et la variance estimée de la population dans `V`.
3. On stocke l'écart-type observé dans l'échantillon dans `R` et l'écart-type estimé de la population dans `S`.
4. Les résultats du calcul sont évidemment identiques aux cas des figure 1 et 2. On a $\hat{\mu} = \bar{x} = 2.0333$, $s = 0.1434$ et $\hat{\sigma} = 0.1498$ (résultats arrondis à la $4^{\text{ème}}$ décimale).

2.5 Rappeler les valeurs calculées précédemment

Une fois un calcul statistique effectué sur des données brutes ou groupées, vos calculatrices vous permettent de récupérer les valeurs calculées ceci est présenté sur la figure 4

FIGURE 4 – Rappel des valeurs calculées après un calcul statistique



De gauche à droite TI en haut, Casio en bas

1. On accède aux variables calculées précédemment avec la touche `VARS` (TI et Casio).
2. Sur TI on demande les variables calculées dans un contexte statistique, pius on choisit les variables que l'on souhaite rappeler.
3. Sur Casio, on choisit `STAT` puis `INPT` et on accède aux valeurs de son choix.

3 Intervalles de confiance, tests de conformité et d'homogénéité

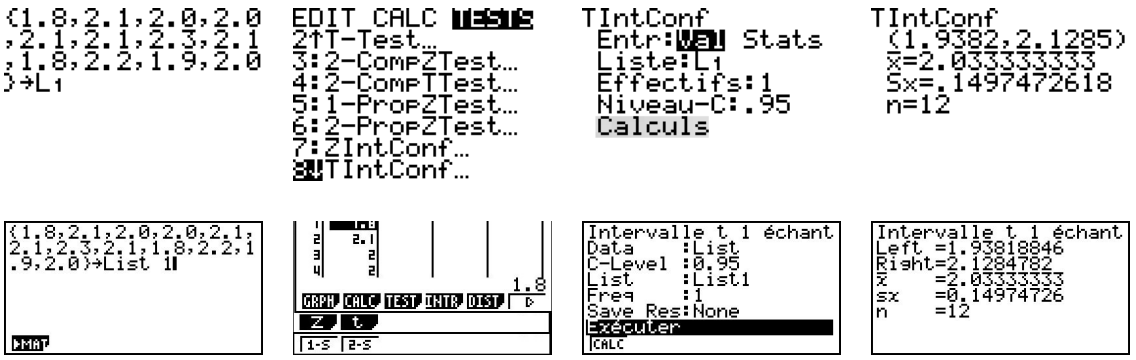
3.1 Un exemple : calcul de l'intervalle de confiance de la moyenne

On va calculer un intervalle de confiance de la moyenne μ à partir du même échantillon présenté des trois façons précédentes. L'échantillon est constitué de 12 observations indépendantes tirées d'une distribution supposée normale dont on veut établir un intervalle de confiance à 95% (risque $\alpha = 5\%$).

3.1.1 Travail sur des données brutes

Les données brutes ont été présentées dans la table 2. Des captures d'ecran du calcul de l'intervalle de confiance sont présentées sur la figure 5.

FIGURE 5 – Intervalle de confiance de la moyenne μ sur des données brutes



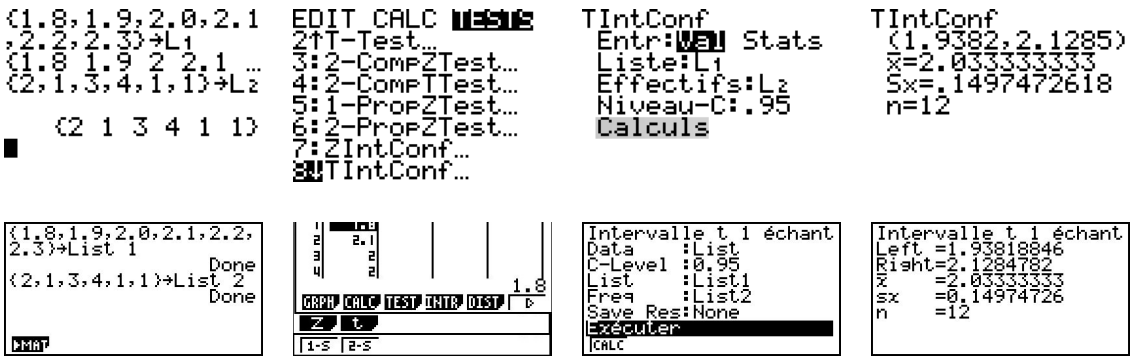
De gauche à droite

1. Les données sont entrées dans une première liste (`L1` ou `List 1`).
2. On choisit un intervalle de confiance de type t (car la variance théorique est inconnue) pour la moyenne à partir d'un échantillon.
3. On effectue les choix qui conviennent : des valeurs sont dans la liste `L1` ou `List 1` , les effectifs sont 1 observation par valeur. On calcule un intervalle de confiance à 95% (risque $\alpha = 5\%$).
4. On lit les résultats du calcul. L'intervalle de confiance à 95% est [1.9382, 2.1285].

3.1.2 Travail sur des données groupées

Le calcul de l'intervalle de confiance de la moyenne sur les données groupées de la table 3 est présenté sur les captures d'écran de la figure 6.

FIGURE 6 – Intervalle de confiance de la moyenne μ sur des données groupées



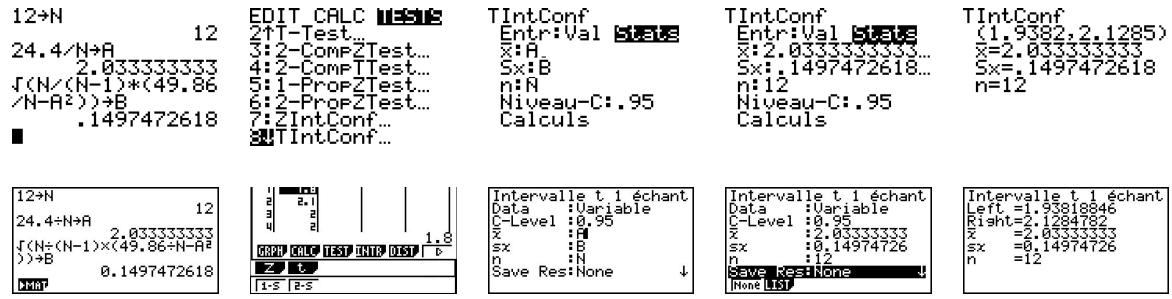
De gauche à droite :

1. Les données sont entrées dans la liste `L1` (ou `List 1`), les effectifs dans la liste `L2` (ou `List 2`).
2. On choisit un intervalle de confiance de type t (car la variance théorique est inconnue) pour la moyenne à partir d'un échantillon.
3. On effectue les choix qui conviennent : des valeurs sont dans la liste `L1`, les effectifs sont dans la liste `L2`. On calcule un intervalle de confiance à 95% (risque $\alpha = 5\%$).
4. On lit les résultats du calcul. L'intervalle de confiance à 95% est [1.9382, 2.1285].

3.1.3 Données résumées

Le calcul de l'intervalle de confiance est présenté sur les captures d'écran de la figure 7.

FIGURE 7 – Intervalle de confiance de la moyenne μ sur des données résumées



De gauche à droite, les étapes du calcul :

1. La taille de l'échantillon ($n = 12$) est stockée dans la variable N. On calcule la moyenne de l'échantillon $\bar{X} = 2.0333333$ et l'estimation de l'écart-type $\hat{\sigma} = \sqrt{\frac{12}{11} \times \left(\frac{1}{12} \times 49.86 - 2.0333333^2\right)} = 0.1497473$, stockées respectivement dans les variables A et B de la calculatrice.
2. On choisit un intervalle de confiance de type t (car la variance théorique est inconnue) pour la moyenne à partir d'un échantillon.
3. On effectue les choix qui conviennent : on va fournir des statistiques, on donne $\bar{X}$, $\hat{\sigma}$ et N en entrant les noms des variables correspondantes qui sont immédiatement remplacées par leurs valeurs. On calcule un intervalle de confiance à 95% (risque $\alpha = 5\%$).
4. On lit les résultats du calcul. L'intervalle de confiance à 95% est [1.9382, 2.1285].

3.2 Un exemple de test d'hypothèses : test d'homogénéité de deux moyennes

On a obtenu un autre échantillon constitué de 10 observations indépendantes dans une seconde population. La figure 8 montre les trois possibilités de présenter les données de cet échantillon.

FIGURE 8 – Un second échantillon

Données brutes

Y_i	2.2	2.2	2.2	2.1	2.3	2.3	2.3	2.4	2.4	2.5

Données regroupées

Y_i	2.1	2.2	2.3	2.4	2.5
n_i	1	3	3	2	1

Données résumées

$$n = 10, \quad \sum Y = 22.9, \quad \sum Y^2 = 52.57$$

On veut comparer les moyennes dans les populations dont proviennent les deux échantillons :

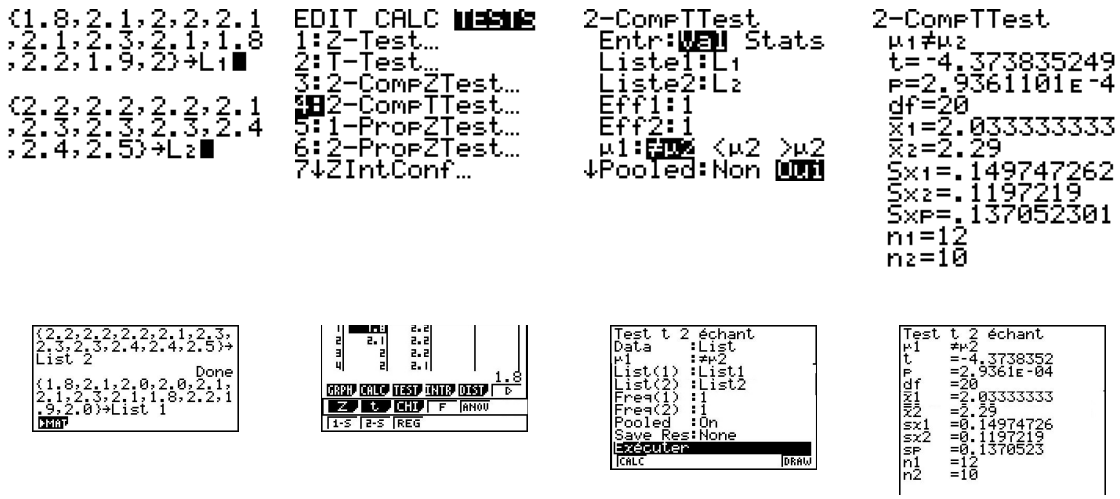
— On utilise un test t d'homogénéité de deux moyennes. On souhaite utiliser un risque de première espèce $\alpha = 0.05$.

— L'hypothèse nulle est $H_0 : \mu_1 = \mu_2$. L'hypothèse alternative est $H_1 : \mu_1 \neq \mu_2$.

— Pour satisfaire les conditions d'application sur des petits échantillons, on doit supposer que les distributions dont sont issus les deux échantillons sont normales et de même variance $\sigma_1^2 = \sigma_2^2 = \sigma_*^2$.

La réalisation du test utilisant les données brutes est illustrée sur la figure 9.

FIGURE 9 – Test t d'homogénéité de deux moyennes sur des données brutes



De gauche à droite

1. Les données brutes sont entrées dans deux listes (`L1` et `L1` ou `List 1` et `List 2`).
2. Le test est un test t (car la variance théorique est inconnue) d'homogénéité de deux moyennes.
3. On effectue les choix qui conviennent : les valeurs dont dans des listes, une seule observation valeur (données brutes), on calcule une variance commune (*pooled* vaut "oui" ou "on"). L'hypothèse alternative est $H_1 : \mu_1 \neq \mu_2$ (test bilatéral).
4. On lit les résultats du calcul. L'écart-type commun estimé est $\hat{\sigma} = 0.137$. La valeur de la statistique est $t = -4.37$ et est associée à une valeur $p = 2.936 \times 10^{-4}$. On a donc $p < \alpha$, ce qui permet de rejeter l'hypothèse nulle avec un risque $\alpha = 0.05$. Les moyennes des populations sont donc différentes.

3.3 Procédures disponibles sur vos calculatrices

3.3.1 Tests et IC sur les moyennes

Les tests et intervalles de confiance sur les proportions disponibles sur vos calculatrices sont présentés dans la table 4. Ils fonctionnent tous à la façon montrée sur la figure 9.

TABLE 4 – Tests et intervalles de confiance sur les moyennes

Un seul échantillon	σ^2 connue	σ^2 estimée

Intervalle de confiance de μ	`ZIntConf`
	`ZInterval`
	`INTR`→ Z→1-S
		`TIntConf`
		`TInterval`
		`INTR`→ t→1-S

Test de conformité $\mu = \mu_0$	`Z-Test` `Z-Test` `TEST`→ Z→1-S	`T-Test` `T-Test` `TEST`→ t→1-S

Deux échantillons	σ_1^2 et σ_2^2 connues	σ_1^2 et σ_2^2 estimées		σ^2 commune estimée	

Intervalle de confiance de $\mu_1 - \mu_2$	`2-CompZIntC`
	`2-SampZInt`
	`INTR`→ Z→2-S
		`2-CompTIntC`
		`2-SampTInt`
		`INTR`→ t→2-S
			~~`pooled`~~	`2-CompTIntC`
				`2-SampTInt`
				`INTR`→ t→2-S
					`pooled`

Test d'homogénéité $\mu_1 = \mu_2$	`2-CompZTest`
	`2-SampZTest`
	`TEST`→ Z→2-S
		`2-CompTTest`
		`2-SampTTest`
		`TEST`→ t→2-S
			~~`pooled`~~	`2-CompTTest`
				`2-SampTTest`
				`TEST`→ t→2-S
					`pooled`

Notation des procédures de tests et intervalles de confiance de la moyenne sur les calculatrices TI-82 Stats.fr (première ligne), TI84 (deuxième ligne) et casio graph 35+ (troisième ligne).

3.3.2 Tests et IC sur les proportions

Les tests et intervalles de confiance sur les proportions disponibles sur vos calculatrices sont présentés dans la table 5. Dans ces tests et intervalles de confiance, on indique les nombres de succès (x ou x_1 et x_2) et le nombre total d'observations (n ou n_1 et n_2). Pour le test de conformité à une proportion théorique (test sur un échantillon), celle-ci est notée p_0. Comme précédemment, on indique l'hypothèse alternative.

TABLE 5 – Tests et intervalles de confiance sur les proportions

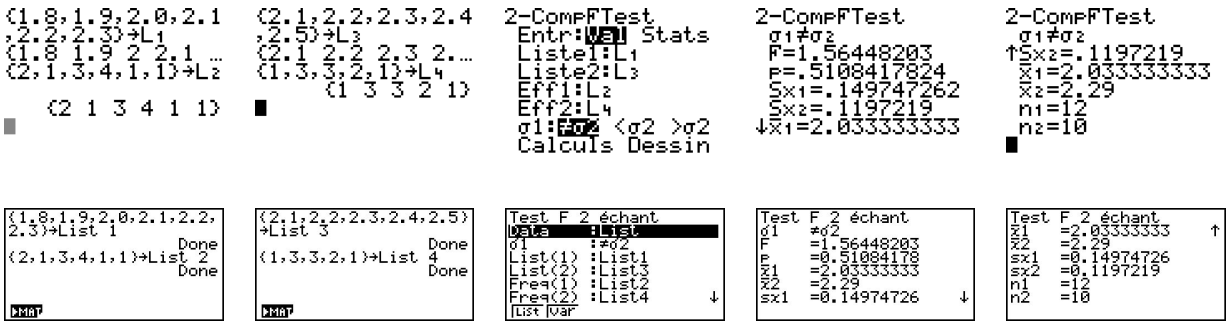
	1 échantillon	2 échantillons
Intervalle de confiance de p ou de $p_1 - p_2$	`1-PropZInt` `1-PropZInt` `INTR`→ Z→1-P	`2-PropZInt` `2-PropZInt` `INTR`→ Z→2-P
Test de conformité $p = p_0$ ou d'homogénéité $p_1 = p_2$	`1-PropZTest` `1-PropZTest` `TEST`→ Z→1-P	`2-PropZTest` `2-PropZTest` `TEST`→ Z→2-P

Notation des procédures de tests et intervalles de confiance de proportions sur les calculatrices TI-82 Stats.fr (première ligne), TI84 (deuxième ligne) et casio graph 35+ (troisième ligne).

3.4 Test d'homogénéité de deux variances

Le test d'homogénéité de deux variances tel qu'implémenté sur vos calculatrices prend en compte des données aux trois formats (brutes, regroupées, résumées). La figure 10 présente la réalisation d'un test d'homogénéité des variances entre les données de la table 3 et de la figure 8 (données groupées).

FIGURE 10 – Test F d'homogénéité de deux variances sur des données groupées



De gauche à droite

1. Les données groupées sont entrées dans deux listes de valeurs (`L1` et `L3` ou `List 1` et `List 3`) et les effectifs dans deux listes d'effectifs (`L2` et `L4` ou `List 2` et `List 4`).
2. Le test est appelé `2-CompFTest` ou `2-SampFTest` (t.i.) on y accède avec les choix `TEST`→`F` (casio).
3. On effectue les choix qui conviennent : les valeurs dont dans des listes, les effectifs sont dans des listes (données groupées). L'hypothèse alternative est $H_1 : \sigma_1 \neq \sigma_2$ (test bilatéral).
4. On lit les résultats du calcul. La valeur de la statistique est $F = 1.56$ et est associée à une valeur $p = 0.51$. On a donc $p > \alpha$, ce qui ne permet pas de rejeter l'hypothèse nulle avec un risque $\alpha = 0.05$. On accepte donc que les variances sont égales avec un risque β inconnu.

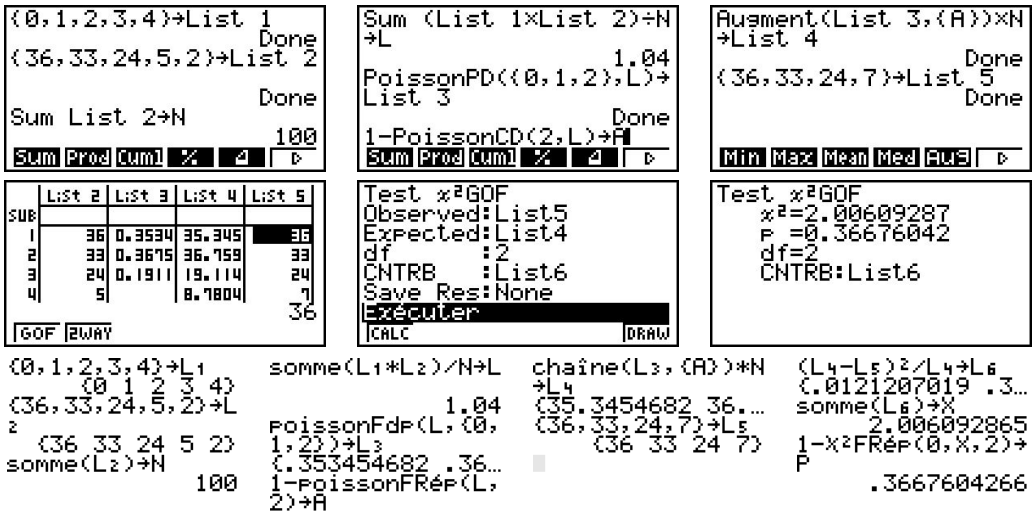
4 Tests de χ^2

4.1 Test χ^2 d'ajustement à une distribution théorique

Sur les machines où il est implémenté, le test χ^2 d'ajustement est nommé `GOF` (goodness of fit). Cette implémentation n'est qu'incomplètement satisfaisante car c'est à l'utlilisateur de calculer et fournir les effectifs attendus théoriques sous H_0 et le nombre de degrés de liberté. La figure 11 montre la réalisation d'un test χ^2 d'ajustement à la loi de Poisson sur une calculatrice disposant d'une fonction spécialisée et sur une calculatrice n'en disposant pas.

FIGURE 11 – Réalisation d'un test χ^2 d'ajustement

Classe (X_i)	0	1	2	3	4
Effectifs observés	36	33	24	5	2



Deux premières lignes (casio) : de gauche à droite et de haut en bas

1. On saisit les données, calcule l'effectif total `N` et la valeur du paramètre λ de la loi de Poisson `L`.
2. Les probabilités théoriques sont calculées pour les classes 0 à 2 avec la loi de probabilité de la loi de Poisson. La dernière classe regroupe toutes les classes supérieures ou égales à 3 (probabilité `A`).
3. Les effectifs attendus sont stockés dans la liste 4 et les effectifs observés regroupant les classes 3 et suivantes sont stockés dans la liste 5.
4. Sur une calculatrice disposant du test, on choisit le test de χ^2 de type "GOF" (notez les contenus des listes 5 = effectifs observés et 4 = effectifs attendus théoriques).
5. Dans les paramètres du test, on indique les bonnes listes pour les effectifs observés et attendus sous H_0 ainsi que le nombre de degrés de liberté (4 classes et un paramètre estimé, d'où deux degrés de liberté). Les contributions calculées pour chaque classe seront stockées dans la liste 6.
6. La calcularice renvoie la valeur de la statistique, $X^2 = 2.006$. La valeur $p = 0.367$ est supérieure au risque $\alpha = 0.05$ classiquement utilisé, on ne peut donc pas rejeter H_0 avec un risque de première espèce $\alpha = 5\%$ et on l'accepte avec un risque β inconnu. La distribution dans la population est conforme à la loi de Poisson.

Dernière ligne (t.i.) : Sur une calculatrice ne disposant pas du test χ^2 d'ajustement ("GOF") le début est identique (trois premières vignettes) et les résultats du test sont obtenus en appliquant la formule de la statistique X^2 et en calculant la $p-$value d'après la fonction de répartition de la loi de χ^2 avec le bon nombre de degrés de liberté (dernière vignette).

4.2 Test χ^2 d'homogénéité (ou d'indépendance)

Un test χ^2 d'homogénéité compare les distributions de p échantillons en k classes. La figure 12 illustre la mise en œuvre sur les données suivantes constituées des répartitions en classes d'âge d'échantillons de 100 individus provenant de quatre pays européens.

FIGURE 12 – Réalisation d'un test χ^2 d'indépendance

Données utilisées

	<20	20-64	66+
France	33	40	27
Belgique	25	45	30
Allemagne	12	52	36
Italie	20	54	26

Réalisation du test

De gauche à droite

1. Les effectifs observés $n_{i,j}$ (classe j de la population i) sont entrés dans une matrice (un tableau) de p lignes (une par échantillon) et k colonnes stockée dans l'une des mémoires de matrice de la calculatrice.
2. Le test est simplement désigné sous le nom `X²-Test` sur la TI82 et sous le nom `2-WAY` sur la Casio Graph 35+.
3. On indique la matrice des effectifs observés que l'on a rentrée et un nom de matrice où seront stockés les effectifs attendus théoriques sous H_0.
4. La calcularice renvoie la valeur de la statistique, $X^2 = 15.01$, *df* est le nombre de degrés de libertés. La valeur $p = 0.02$ est inférieure au risque $\alpha = 0.05$ classiquement utilisé, on peut donc rejeter H_0 avec un risque de première espèce $\alpha = 5\%$. Les distributions en classe d'âge dépendent du pays.

5 Régression linéaire : test de corrélation

On dispose de données organisées par paires de valeurs (X_i, Y_i) (données brutes) ou par triplet avec deux valeurs et un effectif (X_i, Y_i, N_i) (données groupées). Le test est un test de type t nommé `RégLinTTest` (t.i.) ou accessible avec `test→t→REG` (casio). La figure 13 montre la réalisation de ce test sur des données brutes.

FIGURE 13 – Réalisation d'un test de corrélation linéaire

Données utilisées

taille	163	164	170	181	171	166	165
poids	57	51	62	74	60	62	56

Réalisation du test

De gauche à droite

1. Les tailles et les poids observés sont entrés dans les listes 1 et 2.
2. On choisit le test approprié, on indique les listes utilisées pour les données. Il y a une observation par paire.
3. La calculatrice donne la valeur de $t = 4.81$ ainsi que la $p-$value ($p = 0.005$), les coefficients de la droite de régression, le coefficient de corrélation et le coefficient de détermination r^2. Ici, $p < \alpha = 0.05$ donc on peut rejeter l'hypothèse nulle avec un risque $\alpha = 0.05$: la taille et le poids sont positivement corrélés et cette corrélation explique 82% de la variation.

6 Analyse de la variance

On s'intéresse aux données du tableau 6 qui donne les longueurs du tibia de la première paire de pattes d'une espèce de diptère en fonction du sexe et de la richesse du milieu d'élevage des larves.

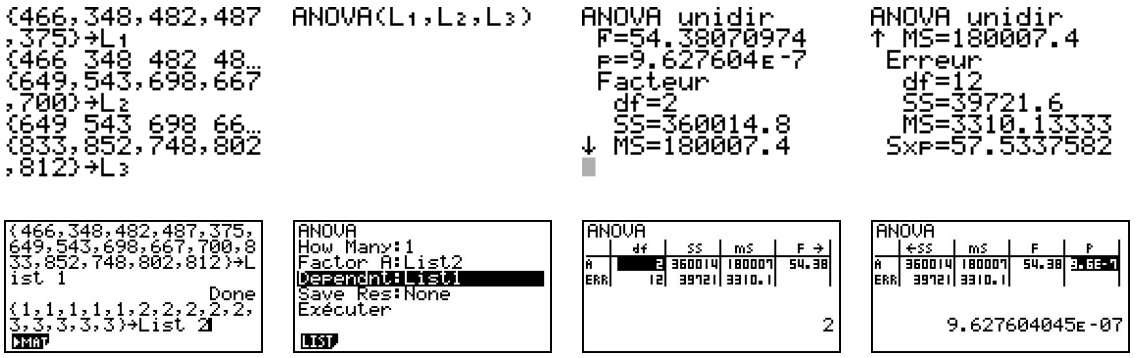
TABLE 6 – Données utilisées pour l'analyse de variance

Très pauvre		Pauvre		Riche	
Mâles	Femelles	Mâles	Femelles	Mâles	Femelles
466	346	649	487	833	585
348	323	543	506	852	527
482	465	698	483	748	549
487	450	667	501	802	530
375	419	700	498	812	556

6.1 ANOVA1

Dans les données de la table 6 on s'intéresse à l'effet du régime alimentaire chez les mâles. La figure 14 montre comment réaliser l'analyse de variance à un facteur.

FIGURE 14 – Réalisation d'une ANOVA à un facteur



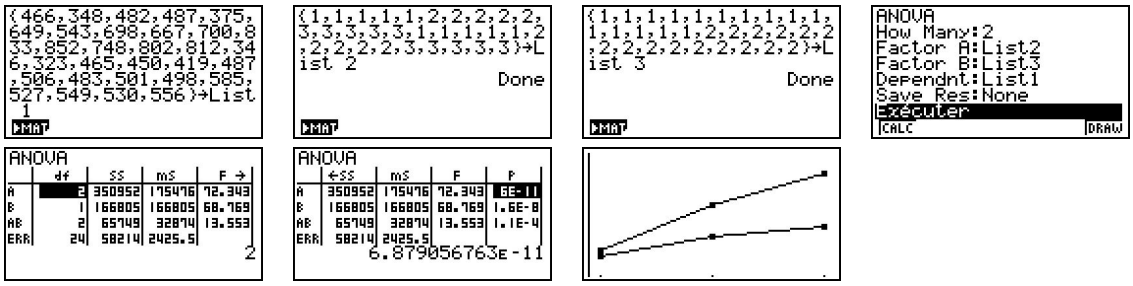
De gauche à droite

1. On entre les données : sur une t.i. les valeurs sont rentrées dans les listes `L1` à `L3` correspondant aux trois modalités de régime larvaire. Sur une casio, on entre deux listes : une liste de valeurs dans `List 1` et une liste de modalités du régime larvaire (arbitrairement 1, 2, ou 3 pour très pauvre, pauvre et riche) dans `List 2`.
2. On choisit le test approprié, il s'agit d'une ANOVA à un facteur. Sur une casio il faut bien distinguer la liste des valeurs de taille (appelée "dependent") et celle des modalités du facteur étudié (Factor A) qui est le régime alimentaire des larves.
3. On lit les résultats tels qu'ils apparaissent dans un tableau d'ANOVA1. df=degrés de libertés, SS=somme des carrés des écarts, ms=carré moyen.
4. La statistique du test est $F = 54.4$ la $p-$value est $p = 9.6 \times 10^{-7}$ on a donc $p < \alpha = 0.05$ et on peut rejeter l'hypothèse nulle selon laquelle la moyenne de la longueur étudiée est indépendante du régime alimentaire des larves avec un risque de première espèce $\alpha = 0.05$.

6.2 ANOVA2

La réalisation d'une ANOVA à deux facteurs : régime alimentaire et sexe sur la longueur moyenne du tibia de la première paire de pattes d'une espèce de diptères (table 6) est illustrée sur la figure 15. Ce test n'est implémenté que sur les calculatrices casio.

FIGURE 15 – Réalisation d'une ANOVA à deux facteurs (casio uniquement)



De gauche à droite

1. On entre les données dans trois listes : une liste de valeurs dans `List 1`, une liste de modalités du régime larvaire (arbitrairement 1, 2, ou 3 pour très pauvre, pauvre et riche) dans `List 2` et une liste de modalités du sexe (arbitrairement 1 et 2 pour mâle et femelle) dans `List 3`.
2. On choisit le test approprié, il s'agit d'une ANOVA à deux facteur. La liste 2 contient les modalités du facteur A (régime alimentaire), la liste 3 les modalités du facteur B (le sexe) et la liste 1 contient les valeurs de taille (appelée "dependent").
3. On lit les résultats du tableau d'ANOVA2. df=degrés de libertés, SS=somme des carrés des écarts, ms=carré moyen. La ligne A correspond à l'effet du régime alimentaire, la ligne B correspond à l'effet du sexe et la ligne AB à l'interaction de ces deux facteurs.
4. Pour chaque facteur et pour l'interaction la $p-$value associée est inférieure à $\alpha = 0.05$ et on peut donc rejeter l'hypothèse nulle selon laquelle la moyenne de la longueur étudiée est indépendante du régime alimentaire des larves, du sexe et de l'interaction de ces deux facteurs à chaque fois avec un risque de première espèce $\alpha = 0.05$.
5. Le graphe d'interaction montre les moyennes estimées pour les couples (sexe, régime alimentaire). Les deux lignes correspondent au facteur B (sexe) et illustrent la réponse différente des deux sexes aux trois régimes alimentaires (l'interaction entre les deux facteurs).