

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ MOHAMED KHIDER, BISKRA

FACULTÉ des SCIENCES EXACTES et des SCIENCES de la NATURE et de la VIE

DÉPARTEMENT DE SNV



Polycopié du TP :

Analyse de données sous SPSS

Estimation et Tests d'hypothèses
(TP 1 et TP 2)

2020/2021

Table des matières

Table des matières	2
Liste des figures	3
1 Estimation sous SPSS	1
1.1 Rappel	1
1.1.1 Estimation ponctuelle	1
1.1.2 Estimation par intervalle de confiance	2
2 Tests d’hypothèses	8
2.1 Principe d’un test	8
2.2 Eléments d’un Test	8
2.2.1 Hypothèses	9
2.2.2 Risque	9
2.3 Test bilatéral et unilatéral	9
2.3.1 Test bilatéral	10
2.3.2 Test unilatéral	10
2.4 La démarche d’un test	10
2.5 Test de Conformité de la Moyenne	11

Table des figures

1.1	Saisie des données sous SPSS.	4
1.2	Réalisation de l'estimation sous SPSS (partie 1).	5
1.3	Etapes de la réalisation de la méthode d'estimation sous SPSS (Partie 2). . .	6
1.4	Fenêtre des résultats de l'estimation sous SPSS.	7
1.5	Intervalle de confiance de la taille moyenne des enfants au risque $\alpha = 1\%$. . .	7
2.1	Etapes de la réalisation d'un test de conformité de la moyenne sous SPSS. . .	12
2.2	Fenêtre des résultats du test de conformité de la moyenne sous SPSS.	13

Chapitre 1

Estimation sous SPSS

Dans ce chapitre, on va présenter un rappel sur la notion d'estimation. Puis, à travers d'un exemple d'application, on va donner les principales étapes à suivre pour faire une estimation de la moyenne à l'aide du logiciel SPSS.

1.1 Rappel

L'estimation consiste à évaluer certaines caractéristiques d'une variable aléatoire grâce à des observations réalisées sur un échantillon. C'est-à-dire, elle consiste à évaluer un paramètre inconnu θ (moyenne, variance, fréquence,...) de la population à partir d'un échantillon représentatif tiré de cette population.

La valeur estimée du paramètre θ est souvent notée $\hat{\theta}$, alors

θ : est la valeur réelle du paramètre.

$\hat{\theta}$: est la valeur estimée de θ .

Un grand nombre de problèmes statistiques consistent à la détermination de la moyenne.

Il existe deux types d'estimation :

1.1.1 Estimation ponctuelle

Lorsqu'un paramètre est estimé par un seul nombre déduit des résultats de l'échantillon, ce nombre est appelé une estimation ponctuelle du paramètre.

- Un estimateur de la moyenne μ est donné par la moyenne de l'échantillon

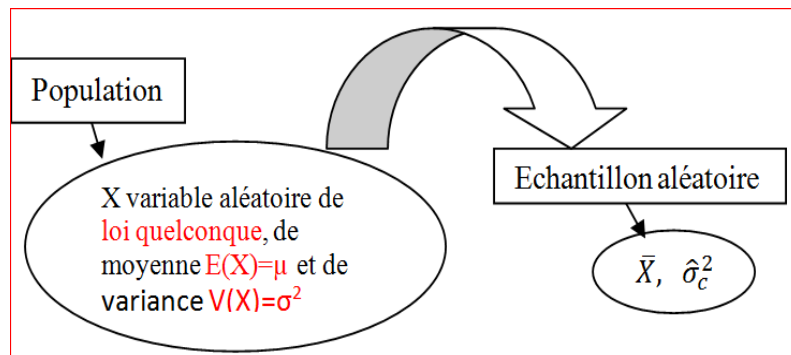
$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

- Un estimateur de la variance σ^2 est donné par la variance corrigée de l'échantillon

$$\hat{\sigma}_c^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Remarque 1.1 Si la moyenne est connue, la variance σ^2 est estimée par

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2.$$



1.1.2 Estimation par intervalle de confiance

Les estimations ponctuelles, bien qu'utiles, ne fournissent aucune information concernant la précision des estimations, c'est-à-dire qu'elles ne tiennent pas compte de l'erreur possible dans l'estimation due aux fluctuations d'échantillonnage. La théorie des intervalles de confiance (IC) consiste à construire, autour de l'estimation ponctuelle, un intervalle qui aura une grande probabilité $(1 - \alpha)$ de contenir la vraie valeur du paramètre.

Typiquement, on cherche deux nombre réels a et b tels que, par exemple

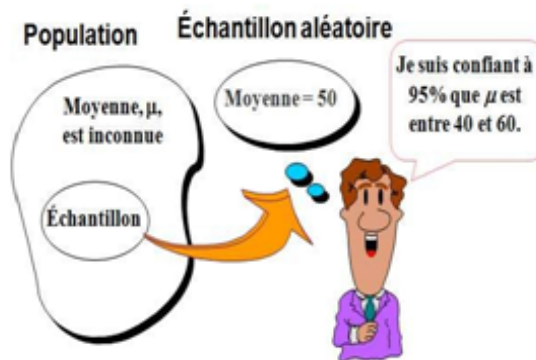
$$P(a \leq \theta \leq b) = 1 - \alpha,$$

où

- L'intervalle $[a, b]$ est appelé intervalle de confiance pour θ et il est noté par

$$IC_{1-\alpha}(\theta) = [a, b].$$

- $(1 - \alpha)$ est le niveau de confiance et α est le risque de se tromper que l'on est prêt à prendre. Dans la pratique, on peut prendre par exemple $\alpha = 5\%$, ce qui nous donne un IC à 95%. Cela signifie qu'il y a 95% de chance que la valeur inconnue θ soit comprise entre a et b . (En général on prend $\alpha : 5\%, 1\%, 10\%$).



Exemple 1.1 La mesure de la taille (en cm) de 10 enfants d'une ville donnée à fourni ce qui suit :

70.5	85.2	93.8	99.1	101	105.8	110.3	121.2	138.6	166.1
------	------	------	------	-----	-------	-------	-------	-------	-------

Questions :

- 1- Déterminer une estimation ponctuelle de la moyenne et de l'écart-type des tailles de ces enfants.
- 2- Déterminer une estimation par intervalle de confiance de la taille moyenne de ces enfants :
 - a- pour un risque $\alpha = 5\%$.
 - b- pour un risque $\alpha = 1\%$.

Pour répondre aux questions de cet exercice sous SPSS, il faut suivre les étapes suivantes

Etape 1. Saisie des données

Entrez les données dans SPSS, dont vous avez une seule variable quantitative (**Tailles des enfants**) à définir dans SPSS : cliquez sur "**Affichages des variables**" puis saisissez tous ses caractéristiques : **Nom, Type, Largeur,...**), ensuite cliquez sur "**Affichage des données**" pour entrer les valeurs des variables où chaque colonne représente une variable. (Voir figure 1.1).

	Taille	var	var	var	var	var	var	var	var	var	var
1	70,50										
2	85,20										
3	93,80										
4	99,10										
5	101,00										
6	105,80										
7	110,30										
8	121,20										
9	138,60										
10	166,10										

FIG. 1.1 – Saisie des données sous SPSS.

Remarque 1.2 *Il faut sauvegarder votre fichier.*

Etape 2. Réalisation de l'estimation sous SPSS (partie 1). Sélectionnez sur la barre de menu Analyse → Comparer les moyennes → Test T pour échantillon unique. (Voir figure 1.2).

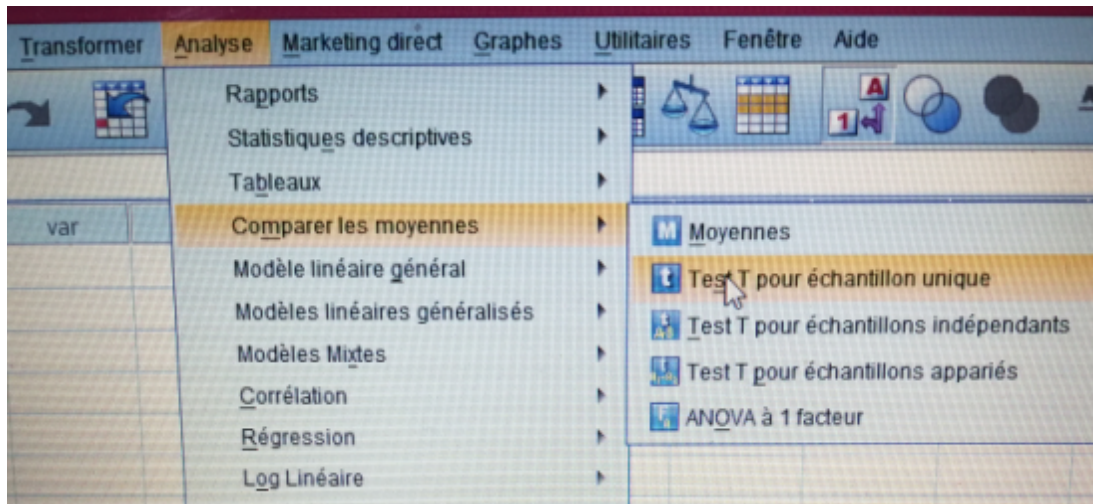


FIG. 1.2 – Réalisation de l'estimation sous SPSS (partie 1).

Etape 3. En cliquant sur " **Test T pour échantillon unique**" (voir Etape 2) une fenêtre sera affichée dont la forme est présentée dans la figure 1.3 à gauche. Alors il faut suivre ce qui suit :

- 1- Mettez la variable **Taille enfant** dans la case "**Variables à tester**".
- 2- Vérifiez dans la case "**Valeur de test**" qu'elle est égale à 0.
- 3- Cliquez sur "**option**" une fenêtre va s'ouvrir (voir la figure 1.3 à droite).
- 4- Choisissez le niveau de confiance, dans la question 2 (a), on a le risque $\alpha = 5\%$, donc le niveau de confiance égale à $1 - \alpha = 1 - 0.05 = 0.95$ (95%).
- 5- Cliquez sur "**poursuivre**".
- 6- Cliquez sur "**OK**".

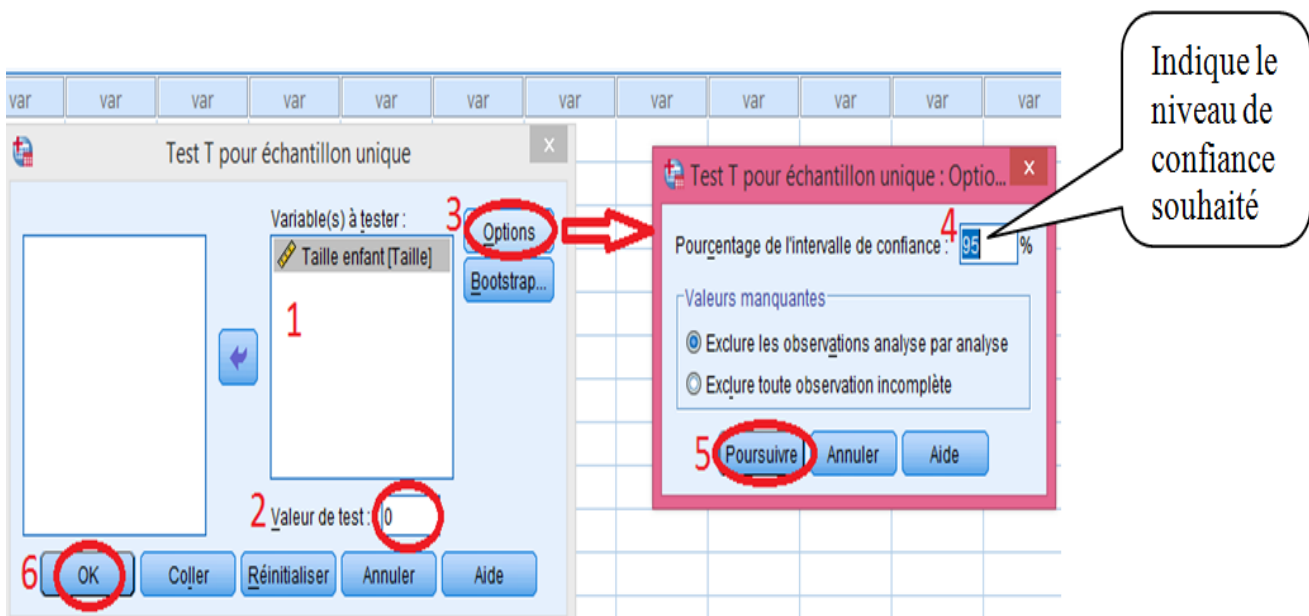


FIG. 1.3 – Etapes de la réalisation de la méthode d’estimation sous SPSS (Partie 2).

Etape 4. Résultats : L’application de ces étapes sur les données de l’exercice permet d’obtenir les résultats (2 tableaux) qui sont présentés dans la figure 1.4. Ces résultats contiennent des réponses sur les questions 1 et 2 (a).

Etape 5. Interprétation des résultats :

Le premier tableau donne :

- La taille de l’échantillon : $n = 10$.
- Une estimation ponctuelle de la taille moyenne des enfants : $\hat{\mu} = \bar{X} = 109.16 \text{ cm}$.
- Une estimation ponctuelle pour l’écart type : $\hat{\sigma}_c = \sqrt{\hat{\sigma}_c^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2} = 27.36345 \text{ cm}$.
- Une estimation ponctuelle pour l’erreur standard moyenne : $\frac{\hat{\sigma}_c}{\sqrt{n}} = 8.65308 \text{ cm}$.

Pour le deuxième tableau, l’information cruciale dans ce tableau est la colonne intitulée "**Intervalle de confiance 95% de la différence**" qui représente l’intervalle de confiance de niveau 95% de la taille moyenne des enfants (les bornes inférieure et supérieure de cet intervalle). Alors

$$IC_{95\%}(\mu) = [89.5854, 128.7346].$$

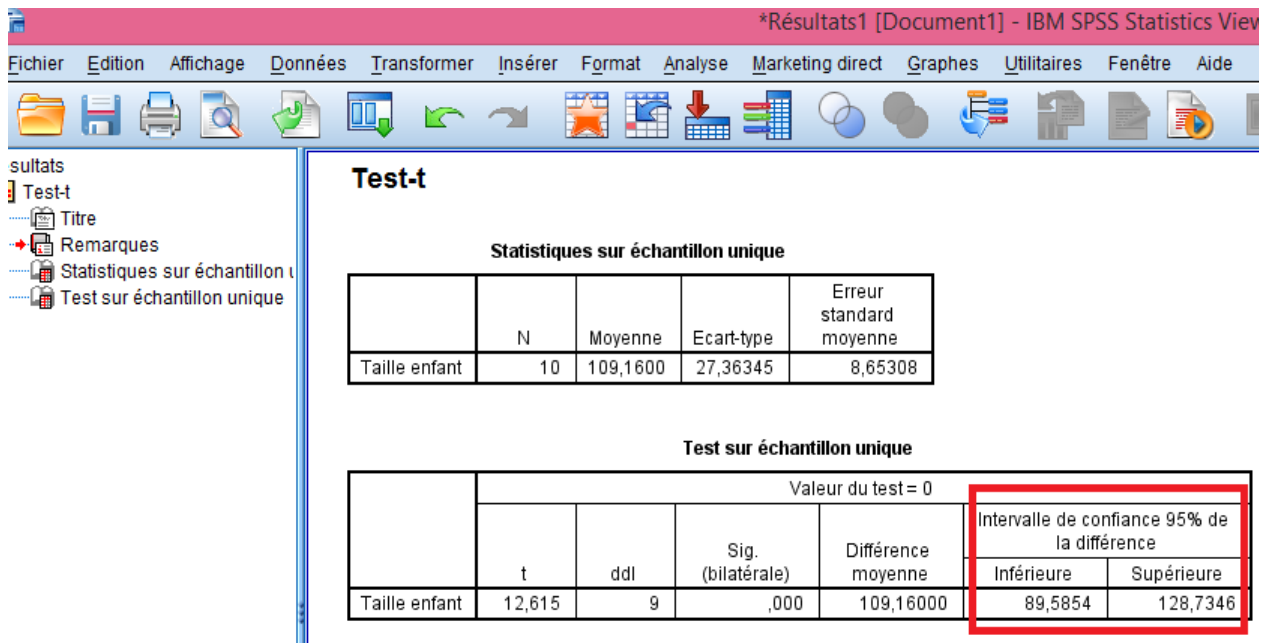


FIG. 1.4 – Fenêtre des résultats de l'estimation sous SPSS.

Remarque 1.3 Pour obtenir un intervalle de confiance de la taille moyenne des enfants au risque $\alpha = 1\%$, il faut suivre les mêmes étapes précédentes avec un petit changement de niveau de confiance (dans ce cas, il est égale à 99%) (voir figure 1.5).

Test sur échantillon unique

	Valeur du test = 0				Intervalle de confiance 99% de la différence	
	t	ddl	Sig. (bilatérale)	Différence moyenne	Inférieure	Supérieure
Taille enfant	12,615	9	,000	109,16000	81,0389	137,2811

FIG. 1.5 – Intervalle de confiance de la taille moyenne des enfants au risque $\alpha = 1\%$.

Chapitre 2

Tests d'hypothèses

Les tests statistiques constituent une approche décisionnelle de la statistique inférentielle. Un tel test a pour objet de décider sur la base d'un échantillon si une caractéristique de la loi mère (ou de la population) répond ou non à une certaine spécification que l'on appelle hypothèse, par exemple : la moyenne de la loi est supérieure à 10.

Dans ce chapitre, on va présenter les notions de base nécessaires à la compréhension de ceux-ci, aussi ses différents types sont donnés.

2.1 Principe d'un test

Un test statistique est une mise à l'épreuve d'une hypothèse concernant une population sur la base de données fournies à partir d'un échantillon (ou plusieurs) représentatif de la population, qui permet de prendre la décision de rejeter ou de ne pas rejeter les hypothèses. Les tests d'hypothèses sont des outils statistiques d'aide à la décision. Ils vont permettre de comparer un ou plusieurs échantillons, et de valider ou d'invalider une hypothèse donnée.

2.2 Eléments d'un Test

On a vu qu'un test d'hypothèses est une règle de décision qui permet, sur la base des données observées et avec des risques d'erreur déterminés, d'accepter ou de refuser une hypothèse

statistique. Alors, il est constitué des éléments suivants :

2.2.1 Hypothèses

Une hypothèse est l'information qu'on veut confirmer ou infirmer concernant les caractéristiques (valeurs des paramètres, forme de la distribution des observations) d'une population. Le principe est de comparer la probabilité d'une hypothèse versus le contraire de cette même hypothèse. On distingue deux types d'hypothèses, dont une et une seule est vraie.

L'hypothèse nulle notée H_0 est celle que l'on considère vraie à priori, c'est l'hypothèse principale.

L'hypothèse alternative notée H_1 on peut choisir pour cette hypothèse n'importe quelle hypothèse compatible avec le problème étudié mais différente de H_0 .

2.2.2 Risque

On ne pourra jamais conclure avec certitude dans un test statistique. Il y aura toujours des erreurs de décision.

Pour effectuer un test statistique, il faudra choisir un certain risque d'erreur qui est la probabilité de se tromper en prenant la décision retenue. Dans la pratique on choisit un risque α qui est la probabilité de rejeter l'hypothèse H_0 alors qu'elle est vraie

$$\alpha = P(\text{rejeter } H_0 / H_0 \text{ est vraie}).$$

Remarque 2.1 - *On appelle α le seuil de risque ou niveau de signification (ou seuil de signification).*

- *Les valeurs usuelles de α sont 1%, 5% et 10%.*

2.3 Test bilatéral et unilatéral

Avant d'appliquer tout test statistique, il s'agit de bien définir le problème posé. En effet, selon les hypothèses formulées, on appliquera soit un test bilatéral, soit un test unilatéral.

2.3.1 Test bilatéral

Un test bilatéral s'applique quand on cherchera une différence entre deux paramètres, ou entre un paramètre et une valeur donnée sans se préoccuper du signe ou du sens de la différence.

$$\left\{ \begin{array}{l} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{array} \right. ; \quad \left\{ \begin{array}{l} H_0 : \theta_1 = \theta_2 \\ H_1 : \theta_1 \neq \theta_2 \end{array} \right. .$$

2.3.2 Test unilatéral

Un test unilatéral s'applique quand on cherchera à savoir si un paramètre est supérieur (ou inférieur) à un autre ou à une valeur donnée (Il y'a un test dit unilatéral à droite et un test dit unilatéral à gauche).

- Test unilatéral à droite : $\left\{ \begin{array}{l} H_0 : \theta = \theta_0 \\ H_1 : \theta > \theta_0 \end{array} \right. ; \quad \left\{ \begin{array}{l} H_0 : \theta_1 = \theta_2 \\ H_1 : \theta_1 > \theta_2 \end{array} \right. .$
- Test unilatéral à gauche : $\left\{ \begin{array}{l} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 \end{array} \right. ; \quad \left\{ \begin{array}{l} H_0 : \theta_1 = \theta_2 \\ H_1 : \theta_1 < \theta_2 \end{array} \right. .$

2.4 La démarche d'un test

On peut résumer la démarche d'un test sous SPSS de la manière suivante

1. Choix des hypothèses H_0 et H_1 .
2. Fixation du niveau α de signification.
3. Choix du type du test qu'on doit l'appliquée (selon le problème posé).
4. Exécution du test sous SPSS.
5. Interprétation des résultats : Prise de décision à l'aide de la valeur "signification (Sig)" donnée par le logiciel SPSS : rejet ou non rejet de H_0 au risque α .

Il existe plusieurs types des tests paramétriques (son objet est de tester certaine hypothèse relative à un ou plusieurs paramètres d'une variable aléatoire de loi spécifiée), parmi les quelles on s'intéresse aux tests de conformité pour la moyenne, tests d'homogénéité ou tests de comparaison de deux moyennes (cas des échantillons indépendants), ces deux tests (tests

de conformité et tests d'homogénéité) sont appelés aussi tests de Student. On s'intéresse aussi au test de comparaison de plusieurs moyennes (plus que deux) (ANOVA 1).

2.5 Test de Conformité de la Moyenne

Dans ce test d'hypothèse, on veut valider l'hypothèse que la moyenne μ de la population entière diffère d'une certaine valeur (μ_0), qui est déterminée par une expérience précédente ou par analogie avec une situation semblable. Par exemple, dans une certaine expérience, on veut voir si le rendement moyen d'un engrais pour la culture du blé (tonnes par hectare) est significativement supérieur à 8 tonnes. Alors, on pose les hypothèses suivantes

$$\begin{cases} H_0 : \mu = 8 \\ H_1 : \mu > 8 \end{cases} .$$

Dans le cas général, les hypothèses à tester sont de la forme suivante

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \quad \text{ou} \quad (H_1 : \mu > \mu_0); (H_1 : \mu < \mu_0) \end{cases} .$$

Remarque 2.2 *Les conditions d'application de ce test sont : il faut que la variable aléatoire X suive une loi de Laplace-Gauss (loi Normale $N(\mu, \sigma^2)$) ou l'échantillon aléatoire est de taille important ($n \geq 30$).*

Exemple 2.1 *Au vu de l'échantillon de l'exemple 1 (exercice 1), peut-on considérer, au seuil de signification 5% que la taille moyenne des enfants est significativement égale à 110 cm ?*

Dans le logiciel SPSS, La procédure de validation d'une hypothèse sur la valeur de la moyenne d'une population lorsqu'on a un échantillon (test de conformité de la moyenne ou "test de Student") est la même procédure de l'estimation, c'est-à-dire, "**Test T pour échantillon unique**", alors dans ce cas, il faut que vous répétez les mêmes étapes 2 et 3 de l'estimation citées précédemment dans l'exemple 1, mais avant de procéder à l'affichage des résultats

c'est-à-dire avant de cliquer sur le bouton "OK", il faut changer "la valeur de tests", en mettant à la place de "0" la valeur " μ_0 " dans notre exemple vous mettez "110" puis cliquez sur "OK". (Voir figure 2.1).

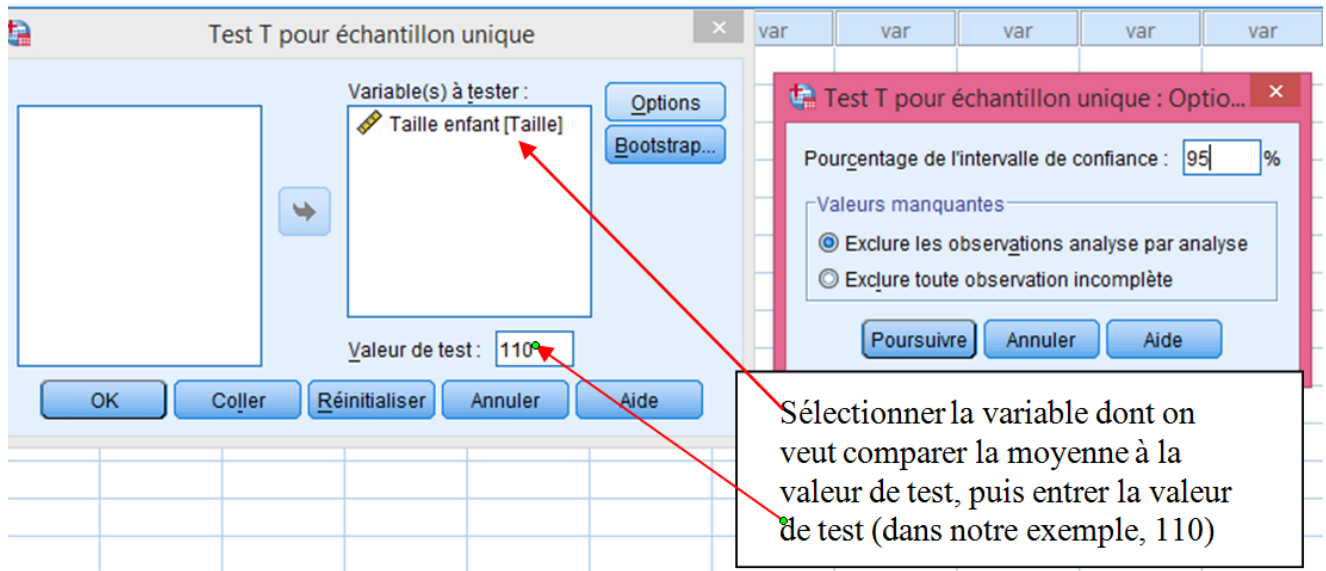


FIG. 2.1 – Etapes de la réalisation d'un test de conformité de la moyenne sous SPSS.

Une fois que vous cliquez sur "OK" vous obtiendrez les résultats qui seront affichés dans la figure 2.2.

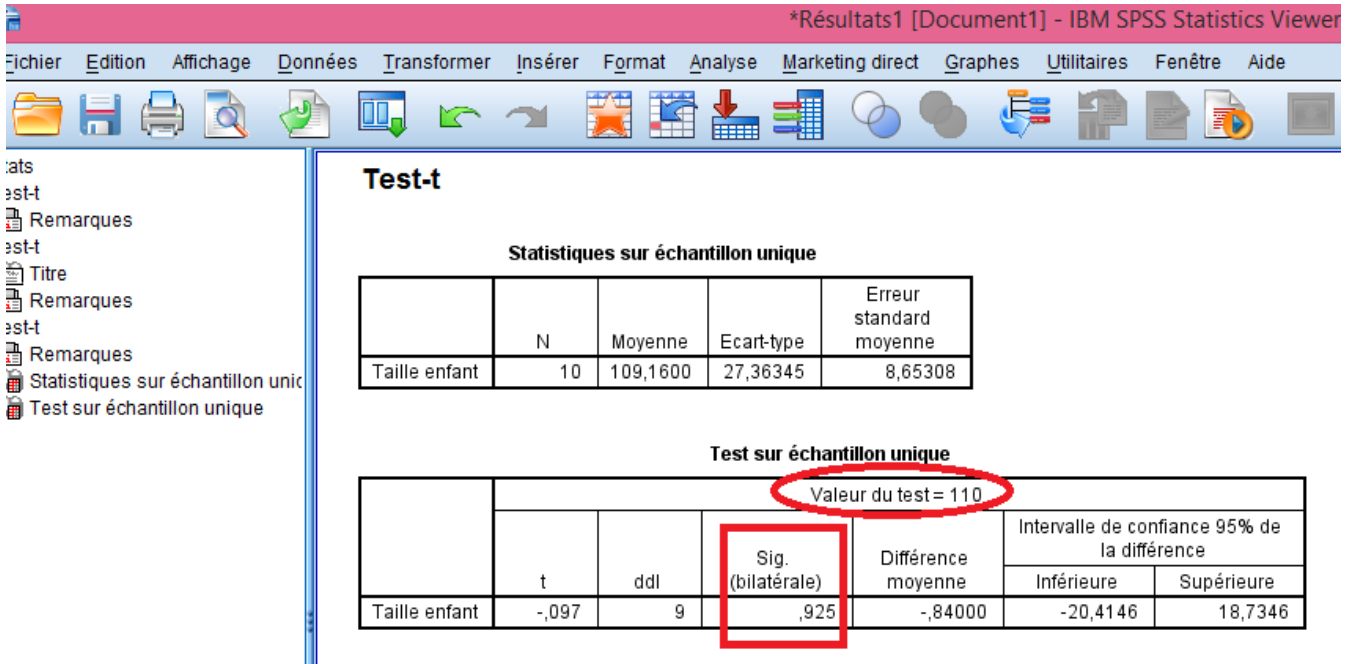


FIG. 2.2 – Fenêtre des résultats du test de conformité de la moyenne sous SPSS.

Interprétation des Résultats :

Afin de prendre une décision si on rejette H_0 ou non on procède comme suit :

$$\begin{cases} \text{non rejet de } H_0 \text{ au risque } \alpha; & \text{si } \alpha < \textit{signification}; \\ \text{rejet de } H_0 \text{ au risque } \alpha; & \text{si } \alpha \geq \textit{signification}; \end{cases}$$

A partir des résultats obtenus sur notre exemple, on constate que pour un seuil de risque $\alpha = 5\%$, qu'on ne peut pas rejeter H_0 , c'est-à-dire, la taille moyenne des enfants est significativement égale à 110 cm car $\alpha < \textit{signification}$ ($0.05 < 0.925$).