



Les Statistiques Descriptives en Sciences de la Santé Descriptive Statistics in Health Sciences

الإحصاء الوصفي في العلوم الصحية

Ahmed Ben Abdelaziz¹, Donia Ben Hassine², Dhekra Chebi³, Sarra Nouira⁴, Asma Ben Abdelaziz⁵, Sarra Melki⁶, Tarek Barhoumi⁷, Kamel Ben Salem⁸, Réseau Maghrébin PRP2S*

* Réseau- Maghrébin: Pédagogie- Recherche- Publication:s en Sciences de Santé (PRP2S)

1. *Professeur de Médecine Préventive et Communautaire. Direction des Systèmes d'Information du CHU Sahloul de Sousse. Laboratoire de Recherche LR19SP01. Université de Sousse (Tunisie). Email: ahmedbenabdelaziz.prp2s@gmail.com*
2. *Résidente de Médecine Préventive et Communautaire. Direction des Systèmes d'Information du CHU Sahloul de Sousse. Laboratoire de Recherche LR19SP01. Université de Sousse (Tunisie).*
3. *Professeur Assistante de Médecine Préventive et Communautaire. Faculté de Médecine de Sousse. Laboratoire de Recherche LR19SP01. Université de Sousse (Tunisie).*
4. *Résidente de Médecine Préventive et Communautaire. Laboratoire de Recherche LR19SP01. Faculté de Médecine de Monastir. Université de Monastir (Tunisie).*
5. *Professeur assistante universitaire. Service de Biochimie du CHU Sahloul de Sousse. Faculté de Pharmacie de Monastir. Université de Monastir (Tunisie).*
6. *Résidente de Médecine Préventive et Communautaire. CHU de Sahloul de Sousse. Laboratoire de Recherche LR19SP01. Université de Monastir (Tunisie).*
7. *Médecin Major de Santé Publique. Direction Régionale de la Santé. Tunisie*
8. *Professeur de Médecine Préventive et Communautaire. Faculté de Médecine de Monastir. Université de Monastir (Tunisie).*

CETTE SÉRIE...

Le Réseau Maghrébin PRP2S et la Rédaction de la revue «La Tunisie Médicale» ont l'honneur de continuer d'une manière régulière, à partir du numéro de mars 2021, et pour la deuxième année successive, la série des fiches techniques en épidémiologie, en bio statistique et en rédaction médicale scientifique. Cette série a eu un grand succès au cours de sa première année d'édition en 2020, comme indique le nombre de téléchargements dépassant significativement celui des articles originaux et illustrant un besoin très manifeste des jeunes chercheurs, au renforcement de leurs capacités en méthodologie de recherche scientifique en sciences de santé, selon une pédagogie centrée sur l'acquisition des compétences pratiques de recherche biomédicale. En effet, nos fiches méthodologiques décrivent, d'une manière standardisée, les modes d'usage des concepts, des outils et des méthodes, utilisés d'une part lors du continuum de la recherche biomédicale scientifique, dès la phase conceptuelle jusqu'à la phase rédactionnelle et d'autre part lors des différentes phases de la rédaction médicale scientifique, depuis l'étape de la recherche documentaire jusqu'à l'étape de la communication médicale scientifique. Cette série est rédigée par les experts du Réseau Maghrébin PRP2S, en méthodologie de recherche, exerçant dans les universités du Grand Maghreb et les facultés sœurs au Nord de la Méditerranée. Chaque fiche répond à trois questions essentielles (Quoi ? Pourquoi ? Comment) du concept étudié, en se basant sur un article publié dans la revue «La Tunisie Médicale».

Le coordinateur de la série «Fiches Méthodologiques»

Professeur Ahmed Ben Abdelaziz (Président du Réseau Maghrébin PRP2S)

Email: ahmedbenabdelaziz.prp2s@gmail.com

Correspondance

Pr. Ahmed Ben Abdelaziz.

Direction des Systèmes d'Information. CHU Sahloul de Sousse. 4054 Tunisie

Email: ahmedbenabdelaziz.prp2s@gmail.com

**Série des Fiches Méthodologiques
Sommaire**

Année 2020

Année 2021

Fiche n°1 (janvier 2020):

Comment calculer la taille d'un échantillon pour une étude observationnelle. Serhier Z, *et al.* (Faculté de Médecine et de Pharmacie de Casablanca. Maroc)

Fiche n°2 (février 2020):

La recherche qualitative: méthodes, outils, analyse. Soulimane A. (Faculté de Médecine, Université Djillali Liabes, Sidi Bel Abbes, Algérie)

Fiche n°3 (mars 2020)

Et Allah ...créa la variabilité. Barhoumi T, *et al* (Réseau Maghrébin PRP2S)

Fiche n°4 (mai 2020)

Réussir votre recherche bibliographique sur *PubMed*. Ben Abdelaziz A, *et al* (Réseau Maghrébin PRP2S)

Fiche n°5 (juin 2020)

Réussir la rédaction de votre «Protocole de Recherche» en sciences de la santé. Ben Abdelaziz A, *et al* (Réseau Maghrébin PRP2S)

Fiche n°6 (juillet 2020)

Analyse multi variée par régression logistique. Ben Salem K, *et al* (Réseau Maghrébin PRP2S)

Fiche n°7 (aout/septembre 2020)

Tests non paramétriques pour comparer deux ou plusieurs moyennes sur des échantillons indépendants. Bezzaoucha A, *et al* (Réseau Maghrébin PRP2S)

Fiche n°8 (septembre 2020)

Comment évaluer la concordance entre deux mesures qualitatives par le test Kappa? Mellakh R, *et al* (Réseau Maghrébin PRP2S)

Fiche n°9 (octobre 2020)

Comment comparer plusieurs moyennes par le test d'Analyse de Variance (ANOVA) ? Khiari H, *et al* (Réseau Maghrébin PRP2S)

Fiche n°10 (novembre 2020)

Tests non paramétriques sur SPSS pour comparer deux ou plusieurs moyennes sur des échantillons appariés. (test de Wilcoxon et test de Friedman). Bezzaoucha A *et al* (Réseau Maghrébin PRP2S)

Fiche n°1 (mars 2021):

Comment calculer et interpréter la valeur de «p» dans une étude épidémiologique. Ladner J *et al.* (Faculté de Médecine et de Pharmacie de Rouen. France)

Fiche n°2 (avril 2021)

La Charge globale de morbidité (Global Burden of Diseases): c'est Quoi? Pourquoi? Comment? Hsairi M *et al* (Faculté de Médecine de Tunis, Tunisie)

Fiche n°3 (mai 2021)

Analyse en Composantes Principales (ACP). Ben Salem K *et al* (Faculté de Médecine de Monastir, Tunisie)

Fiche n°4 (juin 2021)

L'approche 5x5 en Rédaction biomédicale scientifique. Ben Abdelaziz A *et al* (Faculté de Médecine de Sousse, Tunisie)

Fiche n°5 (juillet 2021)

Elaboration du Résumé Structuré d'un article biomédical scientifique. Ben Abdelaziz A *et al* (Faculté de Médecine de Sousse, Tunisie)

Fiche n°6 (aout/septembre 2021)

Comment critiquer un Article Médical Scientifique ? Ben Abdelaziz A *et al* (Faculté de Médecine de Sousse, Tunisie)

Fiche n°7 (octobre)

Réussir l'élaboration d'une liste d'auteurs d'un manuscrit scientifique. Ben Abdelaziz A *et al* (Faculté de Médecine de Sousse, Tunisie)

Fiche n°8 (novembre)

L'Evidence-based Medicine. Ben Abdelaziz A *et al* (Faculté de Médecine de Sousse, Tunisie)

Fiche n°9 : (décembre)

Les Statistiques descriptives. Ben Abdelaziz A *et al* (Faculté de Médecine de Sousse. Tunisie)

ETUDE DE CAS

Dans une publication éditée dans la revue *Tunis Med* [1], à propos de la satisfaction des patients hospitalisés au CHU Sahloul de Sousse (Tunisie), l'équipe de la Direction des Systèmes d'Information (DSI) a conduit une étude transversale dont les objectifs ont été de mesurer les taux de satisfaction, globaux et spécifiques, des patients et d'identifier les déterminants de la satisfaction à l'hôpital universitaire de Sahloul en 2015 et 2016. L'enquête a été menée à l'aide d'un questionnaire explorant quatre dimensions spécifiques de satisfaction: administrative, technique, logistique et relationnelle. Un total de 1897 patients ont été inclus dans l'étude avec un âge moyen de 42,4 ans (ET = 20,5 ans) et un *sex-ratio* de 0,94. Le taux de satisfaction des patients était d'environ 67%. Les éléments de satisfaction étaient principalement en rapport avec la dimension relationnelle: le respect de l'intimité du patient et la qualité de l'information donnée. Ceux d'insatisfaction étaient d'ordre logistique: l'environnement physique dans la chambre, la propreté des toilettes et les délais d'attente. Les déterminants de la satisfaction des patients étaient principalement liés à la qualité de l'accès et de la réception, aux conditions d'hébergement, aux soins techniques, à la qualité de l'information et au respect de l'intimité du patient.

QUIZZ

1. L'âge moyen de la population étudiée a été de 42 ans, ce qui signifie:
 - a. La somme des attributions de la variable «âge» des patients inclus / 1897 patients de l'étude
 - b. La moitié des patients inclus à l'étude avait un âge supérieur à 42 ans
 - c. La moitié des patients inclus à l'étude avait un âge inférieur à 42 ans
 - d. L'âge le plus fréquent des patients inclus à l'étude a été de 42 ans
2. Dans cette étude, l'ET de 20,5 ans, mentionné après la moyenne est:
 - a. L'Ecart Type de la variable «âge» des patients
 - b. La Déviation Standard (*Standard Deviation*: DS) de la variable «âge » des patients
 - c. Un estimateur de la distribution de la variable «âge » des patients, par rapport à sa moyenne
 - d. Un estimateur de la dispersion de la variable âge par rapport à sa moyenne de 42 ans.

3. Avec un âge moyen de 42,4 ans (ET = 20,5 ans), le Coefficient de Variation (CV) est :

- a. 48,35%
- b. 0,48
- c. 21,9
- d. [1,4-83,4]

4. Dans une distribution statistique, la médiane est:

- a. Unique
- b. Peu influencée par les valeurs extrêmes
- c. Plus informative sur la qualité de dispersion
- d. Illustrée par la figure de Box Plot

5. Dans la figure de Box Plot, les paramètres suivants d'une distribution sont illustrés:

- a. La Moyenne
- b. L'Ecart Type
- c. L'Etendue
- d. Le Coefficient de Variation

INTRODUCTION

La variabilité est une caractéristique fondamentale des sciences de la vie. La valeur d'une caractéristique chez les sujets d'une population ou d'un échantillon (taille, poids,..) varie d'un individu à l'autre. En plus des erreurs de mesure, la cause principale de la fluctuation est la variabilité individuelle: inter sujets (ex: tailles différentes,..) et intra sujet (ex: variabilité de la tension artérielle d'un jour à un autre). La Bio Statistique n'est pas le simple dénombrement (ex: le nombre exact des veuves qui ont passé sur le Pont Neuf, au cours de l'année 1860, selon Eugène Labiche). Elle est appliquée à des données biologiques et médicales. Quant à la Statistique, c'est une discipline scientifique qui permet de résumer les données (Statistique descriptive): et d'inférer d'un échantillon à une population (Statistique inférentielle), via l'intervalle de confiance ou les tests d'hypothèse. Après avoir défini l'importance des statistiques descriptives dans la recherche biomédicale (Pourquoi?), clarifié ce concept centré sur le résumé (Quoi?), fondé sur les tendances et la dispersion, cette fiche méthodologique développe la stratégie des études descriptives des bases de données médicales (Comment?), à travers le calcul des paramètres de tendance centrale et de dispersion : les 3M (Moyenne, Médiane et Mode) et les 3E (Etendu, Ecart Type, Ecart Interquartile). Des illustrations par des applications sur le logiciel d'analyse statistique, SPSS, largement utilisé en sciences médicales et sociales, aideront les lecteurs à se familiariser avec les concepts du résumé statistique.

STATISTIQUES DESCRIPTIVES: POURQUOI?

Outre les études épidémiologiques, exclusivement descriptives, toutes les autres études en épidémiologie cliniques ou communautaires, commencent par une phase descriptive du profil de la population d'étude: population exhaustive, échantillon d'étude, cas, témoins, exposés ou non exposés. Ainsi, toutes les publications scientifiques des données épidémiologiques primaires commencent par des tableaux initiaux sur les caractéristiques sociodémographiques et cliniques de la population d'étude. En épidémiologie, l'élaboration de ces tableaux nécessite, obligatoirement, le recours aux statistiques descriptives. Dans une étude bibliométrique conduite sur les méthodes bio statistiques et épidémiologiques employées en recherche biomédicale [2], suite à six mois de publication dans trois revues de référence (*The New England Journal of Medicine*, *The Lancet*, *The Journal of The American Medical Association*), toutes les publications (100%) ont utilisé des statistiques descriptives, versus 83% pour les intervalles de confiance et 57% pour les tests de Chi2 ou le test exact de Fisher.

STATISTIQUES DESCRIPTIVES: C'EST QUOI?

En santé, les statistiques descriptives résument les bases des données épidémiologiques et managériales à travers un «*profiling*» des principales caractéristiques de la population d'étude et de la distribution des phénomènes de santé, en fonction des attributs des personnes, de lieux et de temps. La terminologie bio statistique est formée de neuf concepts et techniques essentielles (encadré 1), dont particulièrement les termes de «variable», «variable quantitative» et «variable qualitative». La méthodologie d'analyse statistique des bases de données, selon une approche descriptive, est décomposée d'une démarche de sept étapes (encadré 2) dont particulièrement la triade d'exploration (tri à plat et révision), de représentation tabulaire/graphique et de résumé par des statistiques de tendance centrale et de dispersion.

Ainsi, une variable est une propriété exclusive des individus. On distingue deux types de variables: qualitative (caractéristique nommée) et quantitative (caractéristique mesurée). La variable quantitative est soit continue (nombre infini des valeurs) soit discrète (nombre limité des valeurs). Les échelles de classification sont des séries des catégories (classes) globalement exhaustives

et mutuellement exclusives. Nous distinguons trois types des échelles de classification: nominale, ordinale, et numérique. Dans les échelles nominales, les classes sont nommées sans ordre (ex: groupe sanguin). Dans les échelles ordinales, les classes sont nommées selon un ordre (ex: niveau de scolarité). Dans l'échelle numérique, les classes sont mesurées soit d'une manière continue (ex: *poids*) soit discrète (ex: *nombre des grossesses*).

Encadré 1: La terminologie bio statistique

1. **Population:** Le plus grand ensemble d'unités concernées par la problématique
2. **Échantillon:** Sous ensemble de cette population
3. **Échantillon aléatoire:** Un échantillon où les sujets sont sélectionnés au hasard
4. **Variable:** Caractéristique observée dont la valeur peut différer à l'intérieur d'une population
5. **Variable quantitative:** Variable comptée. Ex: Taille
6. **Variable qualitative:** Variable nommée. Ex: Sexe
7. **Unité statistique:** Élément de la population ou du groupe étudié
8. **Série statistique:** Ensemble des valeurs observées pour une ou plusieurs variables sur les n sujets ou éléments de la population
9. **Distribution statistique:** Ensemble des couples (x_i, n_i) où x_i est une modalité de la variable et n_i est sa fréquence

Encadré 2 : Les 7 étapes de la stratégie des statistiques descriptives d'une base des données

1. Identifier la variable (type)
2. Identifier l'échelle de mesure
3. Explorer globalement les données
4. Regrouper les données en classes
5. Représentation tabulaire
6. Représentation graphique
7. Résumer les données par les statistiques de tendance centrale et de dispersion

STATISTIQUES DESCRIPTIVES: COMMENT?

A. Données manquantes, erreurs de saisi, distributions des variables

Les objectifs de l'exploration des données sont de quantifier les données manquantes, de détecter les erreurs de saisi et de vérifier la normalité de la distribution.

La figure n°1, résultat de l'analyse des données SPSS, montre que parmi les 1441 enregistrements de la variable «poids» de la base des données, seulement 11 ont été des observations manquantes, soit 0,8% des données manquantes.

	Observations					
	Valide		Manquante		Total	
	N	Pourcent	N	Pourcent	N	Pourcent
Poids(Kg)	1430	99,2%	11	,8%	1441	100,0%

Figure 1. Quantification des données manquantes de la variable «Poids» sur le logiciel d'analyse statistique SPSS

La détection des données manquantes se fait directement sur la base des enregistrements, suite à une opération de triage de la variable (figure n°2).

La détection des erreurs de saisi des enregistrements se fait d'une manière indirecte et globalisante à travers la vérification des données lointaines ou aberrantes dans la figure Box Plot d'une variable particulière (figure n°3).

Dans le répertoire «Données», nous pouvons nous diriger directement vers la valeur saisie de la variable d'étude du dossier mentionné par le Box Plot et effectuer la correction nécessaire (figure n°3).

	n°dos	genre	fumer	fum_jo	atf_avc	par_avc	age_av	atf_ic	paren_ic	age_ic	atp_avc	atp_ic	avoir_ta	avoir_di	taille	poids
1	61	Hommm	non	.	non	.	.	non	.	.	non	non	non	non	.	.
2	62	Hommm	non	.	non	.	.	oui	.	.	non	non	non	non	.	.
3	337	Hommm	non	.	non	.	.	oui	Père /	.	NP	NP	non	non	.	.
4	606	Hommm	non	.	non	.	.	non	.	.	oui	oui	non	oui	.	.
5	664	Hommm	non	.	NP	.	.	NP	.	.	non	non	oui	oui	.	.
6	719	Hommm	oui	oui	NP	.	.	NP	.	.	NP	NP	non	non	.	.
7	772	Hommm	oui	oui	non	.	.	non	.	.	non	non	non	non	.	.
8	364	Femmm	non	.	oui	Père /	77	non	.	.	non	oui	oui	non	166,0	.
9	498	Femmm	non	.	oui	Père /	.	oui	Père /	.	non	non	oui	non	.	.
10	631	Femmm	non	.	non	.	.	non	.	.	non	non	oui	oui	.	.
11	929	Femmm	non	.	non	.	.	non	.	.	non	non	non	non	.	.
12	915	Femmm	non	.	non	.	.	non	.	.	non	non	non	non	150,0	38,6
13	1375	Femmm	non	.	non	.	.	non	.	.	non	non	non	non	158,0	39,2
14	290	Hommm	non	.	non	.	.	non	.	.	non	non	oui	non	147,0	41,9
15	1020	Femmm	non	.	non	.	.	oui	Père /	.	non	non	non	non	151,0	42,1
16	93	Hommm	oui	oui	non	.	.	non	.	.	non	non	non	non	162,0	43,0
17	24	Femmm	non	.	non	.	.	non	.	.	non	non	oui	non	144,0	43,4
18	482	Femmm	non	.	non	.	.	non	.	.	non	non	non	non	143,0	43,6
19	623	Femmm	non	.	non	.	.	non	.	.	non	non	non	non	155,0	44,0
20	557	Femmm	non	.	oui	Père /	.	oui	Frère /	.	non	non	non	non	149,5	44,1

Figure 2. Détection des données manquantes de la variable «Poids» sur le logiciel d'analyse statistique SPSS

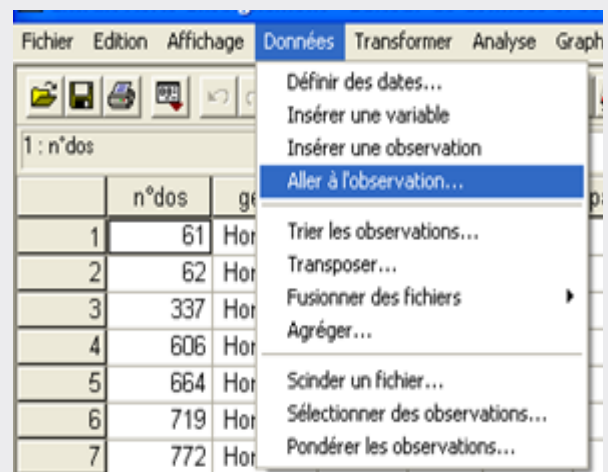
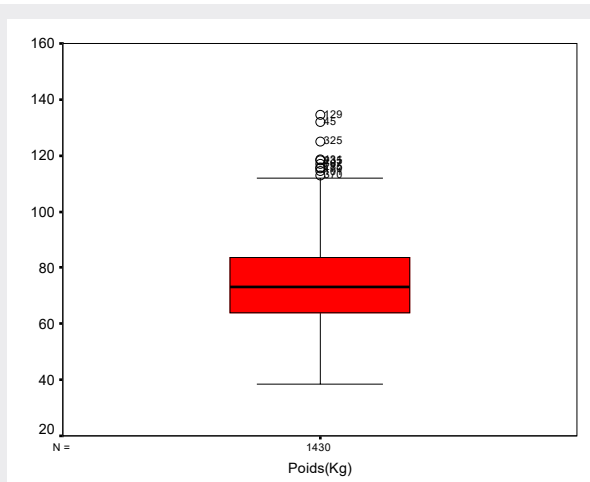


Figure 3. Détection des erreurs de saisi, selon la technique des valeurs exceptionnelles (lointaines ou aberrantes) du Box Plot d'une variable d'étude.

Cette détection pourrait se voir indirectement dans la représentation de la distribution des données d'une variable par la technique des «branches et feuilles» (figure n°4). Enfin, la phase préalable à l'analyse de la base de données (dont l'application des statistiques descriptives),

se termine par l'étude de la normalité des distributions des variables, soit directement, à travers le calcul du test de normalité d'une variable quantitative par le test de Kolmogorov-Smirnov (figure n°5), soit indirectement à travers la représentation graphique de la normalité de

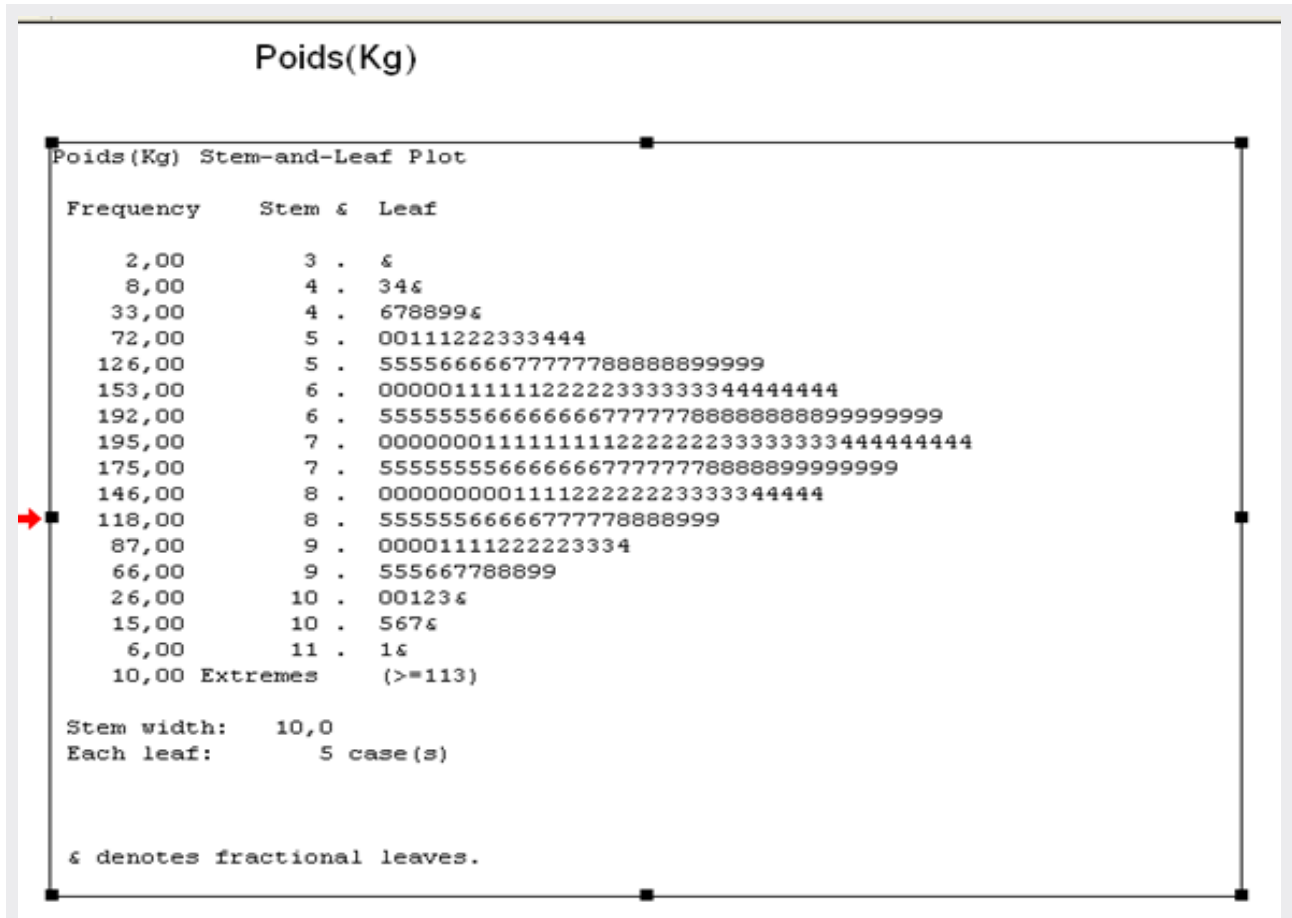


Figure 4. Exemple de la figure « Feuille et Branche » de la représentation d'une variable quantitative, par le logiciel SPSS

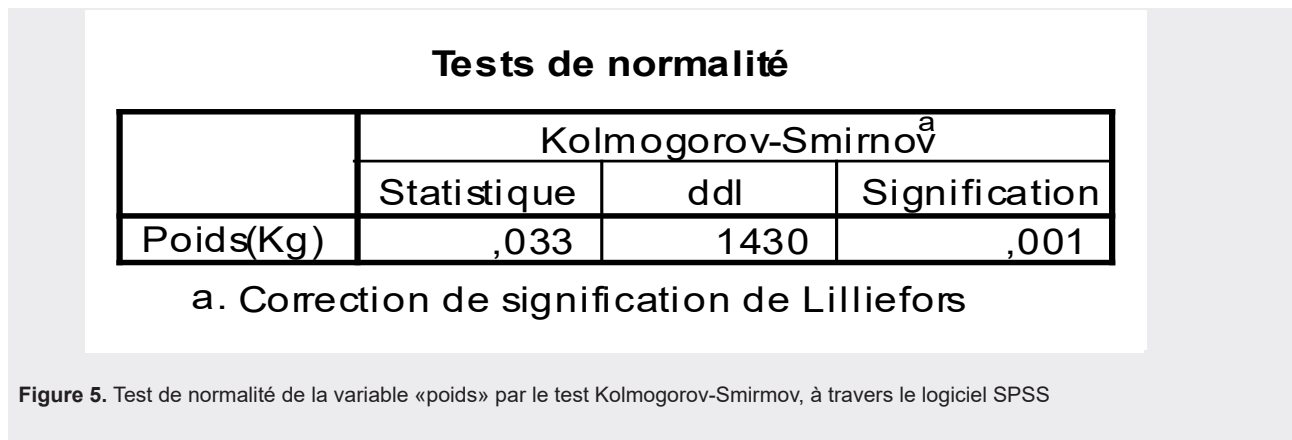


Figure 5. Test de normalité de la variable «poids» par le test Kolmogorov-Smirnov, à travers le logiciel SPSS

la distribution d'une variable et sa superposition avec la courbe gaussienne (figure n°6).

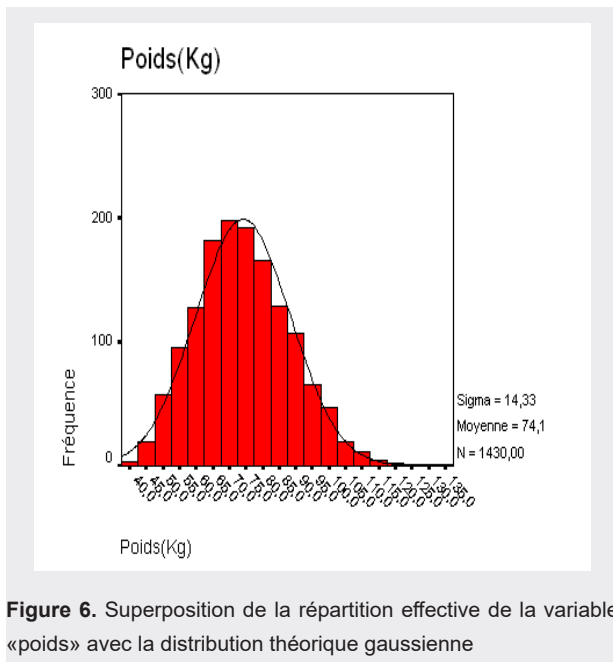


Figure 6. Superposition de la répartition effective de la variable « poids » avec la distribution théorique gaussienne

A. Regroupements / Représentations graphique et tabulaire

Au cours de la statistique descriptive, les données sont **regroupées** à l'aide d'une répartition en classes, en un nombre approprié (ni trop grand, ni trop petit), avec des limites en rapport avec le degré de précision (les intervalles sans limites sont à éviter), en préférant une amplitude égale, en assurant l'exclusivité des classes et en déterminant leurs valeurs centrales. Ainsi, les valeurs de la variable « âge » peuvent être regroupées dans une nouvelle variable des « classes d'âge ».

La représentation tabulaire est une méthode de présentation des données soit à l'aide du tableau des fréquences de distribution des fréquences et de la répartition en classes, soit à l'aide du tableau de contingence de la distribution des fréquences: absolue (effectif, nombre d'observations), relative (pourcentage, rapport entre l'effectif de la classe et l'effectif total) et cumulée (somme de la fréquence d'une classe et de toutes les classes qui la précèdent).

Dans un tableau de contingence, nous pouvons déterminer des pourcentages lignes (par rapport aux totaux des lignes), les pourcentages colonnes (par rapport aux totaux des colonnes), ou les pourcentages lignes et colonnes. (Figure 7)

			Genre Humain		Total
			Homme	Femme	
Classes d'âge	20 à 39 ans	Effectif	139	345	484
		% dans Genre Humain	28,9%	35,9%	33,6%
40 à 59 ans	Effectif	220	395	615	
	% dans Genre Humain	45,7%	41,1%	42,7%	
60 à 79 ans	Effectif	100	194	294	
	% dans Genre Humain	20,8%	20,2%	20,4%	
80 à 99 ans	Effectif	22	26	48	
	% dans Genre Humain	4,6%	2,7%	3,3%	
Total	Effectif	481	960	1441	
	% dans Genre Humain	100,0%	100,0%	100,0%	

Figure 7. Image du tableau de contingence de la variable sexe selon la classe d'âge, avec des pourcentages colonnes, selon le logiciel SPSS

Quant à la **représentation graphique**, elle a trois objectifs:

1. Visualiser la distribution des données, 2. Comparer plusieurs séries des données, 3. Choisir les paramètres de réduction. Les règles de la méthode graphique sont de donner un titre informatif à la figure, préciser les totaux généraux, expliciter les unités de mesure de la variable, marquer les axes en commençant par la valeur zéro et ne pas surcharger les graphiques, par des classes ou des catégories, pour plus de lisibilité. Le choix de graphique dépend de types des variables.

En ce qui concerne les variables quantitatives, nous pouvons sélectionner soit un **histogramme** (rectangles contigus, des aires proportionnelles aux fréquences avec une somme des aires=1), soit un **polygone de fréquence**. Quant aux variables qualitatives, nous optons souvent soit pour un **diagramme en bâtons**, soit pour un **diagramme en secteurs circulaires**.

B. Résumé statistique

Le résumé statistique d'une distribution des données est similaire à la lecture d'un cliché de radiographie de thorax ou d'un tableau de peinture. Il est basé sur le calcul des **statistiques de tendance centrale** (Moyenne, Médiane, Mode) et les **statistiques de dispersion** (Étendue, Quartiles, Variance, Ecart type et CV).

Statistiques de tendance centrale

La tendance centrale est calculée par le calcul des trois paramètres appelés les 3M (Moyenne, Médiane, Mode).

Moyenne. Dans les observations non groupées, la moyenne est la somme des observations / le nombre des enregistrements.

Dans les observations groupées, la moyenne est la somme de (valeur centrale de la classe X nombre des cas) / nombre des observations.

Médiane: Lorsque les valeurs d'une variable sont ordonnées, la médiane correspond à la valeur centrale. La médiane est la valeur de l'observation qui sépare une distribution en deux parties égales: 50 % inférieure et 50 % supérieure. La détermination de la médiane passe par l'identification de la *position* de la médiane: $(n+1)/2$. Contrairement à la MOYENNE, la MEDIANE est peu affectée par les valeurs extrêmes. Le Box Plot représente cinq paramètres dont la médiane, en plus des valeurs minimale et maximale, le premier (25^{ème} percentile) et le troisième (75^{ème} percentile) quartile.

Mode: Le mode d'un ensemble de valeurs correspond à la valeur de la variable, la plus fréquente. S'il n'y a pas de mode, lorsque toutes les valeurs sont différentes, il peut y avoir plusieurs modes dans une distribution. Le Mode est la valeur dominante. C'est la valeur de fréquence maximale, dans la distribution non groupée, et la *classe de fréquence maximale*, dans une distribution groupée.

Statistiques de dispersion

L'Étendue est définie par l'Écart entre les valeurs extrêmes. C'est la Valeur maximale – Valeur minimale (Maxi-Min).

Les quartiles. Un quartile est la valeur de la variable qui sépare une distribution en quatre segments égaux. Un quartile est le 25^{ème} percentile. On distingue: le premier quartile Q1 (0% - 25%), Le deuxième quartile Q2: (médiane) et le troisième quartile Q3 (75% - 100%). L'Ecart Inter Quartiles (EIQ) est la différence entre Q3 et Q1. Ces deux quartiles peuvent se mettre dans un Intervalle Inter Quartiles (IIQ).

La **variance** permet d'évaluer l'importance de la dispersion des valeurs de part et d'autre de la moyenne (écart entre chaque observation et la moyenne). C'est la $(\text{somme des écarts})^2 / n-1$ (n-1: degrés de liberté). L'Écart Type (ET), qui correspond à la racine carrée de la variance, permet d'exprimer la mesure de la dispersion dans les unités de mesure originales de la variable. **L'Écart type**, appelé aussi Déviation Standard (Standard Deviation: SD), est la racine carrée de la variance des observations. Quant au **Coefficient de Variation (CV)**, c'est l'écart type exprimé en pourcentage de la moyenne: $CV = (\text{Ecart Type} / \text{Moyenne}) \times 100$.

CONCLUSION

La description statistique d'une base des données relative à la distribution d'une série des variables relatives à un phénomène de santé, consiste au résumé de leur tendance centrale et de leur dispersion, successivement par les 3 M (Moyenne, Médiane, Mode) et les 3 E (Étendue, Ecart Type, Ecart inter quartiles), via des logiciels appropriés dont la plus utilisés en sciences médico-sociales: SPSS. Cette technique statistique est à différentier du raisonnement statistique qui est un outil d'aide au développement d'une attitude critique à l'égard de l'information. Ce raisonnement consiste à s'inquiéter de la fiabilité de la collecte des données et de quantifier la probabilité, due au hasard, dans son explication.

RÉPONSES AUX QUIZZ

1. L'âge moyen de la population étudiée a été de 42 ans, ce qui signifie :
 - a. La sommes des attributions de la variable «âge» des patients inclus / 1897 patients de l'étude
2. Dans cette étude, l'ET de 20,5 ans, mentionné après la moyenne est:
 - a. L'Ecart Type de la variable «âge» des patients
 - b. La Déviation Standard (*Standard Deviation*: DS) de la variable «âge » des patients
 - c. Un estimateur de la distribution de la variable «âge » des patients, par rapport à sa moyenne
 - d. Un estimateur de la dispersion de la variable âge par rapport à sa moyenne de 42 ans.
3. Avec un âge moyen de 42,4 ans (ET = 20,5 ans), le Coefficient de Variation (CV) est :
 - a. 48,35%
4. Dans une distribution statistique, la médiane est:
 - a. Unique
 - b. Peu influencée par les valeurs extrêmes
 - c. Plus informative sur la qualité de dispersion
 - d. Illustrée par la figure de Box Plot
5. Dans la figure de Box Plot, les paramètres suivants d'une distribution sont illustrés:

Aucun item

RÉFÉRENCES

1. Zemni I, Safer M, Khelil M, Kacem M, Zoghliami C, Ben Abdelaziz A. Patient satisfaction in a tertiary care center (Tunisia, 2015-2016). *Tunis Med* 2018; 96 (10): 737–45.

2. Picat M.-Q, Savès M, Asselineau J, Dumoulin M, Coureau G, Salmi L.-R et al. Méthodes biostatistiques et épidémiologiques employées pour la recherche biomédicale: implications pour la formation médicale initiale. Rev Epidemiol Sante Publique 2013; 61: 261-268.