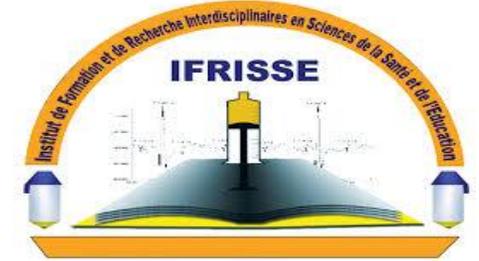


MODULE SPSS



Année académique_2021-2022

KABORE T.A. Geraude

Démographe

Mail: kaboralimata@gmail.com

Tel: (+226) 73382412/ (+226) 66757694

OUEDRAOGO Houdou

Statisticien Economiste

Mail: houdbarack@gmail.com

Tel: (+226) 70129773 /

COULIBALI Moussa

Démographe

Mail: coulibalymoussa330@gmail.com

Tel: (+226) 72555573 / (+226) 74066271

Analyse bivariée

Plan

Introduction

Mesures d'association entre deux variables qualitatives (test de χ^2)

Mesures d'association entre deux variables quantitatives (recours au coefficient de corrélation Pearson)

Introduction 1/2

Les leçons précédentes traitaient de la statistique descriptive univariée, c'est-à-dire de la description d'une série statistique selon un seul caractère. En plus de l'étude séparée de chaque caractère, on peut mesurer les liens existant entre les variables deux à deux : c'est l'objet de la statistique descriptive bivariée.

Considérons, par exemple, les caractéristiques suivantes des enfants Burkinabés âgés de moins de cinq ans et leur mère ayant participé à l'Enquête sur les indicateurs du paludisme (EIP) réalisée en 2014 : La région de résidence, le milieu de résidence, groupe ethnique, religion, niveau de vie du ménage, niveau d'instruction de la femme, exposition aux messages sur le paludisme, connaissance du paludisme, utilisation de la moustiquaire imprégnée, assainissement du ménage, Âge de l'enfant.

Introduction 2/2

La morbidité palustre des enfants varie-t-elle significativement selon la nature du milieu de résidence ? Varie-t-elle significativement selon le niveau d'instruction de la femme ? Cette dernière est-elle associée à l'utilisation de la MII (Moustiquaire imprégnée d'insecticide)? etc,

C'est à ces genres de questions qu'on apportera des éléments de réponse, en recourant aux méthodes statistiques bivariées. L'analyse de la relation entre deux variables X et Y dépend entre autres de leur nature (quantitative ou qualitative).

Mesures d'association entre deux variables qualitatives (test de chi2)

Le chi-deux renseigne sur le degré d'association entre deux variables qualitatives et non sur la manière dont diffèrent les catégories. On peut le calculer aussi pour des variables quantitatives rendues catégorielles. Le V de Cramer nous indique la force de la relation

Exemple: Existe-t-il un lien statistique entre le paludisme et le milieu de résidence ?

Quelle est la force de cette relation ?

Applications sous SPSS (Passant par la boîte de dialogue)

Étape 1: on fait analyse → statistiques descriptives → tableaux croisés

The screenshot shows the IBM SPSS Statistics interface. The 'Analyse' menu is open, and the path 'Statistiques descriptives' > 'Tableaux croisés' is highlighted. The background shows a list of variables with columns for 'Nom', 'Type', 'Largeur', and 'Définition'. The 'Tableaux croisés' dialog box is partially visible in the background, showing options for 'Align', 'Mesure', and 'Rôle'.

Nom	Type	Largeur	Définition
1	hhid	Chaîne	351 0
2	hvidx	Numérique	2 0
3	hv000	Chaîne	24 0
4	hv001	Numérique	3 0
5	hv002	Numérique	5 0
6	hv003	Numérique	2 0
7	hv004	Numérique	3 0
8	hv005	Numérique	7 0
9	hv006	Numérique	2 0
10	hv007	Numérique	4 0
11	hv008	Numérique	4 0
12	hv008a	Numérique	5 0
13	hv009	Numérique	2 0
14	hv010	Numérique	2 0
15	hv011	Numérique	1 0
16	hv012	Numérique	2 0
17	hv013	Numérique	2 0
18	hv014	Numérique	1 0
19	hv015	Numérique	1 0
20	hv016	Numérique	2 0
21	hv017	Numérique	1 0
22	hv018	Numérique	3 0
23	hv019	Numérique	2 0
24	hv020	Numérique	1 0
25	hv021	Numérique	2 0

Étape 2: on fait entrer la variable dépendante en ligne et la variable explicative en colonne

The screenshot shows the IBM SPSS Statistics interface. The main window displays a list of variables with the following columns: Nom, Type, Largeur, Décimales, Libellé, Valeurs, Manquant, Colonnes, Align, Mesure, and Rôle. The 'Tableaux croisés' dialog box is open, showing the following configuration:

- Ligne(s):** paludisme
- Colonne(s):** milres
- Couche 1 de 1:** (Empty)
- Afficher les variables de couche dans les couches du tableau:**
- Afficher les graphiques à barres en cluster:**
- Supprimer les tableaux:**

The dialog box also includes buttons for 'Exact...', 'Statistiques...', 'Cellules...', 'Format...', 'Style...', 'Bootstrap...', 'OK', 'Coller', 'Réinitialiser', 'Annuler', and 'Aide'.

Étape 3: on clique sur statistique → on choisit la statistique khi-carré ensuite on fait poursuivre

Ensemble des statistiques

The screenshot displays the IBM SPSS Statistics interface. The main window shows a list of variables with columns for 'Nom', 'Type', 'Largeur', 'Décimales', 'Libellé', 'Valeurs', 'Manquant', 'Colonnes', 'Align', 'Mesure', and 'Rôle'. A dialog box titled 'Tableaux croisés : Statistiques' is open, showing a list of variables on the left and a list of statistical tests on the right. The 'Khi-carré' checkbox is checked and circled in red. Other tests include 'Corrélations', 'Nominales', 'Ordinales', 'Données nominales / intervalle', and 'Statistiques de Cochran et de Mantel-Haenszel'. The 'Poursuivre' button is visible at the bottom of the dialog box.

Nom	Type	Largeur	Décimales	Libellé	Valeurs	Manquant	Colonnes	Align	Mesure	Rôle	
1	hhid	Chaîne	351	0	case identificati...	Aucun	Aucun	26	Gauche	Nominales	Entrée
2	hvidx	Numérique	2	0	line number	Aucun	Aucun	7	Droite	Nominales	Entrée
3	hv000	Chaîne	24								
4	hv001	Numérique	3								
5	hv002	Numérique	5								
6	hv003	Numérique	2								
7	hv004	Numérique	3								
8	hv005	Numérique	7								
9	hv006	Numérique	2								
10	hv007	Numérique	4								
11	hv008	Numérique	4								
12	hv008a	Numérique	5								
13	hv009	Numérique	2								
14	hv010	Numérique	2								
15	hv011	Numérique	1								
16	hv012	Numérique	2								
17	hv013	Numérique	2								
18	hv014	Numérique	1								
19	hv015	Numérique	1								
20	hv016	Numérique	2								
21	hv017	Numérique	1								
22	hv018	Numérique	3	0	interviewer iden...	Aucun	Aucun	7	Droite	Echelle	Entrée
23	hv019	Numérique	2	0	keyer identificat...	Aucun	Aucun	7	Droite	Nominales	Entrée
24	hv020	Numérique	1	0	ever-married sa...	{0, 0. all wo...	Aucun	7	Droite	Nominales	Entrée
25	hv021	Numérique	2	0	primary care...	Aucun	Aucun	7	Droite	Echelle	Entrée

Étape 4: on fait OK puis on obtient les résultats

The screenshot shows the IBM SPSS Statistics interface with the 'Tableaux croisés' (Crosstabs) dialog box open. The dialog box is positioned over a list of variables. The 'Ligne(s)' field contains 'paludisme' and the 'Colonne(s)' field contains 'milres'. The 'OK' button is highlighted with a red arrow and the word 'Clic'.

Nom	Type	Largeur	Décimales	Libellé	Valeurs	Manquant	Colonnes	Align	Mesure	Rôle	
1	hhid	Chaîne	351	0	case identificati...	Aucun	Aucun	26	Gauche	Nominales	Entrée
2	hvidx	Numérique	2	0	line number	Aucun	Aucun	7	Droite	Nominales	Entrée
3	hv000	Chaîne	24								
4	hv001	Numérique	3								
5	hv002	Numérique	5								
6	hv003	Numérique	2								
7	hv004	Numérique	3								
8	hv005	Numérique	7								
9	hv006	Numérique	2								
10	hv007	Numérique	4								
11	hv008	Numérique	4								
12	hv008a	Numérique	5								
13	hv009	Numérique	2								
14	hv010	Numérique	2								
15	hv011	Numérique	1								
16	hv012	Numérique	2								
17	hv013	Numérique	2								
18	hv014	Numérique	1								
19	hv015	Numérique	1								
20	hv016	Numérique	2								
21	hv017	Numérique	1								
22	hv018	Numérique	3	0	interviewer iden...	Aucun	Aucun	7	Droite	Echelle	Entrée
23	hv019	Numérique	2	0	keyer identificat...	Aucun	Aucun	7	Droite	Nominales	Entrée
24	hv020	Numérique	1	0	ever-married sa... {0, 0. all wo...	Aucun	Aucun	7	Droite	Nominales	Entrée
25	hv021	Numérique	2	0	primaries complet...	Aucun	Aucun	7	Droite	Echelle	Entrée

Tableaux croisés

Ligne(s): paludisme

Colonne(s): milres

Couche 1 de 1

Afficher les graphiques à barres en cluster

Supprimer les tableaux

OK Cojler Réinitialiser Annuler Aide

Clic

Résultats

NB : En sciences sociales, on accepte généralement un pourcentage d'erreur maximum de 10%, 5% ou de 1% selon les objectifs de l'étude et le degré de certitude ou de précision requis. C'est pourquoi on compare le p-value (significativité) à l'un des seuils suivants : 0,10 , 0,05 ou 0,01.

Le test de khi deux nous montre que le milieu de résidence est significativement associé à la morbidité palustre chez les enfants de moins de cinq ans au seuil de 1%.

Résultats

*Sortie1 [Document1] - IBM SPSS Statistics Viewer

Fichier Edition Affichage Données Transformer Insérer Format Analyse Graphiques Utilitaires Extensions Fenêtre Aide



- Sortie
 - Log
 - Tableaux croisés
 - Titre
 - Remarques
 - Jeu de données a
 - Récapitulatif de tr
 - Tableau croisé pr
 - Tests du khi-carré

prevalence du paludisme	5174	100,0%	0	0,0%	5174	100,0%
* milieu de residence						

Tableau croisé prevalence du paludisme * milieu de residence

Effectif

		milieu de residence		Total
		1. urbain	2. rural	
prevalence du paludisme	non impaludés	740	3547	4287
	impaludés	46	841	887
Total		786	4388	5174

Tests du khi-carré

	Valeur	ddl	Signification asymptotique (bilatérale)	Sig. exacte (bilatérale)	Sig. exacte (unilatérale)
khi-carré de Pearson	83,181 ^a	1	,000		
Correction pour continuité ^b	82,246	1	,000		
Rapport de vraisemblance	102,461	1	,000		
Test exact de Fisher				,000	,000
Association linéaire par linéaire	83,165	1	,000		
N d'observations valides	5174				

- a. 0 cellules (0,0%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de 134,75.
- b. Calculée uniquement pour une table 2x2

Le processeur IBM SPSS Statistics est prêt Unicode:ON

Interprétation du khi carré et du V de cramer

Le khi deux nous indique qu'il existe une liaison entre les deux variables.

Le V de cramer nous indique que cette liaison est très forte par sa valeur élevée.

En d'autre terme, plus le V de cramer est proche de zéro, plus il y a indépendance entre les deux variables étudiées. Il vaut 1 en cas de complète dépendance.

Tableau A2 : Grille d'évaluation du V de cramer

Valeur du coefficient de cramer	Intensité du lien
0,000 – 0,045	Très faible
0,045 – 0,090	Faible
0,090 – 0,180	Moyenne
0,180 – 0,360	Forte
0,360 – 1,00	Très forte

Source : Amyotte L., 2011

Toujours en passant par la boîte de dialogue,

On peut également faire un tableau croisé pour obtenir les distributions en fréquence absolue ou relative en colonne ou en ligne. Pour ce faire, on fait analyse → statistiques descriptives → tableaux croisés → on fait entrer la variable dépendante en ligne et la variable explicative en colonne → on clique sur cellule → on choisit le type de pourcentage → on fait poursuivre puis ok et on obtient les résultats

Feuille syntaxe

Les commandes permettant de faire le khi-deux

Test d'indépendance de Khi-deux (χ^2) + V de Cramer: **Crosstabs paludisme by milres /statistics = chisq phi.**

Test d'indépendance de Khi-deux (χ^2) + test de proportion + V de Cramer: **Crosstabs paludisme by milress /cells=prop row /statistics = chisq phi.**

Feuille syntaxe

Les commandes nous permettant d'obtenir les distributions sont :

crosstabs tables=paludisme by milres /cells=total. (% du total)

crosstabs tables=paludisme by milres /cells=row. (% en ligne)

crosstabs tables= paludisme by milres /cells=column. (% en colonne)

Exercice: le paludisme est-il associé à la région de résidence, aux groupes ethniques, à la religion, au niveau de vie du ménage, au niveau d'instruction de la femme, à l'exposition de la femme aux messages sur le paludisme, aux connaissances du paludisme, à l'utilisation de la moustiquaire imprégnée, à l'assainissement du ménage, à l'âge de l'enfant ?

Préciser le seuil de significativité de l'association ainsi que la proportion des impaludés pour chaque modalité de la variable

NB: la variable dépendante dans la base est « paludisme »

Mesures d'association entre deux variables quantitatives (recours au coefficient de corrélation Pearson)

Le coefficient de corrélation linéaire de Pearson mesure à la fois l'intensité de la relation linéaire entre Y et X et celle entre X et Y . Le coefficient de corrélation est une mesure symétrique qui varie entre -1 et 1 . Il prend la valeur 0 s'il n'y a pas de relation linéaire entre les deux variables. Il est important de se souvenir que le coefficient de corrélation est une mesure linéaire. Si l'absence de relation entre deux variables donne un coefficient de corrélation nul, un coefficient de corrélation nul ne signifie pas qu'il n'y a pas de relation entre les deux variables. Il signifie seulement une absence de relation linéaire.

Existe-t-il un lien statistique entre le nombre de moustiquaire dans le ménage et le nombre de personne dans le ménage ?

PAR CONVENTION

- Corrélation=0 absence de lien.
- Corrélation < 0.25 faible.
- Corrélation < 0.50 moyen.
- Corrélation < 0.75 élevé.
- Corrélation < 1.0

Applications sous SPSS (en passant par la boîte de dialogue)

Étape 1: on fait analyse → statistiques descriptives → tableaux croisés

The screenshot shows the IBM SPSS Statistics interface. The 'Analyse' menu is open, and the path 'Statistiques descriptives' > 'Tableaux croisés' is highlighted. The background shows a data list with columns 'Nom', 'Type', 'Largeur', and 'Définition'. The 'Tableaux croisés' dialog box is partially visible, showing options for 'Align' (Gauche, Droite) and 'Mesure' (Nominales, Echelle). The system tray at the bottom shows the date 05/04/2021 and time 20:38.

Nom	Type	Largeur	Définition
1	hhid	Chaîne	351
2	hvidx	Numérique	2
3	hv000	Chaîne	24
4	hv001	Numérique	3
5	hv002	Numérique	5
6	hv003	Numérique	2
7	hv004	Numérique	3
8	hv005	Numérique	7
9	hv006	Numérique	2
10	hv007	Numérique	4
11	hv008	Numérique	4
12	hv008a	Numérique	5
13	hv009	Numérique	2
14	hv010	Numérique	2
15	hv011	Numérique	1
16	hv012	Numérique	2
17	hv013	Numérique	2
18	hv014	Numérique	1
19	hv015	Numérique	1
20	hv016	Numérique	2
21	hv017	Numérique	1
22	hv018	Numérique	3
23	hv019	Numérique	2
24	hv020	Numérique	1

Étape 3: on clique sur statistique → on choisit la statistique corrélation ensuite on fait poursuivre

Ensemble des statistiques

The screenshot shows the IBM SPSS Statistics interface. The main window displays a data table with columns: Nom, Type, Largeur, Décimales, Libellé, Valeurs, Manquant, Colonnes, Align, Mesure, and Rôle. A dialog box titled 'Tableaux croisés : Statistiques' is open, showing a list of variables on the left and a list of statistical tests on the right. The 'Corrélations' checkbox is checked. The 'Nom' column in the background table lists variables like hv001, hv002, hv003, hv005, hv009, hv024, hv025, b16, âge_mère, palu, use_mosquito, nettoyage_environ, milres, proprfemenviron, PropEnftUtilMil, ethnie, taille_menage, education_mere, message_palu, conaispalu, age_enfant, sexe_enfant, religion_mere, v005, and wealthindex_combined.

Nom	Type	Largeur	Décimales	Libellé	Valeurs	Manquant	Colonnes	Align	Mesure	Rôle
1 hv001	Numérique	3	0	cluster number	Aucun	Aucun	7	☰ Droite	Echelle	Entrée
2 hv002	Numérique	5	0	household number	Aucun	Aucun	7	☰ Droite	Echelle	Entrée
3 hv003	Numérique							☰ Droite	Nominales	Entrée
4 hv005	Numérique							☰ Droite	Echelle	Entrée
5 hv009	Numérique							☰ Droite	Echelle	Entrée
6 hv024	Numérique							☰ Droite	Nominales	Entrée
7 hv025	Numérique							☰ Droite	Nominales	Entrée
8 b16	Numérique							☰ Droite	Nominales	Entrée
9 âge_mère	Numérique							☰ Droite	Nominales	Entrée
10 palu	Numérique							☰ Droite	Nominales	Entrée
11 use_mosquito	Numérique							☰ Droite	Nominales	Entrée
12 nettoyage_environ	Numérique							☰ Droite	Nominales	Entrée
13 milres	Numérique							☰ Droite	Nominales	Entrée
14 proprfemenviron	Numérique							☰ Droite	Nominales	Entrée
15 PropEnftUtilMil	Numérique							☰ Droite	Nominales	Entrée
16 ethnie	Numérique							☰ Droite	Nominales	Entrée
17 taille_menage	Numérique							☰ Droite	Nominales	Entrée
18 education_mere	Numérique							☰ Droite	Nominales	Entrée
19 message_palu	Numérique							☰ Droite	Nominales	Entrée
20 conaispalu	Numérique							☰ Droite	Nominales	Entrée
21 age_enfant	Numérique							☰ Droite	Nominales	Entrée
22 sexe_enfant	Numérique	1	0	sexe de l'enfant	{1, 1. masc...	Aucun	13	☰ Droite	Nominales	Entrée
23 religion_mere	Numérique	1	0	religion de la mère	{1, 1. musul...	Aucun	15	☰ Droite	Nominales	Entrée
24 v005	Numérique	7	0	women's individual sample weight (6 decimals)	Aucun	Aucun	9	☰ Droite	Echelle	Entrée
25 wealthindex_combined	Numérique	1	0	wealth index combined	{1, 1. poor...	Aucun	6	☰ Droite	Nominales	Entrée

En passant par la boîte de dialogue,

On peut également faire un tableau croisé pour obtenir les distributions en fréquence absolue ou relative en colonne ou en ligne. Pour ce faire, on fait analyse → statistiques descriptives → tableaux croisés → on fait entrer la variable dépendante en ligne et la variable explicative en colonne → on clique sur cellule → on choisit le type de pourcentage → on fait poursuivre puis ok et on obtient les résultats

Feuille syntaxe

La commande nous permettant d'obtenir le test de corrélation

Syntaxe : crosstabs tables= nombre_MIII1 by taille_menage1 /statistics=corr.

Les commandes nous permettant d'obtenir les distributions sont :

crosstabs tables=paludisme by milres /cells=total. (% du total)

crosstabs tables=paludisme by milres /cells=row. (% en ligne)

crosstabs tables= paludisme by milres /cells=column. (% en colonne)