Statistiques

Généralités :

1. Population : univers statistique (représenté par **Ω**) : ensembles étudiés par la statistique (45 clowns ont participé à une étude sur leur âge et leur taille) -> des clowns
2. Individus : unités statistiques (représentés par **ω**) : éléments des ensembles (45 clowns ont participé à une étude sur leur âge et leur taille) -> 45
3. Recensement : étude de tous les individus d'une population finie
4. Sondage : observation d'une partie de la population
5. Echantillon (représenter par **E**) : le sous ensemble étudié, lors d'un sondage (**E** $\subset$ **Ω**)
6. Observations : nbr d'unités statistiques x nbr de variables (45 clowns ont participé à une étude sur leur âge et leur taille) -> 2 x 45 = 90
7. Variables, caractères : ensemble de caractéristiques décrivant les individus d'une population. (45 clowns ont participé à une étude sur leur âge et leur taille) -> âge, taille

Types de variables statistiques :

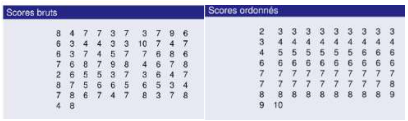
1. Quantitatives (nombres, chiffres, etc…) -> numériques **ATTENTION** (date de naissance)
   - Continues, assimilées : mesure, **quantité** (poids, taille, R = nombres réels)
   - Discrètes : comptage (années, N = entiers naturels, 0 1 2 3 4 5 6 …)
2. Qualitatives (mots, expressions, etc…)
   - Nominale : aucun ordre (nom, prénom, code postal, adresse, tél, email …)
   - Ordinale : qui peut se classer (un peu, bcp, moyen, entre 400 et 500, plus de 500, moins de 400)

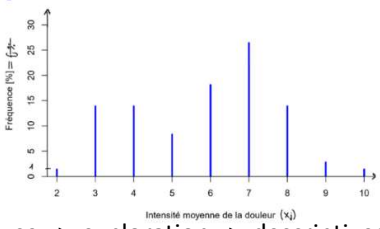
Ex : Votre genre de BD préféré : (a) humour, (b), (c), (d), (e), (f) classer de 1-6 (1=préféré)

- Préférence : (a) rang attribué aux BD d'humour, modalités : 1/2/3/4/5/6, type : numérique (b)… (c)… (d)…

Diagramme en bâtons (variable quantitative/numérique discrète)

1. Tableau des scores bruts (tous les chiffres mélangés sous forme de tableau)
2. Tableau des scores ordonnés (mettre les chiffres dans l'ordre)
3. Tableau des données regroupées
   - ⇨ J = simplement les données
   - ⇨ Xj = intensité de la douleur
   - ⇨ n = la somme des effectifs (72)
   - ⇨ f = fréquence
   - ⇨ nj = effectifs (nbr de fois)
4. Jamovi : entrer les données -> analyse -> exploration -> descriptives -> sélectionner la variable -> cocher Frequency tables -> cocher N et Missing





	A	B	C	D	E	F	G	H	I	J
1	j	1	2	3	4	5	6	7	8	9
2	xj	2	3	4	5	6	7	8	9	10
3	nj	1	10	10	6	13	19	10	2	1
4	fj	1/72	10/72	10/72	6/72	13/72	19/72	10/72	2/72	1/72
5	fj (%)	1,4	13,9	13,9	8,3	18,1	26.4	13.9	2,8	1,4

Histogramme (variable quantitative continue/assimilée)

1. Estimation du nbr de classes K

$$K = \lceil \log_2 n \rceil + 1 = \lceil \log_2 200 \rceil + 1 = \lceil 7.644 \rceil + 1 = 8 + 1 = 9$$

(nbr de résultats ; à arrondir)

2. Calcul de l'empan

$$Empan = \max(X) - \min(X) = 140 - 62 = 78$$

3. Calcul de la largeur des classes a_j

$$\frac{Empan}{K} = 8.667$$

$\log_2(n) = \frac{\log(n)}{\log(2)}$

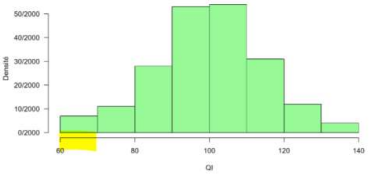
4. Ajustement de la largeur des classes a_j

Comme $5 < 8.667 < 10$, nous donnerons finalement à chaque classe j une largeur égale à $a_j = 10$.

5. Tableau des données regroupées en classe
6. Historigramme
7. Historigramme au tableau des données

j	1	2	3	4
$[e_{j-1};\ e_j[$	[60; 70[	[70; 80[	[80; 90[	[90; 100[
a_j	10	10	10	10
n_j	7	11	28	53
f_j	0.0350	0.0550	0.1400	0.2650
d_j	0.0035	0.0055	0.0140	0.0265

j	5	6	7	8
$[e_{j-1};\ e_j[$	[100; 110[	[110; 120[	[120; 130[	[130; 140]
a_j	10	10	10	10
n_j	54	31	12	4
f_j	0.2700	0.1550	0.0600	0.0200
d_j	0.0270	0.0155	0.0060	0.0020



i	1	2	3	4
$[e_{i-1};\ e_i[$	[350; 450[	[450; 500[	[500; 700[	[700; 850]
c_i	400	475	600	775
a_i	100	50	200	150
d_i	0.0039	0.0046	0.0016	0.0004
f_i	0.39	0.23	0.32	0.06
n_i	390	230	320	60

150·0,0004

0,06 /1000 → car n=1000

Reconstituez, à partir de l'histogramme représenté ci-dessous, le tableau des données regroupées en classes ($n = 1000$).

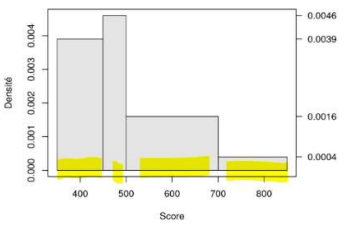


Diagramme en tuyau d'orgue (variable qualitative)

1. Tableau des effectifs
2. Jamovi : analyses -> exploration -> descriptives -> plots -> bar plot
   - La largeur des tuyau n'a pas d'importance (similaires)

Localisation	tête (j=1)	ventre (j=2)	dos (j=3)	autre (j=4)
n_j	18	18	31	5
$f_j = n_j/n$	25%	25%	43,1%	6,9%

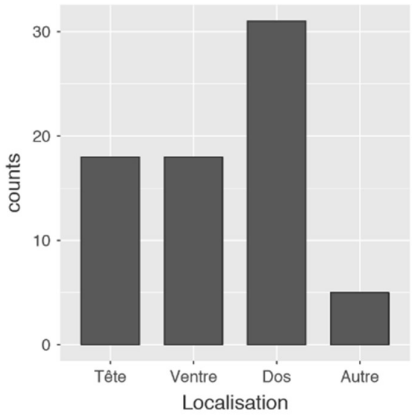
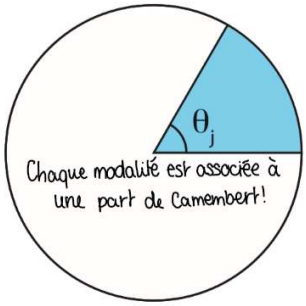


Diagramme en Camembert

	Sexe		
	Fille	Garçon	Total
n_j	59	13	72
f_j	81.9%	18.1%	100%
θ_j	295°	65°	360°

$\theta_{fille} = 360° \times 0,819 = 295°$

$\theta_{garçon} = 360° \times 0,181 = 65°$



$$\theta_j = 360° \times f_j$$

La Médiane (M)

- **JAMOVI :** Analyses -> explorations -> descriptives -> statistiques -> median -> cocher N
- Mettre les résultats dans l'ordre et prendre le résultat du centre (si impair)
- Attention si pair : prendre les deux au centre et faire la moyenne des deux $\frac{X_1 \cdot X_2}{2}$

La Moyenne (⊠) Mean

- Multiplication de tous les scores entre eux, puis diviser par le nbr total de scores multipliés
- **JAMOVI : :** Analyses -> explorations -> descriptives -> statistiques -> mean -> cocher N

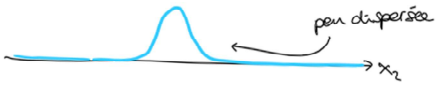
Le Mode (Mo)

- La valeur la + fréquente de la recherche -> exemple :
- Certaines distributions peuvent avoir plusieurs modes, ou pas de mode du tout (tous pareils)

Score	1	2	3	4	5	6	7
Effectif	1	1	2	3	6	5	1

parce que c'est là qu'il y en a le +

Caractéristiques de dispertion



Résumé

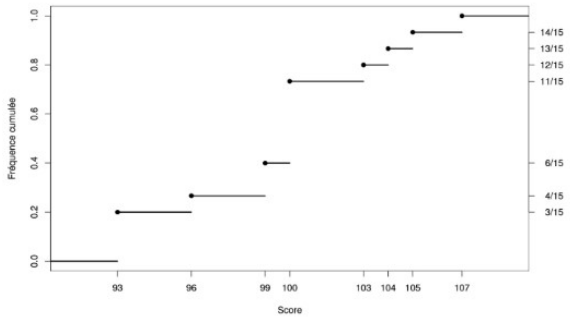
	mode	médiane	moyenne
nominale			
ordinale	X	X	
numérique	X	X	X

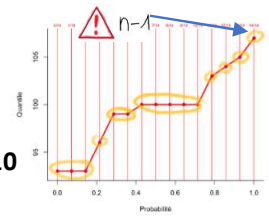
Fonction Quantile (à partir d'une fonction de répartition et de n)

1. Fonction de répartition F(x)

- Chaque point = x_j
- Espace (vertical) entre chaque trait = n_j

j	1	2	3	4	5	6	7	8
x_j	93	96	99	100	103	104	105	107
n_j	3	1	2	5	1	1	1	1
f_j	3/15	1/15	2/15	5/15	1/15	1/15	1/15	1/15
F_j	3/16	4/15	6/15	11/15	12/15	13/15	14/15	15/15



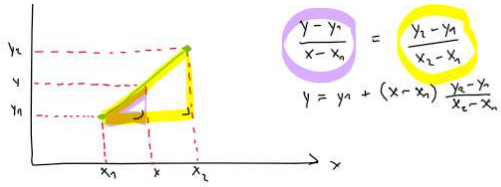


2. Fonction Quantile $F^{-1}(\alpha)$

- Déterminer les quartiles : Q1 = 0,25 / Q2 = 0,5 = Médiane /Q3= 0,75 -> **divided per 4**
- Déterminer les déciles : D1= 0,1 / 0,2 / 0,3 / 0,4 / 0,5/ 0,6 / 0,7 / 0,8/ 0,9 -> **DIvided per 10**
- Déterminer les centiles : C1= 0,01 / C50 = 0,5 = Médiane -> **divided per 100**
- **Jamovi :** Entrer les données, (dans une colonne) -> Mettre variable (continous) -> Analyses -> exploration -> Descriptives -> variable (la colonne)

➔ Cocher : Median, Mode, Mean, Min, Max, Box Plot, Frequency tables

3. (Extrapolation linéaire)



Intervalle interquartile (IQR ou IIQ)

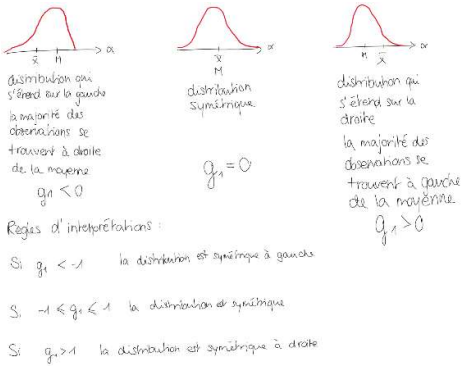
IIQ = Q3 – Q1

Boîte à moustache

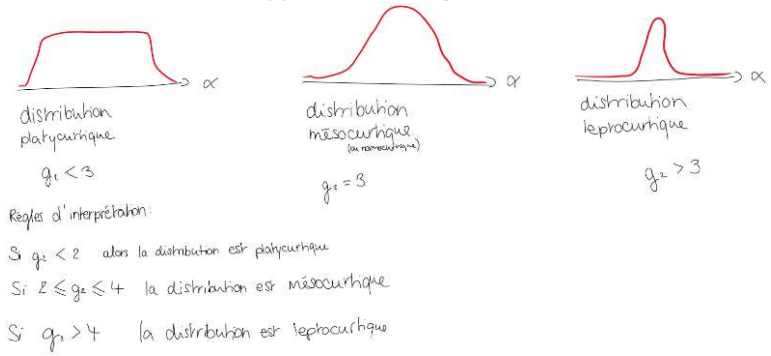
Adjacente inf : plus petite valeur observée supérieure ou égale à Q1 –1,5 (Q3-Q1)
Adjacente sup : plus grande valeur observée inférieur ou égale à Q3 + 1,5 (Q3-Q1)
Ecart type/Standart deviation (s) : Analyse -> exploration -> descriptives -> statistics -> …
Variance (s^2) -> Les valeurs qui augmentent le plus la variante sont les valeurs qui sont les plus éloignées de la moyenne
Données extrêmes : point au-dessus de la moustache

Caractéristiques de forme

Skewness /coeficient d'asymétrie (g1)



Kurtosis /coeficient d'applatissement (g2)



$$g_1 = \frac{m_3}{S^3} \qquad m_3 = (x_i - \bar{x})^3 + (x_i - \bar{x})^3 \ldots$$

$$g_2 = \frac{m_4}{S^4} \qquad m_4 = (x_i - \bar{x})^4 + (x_i - \bar{x})^4 \ldots$$

	Condition	Réponse	Effectif
1	Son	Juste	10
2	Son	Faux	28
3	Vidéo	Juste	17
4	Vidéo	Faux	18
5	Son + Vidéo	Juste	9
6	Son + Vidéo	Faux	23

Identification du gagnant	Son	vidéo	Son + vidéo	
oui	10	17	9	36
non	28	18	23	69
	38	35	32	105

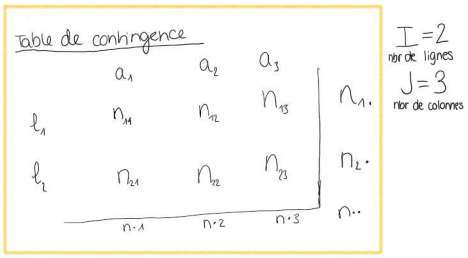


Table de contingence = n_{ij} :

1. Saisir les données
2. Chemin = Analyse – Frequencies – Contingency Tables – Independent Samples
3. Remplir -> Rows (Réponse) Columns (Condition) Counts (Effectif)

Distribution conditionnelle de Y :

1. Faire table de contingence
2. Cocher « Column » (sous Cells – Percentages)

⇨ Si les distributions conditionnelles de Y sont toutes les mêmes = Y ne dépend pas de X, sinon Y dépend de X

Distribution conditionnelle de X :

1. Faire table de contingence
2. Cocher « Row » (sous Cells – Percentages)

⇨ Si les distributions conditionnelles de X sont toutes les mêmes = X ne dépend pas de Y, sinon X dépend de Y

Indépendance des variables X et Y :

a. Les distributions conditionnelles de Y sont les mêmes
b. Les distributions conditionnelles de X sont les mêmes
c. L'effectif associé à (i ; j) = $\frac{n_{i.} \times n_{.j}}{n}$
d. La fréquence associée à la cellule (i ; j) = $f_{ij} = f_i \times f_j$ (Rappel : $f_i = n_i/n$ et $f_j = n_j/n$)

Tableau des effectifs théoriques = e_{ij} (correspond à la situation d'indépendance) :

1. Reprendre les données de la table de contingence (même chemin, remplissage)
2. Cocher « Expected » (sous Cells – Counts)

Tableau des résidus :

1. Permet de comparer le tableau des effectifs théoriques et la table de contingence
2. Contingence (n_{ij}) – Effectif(e^{ij}) -> (10 – 13 = -3)

Tableau des résidus standardisés :

1. ----------→ $\frac{n_{ij} - e_{ij}}{\sqrt{e_{ij}}}$

⇨ Intéressant si la valeur est supérieure ou égale à 2

Règle d'interprétation de Cohen : ----------→

taille d'effet

df^*	small	medium	large
1	.10	.30	.50
2	.07	.21	.35
3	.06	.17	.29
4	.05	.15	.25
5	.04	.13	.22

Valeur du Khi carré (distance entre n_{ij} et e_{ij}):

1. Après avoir fait une table de contingence et un tableau des effectifs théoriques
2. Cocher « X^2 » (sous Statistics)
3. Chiffre recherché = croisement X^2 et Value

⇨ X^2_{max} = n [min (I ; J) -1] => 105 [min (2; 3) -1] => 105 [2 -1] => 105 [1] = 105 (dépendance$_{max}$ fonctionnelle)

⇨ X^2_{min} = 0 (situation d' indépendance)

Coefficient de contingence et V de Cramer :

⇨ Indice ne dépendant plus ni de « n » ni de la dimension de la table de contingence

➔ Cocher « Contingency coefficient et Phi and Cramer's V » sous (Statistics – Nominal)

1. C (coefficient de contingence) = $\sqrt{\frac{X^2}{X^2+n}}$ -> (C < 1) et en situation d'indépendance (C = 0)
2. Φ (phi) = $\sqrt{\frac{X^2}{n}}$ -> situation d'indépendance (Φ = 0), Dépendance fonctionnelle (Φ ≥ 1)
3. V (V de Cramer) = $\sqrt{\frac{X^2}{X^2_{max}}}$ -> X2 = 0, V = 0 / X2max = X2 , V = 1

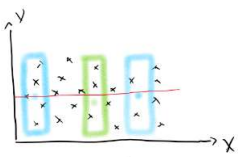
Liaison entre deux variables ordinales (Mesure d'un degré d'accord) :

1. Calculer la concordance (en dessous à droite) -> C
2. Calculer la discordance (en dessous à gauche) -> D
3. Claculer les paires ex-eaquo 1' (lignes) -> Nx
4. Calculer les paires ex-eaquo 2' (colones) -> Ny
5. Gamma$_{Goodman}$ ----------→ $\hat{\gamma} = \frac{C-D}{C+D}$
6. Tau$_b$: ----------→ $\hat{\tau}_b = \frac{C-D}{\sqrt{(C+D+n_x)(C+D+n_y)}}$

Faire diagramme de dispertion avec Jamovi :

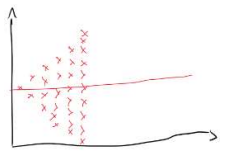
- Tableau des données (attentions variables numériques) -> exploration -> Sactterplot

Etude de diagramme de dispertion (corréaltion graphique) :

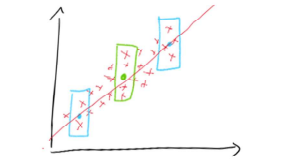


absence de liaison

$E(Y|X=x) = \alpha$

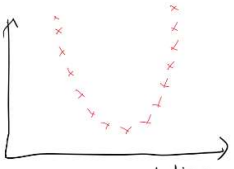


absence de liaison en moyenne mais pas en variance



dépendance linéaire car moyenne sont placés sur une ligne droite

$E(Y|X=x) = a + bx$ (ordonné à l'origine, pente)



dépendance quadratique

$E(Y|X=n) = a + bx + cx^2$

Coefficient de Spearman (r_s) :

1. Calculer les rangs (rg_x et rg_y)

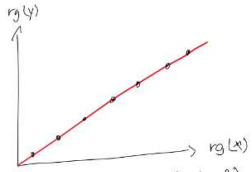
⇨ La donnée la plus basse = 1, puis la suivante = 2, etc. (si plusieurs fois la même valeur, on fait la moyenne)

2. Calculer d_i et d_i^2

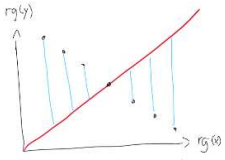
⇨ $d_i = (rg_y - rg_x)$ et $di^2 = di \times di$

3. Analyse -> regression -> correlation Matrix

⇨ Force de corrélation -> 0-0.19 très faible / 0.2-0.39 faible / 0.4-0.59 modéré / 0.6-0.79 forte / 0.8-1 très forte



accord parfait (idéal)



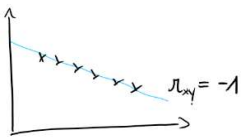
désaccord total !

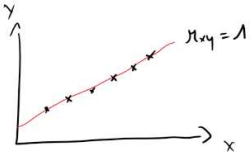
Liaison entre deux variables numériques, Bravais-Pearson (r_{xy})

- Jamovi -> Analyse -> regression -> correlation Matrix

⇨ Interprétation : valeur absolue entre 0-1 (-1) -> 0-0.5 = faible / 0.5-1 = forte / 0 et indé = réciproque fausse !!!

$|r_{xy}| = 1$ si et seulement si tous les points sont placés sur une même droite





Estimation d'une proportion (ponctuelle)

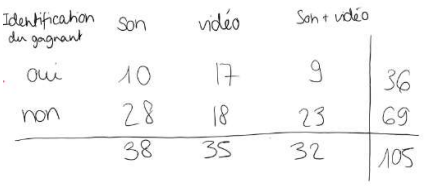
1. Individus statistiques : n = nbr de quartiers (97) / nbr total d'individu (160)
2. n_a = nbr de pépins (26) / nbr d'individu avec tel caractéristique (30)
3. $f_A = 26/97 = 0{,}268$ // $30/160 = 0.188$
4. $\hat{\pi} = f_A = 26{,}8\%$
5. Estimation de $\hat{\pi}$ par intervalle de confiance

⇨ Méthode exacte/binomiale -> Jamovi -> Analyses -> Frequencies -> 2 outcomes (binomial test)

6. Précision -> inversement proportionnelle à la variance F -> accroître la précision = augmenter n (taille d'échantillon
7. SE : (erreur standart -> écart type de la distribution)

Statistiques 6

	Condition	Réponse	Effectif
1	Son	Juste	10
2	Son	Faux	28
3	Vidéo	Juste	17
4	Vidéo	Faux	18
5	Son + Vidéo	Juste	9
6	Son + Vidéo	Faux	23



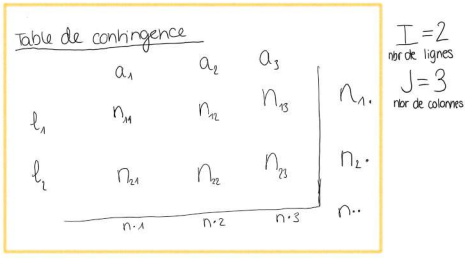


Table de contingence = n_{ij} :

4. Saisir les données
5. Chemin = Analyse – Frequencies – Contingency Tables – Independent Samples
6. Remplir -> Rows (Réponse) Columns (Condition) Counts (Effectif)

Distribution conditionnelle de Y :

3. Faire table de contingence
4. Cocher « Column » (sous Cells – Percentages)

⇨ Si les distributions conditionnelles de Y sont toutes les mêmes = Y ne dépend pas de X, sinon Y dépend de X

Distribution conditionnelle de X :

3. Faire table de contingence
4. Cocher « Row » (sous Cells – Percentages)

⇨ Si les distributions conditionnelles de X sont toutes les mêmes = X ne dépend pas de Y, sinon X dépend de Y

Indépendance des variables X et Y :

e. Les distributions conditionnelles de Y sont les mêmes
f. Les distributions conditionnelles de X sont les mêmes
g. L'effectif associé à (i ; j) = $\frac{n_{i\cdot} \times n_{\cdot j}}{n}$
h. La fréquence associée à la cellule (i ; j) = $f_{ij} = f_i \times f_j$ (Rappel : $f_i = n_i/n$ et $f_j = n_j/n$)

Tableau des effectifs théoriques = e_{ij} (correspond à la situation d'indépendance) :

3. Reprendre les données de la table de contingence (même chemin, remplissage)
4. Cocher « Expected » (sous Cells – Counts)

Tableau des résidus :

3. Permet de comparer le tableau des effectifs théoriques et la table de contingence
4. Contingence (n_{ij}) – Effectif(e^{ij}) -> (10 – 13 = -3)

Tableau des résidus standardisés :

2. ----------→ $\frac{n_{ij} - e_{ij}}{\sqrt{e_{ij}}}$

⇨ Intéressant si la valeur est supérieure ou égale à 2

taille d'effet

*df**	small	medium	large
1	.10	.30	.50
2	.07	.21	.35
3	.06	.17	.29
4	.05	.15	.25
5	.04	.13	.22

Règle d'interprétation de Cohen : ----------→

Valeur du Khi carré (distance entre n_{ij} et e_{ij}):

4. Après avoir fait une table de contingence et un tableau des effectifs théoriques
5. Cocher « X^2 » (sous Statistics)
6. Chiffre recherché = croisement X^2 et Value

⇨ $X^2_{max} = n$ [min (I ; J) -1] => 105 [min (2; 3) -1] => 105 [2 -1] => 105 [1] = 105 (dépendance$_{max}$ fonctionnelle)

⇨ $X^2_{min} = 0$ (situation d' indépendance)

Coefficient de contingence et V de Cramer :

⇨ Indice ne dépendant plus ni de « n » ni de la dimension de la table de contingence

➔ Cocher « Contingency coefficient et Phi and Cramer's V » sous (Statistics – Nominal)

4. C (coefficient de contingence) = $\sqrt{\frac{X^2}{X^2+n}}$ -> (C < 1) et en situation d'indépendance (C = 0)
5. Φ (phi) = $\sqrt{\frac{X^2}{n}}$ -> situation d'indépendance (Φ = 0), Dépendance fonctionnelle (Φ ≥ 1)
6. V (V de Cramer) = $\sqrt{\frac{X^2}{X^2_{max}}}$ -> $X^2 = 0$, V = 0 / $X^2_{max} = X^2$, V = 1