



Année 2010/2011 M1 SHS Sciences de l'éducation

4PAKMOT1 : Méthodes quantitatives

cours de **Jean-Claude Régnier**

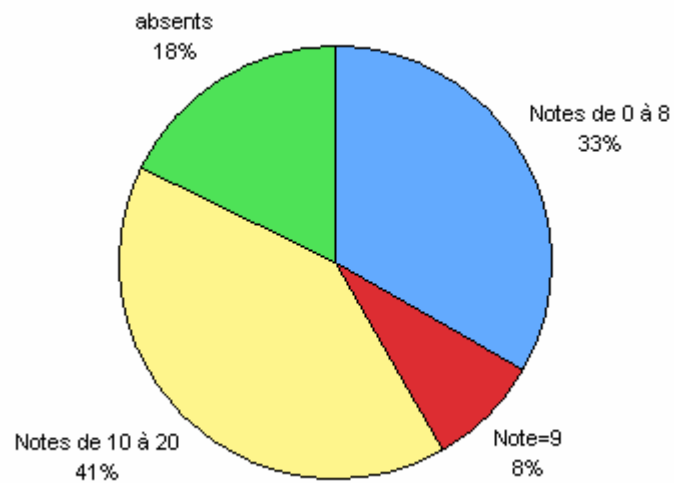
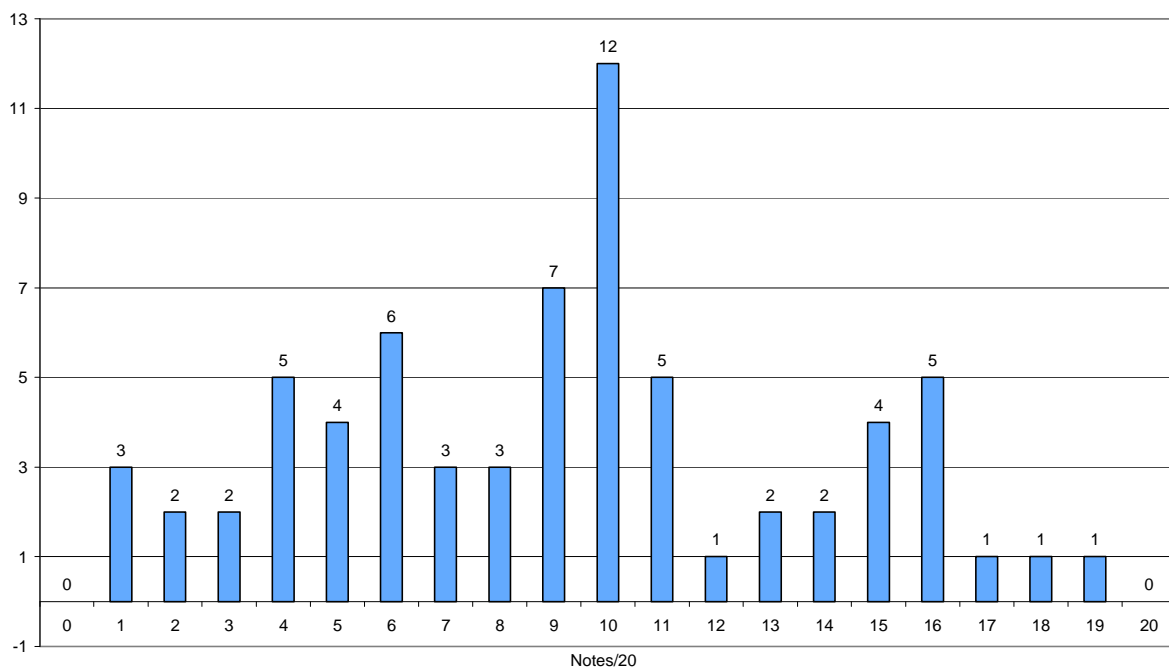
Document autocorrectif

SESSION 1

Synthèse des performances

moyenne	écart-type	min	Q1	Q2	Q3	max	présents	absents
9,16	4,46	1	6	9	11	19	69	15

Performances /20



Document n° 1 : situation de l'étude

La présente situation s'inspire des travaux d'une thèse de doctorat en sciences de l'éducation, intitulée CONSTRUÇÃO DE CONHECIMENTOS EM SITUAÇÕES DE EXCLUSÃO SOCIAL E QUESTÕES DE GÊNERO, préparée en cotutelle par Valdir Pretto (*voir* site cyberthèses Lyon2) et soutenue en août 2009 au Brésil. La thématique centrale porte les connaissances alternatives construites par des sujets confrontés à des situations d'exclusion sociale, pour survivre. Un échantillon de 103 unités a été extrait d'une population habitant dans une zone périphérique d'une ville du Sud du Brésil. Cet échantillon a été constitué en deux étapes : tirage aléatoire des rues, tirage aléatoire des foyers. Une unité est alors le ou les sujets qui ont répondu au questionnaire qui a été administré par le chercheur. Dans un premier temps une enquête par questionnaire a été réalisée afin de recueillir des informations relatives aux individus mais aussi pour identifier les problèmes qu'ils estimaient être ceux qu'ils rencontraient et dont les résolutions nécessitaient des connaissances. D'autres méthodes de construction ont été mises en oeuvre fondées sur des entretiens audio-vidéographiés individuels puis collectifs. En ce qui concerne les problèmes, 14 catégories furent identifiées. Le questionnaire comportait une vingtaine de grandes questions portant sur les caractéristiques individuelles et sociales des sujets, la situation familiale, la composition de la famille, les ressources, le logement, la formation de chacun des membres de la famille, etc.. Dans le cadre de cette étude, nous ne retenons ici que quelques questions que nous présentons ci-dessous. Les tableaux statistiques fournis ensuite contiennent des valeurs réelles mais aussi des valeurs fictives adaptées à la situation d'étude présente.

<i>code de la variable</i>	<i>descriptif succinct</i>
[V1]	Sexe : 1_homme 2_femme
[V2]	Age à la date de l'enquête du sujet qui répond aux questions
[V3]	Lieu de naissance
[V4]	Ancienneté de résidence dans la ville
[V5]	Ancienneté de résidence dans le quartier
[V6]	Etat civil : 1_Célibataire – 2_Marié(e) – 3_Séparé(e) – 4_Divorcé(e) – 5_Vit avec les parents – 6_Veuf(ve)
[V7]	Nombre de personnes qui vivent dans le foyer
[V8]	Ressources du foyer
[V9]	<i>Savez-vous lire et écrire ?</i> 1_Pas du tout ; 2_Un peu ; 3_Moyennement ; 4_Asez facilement ; 5_Très facilement
[V10]	Niveau de scolarité : 1_Analphabète – 2_Semi-analphabète – 3_Enseignement fondamental incomplet – 4_ Enseignement fondamental complet – 5_ Enseignement moyen (lycée) incomplet – 6_ Enseignement moyen (lycée) complet – 7_ Enseignement supérieur incomplet – 8_ Enseignement supérieur complet
[V11]	<i>Quels sont les problèmes que vous rencontrez actuellement là où vous habitez ?</i>
[V12]	<i>Comment vous affrontez, résolvez les problèmes que vous rencontrez là où vous habitez ?</i>
[V13]	Nombre de catégories de problèmes évoqués

Nous rapportons ci-dessous les tableaux statistiques des variables

Tableau statistique de la variable V1

(Tableau n° T1)	Homme	Femme
V1=Sexe		
Echantillon	34	69

Tableau statistique de la variable V4

(Tableau n° T2) V4= Temps de résidence dans la ville (années)	[1 ; 10[[10 ; 19[[19 ; 30[[30; 35[[35; 50[
Échantillon	19	44	22	9	9

Tableau statistique de la variable V6

(Tableau n° T3) V6=Etat civil	1_Célibataire	2_Marié(e)	3_Séparé(e)	4_Divorcé(e)	5_Vit avec les parents	6_Veuf(ve)
Échantillon	6	82	1	8	1	5

Tableau statistique de la variable V9

(Tableau n° T4) V9=Lire et Écrire	1_Pas du tout	2_Un peu	3_Moyennement	4_Asez facilement	5_Très facilement.
Échantillon (valeurs estimées)	8	5	54	24	12

Tableau des catégories de problèmes identifiés

(Tableau n° T5) codages des problèmes	Nombre d'évocations	<i>Catégories</i>	
Pb01	86	Santé	Saúde
Pb02	3	Alimentation	Alimentação
Pb03	3	Économie familiale	Economia familiar
Pb04	2	Situation familiale	Situação familiar
Pb05	8	Travail	Trabalho
Pb06	68	Éducation, formation scolaire	Educação
Pb07	93	Sécurité	Segurança
Pb08	91	Loisir	Lazer
Pb09	8	Questions religieuses	Religiosidade
Pb10	28	Transport	Transporte
Pb11	15	Recyclage	Reciclagem
Pb12	5	Habitation	Habitação
Pb13	79	Assainissement	Saneamento
Pb14	13	Relations de voisinage	Relacionamento com os vizinhos

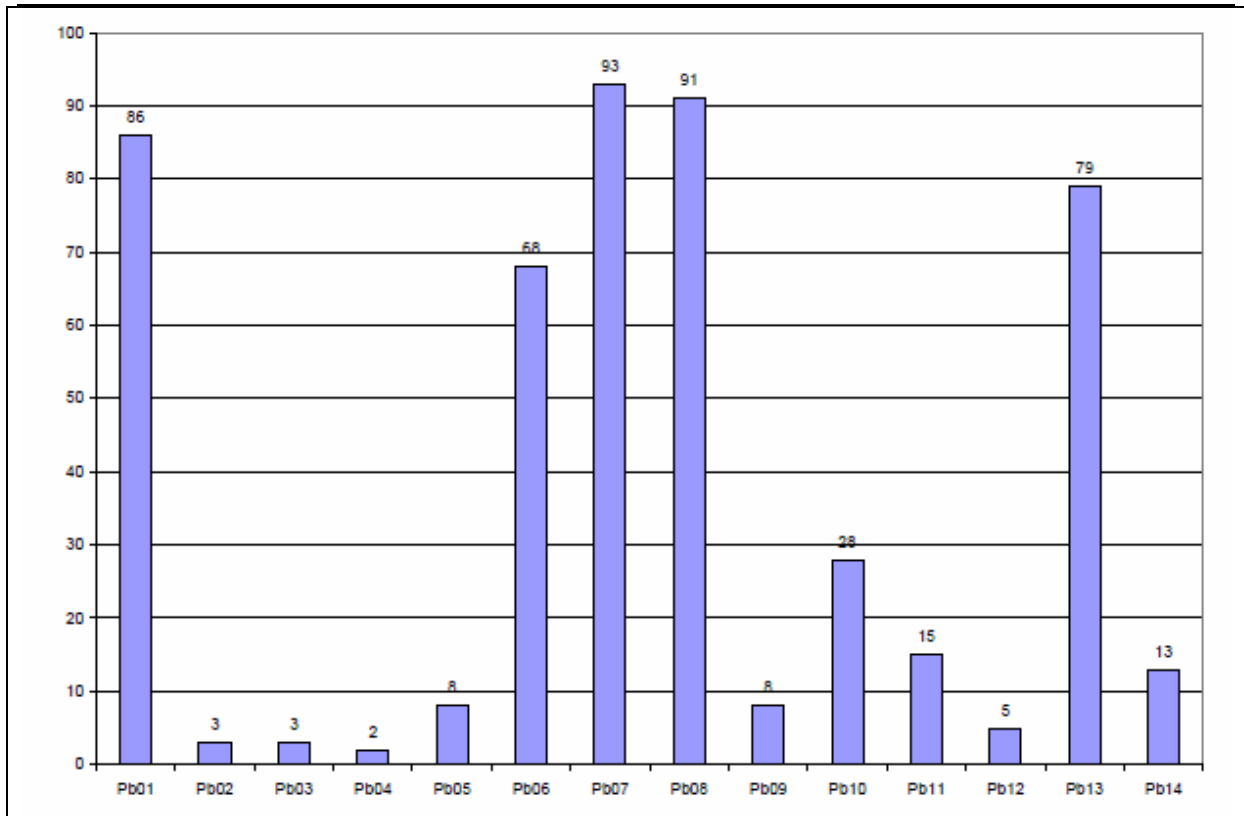


Tableau statistique de la variable V6

(Tableau n° T6) V13=nombre de catégories évoquées	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Échantillon	0	1	5	11	22	31	24	5	1	2	1	0	0	0	0

(Tableau n° T7)	Pb08 cité	Pb08 n'est pas cité	
Homme	27	7	34
Femme	64	5	69
	91	12	103

Document n° 2 : traitement demandé

TQ1- De quoi s'agit-il ?

Q101- Après avoir pris connaissance des informations fournies dans le document n°1, identifier et caractériser "succinctement" mais très clairement, les éléments de cette étude statistique : individus, population(s), échantillon(s), variables.

Population :

Ensemble des habitant dans une zone périphérique d'une ville du Sud du Brésil (**individus**)

Échantillons:

Sous-ensemble des 103 individus extrait de la population

Variables.

Variables quantitatives		Variables qualitatives		
discrètes	continues	ordinales	nominales	Textuelles
[V7]	[V2]	[V9]	[V1]	[V11]
[V8]	[V4]		[V3]	[V12]
[V13]	[V5]		[V6]	
			[V10]	

TQ2-Etude de la variable V4 « Temps de résidence dans la ville (années) »

Q201 Sur l'échantillon, quelle est la valeur m de la durée moyenne de résidence dans la ville et quelle est la valeur de l'écart-type ?

Intervalles	valeurs centrales (xi)	effectifs (ni)	$n_i x_i$	$n_i(x_i-m)^2$
[1 ; 10[5,5	19	104,5	3460,260156
[10 ; 19[14,5	44	638	889,0787068
[19 ; 30[24,5	22	539	666,6752757
[30 ; 35[32,5	9	292,5	1641,429824
[35 ; 50[42,5	9	382,5	4972,30361
		<i>effectif total</i>	<i>somme</i>	<i>somme</i>
		103	1956,5	11629,74757
			<i>moyenne</i>	<i>variance</i>
			18,9951	112,9101706
				<i>écart-type</i>
				10,62591975

Q202- A partir de cette étude, énoncer sous une forme synthétique votre conclusion et votre interprétation.

Ainsi la durée moyenne de résidence sur l'échantillon est d'environ 19 ans avec un fluctuation moyenne donnée par l'écart-type de 10 ans 8 mois environ

TQ3- Étude du lien entre l'évocation de la catégorie Pb08 – Loisir et la variable V01 Sexe

Q301- A partir de l'échantillon des 103 réponses, peut-on considérer qu'il y a ou non un lien entre ces deux variables?

En exploitant les données du tableau T7, tester les hypothèses correspondantes (Ho et H1 qui devront être explicitées) au seuil de risque $\alpha=0.05$ puis $\alpha=0.01$ à l'aide du test d'indépendance du Khi-deux.

Nous mettons en œuvre un test d'indépendance (Test du Khi-deux) pour décider laquelle des deux hypothèses est la plus plausible :

Ho Les hommes et les femmes ont des points de vue homogènes relatifs à la question de l'évocation de la catégorie Pb08 – Loisir

H1 Les hommes et les femmes ont des points de vue hétérogènes relatifs à la question de l'évocation de la catégorie Pb08 – Loisir

Tableau des effectifs observés :

(Tableau n° T7a) Pb08 cité Pb08 n'est pas cité			
Homme	27	7	34
Femme	64	5	69
	91	12	103

Tableau des effectifs théoriques sous l'hypothèse Ho :

(Tableau n° T7b) Pb08 cité Pb08 n'est pas cité			
Homme	30,039	3,961	34
Femme	60,961	8,039	69
	91	12	103

Tableau des contributions absolues :

(Tableau n° T7c) Pb08 cité Pb08 n'est pas cité			
Homme	0,3074	2,3313	2,6387
Femme	0,1515	1,1487	1,3002
	0,4589	3,4800	3,9389

Q302- A partir de cette étude, énoncer sous une forme synthétique votre conclusion et votre interprétation.

Au seuil de $\alpha = 5\%$ la valeur critique lue dans la table du Khi-deux ddl = 1 est de $k = 3,84$. la valeur empirique est de 3,93 et donc supérieure à la valeur théorique sous Ho. On rejette l'hypothèse d'homogénéité/indépendance à ce niveau de risque de 1^{ère} espèce.

Au seuil de $\alpha = 1\%$ la valeur critique lue dans la table du Khi-deux ddl = 2 est de $k = 6,63$. la valeur empirique est de 3,93 et donc inférieure à la valeur théorique sous Ho. On ne rejette pas l'hypothèse d'homogénéité à ce niveau de risque de 1^{ère} espèce. Cette conservation de Ho contre H1 est alors prise avec un risque de seconde espèce de niveau β inconnu

TQ4 - Étude de la variable V9

Q401 Donner la proportion des individus considérés comme ayant atteint un niveau « au moins moyen » pour lire et écrire dans l'échantillon.

Nombre d'individus ayant répondu « 3_Moyennement », ou « 4_Asez facilement » ou « 5_Très facilement » est de $54+24+12=90$

Proportion dans l'échantillon :

$$f = \frac{90}{103} = 0,873786407... \approx 87,38\%$$

Q402 Donner une estimation ponctuelle de la proportion des individus pouvant être considérés comme ayant atteint un niveau « au moins moyen » pour lire et écrire dans la population.

Estimation ponctuelle de la proportion dans la population :

87,38%

Q403 Donner une estimation par intervalle de confiance à 99% de la proportion des individus pouvant être considérés comme ayant atteint un niveau « au moins moyen » pour lire et écrire dans la population.

En ce qui concerne cet intervalle de confiance à 99%, nous l'obtenons en utilisant la formule ci-dessous :

$$\left[f - \kappa \sqrt{\frac{f(1-f)}{n-1}}; f + \kappa \sqrt{\frac{f(1-f)}{n-1}} \right]$$
$$\left[0,8738 - 2,57 \sqrt{\frac{0,8738(1-0,8738)}{103-1}}; 0,8738 + 2,57 \sqrt{\frac{0,8738(1-0,8738)}{103-1}} \right]$$

or $2,57 \left(\sqrt{\frac{0,8738(1-0,8738)}{102}} \right) \approx 2,57(0,032881802) \approx 0,084506233$

d'où [0,789280 ; 0,958292] C'est à dire entre 78,9% et 95,8%

Avec un niveau de confiance de 99%, nous estimons que la proportion se situe entre 78,9% et 95,8%. Ce qui laisse à penser qu'ils sont la majorité.

Q404- A partir de cette étude, énoncer sous une forme synthétique votre conclusion et votre interprétation.

TQ5 - Étude de l'homogénéité des points de vue relatifs à l'ordre d'importance des problèmes rencontrés

Nous rapportons les rangements des 14 catégories de problèmes produits par les 8 sujets qui ont constitué le groupe des sujets auprès desquels ont été réalisés les entretiens. Les résultats qui utilisent les rangs *ex aequo*, sont reportés dans le tableau n°T8 ci-dessous :

Tableau T8		Pb_01	Pb_02	Pb_03	Pb_04	Pb_05	Pb_06	Pb_07	Pb_08	Pb_09	Pb_10	Pb_11	Pb_12	Pb_13	Pb_14
individus	Sujet_01	4	10	10	10	10	5	2	3	10	10	10	10	1	10
	Sujet_02	4	10,5	10,5	10,5	10,5	6	1	10,5	10,5	2	10,5	5	3	10,5
	Sujet_03	4	10,5	10,5	10,5	10,5	6	1	10,5	10,5	2	10,5	5	3	10,5
	Sujet_04	5	11,5	11,5	11,5	11,5	7	2	6	11,5	8	4	11,5	3	1
	Sujet_05	1	3	9	9	9	9	9	9	2	9	9	9	9	9
	Sujet_06	1	3	9	9	9	9	9	9	2	9	9	9	9	9
	Sujet_07	2	1	9,5	9,5	9,5	9,5	3	4	9,5	9,5	9,5	9,5	9,5	9,5
	Sujet_08	2	1	9,5	9,5	9,5	9,5	3	4	9,5	9,5	9,5	9,5	9,5	9,5
	Totaux	23	50,5	79,5	79,5	79,5	61	30	56	65,5	59	72	68,5	47	69
Ho	60	60	60	60	60	60	60	60	60	60	60	60	60	60	
Ecart	1369	90,25	380,3	380,3	380,3	1	900	16	30,25	1	144	72,25	169	81	

Q501 - Quelle est la valeur empirique du coefficient W de concordance de Kendall sans faire la correction liée à l'existence des ex aequo ?

Q502- Peut-on considérer qu'il existe un classement ordonné significatif correspondant à une certaine homogénéité des points de vue de ces étudiants? Réaliser le test W de Kendall de confrontation des deux hypothèses Ho et H1 qui devront être explicitées, à niveau de risque de 1^{ère} espèce $\alpha=0.01$.

Ho Les rangements déterminés par l'ordre d'évocation des problèmes sont hétérogènes au sein de l'échantillon,

H₁ Il existe une tendance à l'homogénéité des rangements déterminés par l'ordre d'évocation des problèmes au sein de l'échantillon,

Nous utilisons le test W de Kendall.

$S_{Kmax} =$	$S_K =$	$W = S_K / S_{Kmax}$	$W_{critique}$ seuil 1%
14560	4014,5	0,27572115	0,26623254

Ho = hypothèse d'hétérogénéité des rangements et d'absence de concordance est rejetée au seuil de 1%. Nous acceptons à ce niveau de risque le rangement suivant comme représentant celui avec lequel l'ensemble des individus serait plutôt en accord.

Q503- Donner le rangement qui serait celui le plus concordant.

objectif	n°1	n°2	n°3	n°4	n°5	n°6	n°7	n°8	n°9	n°10	n°11	n°12	n°13	n°14
RANGEMENT significatif	1	4	12	12	12	7	2	5	8	6	11	9	3	10

Q504- A partir de cette étude, énoncer sous une forme synthétique votre conclusion et votre interprétation.

TQ6- Comparaison des rangements réalisés par les hommes et les femmes relativement aux catégories de problèmes évoqués.

Nous considérons les deux rangements obtenus à partir du test de concordance W de Kendall explicitant les ordres concordants respectivement du groupe des femmes et du groupe des hommes. Les résultats sont consignés dans le tableau T9 ci-dessous (avec présence d'ex aequo comme cela se présente dans la thèse).

Tableau T9	Pb_01	Pb_02	Pb_03	Pb_04	Pb_05	Pb_06	Pb_07	Pb_08	Pb_09	Pb_10	Pb_11	Pb_12	Pb_13	Pb_14
Rangement concordant du groupe Hommes	1	12	13,5	13,5	9	3	2	5	8	6	11	10	4	7
Rangement concordant du groupe Femmes	2	13	12	14	9	5	1	3	10	6	7	11	4	8
Somme des écarts au carré	1	1	2,25	0,25	0	4	1	4	4	0	16	1	0	1

$$\Delta^2 = 35,5$$

Q601-Peut-on considérer qu'il existe une tendance à l'accord significative entre ces deux groupes ?

Tester les hypothèses correspondantes (Ho et H1) au seuil de 5% à l'aide du test de Rhô de Spearman.

Ho Les rangements déterminés par l'ordre d'évocation des problèmes sont indépendants au sein de l'échantillon entre les hommes et les femmes

H₁ Il existe un lien entre les rangements déterminés par l'ordre d'évocation des problèmes au sein de l'échantillon par les hommes et par les femmes.

Calcul de la valeur empirique du coefficient ρ de Spearman

$$R_s = 1 - \frac{\Delta^2}{n(n^2 - 1)} = 1 - \frac{35,5}{14(14^2 - 1)} = 1 - \frac{35,5}{455} = \frac{419,5}{455} \approx 0,922$$

La valeur critique de Rs au seuil $\alpha=0.05$ (test bilatéral) est de 0,538.

Q602- A partir de cette étude, énoncer sous une forme synthétique votre conclusion et votre interprétation.

Nous rejetons l'hypothèse Ho au profit de H1. C'est-à-dire que nous retenons l'idée d'un lien entre l'ordre déterminé par les évocations des problèmes chez les hommes et celui chez les femmes.

GRILLE d'autoévaluation

Nom de l'étudiant:

n° :

code	items repérés et sur lesquels le jugement est porté	validité du résultat				validité du raisonnement			
		R+	R-	E	N	R+	R-	E	N
Q101	- analyse "succincte" de l'étude statistique	R+	R-	E	N	R+	R-	E	N
Q201.1	- moyenne de V05	R+	R-	E	N	R+	R-	E	N
Q201.2	- variance, écart-type de V05	R+	R-	E	N	R+	R-	E	N
Q202	- conclusion et interprétation personnelle	R+	R-	E	N	R+	R-	E	N
Q301.1	- énoncé explicite de Ho et H1	R+	R-	E	N	R+	R-	E	N
Q301.2	- mise en place du test	R+	R-	E	N	R+	R-	E	N
Q301.3	- mise en œuvre du test niveau 5%	R+	R-	E	N	R+	R-	E	N
Q301.4	- mise en œuvre du test niveau 1%	R+	R-	E	N	R+	R-	E	N
Q302	- conclusion et interprétation personnelle	R+	R-	E	N	R+	R-	E	N
Q401	- proportion sur l'échantillon	R+	R-	E	N	R+	R-	E	N
Q402	- estimation ponctuelle	R+	R-	E	N	R+	R-	E	N
Q403	- estimation par intervalle 99%	R+	R-	E	N	R+	R-	E	N
Q404	- conclusion et interprétation personnelle	R+	R-	E	N	R+	R-	E	N
Q501	- valeur du W	R+	R-	E	N	R+	R-	E	N
Q502.1	- énoncé explicite de Ho et H1	R+	R-	E	N	R+	R-	E	N
Q502.2	- mise en place du test	R+	R-	E	N	R+	R-	E	N
Q502.3	- mise en œuvre du test niveau 1%	R+	R-	E	N	R+	R-	E	N
Q503	- conclusion et interprétation personnelle	R+	R-	E	N	R+	R-	E	N
Q601.1	- énoncé explicite de Ho et H1	R+	R-	E	N	R+	R-	E	N
Q601.2	- mise en place du test	R+	R-	E	N	R+	R-	E	N
Q601.3	- mise en œuvre du test au seuil 5%	R+	R-	E	N	R+	R-	E	N
Q602	- conclusion et interprétation personnelle	R+	R-	E	N	R+	R-	E	N

	Tous les items repérés et sur lesquels le jugement est porté	validité du résultat				validité du raisonnement			
		R+	R-	E	N	R+	R-	E	N
	modalités du jugement	R+	R-	E	N	R+	R-	E	N
S01	nombre de...								
	résultat & raisonnement	R+	1	R-	0,5	E	0	N	0
S02	nombre de...								
S03	scores...								

Score total obtenu $S = (0,5) \times n_{R-} + (1) \times n_{R+} =$

Score maximum total $S_{max} = 22 + 22 = 44$

ANNEXE : Explication de la grille d'évaluation

<i>code des modalités du jugement</i>	<i>modalités du jugement</i>	<i>Caractérisation</i>	<i>pondération</i>
R+	réussite forte	- lorsque le résultat est celui attendu par l'évaluateur - lorsque le raisonnement est logiquement vrai	+1
R-	réussite faible	Il s'agit de repérer le déjà-réussi sans laisser la place à l' ancrage de sous-compréhension. - lorsque le résultat n'est pas celui attendu par l'évaluateur sans erreur manifeste de raisonnement - lorsque le raisonnement ne comporte que des erreurs "d'esprit fatigué"	+0,5
E	échec par erreur	- lorsque le résultat est faux mathématiquement - lorsque le raisonnement est logiquement faux	0
N	échec par non-réponse	- lorsque le résultat n'est pas fourni - lorsque le raisonnement n'est pas explicite - lorsque le résultat ou le raisonnement ne sont pas interprétables par l'évaluateur	0

Évaluation des items "CONCLUSION-INTERPRETATION"

Conformément aux objectifs visés en cours, l'item "conclusion" correspond à deux niveaux :

- *niveau 1* (représenté par le point de vue "**résultat**") correspond à une **description** lisible et "bien" rédigée des résultats obtenus dans les traitements abordés. Ces résultats sont rapportés dans le contexte de la situation problème. (Par exemple : les nombres sont donnés avec un ordre de grandeur compatible avec le contexte, avec leurs unités, ou les modalités sont explicitées en clair...)
- *niveau 2* (représenté par le point de vue "**raisonnement**") correspond à l'énoncé d'une **interprétation** personnelle fondée sur les résultats de l'étude et exprimée de façon pertinente.