

Document autocorrectif et auto-évaluatif

Considérons l'étude suivante :

Étude des difficultés et des facilités d'apprentissage de la statistique au travers des méthodes quantitatives et qualitatives dans une formation à distance en vue de l'obtention d'une licence de sciences de l'éducation.

Un questionnaire comportant 45 questions a été soumis à 132 étudiants ayant passé les épreuves d'évaluation du cours de méthodes quantitatives lors de la session 2002 d'examen de licence de sciences de l'éducation, préparation à distance. 297 étudiants sont inscrits dans cette licence. 86 étudiants ont rendu ce questionnaire. Nous n'avons rapporté ici que 7 questions.

Nous rapportons ci-dessous les tableaux statistiques des variables :

(Tableau n° T1) Tableau statistique de la variable V00 = sexe

Ensemble des inscrits			étudiants (au moins une épreuve de l'UE Méthodologie°)		
Modalités de V00	homme	femme		homme	femme
effectifs	49	248		23	109

(Tableau n° T2) Tableau statistique de la variable V01 = âge à la date de l'épreuve

Valeurs de V01	[20 ; 25[[25 ; 30[[30 ; 35[[35 ; 40[[40 ; 45[[45 ; 50[[50 ; 55[
effectifs	22	19	22	8	7	5	3

(Tableau n° T4) Tableau statistique de la variable V03 :

La situation professionnelle comporte-t-elle des activités qui requièrent des connaissances en statistique ?

Modalités de V03	1_Pas du tout	2_Rarement	3_Quelquefois	4_Souvent	5_Très souvent	6_Toujours
effectifs	48	15	18	4	0	0

(Tableau n° T5) Tableau statistique de la variable V18 =

Évaluation de l'efficacité du dossier méthodologique pour l'apprentissage

modalités de V18	1 le plus faible	2	3	4	5	6	7	8	9	10 le plus fort
effectifs	1	0	2	3	9	9	24	18	9	9

(Tableau n° T6) Tableau statistique de la variable V20 =

évaluation de la durée pour l'élaboration du dossier

Valeurs de V20	[0 ; 50[[50 ; 100[[100 ; 150[[150 ; 200[[200 ; 250[
effectifs	22	28	13	3	6
Valeurs de V20	[250 ; 300[[300 ; 350[[350 ; 400[[400 ; 450[[450 ; 675[
effectifs	2	0	0	1	3

**(Tableau n° T7) Tableau statistique de la variable V24 :
évaluation de la fréquence de recours à liste de diffusion**

Modalités de V24	1_Pas du tout	2_Rarement	3_Quelquefois	4_Souvent	5_Très souvent	6_Toujours
effectifs	39	12	20	4	3	4

**(Tableau n° T8) Tableau statistique de la variable V47B=
note obtenue à l'examen à l'épreuve de statistique**

notes/20	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
effectifs	0	1	0	2	8	4	6	7	1	3	5	4	7	9	9	6	2	5	2	4	0

Document n°2 : traitement demandé

TQ2.1- De quoi s'agit-il ?

Q201- Après avoir pris connaissance des informations fournies ci-dessus, identifiez et caractérisez "succinctement" mais clairement, les éléments du modèle statistique : **population, échantillon, individus, variables.**

Population :

Ensemble des étudiants (**individus**) préparant la licence de sciences de l'éducation dans un dispositif de formation à distance

Échantillons :

Sous-ensemble des individus ayant passé l'épreuve de méthodes quantitatives constitué de 132 individus. Partie (**sous-échantillon**) constituée des 86 étudiants ayant rendu le questionnaire de l'enquête issue des 132 étudiants de l'échantillon.

Variables.

Variables.

Variables quantitatives		Variables qualitatives		
discrètes	continues	ordinales	nominales	Textuelles
[V47B]	[V01]→âge [V20]→durée de réalisation du dossier	[V03] [V18] [V24]	[V00] Sexe	

TQ3- Étude du lien entre le jugement d'efficacité du dossier méthodologique et le niveau de réussite à l'épreuve d'examen

Q501- A partir du sous-échantillon des 83 répondants, peut-on considérer qu'il y a une dépendance entre le jugement d'efficacité et le niveau de réussite à l'examen?

Nous avons réduit le nombre de modalités de la variable V18 à deux niveaux.

		<i>Niveau d'efficacité Dossier</i>		
		N1 Faible 1,2,3	N2 Moyen 4,5,6,7	N3 Fort 8,9,10
<i>Niveau de réussite à l'épreuve finale</i>	Faible	16	11	
	Moyen	6	6	
	Fort	26	18	

Tester les hypothèses correspondantes (Ho et H1) au seuil de 5% à l'aide du test d'indépendance du Khi-deux.

Tableau des effectifs observés

Variable 1	Variable 2		Totaux
	modalité 2.1	modalité 2.2	
modalité 1.1	16	11	27
modalité 1.2	6	6	12
modalité 1.3	26	18	44
Totaux	48	35	83

Tableau des fréquences

Variable 1	modalité 2.1	modalité 2.2	Totaux
modalité 1.1	0,19277108	0,13253012	0,3253012
modalité 1.2	0,07228916	0,07228916	0,14457831
modalité 1.3	0,31325301	0,21686747	0,53012048
Totaux	0,57831325	0,42168675	1

Tableau des profils lignes

Variable 1	modalité 2.1	modalité 2.2	Totaux
modalité 1.1	0,59259259	0,40740741	1
modalité 1.2	0,5	0,5	1
modalité 1.3	0,59090909	0,40909091	1
Profil marginal	0,57831325	0,42168675	1

Tableau des profils colonnes

Variable 1	modalité 2.1	modalité 2.2	Profil marginal
modalité 1.1	0,33333333	0,31428571	0,3253012
modalité 1.2	0,125	0,17142857	0,14457831
modalité 1.3	0,54166667	0,51428571	0,53012048
Totaux	1	1	1

Tableau des effectifs théoriques sous Ho

Variable 1	modalité 2.1	modalité 2.2	Totaux
modalité 1.1	15,6144578	11,3855422	27
modalité 1.2	6,93975904	5,06024096	12
modalité 1.3	25,4457831	18,5542169	44
Totaux	48	35	83

Tableau des contributions absolues

Variable 1	modalité 2.1	modalité 2.2	Totaux
modalité 1.1	0,00951956	0,0130554	0,02257496
modalité 1.2	0,12725904	0,17452668	0,30178571
modalité 1.3	0,01207101	0,01655453	0,02862554
Totaux	0,14884961	0,2041366	0,35298621

Tableau des contributions relatives

Variable 1	modalité 2.1	modalité 2.2	Totaux
modalité 1.1	2,6969%	3,6986%	6,3954%
modalité 1.2	36,0521%	49,4429%	85,4950%
modalité 1.3	3,4197%	4,6899%	8,1095%
Totaux	42,1687%	57,8313%	1

Q502- 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 = 2 est de $k = 5,99$. la valeur empirique est de 0,35 et donc inférieure à la valeur théorique sous H_0 . On ne rejette pas l'hypothèse d'homogénéité à ce niveau de risque de 1^{ère} espèce. Cette conservation de H_0 contre H_1 est alors prise avec un risque de seconde espèce de niveau $\hat{\alpha}$ inconnu

TQ6- Étude de l'intérêt personnel pour les TIC en formation

Une question concernant l'intérêt personnel pour les TICE a été posée sous le format suivant : *Quel est votre degré d'intérêt personnel relatif à l'intégration des TIC en formation ? (cocher la case correspondante)*

Très faible (f1) Plutôt faible (f2) Plutôt fort (F3) Très fort (F4) Sans avis (SA5)

(Tableau n° T9) Tableau statistique de la variable V07

Modalités de V09	f1	f2	F3	F4	SA5
Effectifs	15	13	27	15	10

Q601 - A partir des informations obtenues sur l'échantillon et consignées dans le tableau L'hypothèse H_0 peut être ainsi formulée "la distribution des fréquences des diverses opinions f1, f2, F3, F4 et SA5 est uniforme" contre l'hypothèse alternative H_1 "cette distribution n'est pas uniforme et comporte une opinion dominante F3". Pour résoudre ce problème, nous recourons à la procédure suivante :

Un test d'adéquation consiste à expliciter une règle de décision relative à une hypothèse portant sur la **comparaison d'une distribution empirique à une distribution théorique précisée a priori ou ajustée à partir des observations recueillies.**

Conditions d'utilisation:

- Prélever aléatoirement dans la population P inconnue un échantillon de n observations indépendantes réparties en s classes,
- Choisir une loi de distribution théorique dont les r paramètres sont soit fixés a priori, soit estimés à partir de l'échantillon.
- Aucun effectif théorique n'est inférieur à 5.

Critère. variable de décision:

On choisit la variable $D^2 = \sum_{i=1}^{i=s} \frac{(O_i - T_i)^2}{T_i}$ où O_i = effectif observé et T_i = effectif théorique sous la

contrainte de la loi de référence choisie pour chacune des s classes. Si aucun des effectifs théoriques n'est inférieur à 5, la variable D^2 suit approximativement une loi du Khi-Deux à s-1-e degrés de liberté (où e désigne le nombre de paramètres estimés parmi les r paramètres). Si un des effectifs est inférieur à 5, on procède à des regroupements de classes pour se mettre dans les conditions d'application.

Conclusion: Choisir la valeur α du risque de première espèce. Rechercher dans une table du Khi-Deux à $s-1-e$ degrés de liberté, la valeur critique k telle que $\text{Prob}\{\chi^2_{(s-1-e)} > k\} = \alpha$. Calculer d^2 la valeur empirique de D^2 avec les s valeurs O_i observées sur l'échantillon extrait. Si $d^2 < k$, c'est à dire si $d^2 \in A = [0, k]$, région d'acceptation, alors on ne rejette pas l'hypothèse nulle H_0 , on accepte alors H_0 en prenant un risque β d'erreur de seconde espèce. Si $d^2 > k$, c'est à dire si $d^2 \in K = [k, +\infty[$, région critique, alors on rejette l'hypothèse nulle H_0 et on accepte H_1 .

Pour ce qui concerne notre cas, le dépouillement fait apparaître les résultats suivants que nous consignons dans le tableau statistique :

(Tableau n° T9) Tableau statistique de la variable V07

Modalités de V09 = opinions	f1	f2	F3	F4	SA5	
Effectifs observés	15	13	27	15	10	80
Effectifs théoriques sous H_0	16	16	16	16	16	80

Calcul de la valeur d^2 prise par la variable $D^2 = \sum_{i=1}^{i=s} \frac{(O_i - T_i)^2}{T_i}$ où ici $s = 5$.

$$d^2 = \frac{(15-16)^2}{16} + \frac{(13-16)^2}{16} + \frac{(27-16)^2}{16} + \frac{(15-16)^2}{16} + \frac{(10-16)^2}{16} = \frac{168}{16} = 10,5$$

Comme nous avons choisi la valeur $\alpha = 0,05$ du risque de première espèce, nous recherchons alors dans une table du Khi-Deux à $s-1-e = 5-1-0 = 4$ degrés de liberté, la valeur critique k telle que $\text{Prob}\{\chi^2_{4} > k\} = 0,05$, nous trouvons $k = 9,48$.

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

Nous constatons que d^2 est dans la région de rejet de H_0 , $d^2 = 10,5 > 9,48$, alors on rejette l'hypothèse nulle H_0 , avec un risque de première de $\alpha = 0,05$

Si nous choisissons un niveau de risque $\alpha = 0,01$ alors la valeur critique k telle que $\text{Prob}\{\chi^2_{4} > k\} = 0,01$, nous trouvons $k = 13,27$. Comme nous constatons alors que d^2 est dans la région $A = [0, k]$, région d'acceptation, $d^2 = 10,5 < 13,27$, alors on ne rejette pas l'hypothèse nulle H_0 , et l'on accepte en prenant un risque β d'erreur de seconde espèce dont la valeur nous est inconnue.

TQ5- Étude de l'intérêt personnel pour le dispositif Formation Ouverte à Distance FOAD

Une question concernant l'intérêt personnel pour le dispositif a été posée sous le format suivant : *Quel est votre degré d'intérêt personnel pour le dispositif FOAD ? (cocher la case correspondante)*

Très faible (f1) Plutôt faible (f2) Plutôt fort (F3) Très fort (F4) Sans avis (SA5)

(Tableau n° T9) Tableau statistique de la variable V07

Modalités de V09bis	f1	f2	F3	F4	SA5
Effectifs	5	10	24	40	1

Q701 Donner une estimation ponctuelle de la proportion des individus exprimant un intérêt « Plutôt FORT » ou « Très FORT » dans la population.

Q702 Donner une estimation par intervalle de confiance à 95% de la proportion des individus exprimant un intérêt « Plutôt FORT » ou « Très FORT » dans la population.

Estimation ponctuelle et par intervalle de confiance d'une proportion p sur une

population

Estimation ponctuelle

Elle est le résultat d'une réalisation de la fréquence empirique sur l'échantillon choisi. Elle est tout simplement obtenue à partir de la proportion calculée avec les valeurs observées sur l'échantillon : $\pi = f_n$ où f_n est la valeur de F_n calculée à partir de l'échantillon obtenu à l'issue

de n tirages. Notons que $\frac{\pi(1-\pi)}{n}$ la variance de F_n est estimée ponctuellement soit par

$$\frac{f_n(1-f_n)}{n-1}, \text{ soit par sa valeur maximale } \frac{1}{4(n-1)}$$

Estimation par intervalle de confiance

Les deux situations usuelles sont:

- Intervalle de confiance à 95% : $k=1,9600$
- Intervalle de confiance à 99% : $k=2,5758$

Ensuite on utilise les résultats obtenus précédemment dans le cadre de l'estimation ponctuelle à savoir f_n . La fourchette d'estimation de la proportion π s'obtient avec la formule suivante :

$$f_n - \kappa \sqrt{\frac{f_n(1-f_n)}{n-1}} < \pi < f_n + \kappa \sqrt{\frac{f_n(1-f_n)}{n-1}}$$

Ainsi construit, cet intervalle $\left[f_n - \kappa \sqrt{\frac{f_n(1-f_n)}{n-1}}; f_n + \kappa \sqrt{\frac{f_n(1-f_n)}{n-1}} \right]$ a une probabilité de

$(1-\alpha)$ de contenir la vraie valeur π qui nous est inconnue. La valeur critique κ est à chercher dans la table de Laplace-Gauss, si n est supérieur à 100

Bien que l'échantillon ne soit que $n = 80$, nous utiliserons malgré tout l'approximation gaussienne pour trouver la valeur critique qui sera donc $k=1,96$ puisque nous avons choisi un niveau de confiance à 95%.

Ainsi à partir de cet échantillon les individus qui répondent «Plutôt FORT » ou «Très FORT » sont au nombre de $24+40=64$. La proportion sur l'échantillon est de

$$p = \frac{64}{80} = 0.8 = 80\% \text{ Cela donne donc une estimation ponctuelle de } \pi \text{ sur la population :}$$

$$\pi = 80\%$$

En ce qui concerne un intervalle de confiance à 95%, nous obtenons en utilisant la formule ci-dessus :

$$\left[0.8 - 1,96 \sqrt{\frac{0,8(1-0,8)}{80-1}} ; 0.8 + 1,96 \sqrt{\frac{0,8(1-0,8)}{80-1}} \right]$$

$$\text{or } 1,96 \left[\sqrt{\frac{0,8(1-0,8)}{79}} \right] \approx 1,96(0,45003516) \approx 0,88206891$$

$$\text{d'où } [0,7117931 ; 0,88206891]$$

Q703- A partir de cette étude, énoncer sous une forme synthétique votre **conclusion** et votre **interprétation**.

On peut donc estimer la proportion des individus qui exprimeraient un intérêt «Plutôt FORT » ou « Très FORT » dans la population, à 80%. Si on souhaite accorder un degré de confiance de niveau 95% à une estimation nous trouvons l'intervalle de confiance suivant :

$$[71,17\% ; 88,83\%].$$

Nom de l'étudiant:

n° :

code	items repérés et sur lesquels le jugement est porté	validité du résultat				validité du raisonnement			
Q201_1	- Identification : population, individus	R+	R-	E	N	R+	R-	E	N
Q201_2	- Identification : échantillon, individus	R+	R-	E	N	R+	R-	E	N
Q201_3	- Identification : variable 1 = sexe = qualitative nominale (T1)	R+	R-	E	N	R+	R-	E	N
Q201_4	- Identification : variable 2 = âge = quantitative continue (T2)	R+	R-	E	N	R+	R-	E	N
Q201_5	- Identification : variable 3 = V03 = qualitative ordinale (T4)	R+	R-	E	N	R+	R-	E	N
Q201_6	- Identification : variable 4 = V18 = qualitative ordinale (T5) ou quantitative discrète	R+	R-	E	N	R+	R-	E	N
Q201_7	- Identification : variable 5 = V20 = quantitative continue (T6)	R+	R-	E	N	R+	R-	E	N
Q201_8	- Identification : variable 6 = V24 = qualitative ordinale (T7)	R+	R-	E	N	R+	R-	E	N
Q201_9	- Identification : variable 7 = V47B = note = quantitative discrète (T8)	R+	R-	E	N	R+	R-	E	N

Tous les items repérés et sur lesquels les jugements sont portés		validité du résultat				validité du raisonnement			
	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 /18

S1=

Q501.1	- énoncé explicite de Ho et H1	R+	R-	E	N	R+	R-	E	N
Q501.2	- mise en place du test	R+	R-	E	N	R+	R-	E	N
Q501.3	- mise en œuvre du test	R+	R-	E	N	R+	R-	E	N
Q502	- 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	R+	R-	E	N	R+	R-	E	N
Q602	- conclusion et interprétation personnelle	R+	R-	E	N	R+	R-	E	N
Q701	- estimation ponctuelle de la proportion	R+	R-	E	N	R+	R-	E	N
Q702	- estimation par intervalle 95% de la proportion	R+	R-	E	N	R+	R-	E	N
Q703	- conclusion et interprétation personnelle	R+	R-	E	N	R+	R-	E	N

Tous les items repérés et sur lesquels les jugements sont portés		validité du résultat				validité du raisonnement			
	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 /22

S2=

$$\text{Note sur 20} = \frac{S1 + 2S2}{62} \cdot 20 =$$