

## Dossier PSYCHO – Analyse

### Type des variables

1/ Indiquer le type de chacune des variables du dossier PSYCHO, (Nominale / Ordinale / Numérique) :

- Nombre d'années d'études de psychologie (ANPSY) : *Numérique*
- Indice de charge mentale (CHARG) : *Numérique*
- Sexe (SEXE) : *Nominale (binaire)*
- Réponse au questionnaire (QUEST) : *Nominale*
- Série du bac (BAC) : *Nominale*
- Année de naissance (ANNAI) : *Numérique*

2/ Indiquer, parmi les variables numériques précédentes, le nom d'une variable de rapport. Justifier votre réponse :

*ANPSY, pour deux raisons :*

*- le 0 n'est pas arbitraire (0 année d'études de psycho)*

*- le rapport de deux valeurs a un sens (4 années d'études, c'est deux fois plus que 2 années)*

*CHARG peut éventuellement être défini comme une variable de rapport, malgré la présence de valeurs négatives.*

*ANNAI n'est pas une variable de rapport (en particulier le 0 est arbitraire). Par contre l'âge calculé à partir de l'année de naissance est bien une variable de rapport.*

### Liaison ANPSY x CHARG

Objectif : On se demande si le nombre d'année d'études (ANPSY) contribue à faire diminuer la charge mentale induite par la lecture du texte sur la psychanalyse (CHARG).

#### A. Analyse descriptive

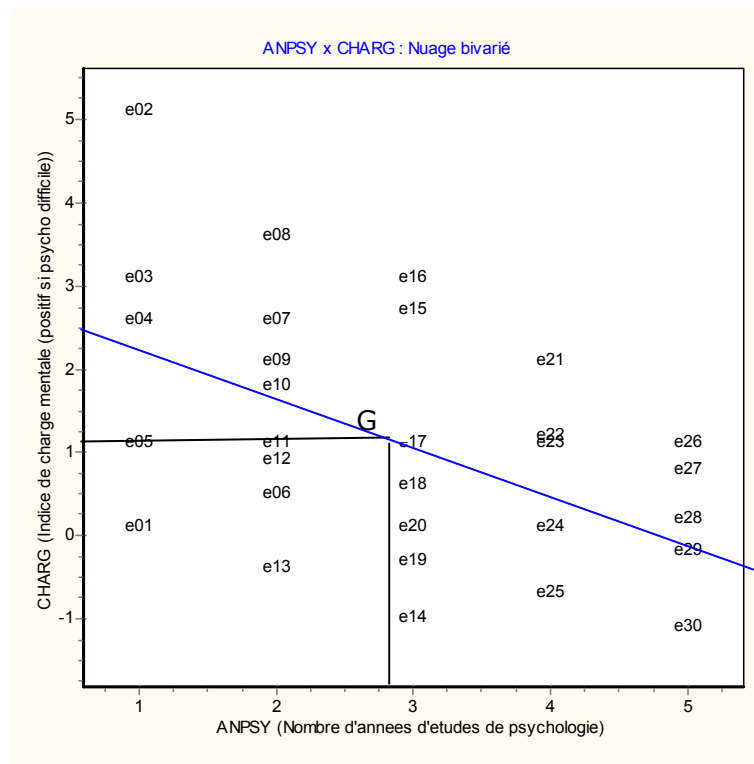
L'analyse séparée de chacune de ces deux variables donne les indications suivantes qui pourront être utilisées pour la suite de l'analyse :

	ANPSY	CHARG
Moy	2.90	1.07
Var	1.76	2.09

1/ L'analyse de la liaison entre les deux variables commence par l'analyse du graphe de corrélation (cf. page suivante).

a/ Quelles sont les coordonnées du point moyen ou centre de gravité du nuage ? On note G ce point. Reporter sa position sur le graphe.

*Coordonnées du point moyen G : (2.90 ; 1.07)*



b/ Indiquer pourquoi, pour analyser la liaison entre deux variables numériques, il est nécessaire d'analyser le graphe de corrélation (indiquer deux raisons) :

*Le graphe permet de voir le sens et l'ampleur de la liaison linéaire entre les variables. Mais cette information est également donnée par le coefficient de corrélation.*

*Plus spécifiquement le graphe permet, par exemple, de détecter :*

- une/des valeur(s) atypique(s)
- l'existence d'une liaison non linéaire
- l'existence de sous-groupes hétérogènes

*Il permet également de situer un individu particulier par rapport au groupe.*

3/ La valeur du coefficient de corrélation linéaire de Bravais-Pearson est -0.48. Indiquer ce que signifie le signe négatif de ce coefficient :

*Ce signe négatif signifie que plus le nombre d'années d'études de psychologie (ANPSY) augmente, plus la charge mentale induite par la lecture du texte de psychanalyse (CHARG) diminue.*

4/ On se demande si le nombre d'années d'études de psychologie (variable ANPSY) permet de prédire, au moins en partie, la charge mentale induite par la lecture du texte de psychologie (variable CHARG).

Un logiciel nous indique l'équation d'une droite qui permet de prédire la charge mentale en fonction du nombre d'années d'études de psychologie.

Cette équation est la suivante : **CHARG = -0.52 × ANPSY + 2.57**

a/ Indiquer pourquoi, compte tenu des résultats précédents, le coefficient -0.52 est nécessairement négatif.

*Le coefficient a de l'équation de régression est toujours de même signe que la corrélation.*

b/ Interpréter la valeur du coefficient +2.57

*C'est le niveau de charge mentale (CHARG) prédit si ANPSY=0, c'est à dire pour un étudiant qui n'aurait fait aucune étude de psychologie.*

c/ Indiquer comment trouver les coordonnées de deux points par lesquels passe la droite sur le graphe de corrélation (cf. page précédente):

*On calcule les coordonnées de deux points (si possible éloignés sur le graphe) de cette droite. par exemple :*

$$\text{si ANPSY} = 1, \text{ CHARG} = -0.52 \times 1 + 2.57 = 2.05$$

$$\text{si ANPSY} = 5, \text{ CHARG} = -0.52 \times 5 + 2.57 = -0.03$$

d/ Calculer la charge mentale prédite par l'équation pour l'étudiant e01 :

*e01 a une année d'études de psychologie. La charge mentale prédite est donc égale à :*  
 $-0.52 \times 1 + 2.57 = 2.05$ .

e/ Calculer l'erreur de prédiction pour ce même étudiant :

*La charge mentale observée pour cet étudiant est égale à 0.*

*L'erreur de prédiction est donc de :*

$$0 - 2.05 = -2.05$$

## **B. Analyse inférentielle**

---

On se demande s'il existe bien un lien entre le nombre d'études de psychologie et la charge mentale induite par la lecture du texte de psychanalyse, dans la population d'où provient l'échantillon. On cherche donc à préciser quelle peut être la valeur de la corrélation dans la population parente. On suppose que cet échantillon a été constitué par tirage au hasard dans cette population.

1/ Préciser la nature de la population parente :

*Il s'agit de l'ensemble des étudiants en psychologie de cette université parisienne.*

2/ Les tests présentés au cours de ce semestre servent à tester ce qui est appelé habituellement une hypothèse nulle (notée  $H_0$ ). Formuler l'hypothèse nulle correspondant à la situation analysée ici :

*$H_0$  : Il n'existe pas de lien entre le nombre d'années d'études de psychologie (ANPSY) et la charge mentale induite par la lecture du texte de psychanalyse (CHARG).*

3/ Pour tester cette hypothèse nulle, on met en œuvre un test  $T$  de Student. On trouve  $t = 2.86$  et  $ddl = 28$ . Indiquer les formules utilisées et le développement des calculs :

$$t = t = \sqrt{n-2} \times \frac{r}{\sqrt{1-r^2}} = \sqrt{30-2} \times \frac{-0.48}{\sqrt{1-(-0.48)^2}} = -2.86$$

*Remarque : on trouve  $t = 2.86$  à partir de la valeur exacte de  $r$  (.4751..)*

$$ddl = n - 2 = 28$$

4/ Un logiciel nous indique  $p = 0.8\%$  et  $p/2 = 0.4\%$  pour ce test. En déduire le résultat du test, en deux étapes :

a/ Test significatif ou non significatif ?

*Le test est significatif car  $p = 0.8\% < 5\%$  (ou  $p/2 = 0.4\% < 2.5\%$ ).*

b/ Rejet ou non rejet de l'hypothèse nulle ?

*Le test étant significatif, on peut rejeter l'hypothèse nulle (au seuil  $p = 0.8\%$ )*

## Liaison SEXE x QUEST

Objectif : On se demande s'il existe des différences entre les garçons et les filles (SEXE) dans les réponses à la question sur les raisons pour lesquelles les gens vont voir un psychologue (QUEST).

On observe la répartition suivante des réponses :

	ANORM	AIDE	ORIEN
G	6	2	0
F	8	8	6

### A. Analyse descriptive

1/ Pourquoi est-il nécessaire de calculer les effectifs d'indépendance (effectifs théoriques) pour analyser la liaison entre deux variables nominales ?

*Les effectifs sous l'indépendance (ou effectifs théoriques) indiquent la répartition que l'on observerait s'il n'y avait aucun lien entre les variables. L'analyse de la liaison entre les variables consiste alors, pour l'essentiel, à comparer les effectifs observés à ces effectifs théoriques.*

2/ L'effectif théorique pour le couple (F, ANORM) est égal à 10.3. Indiquer la formule et développer les calculs :

	ANORM	AIDE	ORIEN
G	3.7	2.7	1.6
F	10.3	7.3	4.4

$$\hat{n}_{jk} = \frac{n_j \cdot n_k}{n} = \frac{22 \times 14}{30} = 10.27$$

3/ Construire ci-dessous le tableau des attractions-répulsions :

	ANORM	AIDE	ORIEN
G	+	-	-
F	-	+	+

4/ Commenter ce tableau des attractions-répulsions :

*Les garçons pensent, plus souvent que les filles, que les gens vont voir les psychologues parce qu'ils se sentent dans un état anormal (ANORM). Les filles pensent, plus fréquemment que les garçons, que les gens vont voir un psychologue par besoin d'aide (AIDE) et pour des problèmes d'orientation (ORIEN).*

On calcule ensuite les contributions absolues au  $\Phi^2$  (cf. tableau ci-dessous). Ces valeurs seront utilisées pour les différentes analyses qui suivent.

	ANORM	AIDE	ORIEN
G	0.0459	0.0056	0.0533
F	0.0167	0.002	0.0194
			0.1429

5/ Indiquer le nom de deux autres statistiques permettant de mesurer, case par case, l'ampleur des écarts à l'indépendance :

*Les écarts bruts à l'indépendance, les taux de liaison (écarts relatifs à l'indépendance), les densités, les contributions relatives au  $\Phi^2$ .*

6/ On souhaite se prononcer sur l'importance de la liaison globale entre ces deux variables, dans l'échantillon.

a/ Indiquer le nom de la statistique permettant de se prononcer sur cette question :

*V<sup>2</sup> de Cramér*

b/ Calculer la valeur de cette statistique après avoir indiqué la formule utilisée :

$$V^2 = \text{Phi}^2 / \text{Phi}^2_{\text{max}} = 0.1429 / 1 = 0.14$$

c/ Concluez brièvement sur l'importance de la liaison observée :

*La liaison observée n'est ni faible ni importante. Elle peut être qualifiée d'intermédiaire ( $0.04 < V^2 < 0.16$ ).*

## **B. Analyse inférentielle**

---

On souhaite évaluer s'il existe un lien entre ces deux variables dans la population parente. On suppose que cet échantillon a été constitué par tirage au hasard dans cette population

1/ Indiquer le nom de la statistique de test appropriée et calculer sa valeur à partir des éléments fournis dans les tableaux précédents :

$$K\text{hi}^2 = n \text{Phi}^2 = 30 \times 0.1429 = 4.29$$

2/ Calculer le nombre de degrés de liberté correspondant à ce test (indiquer la formule et le développement des calculs) :

$$\text{ddl} = (J-1)(K-1) = (2-1) \times (3-1) = 2$$

3/ Utiliser la table appropriée pour déterminer si le test est significatif ou non. Justifier votre réponse.

*La table du  $K\text{hi}^2$  (pour  $\text{ddl}=2$ ) indique  $p = 5\%$  pour  $K\text{hi}^2 = 5.99$ . Le  $K\text{hi}^2$  calculé (4.29) étant inférieur à 5.99, on en déduit que  $p > 5\%$  et que le test est non significatif.*

4/ Rédiger une conclusion détaillée en deux parties, descriptive puis inférentielle, portant uniquement sur la liaison globale entre ces deux variables :

a/ Conclusion descriptive :

*Pour ces 30 étudiants en psychologie de cette université parisienne, on constate qu'il existe un lien entre le sexe et la réponse au questionnaire sur les raisons pour lesquelles les gens consultent un psychologue (ou des différences entre garçons et filles concernant la réponse au questionnaire).  
Cependant la différence (ou la liaison) est seulement d'importance intermédiaire ( $.04 < V^2$  de Cramér =  $.14 < .16$ )*

b/ Conclusion inférentielle :

*Cependant, on ne peut pas conclure à l'existence d'un lien entre ces deux variables chez l'ensemble des étudiants en psychologie de cette université parisienne ( $K\text{hi}^2 [\text{ddl}=2] = 4.29, p > 5\%$ ).*

## Liaison BAC x CHARG

Objectif : On se demande si la charge mentale induite par la lecture du texte de psychanalyse (CHARG) varie selon la série du bac de la personne (BAC).

Afin de comparer les groupes, on a les premiers résultats suivants qui serviront de base à l'analyse et aux calculs suivants :

	Effectif	Moy	Var
L	10	1.01	1.93
S	8	0.79	2.24
ES	6	0.80	0.84
Autres	6	1.82	2.63
<b>Total</b>	<b>30</b>	<b>1.07</b>	<b>2.09</b>

1/ On trouve  $V_{inter} = 0.15$ . et  $V_{intra} = 1.94$ . Indiquer le développement des calculs (sans effectuer) qui permettraient de retrouver la valeur de  $V_{intra}$  :

$$V_{intra} = \frac{(10 \times 1.93) + (8 \times 2.24) + (6 \times 0.84) + (6 \times 2.63)}{30}$$

$$\text{ou } V_{intra} = V_{totale} - V_{inter} = 2.09 - 0.15$$

2/ Indiquer la propriété de décomposition de la variance :

$$V_{totale} = V_{inter} + V_{intra}.$$

3/ Rappeler la formule du rapport de corrélation (noté  $\eta^2$  ou  $Eta^2$ ) et développer les calculs :

$$Eta^2 = V_{inter} / V_{totale} = 0.15 / (0.15 + 1.94) = 0.15 / 2.09$$

4/ Sachant que  $Eta^2 = 7\%$ , que peut-on en conclure sur l'importance de la liaison entre les deux variables dans l'échantillon ?

*Compte tenu de la valeur de  $Eta^2$  (7%) on peut dire que la liaison n'est ni faible, ni importante mais d'importance intermédiaire ( $4\% < Eta^2 < 16\%$ )*

5/ Indiquer quelles configurations des moyennes des groupes et/ou des scores à l'intérieur des groupes auraient conduit à :

- un rapport de corrélation égal à 0 (0%)

*Si les moyennes des groupes avaient été égales (dans ce cas  $V_{inter} = 0$ ).*

- un rapport de corrélation égal à 1 (100%)?

*Si à l'intérieur de chaque groupe, tous les individus ont la même valeur (dans cas  $V_{intra} = 0$  et  $V_{inter} = V_{totale}$ ).*