

TP Analyse des données quantitatives

Résumé de la matière pour l'examen

4 types de questions à l'examen :

- ⇒ 1 sur le recodage d'une variable
- ⇒ 2 sur l'analyse d'un test
- ⇒ 1 « vrai ou faux » sur la matière du cours (ATTENTION à point négatif)

Commentaires :

- ⇒ Pas d'AFCP et de Cluster à l'examen mais il faut être capable de comprendre et lire les données
- ⇒ « Paste » permet d'avoir une trace de ce que l'on fait pour ne pas se perdre ; on commence par * et on finit par un « point ».
- ⇒ « Sig » = signification ; c'est à dire les résultats à analyser
- ⇒ « Mean » = moyenne
- ⇒ Quand on nous demande de comparer les tests entre eux, il faut comparer les résultats des tests de calibrage (V de Kramer, coefficient de corrélation...)
- ⇒ Changer les labels en belles phrases pour une meilleure compréhension
Ex : 1 = catégorie de personnes de 50 ans et plus → remplacer le 1 par « population âgée de 50 ans et plus »

1. Mettre un filtre

ATTENTION à ne pas l'oublier à l'examen

Lorsqu'on souhaite analyser juste une partie de l'échantillon à notre disposition.

Ex : un pays, une tranche d'âge, un genre...

Data → select cases → « if condition is satisfied »

- ⇒ Pour un pays : on choisit la variable pays que l'on rentre puis, « = XXX » (numéro de référence du pays que l'on souhaite étudier)
XXX = numéro de référence du pays étudié
- ⇒ Pour deux pays : country = XXX | country = XXX
(barre droite = ET)

2. Transformation/recodage de variables

ATTENTION toujours commencer par regarder avant le type de variable (quantitative ou qualitative) sur « variable view » et regarder les colonnes « values » et « measure ».

○ Calcul

= pour faire des moyennes

Transform → « compute variable » → change

- **Substitution/recodage**

= pour créer des classes

Ex : de 0 à 5 ans

Transform → « recode into different variable » → change

ATTENTION comme SPSS est inclusif, il vaut mieux mettre la grande valeur en premier

Ex : 2 classes moins de 50 ans et plus de 50 ans

→ En premier : range value trough the highest 50 = 2

→ En deuxième : range value trough the lowest 49 = 1

- **Comptage**

= pour créer une échelle de score lorsque l'on a deux choix : 0 = echec ; +1 = réussite

Transform → « count values within cases » → « defines values » → change

Il faut choisir les variables que l'on veut analyser.

Ex : si valeur de -4 à 0 = 0 et de 1 à 4 = 1

Ca veut dire qu'on ajoute 1 point à ceux qui sont d'accord (par exemple) ce qui correspond à ceux qui ont répondu de 1 à 4 → du coup on met « range highest to 1 »

3. Comment choisir son test ?

a) Il faut toujours commencer par regarder le type de variable : **quantitative ou qualitative** ; sur « variable view » et regarder les colonnes « values » et « measure ».

⇒ Scale/ordinal = variable quantitative

⇒ Nominal = variable qualitative

b) Il faut maintenant savoir si la variable est **indépendante ou dépendante**

RAPPEL

Variable indépendante (Vi) = celle qui influe sur les autres

Variable dépendante (Vd) = celle qui est affectée par la Vi

Ex : la catégorie socioprofessionnelle (Vi) et l'orientation politique (Vd)

c) En définissant si le type de variable on peut maintenant choisir le test adapté :


	Vd = quantitative		Vd = qualitative
Vi = quantitative	Test de corrélation		Test du Chi-carré
Vi = qualitative	+ de 2 modalités* ANOVA	2 modalités* TEST T	

*Modalités = nombre de catégories/classes à vérifier en allant voir la variable

Ex : moins de 50 ans + plus de 50 ans = 2 modalités → Test T

Moins de 25 ans + entre 25 et 50 ans + plus de 50 ans = 3 modalités → ANOVA

4. Propriétés des tests

	TEST DE RELATION ENTRE VARIABLES		TEST D'HOMOGENEITE DES MOYENNES	
	Chi carré	Corrélation	Anova	Test T
Variables	Vd = quali	Vd = quanti Vi = quanti	Vd = quanti Vi = quali + de 2 modalités	Vd = quanti Vi = quali 2 modalités
HO (hypothèse nulle)	Tester HO qui prévoit une absence de relation entre deux variables en comparant les effectifs théoriques (E.T) et les effectifs observés (E.O)	Tester HO qui prévoit une absence de relation linéaire entre deux variables : mesurer la dépendance ou l'indépendance	Tester HO qui prévoit une absence de différence significative entre plusieurs moyennes	Tester HO qui prévoit une absence de différence significative entre deux moyennes
Conditions <i>Variables indépendantes pour tous les tests</i>	Au plus 20% des E.T peuvent présenter une valeur inférieure à 5 et tous les E.T sont toujours supérieurs à 1		n < 50 (donc petit) on doit vérifier la normalité. Homoscedasticité = la variance des groupes à comparer est égale (cf test Lévène) → si ça ne l'est pas, on va lire la 2 ^{ème} ligne du tableau	n < 50 (donc petit) on doit vérifier la normalité. Homoscedasticité = la variance des groupes à comparer est égale (cf test Lévène) → si ça ne l'est pas, on va lire la 2 ^{ème} ligne du tableau
Test de calibrage ou d'association = pour mesurer l'ampleur de l'effet	V de Kramer = indique la force d'association en valeur absolu 0,1 = faible 0,3 = moyen 0,5 = fort	Coefficient R (en valeur absolu) 0,1 = faible 0,3 = moyen 0,5 = fort	Test Lévène = pour vérifier l'homoscedasticité Si homogène on lit l' ANOVA (tukey) + R*2 Si hétérogène on ne lit pas l'ANOVA, mais le test Post Hoc (dunett)	Test Lévène = pour vérifier l'homoscedasticité N2 = $T2/(T2+n1+n2-2)$

I) Le test du Chi-carré

ATTENTION pour les variables qualitatives

Analyze → descriptive statistiques → crosstabs
+ « statistiques » : cocher chi-square + phi & Cramer's V

Commentaire :

Les variables peuvent être mises dans n'importe quel sens

Interprétations :

On test

HO : il n'y a pas de relation entre les deux variables

H1 : il y a une relation entre les deux variables

a) Analyser dans son ensemble le premier tableau « crosstabulation » = tableau des effectifs

b) Analyser le tableau « chi-square » :

Lire « asymptotic significance »

⇒ $\text{Sig} < 0,5$ = faible probabilité de se tromper → il y a une relation entre les deux variables

⇒ $\text{Sig} > 0,5$ = probabilité élevée de se tromper → il n'y a pas de relation entre les deux variables

c) Analyser le tableau « symmetric measures » (pour le V de Kramer il faut regarder « value ») = analyser l'ampleur de la relation

⇒ 0,1 = faible

⇒ 0,3 = moyen

⇒ 0,5 = fort

II. Le test de corrélation

ATTENTION pour Vd & Vi quantitatives

Analyze → correlate → bivariate
+ « options » : cocher « means and standard deviations »

Interprétations :

On test

HO : il n'y a pas de relation entre les deux variables

H1 : il y a une relation entre les deux variables

Le coefficient de corrélation est compris entre

-1 = corrélation négative

1 = corrélation positive

a) Analyser la corrélation de Pearson :

⇒ 0,1 = faible

⇒ 0,3 = moyen

⇒ 0,5 = fort

b) Analyser le coefficient R → sig. (Bilatéral):

Indique la force d'association entre les deux variables

⇒ Sig < 0,5 = faible probabilité de se tromper → il y a une relation entre les deux variables

⇒ Sig > 0,5 = probabilité élevée de se tromper → il n'y a pas de relation entre les deux variables

Le signe associé à la valeur du coefficient indique le sens de la relation :

Corrélation négative = y diminue si x augmente

Corrélation positive = y augmente si x augmente

III. Le test de l'ANOVA

ATTENTION pour Vd = quanti et Vi = quali
Et + de 2 modalités

Analyze → compare means → Anova
+ « options » : cocher descriptive + homogeneity + means plot + test Brown-Fosythe
+ « post hoc » : cocher Tukey + Dunnett's T3

Commentaire :

Sélectionner les deux variables, attention à la différence entre la dépendante à mettre dans « dependant list » et l'indépendante à mettre dans « factor »

Interprétations :

On test

HO : il n'y a pas au moins une des moyennes des variables qui est différente des autres

H1 : il y a au moins une des moyennes des variables qui est différente des autres

→ On cherche à voir si les réponses des différentes catégories sont les mêmes, ou si une catégorie particulière va avoir un avis différent

a) Analyser le test d'homogénéité des variances (Test Lévène) :

⇒ Sig < 0,05 = inégalité entre les variances → hétérogène

⇒ Sig > 0,05 = égalité entre les variance → homogène

b) Si c'est hétérogène on ne lit pas l'ANOVA, mais le test Post Hoc en regardant le Test « Dunett » :

⇒ Sig < 0,05 = il y a une différence significative entre les moyennes → hétérogène

⇒ Sig > 0,05 = il n'y a pas de différence significative entre les moyennes → homogène

c) Si c'est homogène on regarde ANOVA et le Test « Tukey » :

⇒ Sig < 0,05 = il y a une différence significative entre les moyennes → hétérogène

⇒ Sig > 0,05 = il n'y a pas de différence significative entre les moyennes → homogène

Puis on calcule R^2 (à quel point la moyenne va être différente des autres) = racine carré de « between groups »/total

IV. Le Test T

ATTENTION pour Vd = quanti et Vi = quali
Et uniquement 2 modalités

Analyze → compare means →
Soit : « one sample test t » si seulement 1 variable pour la quali (Ex : la Grèce)
Soit : « independant sample » si au moins 2 choix dans la variable quali (Ex : - de 50 ans ; +de 50 ans)

Commentaires :

Pour « One sample »

- + encadré « Tests variables » : insérer la variable continue à partir de laquelle on veut comparer une moyenne
- + onglet « Test value » : entrer la moyenne observée par ailleurs

Pour « Independant sample »

- + encadré « Tests variables » : insérer la variable continue à partir de laquelle on veut comparer une moyenne
- + encadré « Grouping variable » : insérer la variable à partir de laquelle les groupe sont constitués (la variable qualitative)
- + onglet « Define groups » : spécifier les modalités des deux groupes dans l'onglet Ex : groupe 1 : 1 (= valeur que j'ai donné pour la catégorie)
- + option « Cutpoint »

a) Analyser le test d'homogénéité des variances (Test Lévène) dans le tableau « independant samples test :

- ⇒ Sig < 0,05 = inégalité entre les variances → hétérogène
- ⇒ Sig > 0,05 = égalité entre les variance → homogène

- Si hétérogène on lit la ligne du haut « equal variables assumed » : les moyennes ne sont pas semblables
- Si homogène on lit la ligne du bas « equal variables not assumed » : les moyennes sont semblables

b) Calculer à quel point les moyennes sont différentes l'une de l'autre :

Calculer T^2 (voir T dans tableau independant sample) / $t^2 + n1$ (N à trouver dans le tableau group statistique) + $n2 - 2$ → $t^2 / t^2 + n1 + n2 - 2$
Différent car ne va pas jusqu'à 1 :
0,01 = petit ; 0,06 = effet modéré ; 0,14 = effet forte ampleur

5. Représentation graphique

Si une variable qualitative → tableau de fréquence / barre

Si 2 variables qualitatives → tableau croisé / barre juxtaposé

Si variable qualitative & quantitative → means / bloxplot

Si variable quantitative → means explore / histogramme

Si 2 variables quantitatives → xxx / nuage de points

- Pour décrire des variables nominales/ordinales : **Bar Chart/simple** (diagrammes en barres simple)
- Pour décrire des variables quantitatives : **Histogram**(histogramme)
- Pour décrire deux variables nominales/ordinales : **Bar Chart/Clustered**(Diagramme en barres groupées)
- Pour décrire une variable quantitative avec une variable nominale/ordinaire : **Box-plot**(Boîte à moustache)
- Pour décrire deux variables quantitatives : **Scatter dot** (Nuage de point)