Ch4 : Quantitatif × Qualitatif

→ Etudier deux variables l'une est de type quantitative l'autre est qualitative

but Mesurer un lien éventuel entre deux caractères en utilisant un résumé chiffré qui traduit l'importance de ce lien.

Pour étudier le lien entre une variable qualitative à p modalités et un caractère quantitatif, on partitionne la population P en sous-populations : une sous-populat° pour chaque modalité.

Exp : On étudie les deux variables :
- Sexe : de type qualitatif → F, H
- Taille : de type quantitatif

Donc On divise La population en deux sous groupes : populat° des hommes et population de femmes et pour chaque populat° on peut réaliser des résumés numériques et des représentat° graphiques

Exp:

Sexe	EFF	Moyenne	S = ecart-type
H	23	162.30	14.21
F	35	149.29	10.52
Total	58	154,45	13.61

taille moyen des Hom.

taille moy des fem

taille Moy de la populat°

- On étudie la variable quantitative Y sur chaque sous-population en calculant la moyenne et la variance de Y, on parle de « Variation intra »

Variance intra: Moyenne des variances (variance intra-classe)

$$Var^{intra}(Y) = \frac{1}{n}\sum_{l=1}^{p} n_l \cdot S_l^2 = \frac{1}{n}\sum_{i=1}^{n_l} n_l \times \frac{1}{n_l}\sum_{i=1}(y_{li}-\bar{y}_l)^2$$

eff de La sous-populat

variance de la sous populat°

$$Var^{(intra)}(Y) = \frac{1}{58}\left[(23)(14.21)^2 + (35)(10.52)^2\right]$$

variance intra

eff des hommes

Ecart-type hom

eff fem

ecart-type fem

Scanné avec CamScanner

Décomposition de la moyenne :

$$\bar{Y} = \frac{1}{n} \sum_{l=1}^{p} n_l \cdot \bar{y}_l$$

P = nb des modalités pour la Variable quanti

Moyenne total pour toute la populat°.

eff total

eff de sous pop

Moy de s.s pop.

Dans notre exemple :

$$\bar{Y} = \frac{1}{n} \sum_{l=1}^{p} n_l \cdot \bar{y}_l$$

Ici : P = 2 car on a 2 modalités $\underline{\underline{H}}$ et $\underline{\underline{F}}$

$$\Rightarrow \bar{Y} = \frac{1}{58} \left[\overbrace{(23 \times 162,30)}^{Hommes} + \overbrace{(35 \times 149,29)}^{Femmes} \right]$$

taille moy de toute la populat°.

eff total

eff homme

taille moyen hom

eff femmes

taille moy femmes

$$\boxed{Var^{inter}(Y) = \frac{1}{n} \sum_{l=1}^{p} n_l \cdot (\bar{y}_l - \bar{y})^2}$$

Variance inter-classe : La variance des moyennes (Var inter-classes).

. On peut définir alors trois variances pour la variable quantitative Y :

1. **Var intra** explique les variations de Y dans les sous populations. (Variance Résiduelle)
2. **Var inter** explique les variations de Y entre les sous populations (Variance expliquée)
3. **Var Totale** explique les variation de Y dans toute la populat° (Variance totale)

$$\boxed{\underbrace{Var^{(totale)}(Y)}_{\text{Variance Totale}} = \underbrace{Var^{inter}(Y)}_{\text{Variance expliqué}} + \underbrace{Var^{intra}(Y)}_{\text{Variance Résiduelle}}}$$

$\triangle!$ $Var^{(total)}(Y) = \frac{1}{n} \sum (y_i - \bar{y})^2$

Scanné avec CamScanner

Rapport de corrélat° entre X et Y :

$$\boxed{r = \frac{Var^{inter}(Y)}{Var^{total}(Y)}}$$

$Var^{inter}(Y)$ → Variance expliqué

$Var^{total}(Y)$ → Var total

Ce Rapport représente le pourcentage de variabilité de Y expliquée par X. Il varie entre 0 et 1.

(donne % d'explicat°, on dit pas ici corrélat° forte ou faible comme en quanti × quanti).

- $r = 0$ → la variance expliquée (numérateur) est nul. Donc il y a aucun lien entre X et Y

$r = 1$ → $Var^{inter}(Y) = Var^{total}(Y)$ donc Y entièrement expliqué par X

X = Sexe

Notre exemple : Y = taille

- $Var^{inter}(Y) = \frac{1}{n} \sum n_\ell \times S_\ell^2$

$= \frac{1}{58} \left[(23 \times 14{,}21^2) + (35 \times 10.5_2^2) \right]$

$= 141{,}5$

- $Var^{inter}(Y) = \frac{1}{n} \sum n_\ell (\bar{y}_\ell - \bar{y})^2$

$= \frac{1}{58} \left[(23)(162 - 154{,}4)^2 + 35(149{,}3 - 154{,}4)^2 \right]$

$= 40.6$

. On vérifie la formule de la décomposition de la --
Variance : [on a $V^{(total)} = 182,1$]

$$V_{inter} + V_{intra} = 40,6 + 141,5 = 182,1 = V^{(total)}$$

. Rapport de corrélation : $\frac{V_{inter}}{V_{total}} = \frac{40,6}{182,1} = 0.22$

donc 22 % de la variabilité de la taille est expliquée par le sexe.