IDUP / octobre 23

Master 1 / Analyse statistique (JF Léger)

03. Caractéristiques de tendance centrale

Exercice 4 : Utilisation des fréquences élémentaires, conditionnelles et marginales dans le calcul d'une moyenne pondérée

CORRIGE

Complétez le tableau a en utilisant seulement les données de ce tableau.

Tableau a : notes aux examens selon le sexe et la discipline

Discipline	Garçons		Filles		Ensemble	
	Notes	Effectif	Notes	Effectif	Note	Effectif
Droit	11,0	400	?	400	11,5	800
Sciences humaines	10,0	300	?	600	12,0	900
Total	10,6	700	?	1 000	11,8	1 700

CORRIGE

La note moyenne en droit correspond à la somme des notes obtenues par l'ensemble des étudiants (total de points) divisée par le nombre d'étudiants. A partir des données, on connait le nombre total de points et le nombre de points obtenus par l'ensemble des garçons. On peut donc en déduire le nombre total de points obtenus par les filles. Et comme on connaît le nombre d'étudiantes en droit, on peut déterminer leur note moyenne.

$$\sum notes_{Droit} = note\ moyenne_{Droit} \times N_{Etudiants\ en\ Droit}$$

$$\sum notes_{Droit} = \sum notes_{Garçons\ en\ Droit} + \sum notes_{Filles\ en\ Droit}$$

$$\sum notes_{Droit} - \sum notes_{Garçons\ en\ Droit} = \sum notes_{Filles\ en\ Droit}$$

$$\frac{(note\ moyenne_{Droit} \times N_{Etudiants\ en\ Droit}) - (note\ moyenne_{Garçons\ en\ Droit} \times N_{Garçons\ en\ Droit})}{N_{Filles\ en\ Droit}} = note\ moyenne_{Filles\ en\ Droit}$$

On peut formaliser cette démarche autrement. La note moyenne des étudiants en droit correspond à la moyenne des notes des étudiants en droit de sexe masculin et féminin pondérées par le poids de chacun de ces deux groupes au sein de la population étudiante en droit.

$$\bar{x}_{Droit} = \left(\bar{x}_{Droit,G} \times \frac{n_{Droit,G}}{n_{Droit}}\right) + \left(\bar{x}_{Droit,F} \times \frac{n_{Droit,F}}{n_{Droit}}\right) = \frac{1}{n_{Droit}} \times \left[\left(\bar{x}_{Droit,G} \times n_{Droit,G}\right) + \left(\bar{x}_{Droit,F} \times n_{Droit,F}\right)\right]$$

On peut donc en déduire la note moyenne des étudiantes en droit :

$$\bar{x}_{Droit,F} = \frac{\left(\bar{x}_{Droit} \times n_{Droit}\right) - \left(\bar{x}_{Droit,G} \times n_{Droit,G}\right)}{n_{Droit,F}}$$

Application numérique :

$$\bar{x}_{Droit,F} = \frac{(800 \times 11{,}5) - (400 \times 11)}{400} = 12{,}0$$

On reproduit la même démarche pour les étudiants en sciences humaines et pour l'ensemble des étudiants.

$$\bar{x}_{Sc.Hum.,F} = \frac{(900 \times 12{,}0) - (300 \times 10{,}0)}{600} = 13{,}0$$

$$\bar{x}_{F} = \frac{(1\,700 \times 11{,}8) - (700 \times 10{,}6)}{1\,000} = 12{,}6$$

Complétez le tableau b en utilisant seulement les données de ce tableau.

Tableau b : notes aux examens selon le sexe et la discipline

Discipline	Garçons		Filles		Ensemble	
	Notes	Effectif	Notes	Effectif	Note	Effectif
Droit	11,0	57%	?	40%	11,5	47%
Sciences humaines	10,0	43%	?	60%	12,0	53%
Total	10,6	100%	?	100%	11,8	100%

CORRIGE

Dans le cas présent, on cherche les mêmes informations que dans le tableau précédent (les notes moyennes obtenues par les filles) mais à partir de données différentes. On dispose à nouveau des notes moyennes des garçons et celles de l'ensemble des étudiants. Mais cette fois-ci on connait, pour chaque sexe, la répartition des étudiants selon la discipline. Il s'agit des fréquences conditionnelles « en colonne ». On connait aussi les fréquences marginales « en colonne ». On peut s'aider de la résolution précédente pour comprendre la démarche à mettre en œuvre ici.

On a vu que :

$$\frac{(\bar{x}_{Droit} \times n_{Droit}) - (\bar{x}_{Droit,Garçons} \times n_{Droit,Garçons})}{n_{Droit,Filles}} = \bar{x}_{Droit,Filles}$$

On ne connait pas dans le cas présent le nombre d'étudiants en droit mais leur proportion dans la population totale (47 %). On connaît donc :

$$\frac{n_{Droit}}{N} = f_{Droit}$$

On peut faire apparaître ce terme dans la première formule en multipliant le numérateur et le dénominateur par le nombre total d'étudiants (N) :

$$\frac{\dfrac{(\bar{x}_{Droit} \times n_{Droit}) - (\bar{x}_{Droit,Garçons} \times n_{Droit,Garçons})}{N}}{\dfrac{n_{Droit,Filles}}{N}} = \bar{x}_{Droit,Filles}$$

$$\frac{\left(\bar{x}_{Droit} \times \frac{n_{Droit}}{N}\right) - \left(\bar{x}_{Droit,Garçons} \times \frac{n_{Droit,Garçons}}{N}\right)}{\frac{n_{Droit,Filles}}{N}} = \bar{x}_{Droit,Filles}$$

$$\frac{\left(\bar{x}_{Droit} \times f_{Droit}\right) - \left(\bar{x}_{Droit,Garçons} \times f_{Droit,Garçons}\right)}{f_{Droit,Filles}} = \bar{x}_{Droit,Filles}$$

Il faut donc déterminer les fréquences élémentaires d'étudiants masculins et féminins en droit, à savoir 1) la part des garçons suivant des études de droit au sein de l'ensemble de la population étudiante et 2) la part des filles étudiant le droit au sein de l'ensemble de la population étudiante.

On connaît les probabilités d'étudiants en droit parmi les garçons et les filles. Il faut donc chercher à faire apparaître ces proportions dans la formule ci-dessus. Il faut donc trouver le facteur par lequel multiplier la fréquence conditionnelle « en colonne » pour obtenir la fréquence élémentaire :

$$f_{Droit\ /\ Garçons} \times y = f_{Droit,\ Garçons}$$

$$\frac{n_{Droit,\ Garçons}}{n_{Garçons}} \times y = \frac{n_{Droit,\ Garçons}}{n_{Etudiants}}$$

$$y = \frac{n_{Droit,\ Garçons}}{n_{Etudiants}} \times \frac{n_{Garçons}}{n_{Droit,\ Garçons}} = \frac{n_{Garçons}}{n_{Etudiants}} = f_{Garçons}$$

On a donc :

$$f_{Droit\ /\ Garçons} \times f_{Garçons} = f_{Droit,\ Garçons}$$

On peut également faire la même chose pour les filles :

$$f_{Droit\ /\ Filles} \times f_{Filles} = f_{Droit,\ Filles}$$

La formule s'écrit dorénavant :

$$\frac{\left(\bar{x}_{Droit} \times f_{Droit}\right) - \left(\bar{x}_{Droit,\ Garçons} \times \left(f_{Droit\ /\ Garçons} \times f_{Garçons}\right)\right)}{f_{Droit\ /\ Filles} \times f_{Filles}} = \bar{x}_{Filles\ en\ Droit}$$

Il reste deux inconnues : les proportions de filles et de garçons dans la population étudiante totale. Il faut donc déterminer ces deux proportions. On sait que :

$$f_{Droit,\ Garçons} + f_{Droit,\ Filles} = f_{Droit}$$

car :

$$n_{Droit,\ Garçons} + n_{Droit,\ Filles} = n_{Droit}$$

$$\frac{n_{Droit,\ Garçons}}{n_{Etudiants}} + \frac{n_{Droit,\ Filles}}{n_{Etudiants}} = \frac{n_{Droit}}{n_{Etudiants}}$$

On sait aussi que :

$f_{Garçons} + f_{Filles} = 1$, donc $f_{Garçons} = 1 - f_{Filles}$

Par conséquent :

$$f_{Droit,\,Garçons} + f_{Droit,\,Filles} = f_{Droit}$$
$$\left[f_{Droit\ /\ Garçons} \times f_{Garçons}\right] + \left[f_{Droit\ /\ Filles} \times f_{Filles}\right] = f_{Droit}$$
$$\left[f_{Droit\ /\ Garçons} \times (1 - f_{Filles})\right] + \left[f_{Droit\ /\ Filles} \times f_{Filles}\right] = f_{Droit}$$
$$\left(f_{Droit\ /\ Garçons} \times 1\right) - \left(f_{Droit\ /\ Garçons} \times f_{Filles}\right) + \left(f_{Droit\ /\ Filles} \times f_{Filles}\right) = f_{Droit}$$
$$f_{Filles} \times \left(f_{Droit\ /\ Filles} - f_{Droit\ /\ Garçons}\right) = f_{Droit} - f_{Droit\ /\ Garçons}$$
$$f_{Filles} = \frac{f_{Droit} - f_{Droit\ /\ Garçons}}{f_{Droit\ /\ Filles} - f_{Droit\ /\ Garçons}}$$

Application numérique

$$f_{Filles} = \frac{0{,}47 - 0{,}57}{0{,}40 - 0{,}57} = 0{,}59 = 59\ \%$$
$$f_{Garçons} = 1 - 0{,}59 = 0{,}41 = 41\ \%$$

On peut maintenant calculer les notes moyennes des filles en droit :

$$\bar{x}_{Droit} = \frac{\left[\bar{x}_{Garçons,\,Droit} \times f_{Garçons,\,Droit}\right] + \left[\bar{x}_{Filles,\,Droit} \times f_{Filles,\,Droit}\right]}{f_{Droit}}$$
$$\bar{x}_{Droit} = \frac{\left[\bar{x}_{Garçons,\,Droit} \times \left(f_{Droit\ /\ Garçons} \times f_{Garçons}\right)\right] + \left[\bar{x}_{Filles,\,Droit} \times \left(f_{Droit\ /\ Filles} \times f_{Filles}\right)\right]}{f_{Droit}}$$
$$\bar{x}_{Filles,\,Droit} = \frac{\left[\bar{x}_{Droit} \times f_{Droit}\right] - \left[\bar{x}_{Garçons,\,Droit} \times \left(f_{Droit\ /\ Garçons} \times f_{Garçons}\right)\right]}{\left(f_{Droit\ /\ Filles} \times f_{Filles}\right)}$$
$$\bar{x}_{Filles,\,Droit} = \frac{\left[11{,}5 \times 0{,}47\right] - \left[11{,}0 \times (0{,}57 \times 0{,}41)\right]}{0{,}40 \times 0{,}59} = 12{,}0$$

On adopte la même démarche pour le calcul de la moyenne des notes des étudiantes en sciences humaines. La proportion de garçons et de filles dans l'ensemble de la population étudiante est bien sûr la même.

$$\bar{x}_{Sc.Hum.} = \frac{\left[\bar{x}_{Garçons,\,Sc.Hum.} \times f_{Garçons,\,Sc.Hum.}\right] + \left[\bar{x}_{Filles,\,Sc.Hum.} \times f_{Filles,\,Sc.Hum.}\right]}{f_{Sc.Hum.}}$$
$$\bar{x}_{Sc.Hum.} = \frac{\left[\bar{x}_{Garçons,\,Sc.Hum.t} \times \left(f_{Sc.Hum.\ /\ Garçons} \times f_{Garçons}\right)\right] + \left[\bar{x}_{Filles,\,Sc.Hum.} \times \left(f_{Sc.Hum.\ /\ Filles} \times f_{Filles}\right)\right]}{f_{Sc.Hum.}}$$
$$\bar{x}_{Filles,\,Sc.Hum.} = \frac{\left[\bar{x}_{Sc.Hum.} \times f_{Sc.Hum.}\right] - \left[\bar{x}_{Garçons,\,Sc.Hum.} \times \left(f_{Sc.Hum.\ /\ Garçons} \times f_{Garçons}\right)\right]}{\left(f_{Sc.Hum.\ /\ Filles} \times f_{Filles}\right)}$$
$$\bar{x}_{Filles,\,Sc.Hum.} = \frac{\left[12{,}0 \times 0{,}53\right] - \left[10{,}0 \times (0{,}43 \times 0{,}41)\right]}{0{,}60 \times 0{,}59} = 13{,}0$$

La note moyenne générale des filles la suivante :

$$\bar{x}_{Filles} = \left(\bar{x}_{Filles\,,\,Droit} \times f_{Droit\ /\ Filles}\right) + \left(\bar{x}_{Filles\,,\,Sc.Hum.} \times f_{Sc.Hum.\ /\ Filles}\right)$$

$$\bar{x}_{Filles} = (12{,}0 \times 0{,}40) + (13{,}0 \times 0{,}60)$$

$$\bar{x}_{Filles} = 12{,}6$$