Statistiques à deux variables

Tableau croisé, nuage de points, ajustement affine, moindres carrés et changement de variable

Tableau croisé et fréquences

1re ens. sci. | 1re STI2D

Un tableau croisé donne l'effectif de chaque couple de modalités de deux caractères, avec les totaux.

- Fréquence marginale d'une modalité : $\dfrac{\text{total de la modalité}}{\text{effectif total}}$.
- Fréquence conditionnelle de B parmi A : $\dfrac{\text{effectif de } A \text{ et } B}{\text{effectif de } A}$; « parmi » désigne le groupe de référence, donc le dénominateur.
- Pour voir un lien entre les deux caractères, on compare les fréquences conditionnelles des différents groupes.

	Internes	Externes	Total
Filles	30	90	120
Garçons	24	56	80
Total	54	146	200

Internes : $\dfrac{54}{200} = 0{,}27$ de l'ensemble, $\dfrac{30}{120} = 0{,}25$ des filles et $\dfrac{24}{80} = 0{,}3$ des garçons.

Nuage de points et ajustement affine

1re ens. sci. | Tle compl. | 1re STI2D | Tle STI2D

- Nuage de points $(x_i ; y_i)$; point moyen $G(\bar{x} ; \bar{y})$.
- Ajustement affine : la calculatrice donne la droite des moindres carrés $y = ax + b$, qui passe par G.
- Méthode de Mayer (1re) : on coupe le nuage en deux groupes ; la droite passe par leurs points moyens G_1 et G_2.
- Interpolation (dans la plage des données) ; extrapolation (en dehors : prudence).
- Une corrélation ne prouve pas un lien de cause à effet.

Exemple : les points $(1;2)$, $(2;3)$, $(3;5)$ et $(4;6)$ ont pour point moyen $G(2{,}5;4)$; la droite des moindres carrés est $y = 1{,}4x + 0{,}5$ et passe par G : $1{,}4 \times 2{,}5 + 0{,}5 = 4$.

Moindres carrés et changement de variable

Tle compl. | Tle STI2D

- Tle compl. : $\operatorname{cov}(x,y) = \dfrac{1}{n}\sum (x_i - \bar{x})(y_i - \bar{y})$; droite des moindres carrés : $a = \dfrac{\operatorname{cov}(x,y)}{V(x)}$ et $b = \bar{y} - a\bar{x}$; coefficient de corrélation $r = \dfrac{\operatorname{cov}(x,y)}{\sigma(x)\,\sigma(y)}$, entre -1 et 1 : l'ajustement est pertinent si $|r|$ est proche de 1.
- Changement de variable : si les points $(x_i ; z_i)$ sont presque alignés, avec $z = \ln y$ (Tle compl.) ou $z = \log y$ (Tle STI2D), on ajuste $z = ax + b$, d'où $y = e^b\, e^{ax}$ ou $y = 10^b \times (10^a)^x$: modèle exponentiel.

Exemple (série ci-dessus) : $\operatorname{cov}(x,y) = 1{,}75$ et $V(x) = 1{,}25$, d'où $a = 1{,}4$ et $b = 4 - 1{,}4 \times 2{,}5 = 0{,}5$.