3ème partie: STATISTIQUE A 2 VARIABLES

Il va s'agir de la description et de la quantification de la relation entre les valeurs de 2 caractères X et Y simultanément observés sur les individus d'une population définie.
Les caractères X et Y peuvent être de nature qualitative ou quantitative.

I : Caractères qualitatifs

1°) test d'indépendance

Si les caractères sont qualitatifs, les résultats sont présentes dans un tableau à double entrée noté *tableau de contingence*.

$X/^Y$	y_1	y_2	y_j	y_k	total
x_1	n_{11}	n_{12}	n_{1j}	n_{1k}	$n_{1.}$
x_2	n_{21}	n_{22}	n_{2j}	n_{2k}	$n_{2.}$
.	.	.	.	.	.
.	.	.	.	.	.
x_i	n_{i1}	n_{i2}	n_{ij}	n_{ik}	$n_{i.}$
.	.	.	.	.	.
.	.	.	.	.	.
x_l	n_{l1}	n_{l2}	n_{lj}	n_{lk}	$n_{l.}$
total	$n_{.1}$	$n_{.2}$	$n_{.j}$	$n_{.k}$	$n_{..}$

Les distributions marginales donnent pour l'ensemble des modalités d'une variable, la répartition de l'effectif total selon les modalités de l'autre variable.
Si X et Y sont **indépendants**, alors chaque n_{ij} devrait être égal à :

$$\boxed{n'_{ij} = (n_{i.}\ n_{.j}) / n_{..}}$$

On établit le tableau des effectifs théoriques selon le modèle de l'indépendance et on effectuera le ***test de l'indépendance*** qui consiste à définir une règle de décision concernant la validité de l'hypothèse relative à l'indépendance des 2 caractères X et Y .
hypothèses :
Hypothèse nulle d'indépendance H0 : X et Y sont indépendants
Hypothèse alternative H1 : X et Y sont liés

critère statistique calculé: $\boxed{X^2 \text{ calc} = (n_{ij} - n'_{ij})^2 / n'_{ij}}$

critère statistique théorique pour un niveau de probabilité (1 - α) choisi **à priori : X² théo** interprétation du test:

X^2 calc < **X^2 théo** on ne peut pas mettre en évidence l'existence d'un lien, on garde H0

X^2 calc > **X^2 théo** on rejette H0 avec un risque a de le faire à tort. X et Y sont liés.

Il existe une autre approche si α n'est pas fixé.

conditions d'utilisation: n.. $\geq$ 20 et chaque n'_{ij} calculé $\geq$ 5

2°) test d'adéquation

Il s'agit d'une autre utilisation du critère X^2. On parle de test d'ajustement, il permet de comparer la forme de la distribution de 2 séries statistiques entre elles.

II :Caractères quantitatifs

On a souvent besoin d'examiner la façon dont une variable quantitative est reliée à une autre variable quantitative, il s'agit de la ***régression***.
On étudiera le cas où Y est relié linéairement à une seule variable X, il s'agit de régression simple. Y est la **variable dépendante** à expliquer ou variable de réponse et X est la **variable explicative** ou **variable indépendante** ou encore **régresseur**.

1°) la covariance
La covariance est la moyenne du produit des écarts de 2 variables statistiques à leurs moyennes respectives.
formule de définition: cov (x, y) = SPE (x,y) / n
formule de calcul manuel: Cov (x,y) = E(XY) – E(X)*E(Y)
La covariance peut être positive, négative ou nulle.

2°) nuage de points
L'observation de l'allure du nuage de points nous donne déjà un certain nombre de renseignements.

3°) coefficients de la régression

Les coefficients de la régression sont déterminés par la méthode des moindres carrés ordinaires notée MCO.

$$a = \bar{y} - b\,\bar{x}$$

$$b = SPE(x,y) / SCE\,x \qquad b = cov(x,y) / var\,x$$

Les coefficients a et b sont les coefficients de régression, ils s'interprètent.

La droite de régression passe nécessairement par le centre de gravité du nuage ($\bar{y}$; $\bar{x}$).
La droite de régression de y en fonction de x (droite de y pour x fixé; Dy/x) a pour équation:

$$\hat{y_i} = bx_i + a \qquad \text{ou encore} \qquad y_i = bx_i + a + e_i \qquad \text{avec } \Sigma e_i = 0$$

4°) analyse de la qualité de la régression

☺ coefficient de détermination $\mathbf{r^2}$
Il mesure la part de la variance totale qui est expliquée par la régression.
Il représente le pourcentage des variations deY expliquées par les variations de X.

variance totale des y = variance expliquée par la régression + variance résiduelle

$$\boxed{\mathbf{s^2(y) = s^2(\hat{y}) + s^2(e)}}$$

$$\mathbf{r^2 = s^2(\hat{y}) / s^2(y)}$$

mais c'est aussi :

$$\mathbf{r^2 = cov(x,y)^2 / (s^2x * s^2y)}$$

☺ coefficient de corrélation $\mathbf{r}$
Le carré du coefficient de corrélation est égal au coefficient de détermination.

☺ analyse des résidus

normalité des résidus
graphe des résidus standardisés
indépendance des résidus