

ETUDE STATISTIQUE DES MARIAGES

I- Récupérer les données

Le fichier texte donnees_mariage.txt contient des données sur les âges des époux qui se sont mariés en 2013. (source : INSEE)

Utiliser l'assistant d'importation pour copier le fichier dans une nouvelle feuille intitulée "**Donnees**".

Indiquer que le séparateur de cellule est "TAB"

Modifier les entêtes de lignes et de colonnes pour enlever le terme "**ans**".

On gardera 15 ans pour la première colonne

*On prendra 60 pour la classe **60 et plus***

II- Etude statistique

Le but est de réaliser une régression de cette série double.

La présence des effectifs empêche d'utiliser l'assistant d'Excel. Il faut alors faire les calculs "à la main".

On notera **X** la série statistique des âges du marié et **Y** la série statistique des âges de la mariée.

1) Définir un nom aux plages suivantes : **valeurs_X** , **valeurs_Y** , **effectifs_X** , **effectifs_Y** , **effectif_total** , **tableau_donnees**

2) Dans une nouvelle feuille de calcul intitulée **Regression**, calculer la liste des **ni*xi**.

*Pour cela, sélectionner une plage (ici une colonne) de la même dimension que **valeurs_X**,*

*Saisir la formule **=valeurs_X*effectifs_X** et valider par **Ctrl+Shift+Entrée**.*

*La formule devient automatiquement **{=valeurs_X*effectifs_X}**.*

3) Calculer les espérances, variances et écart-type des variables **X** et **Y**.

On utilisera la calcul matriciel

{=somme(valeurs_Y*effectifs_Y)/effectif_total}

4) Calculer E(XY).

On utilisera la formule matricielle

{=Somme(valeurs_Y*tableau_donnees*valeurs_X)}

5) Calculer finalement la covariance et le coefficient de corrélation.

Qu'indique le coefficient de corrélation ?

6) Donner l'équation de la droite de regression et tracer le graphique de la droite.

On rappelle les formules de probabilité suivantes.

$$\mathbb{E}(X) = \frac{1}{n} \sum_i n_i x_i \quad , \quad \mathbb{E}(Y) = \frac{1}{n} \sum_j n_j y_j$$

$$Var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \quad , \quad Var(Y) = \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2$$

$$\mathbb{E}(XY) = \frac{1}{n} \sum_{i,j} n_{i,j} x_i y_j$$

$$Cov(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

$$Cor(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

droite de regression : $y = ax + b$ avec

$$a = \frac{Cov(X, Y)}{Var(X)} \quad \text{et} \quad b = \mathbb{E}(Y) - a\mathbb{E}(X)$$

III- Recherches dans les données

Insérer une nouvelle feuille intitulée "Recherche".

- 1) On s'intéresse aux âges des époux avec le plus grand nombre de mariages.
 - a) Recopier les âges des époux et indiquer pour chacun quel est le plus grand nombre de mariages.
 - b) Indiquer alors, pour chaque âge d'époux, l'âge des épouses correspondant au plus grand nombre de mariages.

On utilisera la formule =equiv avec type=0.

- 2)
 - a) Tester la formule matricielle {=(valeurs_X=valeurs_Y)}.
 - b) Tester la formule matricielle {=SI(valeurs_X=valeurs_Y;1;0)}.
 - c) En déduire le nombre de mariages entre époux de même âge.
- 3)
 - a) Ecrire en colonne les écarts d'âge possibles de 0 à 45.
 - b) Calculer pour chaque écart d'âge, le nombre de mariages correspondant.
 - c) Représenter l'histogramme des écarts d'âge.

mot de passe : mariages