U.F.R. SPSE Master 1
PMP STA 21 Méthodes statistiques pour l'analyse des données en psychologie

TD4 Corrélation – Tests d'indépendance de deux variables quantitatives

Données de l'exercice

L'exercice traité porte sur les données [1] observées sur un échantillon de 474 employés tirés au sort dans une entreprise canadienne. Les variables étudiées sont les suivantes :

- `salary` : salaire brut actuel, en $/an ;
- `salbegin` : salaire de départ, en $/an ;
- `jobtime` : nombre de mois depuis l'entrée dans l'entreprise ;
- `prevexp` : expérience professionnelle antérieure (nombre de mois de travail avant l'entrée dans l'entreprise) ;
- `educ` : nombre d'années d'étude ;
- `minority` : appartenance à une minorité (0 = Non, 1 = Oui) ;
- `sex` : sexe (0 = Homme, 1 = Femme).

Exercice

1re partie. Etude descriptive des données pour un couple de variables quantitatives.

1. On a déterminé la matrice des corrélations pour l'ensemble des variables quantitatives.
   Indiquer pour quels couples de variables la corrélation linéaire observée est la plus forte, la plus faible.
2. On a tracé le nuage de points représentant les observations conjointes des deux variables *salaire de départ* et *salaire actuel*.
   Relever le coefficient de corrélation linéaire observé. Que peut-on dire de l'intensité de ce coefficient ?

2e partie. On étudie les salaires de départ et salaires actuels des employés de l'entreprise.

1. Indiquer la population et le couple de variables étudiées.
2. Tester la normalité de chacune des deux variables.
   Peut-on en conclure que le couple de variables suit une loi binormale ?
3. On veut tester l'existence d'une liaison positive entre le salaire de départ et le salaire actuel chez les employés de l'entreprise.
   (a) Peut-on utiliser le test d'indépendance basé sur le coefficient de corrélation linéaire ?
   (b) Ecrire les hypothèses du test.
   (c) Relever la valeur observée de la statistique de test et sa p-valeur.
   Prendre la décision, en précisant le risque d'erreur associé à la décision.

3e partie

On s'intéresse ici aux employés hommes qui ont plus de 8 ans d'ancienneté dans l'entreprise.
On veut tester l'existence d'une liaison positive entre le salaire de départ et le salaire actuel.

1. Définir la sous-population et les variables étudiées.
2. Dans l'échantillon initial, on a sélectionné le sous-échantillon d'individus issu de la sous-population étudiée. Préciser la taille de ce sous-échantillon.
3. Quel test d'indépendance peut-on utiliser ?
4. Ecrire les hypothèses du test.
5. Relever la valeur observée de la statistique de test et sa p-valeur.
   Prendre la décision, en précisant le risque d'erreur associé à la décision.

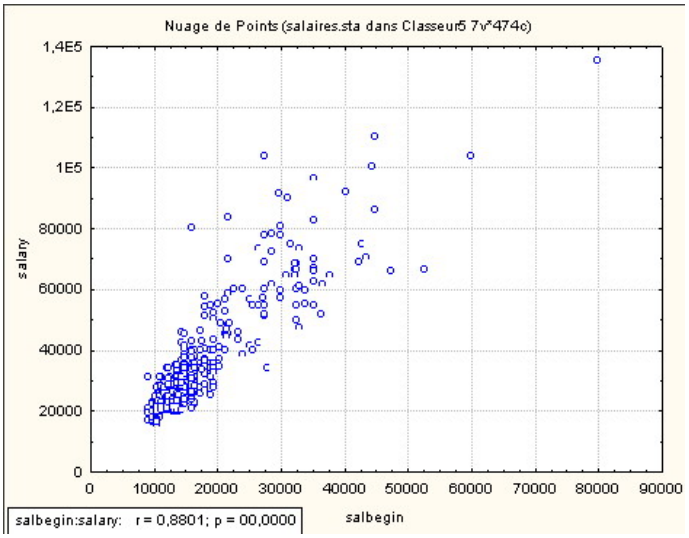
1. fichier de données du TD1

Graphiques et tableaux de résultats obtenus avec Statistica

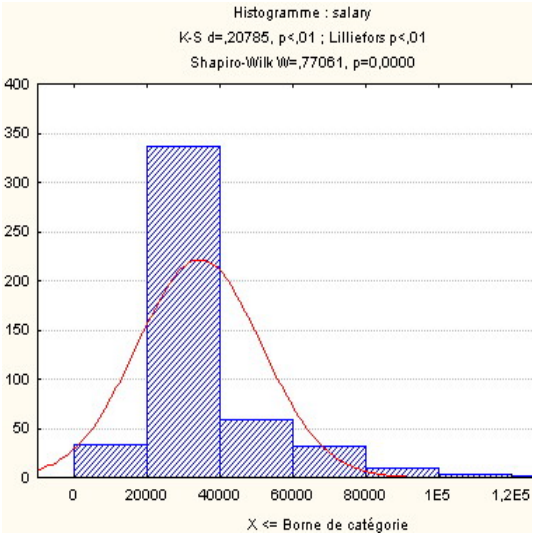
1re partie. *1. Matrice des corrélations :*

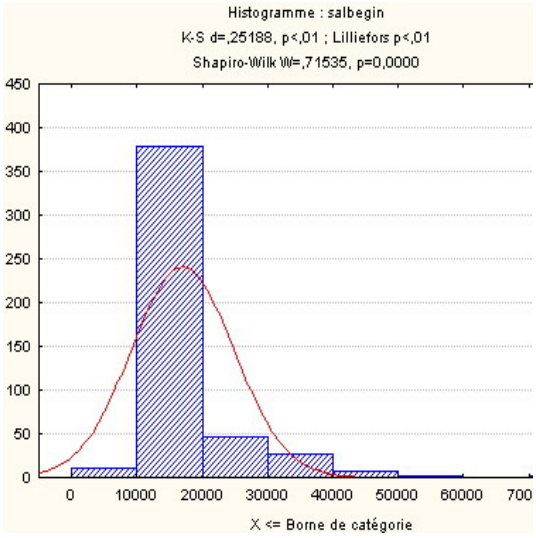
Variable	Corrélations (salaires.sta dans Classeur5) Corrélations significatives marquées à p < ,05000 N=474 (Observations à VM ignorées) salary	salbegin	jobtime	prevexp	educ
salary	1,00	0,88	0,08	-0,10	0,66
salbegin	0,88	1,00	-0,02	0,05	0,63
jobtime	0,08	-0,02	1,00	0,00	0,05
prevexp	-0,10	0,05	0,00	1,00	-0,25
educ	0,66	0,63	0,05	-0,25	1,00

2. Nuage de points :



2e partie. *2. Tests de normalité :*





3e partie. *2. Sous-échantillon :*

	1 salary	2 salbegin
1	57000	27000
2	40200	18750
3	45000	21000
4	32100	13500
5	36000	18750
6	28350	12000
7	27750	14250
8	27300	13500
9	40800	15000
10	46000	14250
11	103750	27510
12	42300	14250
13	21750	12750

5. Résultats du test de Spearman :

Couples de variables	Coeffs de Corrélations de Rangs de Spearman (F Cellules à VM ignorées Corrélations significatives marquées à p <,05000 N Actifs	R de Spearman	t(N-2)	niv. p
salary & salbegin	13	0,786731	4,226967	0,001420