

Examen de Logiciels Statistiques

1. Analyse de variables QT

On s'intéresse aux variables quantitatives du dossier SALARY qui sont des données correspondant à une étude sur les salaires d'employés en informatique. On trouvera ces données et leur descriptif à l'endroit habituel K:\STAT_AD\ . Les données sont dans le fichier SALARY.DAR et le descriptif dans le fichier SALARY.TXT.

Effectuez l'étude statistique *univariée* des variables quantitatives SAL et EX. Vous fournirez un tableau résumé **court** de cette analyse et vous rédigerez ce qu'on peut en conclure. On ne demande aucun graphique.

2. Analyse de variables QL

Etudiez maintenant les variables EDU et RESP issues de ces mêmes données, à la fois en univarié et bivarié. Que peut-on en conclure ?

Vous fournirez, parmi tous les tableaux résumés possibles, celui qui est à la fois le plus concis et le plus explicite possible en termes de calculs absolus et relatifs.

3. Une analyse de variance

On voudrait maintenant savoir s'il y a une différence au niveau des salaires entre les personnes qui exercent des responsabilités et celles qui n'en exercent pas. Réalisez le ou les tests correspondants. Vous fournirez l'interprétation statistique traditionnelle (p-value, hypothèse nulle...) et l'interprétation métier ("différence [*non*] significative au seuil de ...").

On ne demande, pour cette analyse de variance, aucun graphique, mais vous choisirez soigneusement ce que vous fournirez comme résultats numériques.

4. Une régression linéaire

Enfin, on veut savoir s'il y a une relation linéaire entre le nombre d'années d'expérience et le salaire. Vous réaliserez la régression linéaire correspondante en prenant bien soin du sens de la relation. Vous discuterez d'une causalité éventuelle et de la présence d'une relation d'un type autre que linéaire.

On ne demande aucun graphique.

5. Discussion

Essayer de répondre en une dizaine de lignes à la question suivante :

Est-il préférable d'utiliser le logiciel SAS ou le logiciel R pour une initiation aux statistiques ?

Pour faire « *évolué(e)* », on utilisera au moins 3 mots de 4 syllabes ou plus.

ANNEXE 1 : DESCRIPTIF DES DONNEES *WAGES*

Les données initiales sont issues du site STATLIB, dans la rubrique DATA-SET. Le dossier correspondant a pour nom "CPS_85_Wages" :

http://lib.stat.cmu.edu/datasets/CPS_85_Wages

Ces données proviennent de la CPS "Current Population Survey", datant de 1985. Elles complètent les informations entre les différentes années de recensement. Le dossier correspond à un échantillonnage de 534 personnes, pris au hasard, sur 11 variables qui sont les suivantes :

Variable	Description
IDEN	Cette colonne de données est un identificateur de la personne interrogée.
EDUC	Nombre d'années de scolarité.
EXP	Nombre d'années d'expérience professionnelle.
SALAIRE	Salaire de la personne interrogée. Il est exprimé en dollars par heure.
AGE	Age de la personne interrogée (en années).
SUD	Région de résidence principale : 1=Sud ; 0=Ailleurs.
SEXE	Sexe de la personne interrogée : 0=Homme ; 1=Femme.
SYNDI	Membre d'un syndicat : 1=membre ; 0=non membre.
ETHNIE	Ethnie de la personne interrogée : 3=blanc ; 2=hispanique ; 1=autre.
OCCUP	Statut professionnel de la personne interrogée : 6=autre ; 5=profession libérale, 4=militaire ; 3=employé ; 2=vente ; 1=direction.
SECTEUR	Secteur professionnel : 2=construction ; 1=manufacture ; 0=autre.
MARR	Situation maritale de la personne interrogée : 0=non marié ; 1=marié.

ANNEXE 2 : EXTRAIT DES DONNEES *WAGES*

IDEN	EDUC	EXP	SALAIRE	AGE	SUD	SEXE	UNION	RACE	OCCUP	SECTEUR	MARR
P001	8	21	5	35	0	1	0	2	6	1	1
P002	9	42	5	57	0	1	0	3	6	1	1
P003	12	1	7	19	0	0	0	3	6	1	0
P004	12	4	4	22	0	0	0	3	6	0	0
P005	12	17	8	35	0	0	0	3	6	0	1
P006	13	9	13	28	0	0	1	3	6	0	0
P007	10	27	4	43	1	0	0	3	6	0	0
P008	12	9	19	27	0	0	0	3	6	0	0
P009	16	11	13	33	0	0	0	3	6	1	1
P010	12	9	9	27	0	0	0	3	6	0	0
P011	12	17	11	35	0	0	1	3	6	0	1
P012	12	19	12	37	0	0	1	3	6	1	0
P013	8	27	7	41	1	0	0	3	6	0	1
P014	9	30	6	45	1	0	1	3	6	0	0
P015	9	29	20	44	1	0	0	3	6	0	1
P016	12	37	7	55	0	0	0	3	6	2	1
P017	7	44	8	57	1	0	0	3	6	0	1
P018	12	26	22	44	0	0	1	3	6	1	1
P019	11	16	4	33	0	0	0	3	6	0	0
P020	12	33	21	51	0	0	0	3	6	0	1
...											
P522	16	17	23	39	0	0	0	3	5	1	1
P523	16	10	8	32	0	0	1	3	5	0	1
P524	17	2	7	25	1	0	0	3	5	0	1
P525	9	34	6	49	1	1	1	1	5	0	1
P526	15	11	8	32	0	1	0	3	5	0	1
P527	15	10	13	31	0	0	0	3	5	0	0
P528	12	12	16	30	1	0	0	3	5	0	1
P529	16	6	12	28	0	1	1	3	5	0	0
P530	18	5	11	29	0	0	0	3	5	0	0
P531	12	33	6	51	0	1	0	1	5	0	1
P532	17	25	23	48	0	1	1	1	5	0	1
P533	12	13	20	31	1	0	1	3	5	0	1
P534	16	33	15	55	0	0	0	3	5	1	1