

# Statistique à deux variables quantitatives

## Exercice

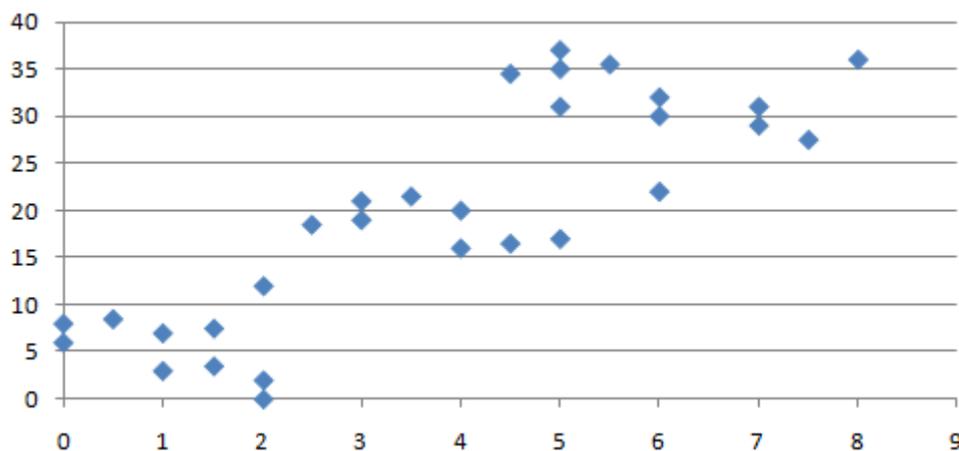
Le ministère du travail cherche à savoir si les salariés du privé sont satisfaits de leur travail et surtout à comprendre les raisons pour lesquelles ils le sont ou non.

Une enquête est menée par un statisticien auprès d'un échantillon de salariés et elle donne les résultats suivants.

responsabilités	0	1	2	0	1,5	2	0,5	1	1,5	2	2,5	3	4	3	4,5
satisfaction	6	3	0	8	4	2	9	7	8	12	19	19	16	21	17
responsabilités	3,5	4	5	6	5	4,5	6	5	6	7	7,5	5	5,5	7	8
satisfaction	22	40	17	22	31	35	30	35	32	29	28	37	36	31	36

Pour  $i$  allant de 1 à 30,  $x_i$  est un indice (en rouge) proportionnel au niveau de responsabilité du salarié  $i$ ,  $y_i$  est un indice (en bleu) proportionnel au niveau de satisfaction au travail du salarié  $i$ .

Le nuage de points correspondant à la série des  $(x_i; y_i)$  pour  $i$  allant de 1 à 30 est le suivant.



Il semblerait donc que, plus les responsabilités augmentent, plus la satisfaction au travail augmente...

1. Le statisticien a mélangé ses papiers. Il sait que le coefficient de corrélation linéaire  $r$  de la série double et la droite de régression de  $y$  en  $x$  d'équation  $y = ax + b$  vérifient:

Soit:  $a \approx 4,32$  et  $b \approx 3,27$  et  $r \approx -0,847$

Soit:  $a \approx 4,29$  et  $b \approx 3,35$  et  $r \approx 0,84$

Soit:  $a \approx -3,43$  et  $b \approx 37,51$  et  $r \approx 0,39$

(les coefficients sont arrondis à 0,01 près).

Quelle est la bonne réponse?

2. En sciences humaines, on considère que la corrélation est forte si  $|r| > 0,5$ . Y a-t-il une corrélation affine entre les niveaux de responsabilité et les niveaux de satisfaction?
3. Tout laisse donc à penser qu'effectivement, plus les responsabilités augmentent, plus la satisfaction au travail augmente...

En supposant que la corrélation ne soit pas due au hasard, le ministère du travail se demande s'il y a vraiment un lien de cause à effet entre les 2 variables, où s'il existe une "variable de confusion" (voir exercice 3 (menu-exercices-statistiques-ex3.html)).

Le ministère soupçonne que la variable "cachée" (dont dépendraient les 2 autres) serait potentiellement le salaire.

Or, l'enquête précédente a permis aussi de collecter les rémunérations des salariés.

Les salariés sont alors répartis en 3 groupes suivants leurs rémunérations (annuelles en milliers d'euros), et on obtient les 3 séries qui suivent.

salaires	20	20	20	21	21	21	22	22	23	26
responsabilités	0	1	2	0	1,5	2	0,5	1	1,5	2
satisfaction	6	3	0	8	4	2	9	7	8	12
salaires	30	31	31	32	32	33	33	33	37	
responsabilités	2,5	3	4	3	4,5	3,5	4	5	6	
satisfaction	19	19	16	21	17	22	40	17	22	
salaires	40	41	41	42	42	42	42	43	43	43
responsabilités	5	4,5	6	5	6	7	7,5	5	5,5	7
satisfaction	31	35	30	35	32	29	28	37	36	31

Dans chacune des séries, les salaires sont proches. Pour obtenir ce résultat, 3 salariés (en jaune) ne font partie d'aucune des 3 séries.

Les nuages de points correspondant à chacune des 3 séries sont donnés ci-dessous.

On y représente toujours l'indice de satisfaction en fonction de l'indice de responsabilité.



De quel signe seront les coefficients de corrélation linéaire des ces 3 séries?

Déterminer à l'aide de votre calculatrice les coefficients de corrélation linéaire  $r_1$ ,  $r_2$  et  $r_3$  de chacune des 3 séries (arrondis à 0,01 près).

4. Conclure.

5. Pour confirmer votre conclusion, déterminer à l'aide de votre calculatrice le coefficient de corrélation linéaire  $r_s$  de la série double (arrondi à 0,01 près) donnant l'indice de satisfaction en fonction du salaire.

### Corrigé

1. Graphiquement, on constate que le nuage de point peut être ajusté par une droite croissante. Donc nécessairement:  $a > 0$  et  $r > 0$ .

Donc, par élimination, la bonne réponse est:

$$a \approx 4,29 \text{ et } b \approx 3,35 \text{ et } r \approx 0,84$$

2. Le coefficient de corrélation linéaire  $r$  de la série double vérifie:  $r \approx 0,84$ .

C'est très correct! On a largement  $|r| > 0,5$ . L'ajustement est donc très satisfaisant.

Il y a effectivement **une corrélation affine entre les niveaux de responsabilité et les niveaux de satisfaction**.

3. La forme des nuages de points laisse penser qu'on peut les ajuster par des droites, et que ces droites seront toutes "décroissantes", c'est à dire de coefficients directeurs négatifs. Et par là, les coefficients de corrélation linéaire des ces 3 séries devraient être **négatifs**.

A l'aide de la calculatrice, on obtient:  $r_1 \approx -0,72$ ,  $r_2 \approx -0,54$  et  $r_3 \approx -0,79$ .

Les coefficients sont bien négatifs.

4. On constate que:  $|r_i| > 0,5$  pour  $i$  allant de 1 à 3.

Les 3 ajustements sont donc satisfaisants (voire très satisfaisants pour 2 d'entre eux).

Il y a, à nouveau, une corrélation affine entre les niveaux de responsabilité et les niveaux de satisfaction.

Le problème, c'est que, cette fois-ci, la corrélation est négative!

A salaire proche, plus les responsabilités augmentent, plus la satisfaction au travail diminue!

Finalement, il semble que **la satisfaction au travail augmente avec le salaire, mais que les responsabilités la font diminuer.**

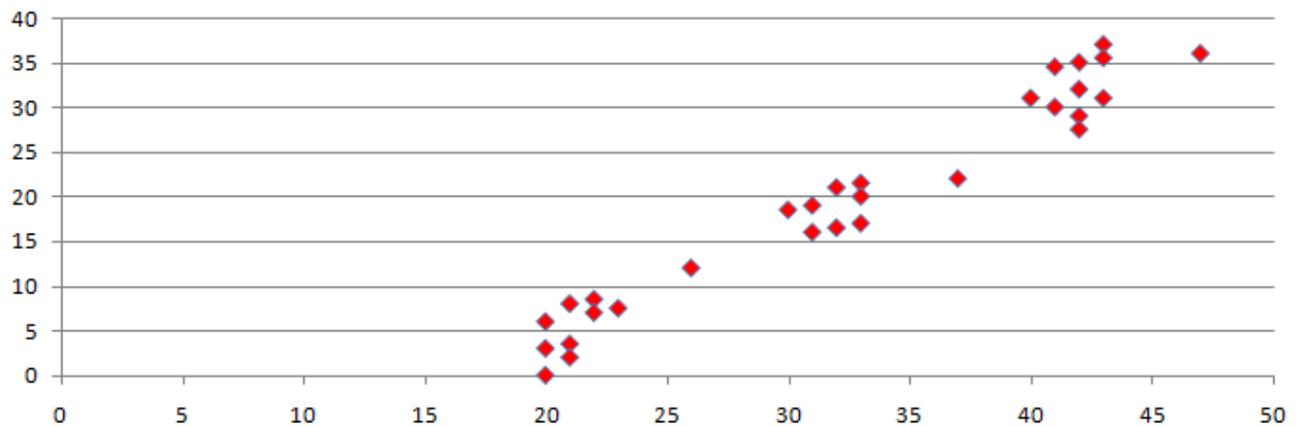
5. A l'aide de la calculatrice, on trouve que le coefficient de corrélation linéaire  $r_s$  de la série double donnant l'indice de satisfaction en fonction du salaire vérifie:

$$r_s \approx 0,98.$$

C'est très proche de 1. La corrélation est remarquable.

Conclusion: **la satisfaction au travail augmente surtout grâce au salaire, et ce malgré la hausse des responsabilités qui y est liée.**

Pour information, voici le nuage de points donnant l'indice de satisfaction en fonction du salaire.



Réduire...