

TD6 : Paramètres d'association (paramètres d'association (suite) et droite de régression)

Questions de réflexion

1/ En cours, nous avons notamment parlé de corrélation, de variance expliquée et de relation de causalité. Définissez brièvement ces différentes notions. Certaines d'entre elles sont-elles synonymes ?

2/ Quels sont les points communs entre la régression linéaire et la corrélation ? Et les différences ? Justifiez vos réponses.

3/ La covariance est-elle suffisante pour estimer la relation existant entre deux variables ? Justifiez vos réponses.

4/ Pour estimer la relation existant entre deux variables, plusieurs outils sont à votre disposition. Lesquels, et qu'est-ce que chacun vous permet de conclure quant à la relation entre les variables ?

5/ On s'intéresse à la relation entre le temps de révision avant de dormir la veille d'un examen et la note finale à cet examen. Voici nos données recueillies sur six étudiants :

Numéro de l'étudiant	Temps de révision (en mn)	Note à l'examen (sur 20)
1	2	8
2	240	9
3	30	17
4	60	15
5	20	15
6	10	14

Supposons que l'on vous demande de calculer la covariance entre ces deux variables. Quelle serait la valeur du n ?

6/ Quelle est la méthode pour déterminer rapidement si une observation est un « outlier » par rapport à un ensemble de données ? Donnez deux exemples.

Exercices

A) On s'intéresse à la relation entre le nombre de personnes présentes sur un lieu où se déroule une agression et le taux d'assistance à la victime. Dans cette perspective, une expérience en psychologie sociale a été montée, dans laquelle une agression en milieu extérieur était mise en scène par des compères de l'expérimentateur alors que des personnes naïves¹ étaient témoins de la scène. A chaque itération de l'expérience, le nombre de personnes naïves présentes et l'apport ou non d'une aide à la « victime » par un ou plusieurs des témoins naïfs étaient relevés.

L'expérience ayant été conduite sur un nombre de situations important, les situations dans lesquelles le nombre de personnes présentes était identique ont été rassemblées, et un pourcentage de situations où une aide a été apportée à la « victime » a été calculé sur la base des données recueillies par les chercheurs. Les données de cette expérience sont résumées dans le tableau ci-dessous.

Nombre de personnes naïves présentes	% de situations où une aide a été apportée à la victime
1	78%
2	60%
3	40%
4	20%
5	10%
6	5%

Le but des chercheurs dans cette expérience était de prédire le % de situations où une aide serait apportée à la victime en fonction du nombre de personnes présentes. Leur souhait était de généraliser ces résultats à une population plus importante.

- 1/ Calculez le coefficient de corrélation de Bravais-Pearson (r) et la variance expliquée de ces données.
- 2/ Tracez le diagramme de dispersion des données.
- 3/ Déterminez la formule de régression linéaire correspondant à ces données et tracez la droite de régression correspondante sur le diagramme de dispersion.
- 4/ Concluez sur le lien entre les deux variables sur la base de vos résultats aux questions 1, 2 et 3.

¹ En psychologie, on parle de participant naïf pour désigner un individu qui ne sait pas qu'il est inséré dans un dispositif de recherche, que son comportement est observé ou qui se retrouve dans une situation où d'autres personnes présentes sont des compères de l'expérimentateur.

B) On s'intéresse à la relation entre le nombre de critiques négatives faites à des enfants sur leur comportement et le niveau de réussite scolaire de ces derniers (mesuré par la moyenne générale). Pour 8 enfants de niveau CM1, ont été relevés le nombre moyen de critiques négatives sur le comportement notées sur leurs bulletins scolaires, et la moyenne générale (sur 20) obtenue en fin d'année scolaire. Les résultats sont reportés dans le tableau ci-dessous :

Prénom de l'enfant	Nombre moyen de critiques négatives sur le comportement	Moyenne générale
Delphine	2	8
Morgane	1	9
Melody	5	3
Matthieu	3	5
Jessica	4	6
Roman	0	12
Gwladys	5	7
Gabriel	0	13

Le but des chercheurs était de déterminer si l'on peut prédire la moyenne générale en fonction du nombre de critiques négatives. Ici, les chercheurs s'intéressaient uniquement aux données recueillies et ne souhaitaient pas généraliser leurs résultats à la population.

- 1/ Calculez le coefficient de corrélation de Bravais-Pearson (r) et la variance expliquée de ces données.
- 2/ Déterminez la formule de régression linéaire correspondant à ces données.
- 3/ Tracez le diagramme de dispersion et représentez la droite de régression sur ce diagramme.
- 4/ Concluez sur la relation entre les variables sur la base de vos réponses aux questions 1 et 3.

5/ Nous avons rajouté un enfant à nos données initiales. Voici le tableau actualisé :

Prénom de l'enfant	Nombre moyen de critiques négatives sur le comportement	Moyenne générale
Delphine	2	8
Morgane	1	9
Melody	5	3
Matthieu	3	5
Jessica	4	6
Roman	0	12
Gwladys	5	7
Gabriel	0	13
Isabella	12	20

Tracez le diagramme de dispersion sur la base de ces nouvelles données, et calculez le coefficient de corrélation. Que pensez-vous de vos résultats ? Comment qualifieriez-vous les données correspondant au sujet qui a été rajouté, et comment pouvez-vous vous assurer de son statut par rapport au reste de la distribution ?

C) Dans différents centres pour adolescents présentant des troubles du comportement, ont été relevés le pourcentage de familles présentant un fonctionnement de type double bind² (évalué sur la base de plusieurs entretiens avec l'ensemble de la famille et le personnel) et le pourcentage d'adolescents diagnostiqués comme schizophrènes. Les résultats sont reportés dans le tableau ci-dessous :

Numéro du centre pour adolescent	% de familles présentant un fonctionnement double bind	% de diagnostics de schizophrénie
1	12%	6%
2	6%	3%
3	2%	1%
4	15%	10%
5	11%	13%
6	10%	18%

Le but des chercheurs était de déterminer si l'on peut prédire le % de diagnostics de schizophrénie en fonction du % de familles présentant un fonctionnement double bind. L'objectif était de généraliser ces résultats à la population dont sont issues ces données.

1/ Tracez le diagramme de dispersion des données.

2/ Calculez le coefficient de corrélation de Bravais-Pearson (r) et la variance expliquée de ces données.

3/ Interprétez la relation entre les deux variables sur la base de vos résultats aux questions 1 et 2.

4/ Déterminez la formule de régression linéaire correspondant à ces données et tracez la droite de régression correspondante sur le diagramme de dispersion.

² La notion de double bind (ou double contrainte) fait référence à l'utilisation récurrente d'un langage paradoxal. Dans le cadre familial, cela signifie que la communication est pathologique entre les différents membres, et se traduit par l'emploi récurrent d'injonctions paradoxales ou contradictoires (exemple : Un parent qui va dire à son enfant « Viens me faire un câlin » tout en ayant une posture rigide et froide, traduisant le désir que l'enfant reste à distance. L'enfant reçoit donc à la fois un message l'invitant à s'approcher et un message l'invitant au contraire à se tenir à distance).