



La corrélation statistique ^[1]

[Explorable.com](#) ^[2] 11.3K reads

La corrélation statistique est une technique statistique qui s'utilise pour voir si deux variables sont liées.

Par exemple, considérez le revenu et les dépenses d'une famille comme variables. Il est bien connu que l'augmentation ou la diminution des revenus et des dépenses, se passent au même temps. Ainsi, ils sont liés tous les deux dans le sens où le changement dans une de ces deux variables est accompagné d'un changement dans l'autre variable.

^[3]

Encore une fois des prix et de la demande de produits sont des variables connexes; lorsque la demande des hausses de prix ont tendance à diminuer et vice versa.

Si le changement dans une variable est accompagné d'un changement dans l'autre, puis les variables sont dites d'être corrélées. Nous pouvons donc dire que le revenu familial et les dépenses, - le prix et la demande- sont corrélés.

Relation entre les variables

La corrélation peut vous dire quelque chose au sujet de la relation entre les variables ^[4]. Il est utilisé pour comprendre:

1. si la relation est positive ou négative
2. la force de la relation.

La corrélation est un outil puissant qui fournit ces éléments d'information essentiels.

Dans le cas de revenus et les dépenses de la famille, il est facile de voir qu'ils évoluent en augmentant ou tombant ensemble dans la même direction. C'est ce qu'on appelle une corrélation positive.

Dans le cas du prix et de la demande, le changement se produit dans la direction opposée de façon que si une augmentation se produit dans une variable, elle est accompagnée par une diminution de l'autre. C'est ce qu'on appelle une corrélation négative.

Coefficient de Corrélation

La corrélation statistique ^[5] est mesurée par ce qu'on appelle le Coefficient de corrélation

(r). Sa valeur numérique varie de 1,0 à -1,0. Cela nous donne une indication de la force de la relation.

En général, $r > 0$ indique une relation positive, $r < 0$ indique une relation négative tandis que $r = 0$ indique l'absence de relation (ou que les variables sont indépendantes et non liées). Ici, $r = +1,0$ décrit une corrélation positive parfaite et $r = -1,0$ décrit une corrélation négative parfaite.

Si les coefficients sont plus proche de +1,0 et -1,0, plus grande est la force de la relation entre les variables.

En règle générale, les lignes directrices suivantes sur la force de la relation sont souvent utiles (bien que de nombreux experts sont plutôt en désaccord sur le choix des limites).

Valeur de r	Force de la relation
-1,0 À -0,5 ou 1,0 à 0,5	Fort
-0,5 À -0,3 ou 0,3 à 0,5	Modéré
-0,3 À -0,1 ou 0,1 à 0,3	Faible
-0,1 À 0,1	Absente ou très faible

La corrélation est seulement approprié pour examiner la relation entre les données significatives quantifiables (par exemple l'air sous pression, température) plutôt que des données catégoriques tels que le sexe, la couleur préférée, etc

Inconvénients

Alors que 'r' (coefficient de corrélation) est un outil puissant, il doit être manipulé avec soin.

1. Les coefficients de corrélation les plus utilisés ne mesurent relation linéaire [6]. Il est donc parfaitement possible que s'il existe une forte relation non linéaire [7] entre les variables [3], r est proche de 0 ou même est 0. Dans un tel cas, un diagramme de dispersion peut grossièrement indiquer l'existence ou non d'une relation non linéaire.
2. Il faut être prudent dans l'interprétation de la valeur de «r». Par exemple, on pourrait calculer «r» entre la taille de la chaussure et de l'intelligence des individus, des hauteurs et des revenus. Indépendamment de la valeur de «r», cela n'a aucun sens et est donc appelée corrélation hasard ou sans-sens
3. «R» ne doit pas être utilisé pour dire quelque chose à propos de relation de cause à effet [8]. Autrement dit, en examinant la valeur de «r», on pourrait conclure que les variables X et Y sont liées. Cependant, la même valeur de «r» ne nous dit pas si X influence Y ou l'inverse. Corrélation statistique ne doit pas être le principal outil utilisé pour étudier la causalité [9], à cause du problème avec des tierces variables [10].

URL source: <https://staging.explorables.com/fr/la-correlation-statistique>

Liens

[1] <https://staging.explorables.com/fr/la-correlation-statistique>

[2] <https://staging.explorables.com/en>

[3] <https://staging.explorables.com/research-variables>

[4] <https://staging.explorables.com/relationship-between-variables>

- [5] <http://www.surveysystem.com/correlation.htm>
- [6] <https://staging.explorable.com/linear-relationship>
- [7] <https://staging.explorable.com/non-linear-relationship>
- [8] <https://staging.explorable.com/cause-and-effect>
- [9] <https://staging.explorable.com/fr/corr%C3%A9lation-et-causalit%C3%A9>
- [10] <https://staging.explorable.com/confounding-variables>