Лекція 13.

Кореляція. Коефіцієнт кореляції

1. Кореляційний зв'язок.

2. Коефіцієнт кореляції.

3. Лінія регресії.

1. Кореляційний зв'язок.

Кореляція (від лат. *correlation* – відповідність) − статистична залежність між величинами, яка не має, взагалі кажучи, строго функціонального характеру. Кореляційна залежність виникає тоді, коли одна з величин залежить не тільки від заданої другої, а й від деяких випадкових факторів; або, коли серед умов, від яких залежать обидві величини, є загальні для них обох.

Озн. 1: Кореляційний зв'язок – це не точна залежність однієї величини від іншої. Числовим значенням однієї змінної ставиться у відповідність середнє декількох значень інших. Наприклад, між кількістю внесених на поле добрив і врожайністю пшениці існує незаперечна залежність. Але це не означає, що конкретній кількості добрив відповідає визначена величина урожаю. На урожай впливає багато інших факторів: склад і структура ґрунту, різні методи посіву і таке інше.

Кореляційний зв'язок виявляється у середньому для усієї сукупності спостережень. По відношенню ж до окремих спостережень цей зв'язок є дуже неповним і неточним. Відомо, наприклад, що існує кореляція між вагою тварини і її висотою. Це означає, що більш високі тварини звичайно важчі за низьких. Та в деяких випадках низька тварина може виявитися важчою за високу.

Кореляційний зв'язок може мати різну степінь – від повної незалежності до функціональної залежності. Крім того, характер зв'язку між різними величинами може бути різний. Тому виникає необхідність визначити форму, напрям і степінь кореляційних зв'язків.

За формою кореляція може бути *прямолінійною* і *криволінійною*, за напрямком – *прямою* і *оберненою*.

При додатній кореляції залежність між величинами буде прямою: при збільшенні однієї величини, збільшується й інша. При від'ємній кореляції залежність обернена: збільшення однієї величини пов'язано зі зменшенням другої. Степінь кореляції вимірюється різними показниками зв'язку. Такими показниками є коефіцієнт кореляції, кореляційне відношення та ін.

2. Коефіцієнт кореляції.

Озн 2: Кореляційним моментом називають математичне сподівання я добутків відхилень випадкових величин X та Y від їх середніх:

$$\mu = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\cdot\left(Y_i - \overline{Y}\right)}{n} \tag{1}$$

Числове значення μ не може бути мірою тісноти зв'язку, бо залежить від одиниць вимірювання величини X та Y. Тому вводиться поняття коефіцієнта кореляції.

Озн 3: Коефіцієнтом парної кореляції називають відношення кореляційного моменту до добутку середніх квадратичних відхилень:

$$r = \frac{\mu}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\cdot\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \cdot \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}} \tag{2}$$

де X_i, Y_i – числові значення величин, між якими встановлюється кореляційний зв'язок, $\overline{X}, \overline{Y}$ − їх середні арифметичні значення величин.

Для незалежних величин r=0, для функціональних залежностей r=±1. Якщо зростання X призводить до зростання Y, то r − додатне, якщо до зменшення − то r − від'ємне. Щоб нехтувати знаком r, вводять поняття коефіцієнта детермінації $R = r^2$, який завжди додатний.

Кореляція вважається сильною, якщо r<0,75.

<u>Приклад 1.</u> Дві випадково розподілені величини X і У. Знайти коефіцієнт кореляції:

X	40	40	39	40	41	38	42	40	42	38
Y	35	37	35	36	35	36	37	36	38	35

Розв'язання: Обчислимо середні значення $\overline{X}$ і $\overline{У}$:

$$\overline{X} = \frac{40+40+39+40+41+38+42+40+42+38}{10} = 40;$$

$$\overline{Y} = \frac{35+37+35+36+35+36+37+36+38+35}{10} = 36.$$

Для обчислення r необхідно, крім $\overline{X}$ і $\overline{У}$, визначити наступні величини: $\left(X_i - \overline{X}\right)$, $\left(Y_i - \overline{Y}\right)$, $\left(X_i - \overline{X}\right)\cdot\left(Y_i - \overline{Y}\right)$, $\left(X_i - \overline{X}\right)^2$, $\left(Y_i - \overline{Y}\right)^2$. Тому результати обчислень зведемо в таблицю:

X	У	$\left(X_i - \overline{X}\right)$	$\left(Y_i - \overline{Y}\right)$	$\left(X_i - \overline{X}\right)\cdot\left(Y_i - \overline{Y}\right)$	$\left(X_i - \overline{X}\right)^2$	$\left(Y_i - \overline{Y}\right)^2$
40	35	0	−1	0	0	1
40	37	0	1	0	0	1
39	35	−1	−1	1	1	1
40	36	0	0	0	0	0
41	35	1	−1	1	1	1
38	36	−2	0	2	4	1
42	37	2	1	0	0	0
40	36	0	0	0	4	0
42	38	2	2	4	4	4
38	35	−2	−1	2	4	1
$\overline{X}$=40	$\overline{У}$=36	$\sum = 0$	$\sum = 0$	$\sum = 8$	$\sum = 18$	$\sum = 10$

З таблиці легко отримати:

$$\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right) = 8 \sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = 18; \sum_{i=1}^{n}\left(Y_i - \overline{Y}\right) = 10.$$

Тоді $r = \dfrac{8}{\sqrt{18 \cdot 10}} \approx 0{,}596$.

3. Лінія регресії.

Прямолінійний кореляційний зв'язок характеризується рівномірною зміною середнього значення величини *У* під впливом відповідної зміни величини *X*.

Озн. 4: Рівняння парної регресії має вигляд: $У - \overline{У} = r_{xy} \cdot \dfrac{\sigma_y}{\sigma_x} \cdot \left(X - \overline{X}\right)$ (3)

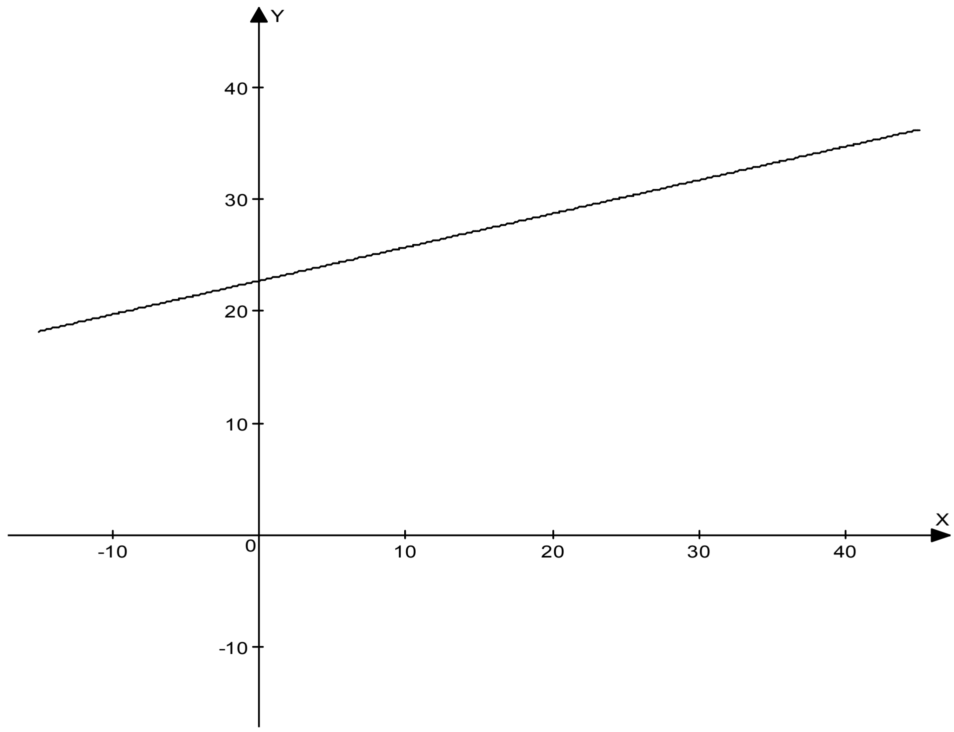
<u>Приклад 2.</u> На основі попереднього прикладу побудувати лінію регресії та позначити експериментальні точки.

<u>Розв'язання:</u> Запишемо рівняння регресії: $У - \overline{У} = r_{xy} \cdot \dfrac{\sigma_y}{\sigma_x} \cdot \left(X - \overline{X}\right)$.

З попередніх обчислень: $\overline{X}$=40, $\overline{У}$=36, $\sigma_x = 18$, $\sigma_y = 10$, $r = 0{,}596 \approx 0{,}6$, тоді:

$У - 36 = 0{,}6 \cdot \dfrac{10}{18} \cdot \left(X - 40\right) \Rightarrow У = 0{,}3X + 22{,}7$.

Побудуємо отриману функціональну залежність:



Д.з.: «Т. й.» Шевченко Р.Л., Ревицька У.С. Розділ Х. § 1–3.