Коэффициент корреляции.

Величина cov(ξ;η) зависит от единиц измерения, в которых выражаются ξ и η. (Например, пусть ξ и η—линейные размеры некоторой детали. Если за единицу измерения принять 1 см, то cov(ξ;η) примет одно значение, а если за единицу измерения принять 1 мм, то cov(ξ;η) примет другое, большее значение (при условии cov(ξ;η)≠0)). Поэтому cov(ξ;η) неудобно принимать за показатель связи.

Чтобы иметь дело с безразмерным показателем, рассмотрим случайные величины

$$\xi^* = \frac{\xi - M\xi}{\sqrt{D\xi}} = \frac{\xi - M\xi}{\sigma_\xi}; \quad \eta^* = \frac{\eta - M\eta}{\sqrt{D\eta}} = \frac{\eta - M\eta}{\sigma_\eta}$$

Такие случайные величины называются **нормированными отклонениями** случайных величин ξ и η.

Каждая из случайных величин ξ* и η* имеет центром (математическое ожидание) нуль и дисперсию, равную единице. Приведём доказательство для случайной величины ξ*.

$$M\xi^* = M\left(\frac{\xi - M\xi}{\sigma_\xi}\right) = \frac{1}{\sigma_\xi}\left(M\xi - M\left(M\xi\right)\right) = \frac{1}{\sigma_\xi}\left(M\xi - M\xi\right) = 0$$

$$D\xi^* = D\left(\frac{\eta - M\eta}{\sigma_\xi}\right) = \frac{1}{\sigma_\xi^{\ 2}} D(\eta - M\eta) = \frac{D\eta}{\sigma_\xi^{\ 2}} = 1$$

Ковариация ξ* и η* называется **коэффициентом корреляции** случайных величин ξ и η (обозначается $\rho_{\xi\eta}$).

$$\operatorname{cov}(\xi,\eta) = \rho_{\xi\eta} = M\left(\frac{\xi - M\xi}{\sigma_\xi}\cdot\frac{\eta - M\eta}{\sigma_\eta}\right) = \frac{M\left((\xi - M\xi)(\eta - M\eta)\right)}{\sigma_\xi\sigma_\eta} =$$

$$= \frac{\operatorname{cov}(\xi;\eta)}{\sigma_\xi\sigma_\eta} = \frac{M(\xi\eta) - M\xi M\eta}{\sigma_\xi\sigma_\eta}; \quad \sigma_\xi = \sqrt{D\xi};\ \sigma_\eta = \sqrt{D\eta}.$$

Для независимых ξ и η $\rho_{\xi\eta}$=0, так как в этом случае cov(ξ;η)=0

Обратного заключения сделать нельзя. Случайные величины могут быть связаны даже функциональной зависимостью (каждому значению одной случайной величины соответствует единственное значение другой случайной величины), но коэффициент их корреляции будет равен нулю.

Примеры:

1. Пусть случайная величина ξ симметрично распределена около нуля. Тогда $M\xi=0$. Пусть $\eta=\xi^2$. Тогда $M(\xi\ \eta)=M(\xi^3)=0$, так ξ^3 тоже симметрично распределена около нуля. С другой стороны $M\xi M\eta=0$, так как $M\xi=0$. Таким образом $\rho_{\xi\eta} = \dfrac{M(\xi\eta) - M\xi M\eta}{\sigma_\xi\sigma_\eta} = 0$.

2. Пусть закон совместного распределения случайных величин ξ и η задан таблицей

η / ξ	1	2	
1	1/5	0	1/5
2	0	3/5	3/5
3	1/5	0	1/5
	2/5	3/5	

Проведём вычисления:

$$M\xi = 1\cdot\frac{1}{5} + 2\cdot\frac{3}{5} + 3\cdot\frac{1}{5} = 2;\ M\eta = 1\cdot\frac{2}{5} + 2\cdot\frac{3}{5} = \frac{8}{5};$$

$$M\xi\eta = 1\cdot1\cdot\frac{1}{5} + 2\cdot2\cdot\frac{3}{5} + 3\cdot1\cdot\frac{1}{5} = \frac{16}{5};\ M\xi\eta - M\xi M\eta = 0.$$

Отсюда следует, что $\rho_{\xi\eta}=0$. При этом очевидно, что имеет место функциональная зависимость случайной величины η от случайной величины ξ.

Коэффициент корреляции $\rho_{\xi\eta}$ не меняет своей величины, если вместо случайной величины ξ рассматривать случайную величину $\xi_1=\xi+$*à* или $\xi_2=k\xi$ (*à* и

k—постоянные числа, $k > 0$), так как при перемене начала координат или при изменении масштаба величины ξ нормированное отклонение не меняется. Сказанное в равной мере относится и к η.

Вставка! Полезно запомнить формулу

$D(\xi\pm\eta)=D\xi+D\eta+2\mathrm{cov}(\xi;\eta)$

Отсюда следует свойство дисперсии для независимых ξ и η:

$D(\xi\pm\eta)=D\xi+D\eta$

<u>Свойства коэффициента корреляции.</u>

1. $-1\le\rho_{\xi\eta}\le1$
2. Если $\rho_{\xi\eta}=1$, то $\eta=k\xi+b$, где k и b—константы, $k>0$.
3. Если $\rho_{\xi\eta}= -1$, то $\eta= k\xi+b$, где $k<0$.
4. Если $\eta=k\xi+b$, $(k\neq0)$ или $\xi=k_1\eta+b_1$, то

$\rho_{\xi\eta}=1$ при $k>0$

$\rho_{\xi\eta}= -1$ при $k<0$.

Коэффициент корреляции $\rho_{\xi\eta}$ достигает своих предельных значений –1 и 1 в том и только в том случае, если совместное распределение ξ и η все концентрируется на некоторой прямой в плоскости ξ; η, то есть между ξ и η имеется такая линейная зависимость.

Если $|\rho_{\xi\eta}|<1$, то такой линейной зависимости нет. Все же по мере приближения $|\rho_{\xi\eta}|$ к единице совместное распределение ξ; η имеет тенденцию концентрироваться вблизи некоторой прямой линии и величину $|\rho_{\xi\eta}|$ можно считать мерой близости к полной линейной зависимости между ξ и η.

Пример. Рассчитаем коэффициент корреляции $\rho_{\xi\eta}$ для случайных величин при заданном законе совместного распределения

η / ξ	1	2	3	
10	1/36	0	0	1/36
20	2/36	1/36	0	3/36
30	2/36	2/36	2/36	6/36
40	1/36	9/36	16/36	26/36
	6/36	12/36	18/36	

$$M\xi = 10\cdot\frac{1}{36} + 20\cdot\frac{3}{36} + 30\cdot\frac{6}{36} + 40\cdot\frac{26}{36} \cong 35{,}83$$

$$M\eta = 1\cdot\frac{6}{36} + 2\cdot\frac{12}{36} + 3\cdot\frac{18}{36} \cong 2{,}3$$

$$D\xi = (10-35{,}83)^2\cdot\frac{1}{36} + (20-35{,}83)^2\cdot\frac{3}{36} + (30-35{,}83)^2\cdot\frac{6}{36} +$$

$$+ (40-35{,}83)^2\cdot\frac{26}{36} \cong 57{,}64$$

$\sigma_\xi \cong 7{,}6$

$$D\eta = (1-2{,}3)^2\cdot\frac{6}{36} + (2-2{,}3)^2\cdot\frac{12}{36} + (3-2{,}3)^2\cdot\frac{18}{36} \cong 0{,}556$$

$\sigma_\eta \cong 0{,}746$

$$M(\xi\eta) = 10\cdot1\cdot\frac{1}{36} + 20\cdot1\cdot\frac{2}{36} + 20\cdot2\cdot\frac{1}{36} + 30\cdot1\cdot\frac{2}{36} + 30\cdot2\cdot\frac{2}{36} +$$

$$+ 30\cdot3\cdot\frac{2}{36} + 40\cdot1\cdot\frac{1}{36} + 40\cdot2\cdot\frac{9}{36} + 40\cdot3\cdot\frac{16}{36} = 86{,}94$$

$$\rho_{\xi\eta} = \frac{86{,}94 - 2{,}3\cdot35{,}83}{7{,}6\cdot0{,}746} \cong 0{,}8$$

Введем понятие корреляционной зависимости между ξ и η.

Пусть задан закон совместного распределения двух случайных величин ξ и η (как в вышеприведенном примере), и условное математическое ожидание ξ меняется в зависимости от значения η. Тогда принято говорить о корреляционной зависимости ξ от η. Если условное математическое ожидание ξ есть линейная функция от η, то между ξ и η имеется линейная корреляционная связь или зависимость.

Как правило, говоря о корреляционной зависимости, имеют в виду линейную корреляционную зависимость. Если имеется в виду нелинейная корреляционная зависимость, то это особо оговаривают.

Можно дать определение корреляционной зависимости двух случайных величин ξ и η как связи между тенденциями роста ξ и η. Например, между ξ и η существует прямая корреляционная зависимость, если с ростом ξ случайная величина η имеет тенденцию возрастать. (Это означает, что при больших значениях ξ с большей вероятностью встречаются большие значения η). Если большим значениям ξ ñ большей вероятностью соответствуют меньшие значения η, то есть с ростом ξ случайная величина η имеет тенденцию убывать, говорят, что между ξ и η существует обратная корреляционная зависимость.

Глубина (или теснота) корреляционной зависимости (или связи) характеризуется коэффициентом $\rho_{\xi\eta}$. Чем ближе $|\rho_{\xi\eta}|$ к единице, тем теснее глубина корреляционной зависимости.

Чем ближе зависимость между условным математическим ожиданием ξ и случайной величиной η к линейной, и чем теснее значения ξ группируются около условных математических ожиданий, тем глубже (теснее) корреляционная связь.

Можно говорить о совместном распределении двух непрерывных случайных величин. В большинстве случаев возможен переход от непрерывных случайных величин к совместному распределению двух дискретных случайных величин следующим образом.

Нужно разбить отрезок $[a; b]$ изменения случайной величины ξ на равные отрезки $[c_0=a; c_1]$; $[c_1; c_2]$; $[c_2; c_3]$,…,$[c_{n-1}; c_n=b]$. За значение случайной величины ξ принять середину каждого отрезка.

Также надо поступить со случайной величиной η, разбив ее область значений $[e; f]$ на равные отрезки $[g_0 = e; g_1]$; $[g_1; g_e]$…$[g_{k-1}; g_k=f]$, и приняв за возможные значения η середины отрезков $[g_{k-1}; g_k]$. Таким образом мы получили дискретные случайные величины $\xi^*=\{x_1; x_2; \ldots x_n\}$ и $\eta^*=\{y_1; y_2; \ldots y_k\}$, причем каждой паре $(x_i; y_j)$ ставится в соответствие вероятность

$$P_i^j = P((\xi \in [c_{i-1}; c_i]) \cap (\eta \in [g_{i-1}; g_i]))$$

Таким образом мы придем к уже изученному материалу.