



КОРРЕЛЯЦИЯ

Авторы: А. В. Прохоров

КОРРЕЛЯЦИЯ в математике, зависимость между случайными величинами, не имеющая, вообще говоря, строго функционального характера. В отличие от функциональной зависимости K ., как правило, рассматривается тогда, когда одна из величин зависит не только от данной др. величины, но и от ряда иных случайных факторов.

Зависимость между двумя случайными событиями проявляется в том, что условная вероятность одного из них при условии, что другое произошло, отличается от безусловной вероятности. Аналогично, влияние одной случайной величины на другую характеризуется условными распределениями одной из них при фиксированных значениях другой.

Пусть X и Y – случайные величины с заданным совместным распределением вероятностей, a_X и a_Y – математические ожидания, σ^2_X и σ^2_Y – дисперсии и ρ – корреляции коэффициент случайных величин X и Y . Если для каждого возможного значения x случайной величины X определено условное математич. ожидание $y(x) = \mathsf{E}(Y|X=x)$, то функция $y(x)$ называется регрессией величины Y по X . Для оценки того, насколько точно регрессия передаёт изменение Y при изменении X , используется условная дисперсия Y при данном значении $X=x$ или её ср. величина (мера рассеяния Y около линии регрессии), равная $\sigma^2_{Y|X} = \mathsf{E}(Y - \mathsf{E}(Y|X))^2$.

Если X и Y независимы, то условные математич. ожидания Y не зависят от x и совпадают с безусловным, т. е. $y(x) = a_Y$, при этом $\sigma^2_{Y|X} = \sigma^2_Y$. При функциональной связи между Y и X величина Y при каждом данном $X=x$ принимает одно значение и $\sigma^2_{Y|X} = 0$. Аналогично определяется $x(y) = \mathsf{E}(X|Y=y)$ – регрессия X по Y . Показателем концентрации распределения вблизи линии регрессии $y(x)$ служит корреляционное отношение $\eta^2_{Y|X} = (\sigma^2_Y - \sigma^2_{Y|X}) / \sigma^2_Y = 1 - \sigma^2_{Y|X} / \sigma^2_Y$. Величина $\eta^2_{Y|X}$ равна нулю тогда и только тогда, когда регрессия имеет вид $y(x) = a_Y$, в этом случае коэф. $K. \rho$ равен нулю и величина Y не коррелирована с X . Если регрессия Y по X линейна, т. е. линия регрессии – прямая, имеющая вид $y(x) = a_Y + \rho \frac{\sigma_Y}{\sigma_X} (x - a_X)$, то $\sigma^2_{Y|X} = \sigma^2_Y (1 - \rho^2)$ и $\eta^2_{Y|X} = \rho^2$. Если, кроме того, $|\rho| = 1$, то Y связана с X точной линейной зависимостью, если же $\eta^2_{Y|X} = \rho^2 < 1$, то между Y и X нет точной функциональной зависимости. Точная функциональная зависимость Y от X , отличная от линейной, имеет место тогда и только тогда, когда $\rho^2 < 1$ и $\eta^2_{Y|X} = 1$. Практич. использование коэф. $K.$ в качестве меры отсутствия зависимости оправданно (за редким исключением) лишь тогда, когда совместное распределение X и Y нормально (или близко к нормальному распределению), т. к. в этом случае из равенства $\rho = 0$ следует независимость X и Y . Для произвольных случайных величин X и Y использование ρ как меры зависимости часто приводит к ошибочным выводам, т. к. ρ может равняться нулю даже при функциональной связи между величинами. Если совместное распределение X и Y нормально, то обе линии регрессии $y(x)$ и $x(y)$ суть прямые, при $|\rho| = 1$ прямые регрессии сливаются в одну, что соответствует линейной зависимости между X и Y , при $\rho = 0$ величины X и Y независимы.

При изучении связи между несколькими случайными величинами X_1, \dots, X_n с заданным совместным распределением используется корреляционная матрица, элементами которой являются обычные коэффициенты $K. \rho_{\{ij\}}$ между X_i и X_j , $i, j = 1, \dots, n$. Мерой линейной $K.$ между X_1 и совокупностью остальных величин X_2, \dots, X_n служит множественный коэф. $K.$, который определяется как обычный коэф. $K.$ между X_1 и наилучшим линейным приближением X_1 по X_2, \dots, X_n , т. е. между X_1 и $\beta_1 + \beta_2 X_2 + \dots + \beta_n X_n$, где числа β_1, \dots, β_n определяются так, чтобы дисперсия величины $X_1 - (\beta_1 + \beta_2 X_2 + \dots + \beta_n X_n)$ была минимальной. Множественный коэф. $K.$ выражается через элементы корреляционной матрицы, напр. при $n=3$ он равен $\rho_{\{1\} \cdot (23)} = \sqrt{\frac{\rho_{\{12\}}^2 + \rho_{\{13\}}^2 - 2\rho_{\{12\}}\rho_{\{13\}}\rho_{\{23\}}}{1 - \rho_{\{23\}}^2}}$.

Если предполагается, что изменение величин X_1 и X_2 определяется в какой-то мере изменением остальных величин X_3, \dots, X_n , то показателем линейной связи между X_1 и X_2 при исключении влияния X_3, \dots, X_n является частный коэф. $K.$ между X_1 и X_2 относительно X_3, \dots, X_n , который определяется как обычный коэф. $K.$ между $X_1 - X_1^*$ и $X_2 - X_2^*$, где X_1^*, X_2^* – соответственно наилучшие линейные приближения X_1 и X_2 по X_3, \dots, X_n . Напр., в случае $n=3$ этот коэф. равен $\rho_{\{12\} \cdot 3} = \frac{\rho_{\{12\}} - \rho_{\{12\}}\rho_{\{23\}}}{\sqrt{(1 - \rho_{\{13\}}^2)(1 - \rho_{\{23\}}^2)}}$. В математич. статистике разработаны методы оценки $K.$ между случайными величинами и методы проверки гипотез о значениях $K.$, использующие их выборочные аналоги. См. [Корреляционный анализ](#).

Литература

Лит.: Кендалл М., Стьюарт А. Статистические выводы и связи. М., 1973.

Processing math: 0%