



КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Авторы: А. В. Прохоров

КОРРЕЛЯЦИОННЫЙ АНАЛИЗ, раздел [математической статистики](#), предназначенный для оценки [корреляции](#) между случайными величинами и проверки гипотез о значимости связи между ними. К. а. статистич. данных использует следующие осн. практич. приёмы: построение корреляционного поля и составление корреляционной таблицы; вычисление выборочных коэффициентов корреляции; проверку статистич. гипотез значимости связи. Дальнейшее исследование может заключаться в установлении конкретного вида зависимости между величинами (см. [Регрессионный анализ](#)).

Вспомогат. средствами при анализе выборочных двумерных данных являются корреляционное поле и корреляционная таблица. Корреляционное поле получают, нанося выборочные точки на координатную плоскость. По характеру расположения точек поля можно составить предварит. представление о форме зависимости случайных величин (напр., о том, что одна величина в ср. возрастает или убывает при возрастании другой). Для численной обработки результаты обычно группируют и представляют в форме корреляционной таблицы. В каждой клетке этой таблицы приводятся численности n_{ij} тех пар (x, y) , компоненты которых попадают в соответствующие интервалы группировки по каждой переменной. Обычно длины интервалов группировки (по каждой из переменных) выбирают равными между собой, и центры x_i (соответственно y_j) этих интервалов, и числа n_{ij} используют в качестве основы для расчётов.

Корреляционная таблица позволяет, в частности, вычислить выборочный коэф. корреляции и выборочное корреляционное отношение. Выборочный коэф. корреляции определяется по формуле
$$\hat{r} = \frac{\sum_i \sum_j (x_i - \bar{x})(y_j - \bar{y})n_{ij}}{\sqrt{\sum_i n_{i\cdot} (x_i - \bar{x})^2} \sqrt{\sum_j n_{\cdot j} (y_j - \bar{y})^2}},$$
 где $n_{i\cdot} = \sum_j n_{ij}$, $n_{\cdot j} = \sum_i n_{ij}$ и $\bar{x} = \sum_i n_{i\cdot} x_i / n$, $\bar{y} = \sum_j n_{\cdot j} y_j / n$. При большом числе независимых наблюдений, подчинённых одному и тому же распределению, близкому к нормальному, \hat{r} близок к истинному [корреляции коэффициенту](#) ρ . В др. случаях в качестве характеристики связи между X и Y рекомендуется использовать корреляционное отношение $\eta^2_{Y|X}$, интерпретация которого не зависит от вида исследуемой зависимости. Выборочное значение $\hat{\eta}^2_{Y|X}$ вычисляется по данным корреляционной таблицы:
$$\hat{\eta}^2_{Y|X} = \frac{\sum_i n_{i\cdot} (\bar{y}_i - \bar{y})^2 / n}{\sum_j n_{\cdot j} (y_j - \bar{y})^2 / n},$$
 где числитель характеризует рассеяние условных средних значений \bar{y}_i около безусловного среднего \bar{y} (аналогично определяется выборочное значение $\hat{\eta}^2_{X|Y}$). Величина $\hat{\eta}^2_{Y|X}$ - \hat{r}^2 используется в качестве индикатора отклонения регрессии от линейной.

Один из методов проверки гипотезы о значимости связи между X и Y основывается на распределении выборочного коэф. корреляции. В случае нормального распределения величина выборочного коэф. корреляции \hat{r} считается значимо отличной от нуля, если выполняется неравенство $\hat{r}^2 > (1 + (n-2)t_{\alpha}^2)^{-1}$, где t_{α} есть квантиль порядка α [Стюдента распределения](#) с $n-2$ степенями свободы, соответствующая выбранному [значимости уровню](#) α . В случае $\rho \neq 0$ часто используют т. н. z -преобразование Фишера, заменяя величину \hat{r} на $z = \frac{1}{2} \ln \frac{1 + \hat{r}}{1 - \hat{r}}$. Уже при сравнительно небольших n распределение величины z

хорошо приближается нормальным распределением с математич. ожиданием, равным $\frac{1}{2} \ln \frac{1+\rho}{1-\rho} + \frac{\rho}{2(n-1)}$, и дисперсией, равной $1/(n-3)$. Из этого можно определить интервалы (доверительные границы) для истинного коэф. корреляции ρ .

Литература

Лит.: Ван дер Варден Б. Л. Математическая статистика. М., 1960; Айвазян С. А. Статистическое исследование зависимостей. М., 1968; Кендалл М., Стьюарт А. Статистические выводы и связи. М., 1973; Крамер Г.

Математические методы статистики. 3-е изд. М.; Ижевск, 2003.

Processing math: 0%