Оценка значимости коэффициента корреляции

Сальников Владислав
студент ФГОУ ВО
Курский государственный университет,
колледж коммерции, технологий и сервиса,
Россия, г. Курск

Ефимцева Ирина Борисовна,
Научный руководитель, канд. тех. наук.
преподаватель ФГОУ ВО
Курский государственный университет,
колледж коммерции, технологий и сервиса,
Россия, г. Курск

Если выборочный коэффициент корреляции получается близким к нулю, то возможно, что генеральный коэффициент корреляции равен нулю, а отклонения от нуля произошли за счёт случайности наблюдений. Выдвинем нулевую гипотезу H_o: выборочный коэффициент корреляции не значимо отличается от нуля, или генеральный коэффициент корреляции равен нулю. Для проверки данной гипотезы часто используют критерий Стьюдента, который заключается в следующем:

Сначала вычисляется наблюдаемое значение критерия $T_{набл}$ по формуле

$$T_{набл} = \left|r_{xy}\right| \cdot \sqrt{\frac{N-2}{1-r_{xy}^2}}\,. \qquad (1.1)$$

Далее выбирается уровень значимости α или доверительная вероятность $P = 1 - \alpha$. Уровень значимости выбирается в зависимости от важности и степени ответственности решаемой задачи. Если нас не пугают последствия ошибочных выводов, можно брать уровень значимости меньше. Для выбранного α и $k = N - 2$ находят соответствующий квантиль распределения Стьюдента или критическое значение $T_{табл} = t(\alpha,N)$.

Теперь, если $T_{набл} \leq T_{табл}$, то гипотеза H_o принимается, иначе, выборочный коэффициент корреляции незначимо отличен от нуля, или генеральный коэффициент равен нулю, или X и Y не коррелированы с вероятностью вывода P. В противном случае гипотеза H_o отвергается.

Если гипотеза о равенстве нулю коэффициента корреляции не подтвердилась, важно знать насколько может отличаться генеральный коэффициент корреляции от полученной для данной конкретной выборки точечной оценки выборочного коэффициента. Задают доверительную вероятность P (P = 0.9, 0.95, 0.99, …) и находят для неё интервал, в котором с вероятностью P находится генеральный коэффициент. Этот интервал называется доверительным.

Для визуального представления о линейной корреляционной связи между случайными величинами изображают две прямые – прямые регрессии Y на X и X наY. Их уравнения записываются в виде:

Y на X: $$y - M[Y] = r_{xy} \cdot \frac{S_y}{S_x} \cdot (x - M[X]), \qquad (1.2)$$

X на Y: $$x - M[X] = r_{xy} \cdot \frac{S_x}{S_y} \cdot (y - M[Y]). \qquad (1.3)$$

Прямые пересекаются в точке $M_o(x_o,y_o)$, где $x_o = M[X]$, $y_o = M[Y]$.

Обычно на графике изображают обе прямые регрессии и точки (x_i,y_i), i = 1,…,N, взятые из выборки:

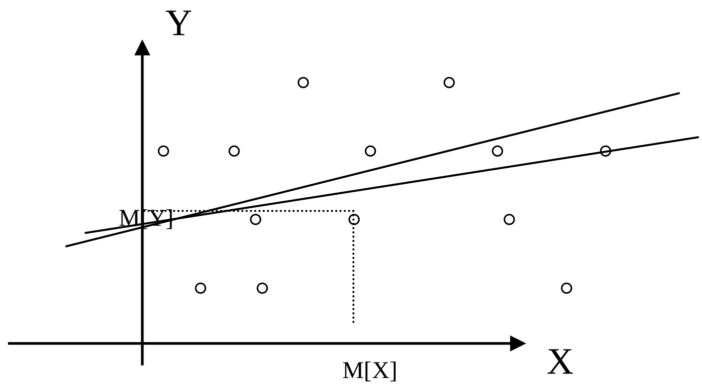

Рис.1.1. Прямые регрессии

Чем меньше угол между прямыми, тем теснее корреляционная связь между X и Y. Обе прямые регрессии могут быть получены по точкам (x_i, y_i) методом наименьших квадратов.

Чтобы найти прямую регрессии Y на X, нужно записать прямую в виде y = ax + b и найти a и b из условия минимума суммы:

$$S(a,b) = \sum_{i=1}^{N} (y_i - ax_i - b)^2 . \tag{1.4}$$

Чтобы найти прямую регрессии X на Y, нужно записать прямую в виде x = ay + b и найти a и b из условия минимума суммы:

$$S(a,b) = \sum_{i=1}^{N} (x_i - ay_i - b)^2 . \tag{1.5}$$

Если рассматриваются в совокупности сразу несколько случайных величин и выясняются линейные связи между ними, то рассматривают матрицу:

$$\begin{pmatrix} 1 & R_{12} & \dots & R_{1n} \\ R_{21} & 1 & \dots & R_{2n} \\ . & . & . & . \\ R_{n1} & R_{n2} & \dots & 1 \end{pmatrix},$$

где $R_{i\,j}$ – коэффициент корреляции между случайными величинами X_i и X_j . Эта матрица называется корреляционной матрицей системы случайных величин X_1, X_2, …, X_n. Эта матрица – симметрическая, так как $R_{i\,j} = R_{j\,i}$.

Если представляет интерес линейная зависимость одной из случайных величин от других, т.е. например, зависимость

$$X_1 = A_2X_2 + A_3X_3 + \ldots + A_nX_n + B_0 ,$$

то рассматривается коэффициент множественной корреляции

$$R_{X_1,(X_2,X_3,\ldots,X_n)} .$$

Формула для вычисления коэффициента множественной корреляции в случае трёх случайных величин X, Y, Z имеет вид:

$$R_{X,(Y,Z)} = \sqrt{\frac{R_{xy}^2 + R_{xz}^2 - 2 \cdot R_{xy} \cdot R_{xz} \cdot R_{yz}}{1 - R_{yz}^2}} \qquad (1.6)$$

Список литературы

1. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. М.: Высш. шк., 1979. 400с.
2. Пискунов Н.С. Дифференциальное и интегральное исчисление. Т. 2. М.: Наука, 1985. 576с.
3. Очков В.Ф. Mathcad 7 Pro для студентов и инженеров. М.: КомпьютерПресс, 1998. 384с.
4. MATHCAD 6.0 PLUS. Финансовые, инженерные и научные расчёты в среде WINDOWS 95. / Перевод с анг. М.: Информационно-издательский дом “Филин”, 1996. 712с.