

РЕГРЕССИЯ ЛИНЕЙНАЯ МНОЖЕСТВЕННАЯ - причинная модель статистической связи линейной (см.) между переменной зависимой (см.) y и переменными независимыми (см.) $x_1, x_2, \dots, x_k$, представленная уравнением $y = b_1x_1 + b_2x_2 + \dots + b_kx_k + a = \sum b_ix_i + a$ (см. Анализ регрессионный). Коэффициенты $b_1, b_2, \dots, b_k$ называются нестандартизированными коэффициентами, a - свободным членом уравнения регрессии. Уравнение регрессии существует также в стандартизированном виде, когда вместо исходных переменных используются их z -оценки (см. Переменная стандартизированная): $z_y = \sum \beta_i z_{x_i}$. Здесь z_y - z -оценка переменной y ; $z_1, z_2, \dots, z_k$ - z -оценки переменных $x_1, x_2, \dots, x_k$; $\beta_1, \beta_2, \dots, \beta_k$ - стандартизированные коэффициенты регрессии (свободный член отсутствует).

Для того чтобы найти стандартизированные коэффициенты, необходимо решить систему линейных уравнений:

$$\beta_1 + r_{12}\beta_2 + r_{13}\beta_3 + \dots + r_{1k}\beta_k = r_{1y},$$

$$r_{21}\beta_1 + \beta_2 + r_{23}\beta_3 + \dots + r_{2k}\beta_k = r_{2y},$$

$$r_{31}\beta_1 + r_{32}\beta_2 + \beta_3 + \dots + r_{3k}\beta_k = r_{3y},$$

...

$$r_{k1}\beta_1 + r_{k2}\beta_2 + r_{k3}\beta_3 + \dots + \beta_k = r_{ky},$$

в которой r_{ij} - коэффициенты линейной корреляции Пирсона (см.) для переменных x_i и x_j ; r_{iy} - коэффициент корреляции Пирсона для переменных x_i и y .

Нестандартизированные коэффициенты регрессии вычисляются по формуле $b_i = \beta_i \cdot s_y / s_{x_i}$, где s_y - стандартное отклонение (см.) переменной y ; s_{x_i} - стандартное отклонение переменной x_i . Свободный член уравнения регрессии находится по формуле $a = \bar{y} - \sum b_i \bar{x}_i$, где $\bar{y}$ - среднее арифметическое (см.) переменной y , $\bar{x}_i$ - средние арифметические для переменных x_i .

В настоящее время используются два подхода к интерпретации нестандартизированных коэффициентов линейной регрессии b_i . Согласно первому из них, b_i представляет собой величину, на которую изменится предсказанное по модели значение $\hat{y} = \sum b_ix_i$ при увеличении значения независимой переменной x_i на единицу измерения; согласно второму - величину, на которую в среднем изменяется значение переменной y при увеличении независимой переменной x_i на единицу. Значения коэффициентов b_i существенно зависят от масштаба шкал, по которым измеряются переменные y и x_i , поэтому по ним нельзя судить о степени влияния независимых переменных на зависимую. Свободный член уравнения регрессии a равен предсказанному значению зависимой переменной $\hat{y}$ в случае, когда все независимые переменные $x_i = 0$.

Стандартизированные коэффициенты β_i являются показателями степени влияния независимых переменных x_i на зависимую переменную y . Они интерпретируются как "вклад" соответствующей независимой переменной в дисперсию (см.) (изменчивость) зависимой переменной.

Качество (объясняющая способность) уравнения множественной линейной регрессии измеряется коэффициентом множественной детерминации (см.), который равен квадрату коэффициента корреляции множественной (см.) R^2 .

Предполагается, что все переменные (см.) в уравнении множественной линейной регрессии являются количественными. При необходимости включить в модель номинальные переменные используется техника думми-кодирования (см.).

О.В. Терещенко