

ДИСПЕРСИЯ ОБЪЯСНЕННАЯ (доля объясненной дисперсии) - универсальная мера качества (объясняющей способности) причинных статистических моделей, применяется, главным образом, в анализе дисперсионном (см.) и анализе регрессионном (см.). В некоторых случаях используется также в качестве меры связи (см. Корреляционное отношение). Интерпретируется как доля дисперсии зависимой переменной  $y$ , объясненная изменениями независимой переменной  $x$ .

Основная идея разделения дисперсии зависимой переменной  $y$  на Д.О. и дисперсию ошибки состоит в предположении, что полная вариация зависимой переменной  $y$  определяется двумя группами факторов: учтенными факторами, зафиксированными в виде независимых переменных, и неучтенными факторами, представленными "ошибками модели"  $\epsilon_i = y_i - \hat{y}_i$ , где  $\epsilon_i$  - ошибка модели для объекта с номером  $i$ ;  $y_i$  - измеренное значение переменной  $y$  для объекта с номером  $i$ ;  $\hat{y}_i$  - значение переменной  $y$  для объекта с номером  $i$ , предсказанное по модели. Заметим, что традиционно речь идет о доле Д.О., хотя вместо дисперсии используется сумма квадратов (см.).

Согласно этой концепции, полная (общая) сумма квадратов может быть разложена на сумму квадратов объясненную и сумму квадратов ошибки  $SS_{\text{ошибки}} = \sum \epsilon_i^2 = \sum (y_i - \hat{y}_i)^2$  :

$SS_{\text{общая}} = SS_{\text{объясненная}} + SS_{\text{ошибки}}$  .

Если разделить обе части этого выражения на величину  $SS_{\text{общая}}$  , то доля Д.О. может быть представлена следующим образом:

$SS_{\text{объясненная}} / SS_{\text{общая}} = 1 - SS_{\text{ошибки}} / SS_{\text{общая}} = 1 - \sum (y_i - \hat{y}_i)^2 / \sum (y_i - \bar{y})^2$  .

Вычисление доли Д.О. для разных моделей различается способами получения предсказанного значения  $\hat{y}_i$ . С этим связаны и различия в названиях составляющих общей суммы квадратов. Для регрессионных моделей предсказанное значение вычисляется по уравнению регрессии:  $\hat{y}_i = f(x_i)$ , поэтому  $SS_{\text{объясненная}}$  называется для них  $SS_{\text{регрессии}}$  . В частности, для уравнения регрессии линейной парной (см.) доля Д.О. (коэффициент детерминации - см.) равна квадрату коэффициента линейной корреляции Пирсона  $r^2$ ; для регрессии линейной множественной (см.) - квадрату коэффициента корреляции множественной  $R^2$ .

В анализе дисперсионном (см.) и корреляционном отношении  $\hat{y}_i$  определяется как среднее арифметическое (см.) переменной  $y$  для той группы объектов, в которую входит объект с номером  $i$ . Поэтому  $SS_{\text{ошибки}}$  в этих методах носит название  $SS_{\text{внутригрупповая}}$ , а  $SS_{\text{объясненная}}$  - название  $SS_{\text{межгрупповая}}$ .

О.В. Терещенко