Лекция 20

ҚАЗАҚСТАННЫҢ
АШЫҚ
УНИВЕРСИТЕТІ

ОСНОВЫ СТАТИСТИКИ

Множественная регрессия и корреляция. Проверка существенности факторов и показатели качества регрессии

В данной лекции мы рассмотрим основные показатели оценки показателей модели множественной регрессии. Освоим методы ручного расчета предложенных понятий для более глубокого понимания проблемы.

Практическая значимость уравнения множественной регрессии оценивается с помощью показателя множественной корреляции и его квадрата – показателя детерминации.

Показатель множественной корреляции характеризует тесноту связи рассматриваемого набора факторов с исследуемым признаком или, иначе, оценивает тесноту совместного влияния факторов на результат.

Независимо от формы связи показатель множественной корреляции может быть найден как индекс множественной корреляции:

Формула 20.1

$$R_{yx_1x_2\ldots x_m} = \sqrt{1 - \frac{\sigma^2_{\text{ост}}}{\sigma^2_y}}$$

где σ^2_y – общая дисперсия результативного признака; $\sigma^2_{\text{ост}}$ – остаточная дисперсия.

Границы изменения индекса множественной корреляции от 0 до 1. Чем ближе его значение к 1, тем теснее связь результативного признака со всем набором исследуемых факторов. Величина индекса множественной корреляции должна быть больше или равна максимальному парному индексу корреляции:

$$R_{yx_1x_2\ldots x_m} \geq r_{yx_i(\max)} \quad \left(i = \overline{1,m}\right)$$

При правильном включении факторов в регрессионную модель величина индекса множественной корреляции будет существенно отличаться от индекса корреляции парной зависимости. Если же дополнительно включенные в уравнение множественной регрессии факторы третьестепенны, то индекс множественной корреляции может практически совпадать с индексом парной корреляции (различия в третьем, четвертом знаках). Отсюда ясно, что сравнивая индексы множественной и парной корреляции, можно сделать вывод о целесообразности включения в уравнение регрессии того или иного фактора.

Расчет индекса множественной корреляции предполагает определение уравнения множественной регрессии и на его основе остаточной дисперсии:

Формула 20.2

$$\sigma^2_{\text{ост}} = \frac{1}{n}\sum\left(y - \hat{y}_{x_1x_2\ldots x_m}\right)^2$$

Можно пользоваться следующей формулой индекса множественной детерминации:

Формула 20.3

$$R^2_{yx_1x_2\ldots x_m} = 1 - \frac{\sum\left(y - \hat{y}_{x_1x_2\ldots x_m}\right)^2}{\sum\left(y - \bar{y}\right)^2}$$

При линейной зависимости признаков формула индекса множественной корреляции может быть представлена следующим выражением:

Формула 20.4

$$R_{yx_1x_2\ldots x_m} = \sqrt{\sum \beta_i \cdot r_{yx_i}}$$

где β_i – стандартизованные коэффициенты регрессии; r_{yxi} – парные коэффициенты корреляции результата с каждым фактором.

Формула индекса множественной корреляции для линейной регрессии получила название линейного коэффициента множественной корреляции, или, что то же самое, совокупного коэффициента корреляции.

Возможно также при линейной зависимости определение совокупного коэффициента корреляции через матрицу парных коэффициентов корреляции:

Формула 20.5

$$R_{yx_1x_2,\ldots,x_p} = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}}$$

где

$$\Delta r = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & \dots & r_{yx_p} \\ r_{yx_1} & 1 & r_{x_1x_2} & \dots & r_{x_1x_p} \\ r_{yx_2} & r_{x_2x_1} & 1 & \dots & r_{x_2x_p} \\ \dots & \dots & \dots & \dots & \dots \\ r_{yx_p} & r_{x_px_1} & r_{x_px_2} & \dots & 1 \end{vmatrix}$$ – определитель матрицы парных коэффициентов корреляции;

$$\Delta r_{11} = \begin{vmatrix} 1 & r_{x_1x_2} & \dots & r_{x_1x_p} \\ r_{x_2x_1} & 1 & \dots & r_{x_2x_p} \\ \dots & \dots & \dots & \dots \\ r_{x_px_1} & r_{x_px_2} & \dots & 1 \end{vmatrix}$$ – определитель матрицы межфакторной корреляции.

Как видим, величина множественного коэффициента корреляции зависит не только от корреляции результата с каждым из факторов, но и от межфакторной корреляции. Рассмотренная формула позволяет определять совокупный коэффициент корреляции, не обращаясь при этом к уравнению множественной регрессии, а используя лишь парные коэффициенты корреляции.

В рассмотренных показателях множественной корреляции (индекс и коэффициент) используется остаточная дисперсия, которая имеет систематическую ошибку в сторону преуменьшения, тем более значительную, чем больше параметров определяется в уравнении регрессии при заданном объеме наблюдений n. Если число параметров при x_i равно m и приближается к объему наблюдений, то остаточная дисперсия будет близка к нулю и коэффициент (индекс) корреляции приблизится к единице даже при слабой связи факторов с результатом. Для того чтобы не допустить возможного преувеличения тесноты связи, используется скорректированный индекс (коэффициент) множественной корреляции.

Скорректированный индекс множественной корреляции содержит поправку на число степеней свободы, а именно остаточная сумма квадратов $\sum\left(y-\hat{y}_{x_1x_2\dots x_m}\right)^2$ делится на число степеней свободы остаточной вариации (n-m-1), а общая сумма квадратов отклонений $\sum\left(y-\bar{y}\right)^2$ на число степеней свободы в целом по совокупности (n-1).

Формула скорректированного индекса множественной детерминации имеет вид:

Формула 20.6

$$R^2_{\text{ск}} = 1 - \frac{\sum\left(y-\hat{y}\right)^2 / (n-m-1)}{\sum\left(y-\bar{y}\right)/(n-1)}$$

где m – число параметров при переменных x; n – число наблюдений.

Поскольку $\frac{\sum\left(y-\hat{y}_{x_1x_2\dots x_m}\right)^2}{\sum\left(y-\bar{y}\right)^2} = 1-R^2$, то величину скорректированного индекса детерминации можно представить в виде:

Формула 20.7

$$R^2_{\text{ск}} = 1-\left(1-R^2\right)\cdot\frac{n-1}{n-m-1}$$

Чем больше величина m, тем сильнее различия $R^2_{\text{ск}}$ и R^2.

Как было показано выше, ранжирование факторов, участвующих во множественной линейной регрессии, может быть проведено через стандартизованные коэффициенты регрессии (β-коэффициенты). Эта же цель может быть достигнута с помощью частных коэффициентов корреляции (для линейных связей). Кроме того, частные показатели корреляции широко используются при решении проблемы отбора факторов: целесообразность включения того или иного фактора в модель можно доказать величиной показателя частной корреляции.

Частные коэффициенты корреляции характеризуют тесноту связи между результатом и соответствующим фактором при элиминировании (устранении влияния) других факторов, включенных в уравнение регрессии.

Показатели частной корреляции представляют собой отношение сокращения остаточной дисперсии за счет дополнительного включения в анализ нового фактора к остаточной дисперсии, имевшей место до введения его в модель.

В общем виде при наличии m факторов для уравнения

$$y=a+b_1x_1+b_2x_2+...+b_mx_m+\varepsilon$$

коэффициент частной корреляции, измеряющий влияние на y фактора x_i, при неизменном уровне других факторов, можно определить по формуле:

Формула 20.8

$$r_{yx_i \cdot x_1x_2...x_{i-1}x_{i+1}...x_m} = \sqrt{1-\frac{1-R^2_{yx_1x_2...x_i...x_m}}{1-R^2_{yx_1x_2...x_{i-1}x_{i+1}...x_m}}}$$

где $R^2_{yx_1x_2...x_i...x_m}$ – множественный коэффициент детерминации всех m факторов с результатом; $R^2_{yx_1x_2...x_i...x_m}$ – тот же показатель детерминации, но без введения в модель фактора x_i.

При двух факторах формула (2.18) примет вид:

Формула 20.8а

$$r_{yx_1 \cdot x_2} = \sqrt{1-\frac{1-R^2_{yx_1x_2}}{1-r^2_{yx_2}}} \qquad r_{yx_2 \cdot x_1} = \sqrt{1-\frac{1-R^2_{yx_1x_2}}{1-r^2_{yx_1}}}$$

Порядок частного коэффициента корреляции определяется количеством факторов, влияние которых исключается. Например, r_{yx1x2} – коэффициент частной корреляции первого порядка. Соответственно коэффициенты парной корреляции называются коэффициентами нулевого порядка. Коэффициенты частной корреляции более высоких порядков можно определить через коэффициенты частной корреляции более низких порядков по рекуррентной формуле:

Формула 20.9

$$r_{yx_i \cdot x_1x_2...x_{i-1}x_{i+1}...x_m} = \frac{r_{yx_i \cdot x_1x_2...x_{i-1}x_{i+1}...x_{m-1}} - r_{yx_m \cdot x_1x_2...x_{m-1}} \cdot r_{x_ix_m \cdot x_1x_2...x_{i-1}x_{i+1}...x_{m-1}}}{\sqrt{\left(1-r^2_{yx_m \cdot x_1x_2...x_{m-1}}\right)\cdot\left(1-r^2_{x_ix_m \cdot x_1x_2...x_{i-1}x_{i+1}...x_{m-1}}\right)}}$$

При двух факторах данная формула примет вид:

Формула 20.9а

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{\sqrt{\left(1-r^2_{yx_2}\right)\cdot\left(1-r^2_{x_1x_2}\right)}} \qquad r_{yx_2 \cdot x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{\sqrt{\left(1-r^2_{yx_1}\right)\cdot\left(1-r^2_{x_1x_2}\right)}}$$

Для уравнения регрессии с тремя факторами частные коэффициенты корреляции второго порядка определяются на основе частных коэффициентов корреляции первого порядка. Так, по уравнению $y=a+b_1x_1+b_2x_2+b_3x_3+\varepsilon$ возможно исчисление трех частных коэффициентов корреляции второго порядка:

$$r_{yx_1 \cdot x_2x_3},\ r_{yx_2 \cdot x_1x_3},\ r_{yx_3 \cdot x_1x_2},$$

каждый из которых определяется по рекуррентной формуле. Например, при i=1 имеем формулу для расчета $r_{yx_1 \cdot x_2x_3}$:

Формула 20.10

$$r_{yx_1 \cdot x_2x_3} = \frac{r_{yx_1 \cdot x_2} - r_{yx_3 \cdot x_2} \cdot r_{x_1x_3 \cdot x_2}}{\sqrt{\left(1-r^2_{yx_3 \cdot x_2}\right)\left(1-r^2_{x_1x_3 \cdot x_2}\right)}}$$

Рассчитанные по рекуррентной формуле частные коэффициенты корреляции изменяются в пределах от –1 до +1, а по формулам через множественные коэффициенты детерминации – от 0 до 1. Сравнение их друг с другом позволяет ранжировать факторы по тесноте их связи с результатом. Частные коэффициенты корреляции дают меру тесноты связи каждого фактора с результатом в чистом виде. Если из стандартизованного уравнения регрессии $ty=\beta_1 t_{x1}+\beta_2 t_{x2}+\beta_3 t_{x3}+\varepsilon$ следует, что $\beta_1>\beta_2>\beta_3$, т.е. по силе влияния на результат порядок факторов таков: x_1, x_2, x_3, то этот же порядок факторов определяется и по соотношению частных коэффициентов корреляции, $r_{yx_1\cdot x_2x_3} > r_{yx_2\cdot x_1x_3} > r_{yx_3\cdot x_1x_2}$.

В статистике частные коэффициенты корреляции обычно не имеют самостоятельного значения. Их используют на стадии формирования модели. Так, строя многофакторную модель, на первом шаге определяется уравнение регрессии с полным набором факторов и рассчитывается матрица частных коэффициентов корреляции. На втором шаге отбирается фактор с наименьшей и несущественной по t-критерию Стьюдента величиной показателя частной корреляции. Исключив его из модели, строится новое уравнение регрессии. Процедура продолжается до тех пор, пока не окажется, что все частные коэффициенты корреляции существенно отличаются от нуля. Если исключен несущественный фактор, то множественные коэффициенты детерминации на двух смежных шагах построения регрессионной модели почти не отличаются друг от друга, $R^2_{m+1} \approx R^2_m$, где m – число факторов.

Из приведенных выше формул частных коэффициентов корреляции видна связь этих показателей с совокупным коэффициентом корреляции. Зная частные коэффициенты корреляции (последовательно первого, второго и более высокого порядка), можно определить совокупный коэффициент корреляции по формуле:

Формула 20.11

$$R_{yx_1x_2\ldots x_m} = \sqrt{1-\left(1-r^2_{yx_1}\right)\cdot\left(1-r^2_{yx_2\cdot x_1}\right)\cdot\left(1-r^2_{yx_3\cdot x_1x_2}\right)\cdot\ldots\cdot\left(1-r^2_{yx_m\cdot x_1x_2\ldots x_{m-1}}\right)}$$

В частности, для двухфакторного уравнения формула (2.11) принимает вид:

Формула 20.12

$$R_{yx_1x_2\ldots x_m} = \sqrt{1-\left(1-r^2_{yx_1}\right)\cdot\left(1-r^2_{yx_2\cdot x_1}\right)}$$

При полной зависимости результативного признака от исследуемых факторов коэффициент совокупного их влияния равен единице. Из единицы вычитается доля остаточной вариации результативного признака ($1-r^2$), обусловленная последовательно включенными в анализ факторами. В результате подкоренное выражение характеризует совокупное действие всех исследуемых факторов.

Значимость уравнения множественной регрессии в целом, так же как и в парной регрессии, оценивается с помощью F-критерия Фишера:

Формула 20.13

$$F = \frac{S_{\text{факт}}}{S_{\text{ост}}} = \frac{R^2}{1-R^2}\cdot\frac{n-m-1}{m}$$

где $S_{\text{факт}}$ – факторная сумма квадратов на одну степень свободы; $S_{\text{ост}}$ – остаточная сумма квадратов на одну степень свободы; R^2 – коэффициент (индекс) множественной детерминации; m – число параметров при переменных x (в линейной регрессии совпадает с числом включенных в модель факторов); n – число наблюдений.

Оценивается значимость не только уравнения в целом, но и фактора, дополнительно включенного в регрессионную модель. Необходимость такой оценки связана с тем, что не каждый фактор, вошедший в модель, может существенно увеличивать долю объясненной вариации результативного признака. Кроме того, при наличии в модели нескольких факторов они могут вводиться в модель в разной последовательности. Ввиду корреляции между факторами значимость одного и того же фактора может быть разной в зависимости от последовательности его введения в модель. Мерой для оценки включения фактора в модель служит частный F-критерий, т.е. F_{xi}.

Частный F-критерий построен на сравнении прироста факторной дисперсии, обусловленного влиянием дополнительно включенного фактора, с остаточной дисперсией на одну степень свободы по регрессионной модели в целом. В общем виде для фактора x_i частный F-критерий определится как

Формула 20.14

$$F = \frac{S_{\text{факт}}}{S_{\text{ост}}} = \frac{R^2}{1-R^2} \cdot \frac{n-m-1}{m}$$

где $R^2_{yx_1 \ldots x_i \ldots x_m}$ – коэффициент множественной детерминации для модели с полным набором факторов, $R^2_{yx_1 \ldots x_{i-1} x_{i+1} \ldots x_m}$ – тот же показатель, но без включения в модель фактора xi, n – число наблюдений, m – число параметров в модели (без свободного члена).

Фактическое значение частного F-критерия сравнивается с табличным при уровне значимости α и числе степеней свободы: 1 и n-m-1. Если фактическое значение F_{xi} превышает $F_{\text{табл}}(\alpha, k_1, k_2)$, то дополнительное включение фактора x_i в модель статистически оправданно и коэффициент чистой регрессии b_i при факторе статистически значим. Если же фактическое значение меньше табличного, то дополнительное включение x_i в модель фактора не увеличивает существенно долю объясненной вариации признака y, следовательно, нецелесообразно его включение в модель; коэффициент регрессии при данном факторе в этом случае статистически незначим.

Для двухфакторного уравнения частные F-критерии имеют вид:

Формула 20.15

$$F_{x_1} = \frac{R^2_{yx_1x_2} - r^2_{yx_2}}{1 - R^2_{yx_1x_2}} \cdot (n-3) \qquad F_{x_2} = \frac{R^2_{yx_1x_2} - r^2_{yx_1}}{1 - R^2_{yx_1x_2}} \cdot (n-3)$$

С помощью частного F-критерия можно проверить значимость всех коэффициентов регрессии в предположении, что каждый соответствующий фактор x_i вводился в уравнение множественной регрессии последним.

Частный F-критерий оценивает значимость коэффициентов чистой регрессии. Зная величину F_{xi}, можно определить и t-критерий для коэффициента регрессии при i-м факторе, t_{bi}, а именно:

Формула 20.16

$$t_{b_i} = \sqrt{F_{x_i}}$$

Оценка значимости коэффициентов чистой регрессии по t-критерию Стьюдента может быть проведена и без расчета частных F-критериев. В этом случае, как и в парной регрессии, для каждого фактора используется формула:

$$t_{b_i} = \frac{b_i}{m_{b_i}}$$

Формула 20.17

где b_i – коэффициент чистой регрессии при факторе x_i, m_{bi} – средняя квадратическая (стандартная) ошибка коэффициента регрессии b_i.

Для уравнения множественной регрессии $y=a+b_1x_1+b_2x_2+\ldots+b_mx_m$ средняя квадратическая ошибка коэффициента регрессии может быть определена по следующей формуле:

Формула 20.18

$$m_{b_i} = \frac{\sigma_y \sqrt{1 - R^2_{yx_1 \ldots x_m}}}{\sigma_{x_i} \sqrt{1 - R^2_{x_i x_1 \ldots x_m}}} \cdot \frac{1}{\sqrt{n-m-1}}$$

где σ_y – среднее квадратическое отклонение для признака y, σ_{xi} – среднее квадратическое отклонение для признака x_i, $R^2_{yx_1 \ldots x_m}$ – коэффициент детерминации для уравнения множественной регрессии, $R^2_{x_i x_1 \ldots x_m}$ – коэффициент детерминации для зависимости фактора x_i со всеми другими факторами уравнения множественной регрессии; n-m-1 – число степеней свободы для остаточной суммы квадратов отклонений.

Как видим, чтобы воспользоваться данной формулой, необходимы матрица межфакторной корреляции и расчет по ней соответствующих коэффициентов детерминации $R^2_{x_i x_1 \ldots x_m}$.

Так, для уравнения $y=a+b_1x_1+b_2x_2+ b_3x_3$ оценка значимости коэффициентов регрессии b_1, b_2, b_3 предполагает расчет трех межфакторных коэффициентов детерминации: $R^2_{x_1 \cdot x_2 x_3}$, $R^2_{x_2 \cdot x_1 x_3}$, $R^2_{x_3 \cdot x_1 x_2}$.

Взаимосвязь показателей частного коэффициента корреляции, частного F-критерия и t-критерия Стьюдента для коэффициентов чистой регрессии может использоваться в процедуре отбора факторов. Отсев факторов при построении уравнения регрессии методом исключения практически можно осуществлять не только по частным коэффициентам корреляции, исключая на каждом шаге фактор с наименьшим незначимым значением частного коэффициента корреляции, но и по величинам t_{bi} и F_{xi}. Частный F-критерий широко используется и при построении модели методом включения переменных и шаговым регрессионным методом.

Пример. Оценим качество уравнения, полученного в предыдущем параграфе. Сначала найдем значения парных коэффициентов корреляции:

$$r_{yx_1}=\frac{\overline{y\cdot x_1}-\bar{y}\cdot\bar{x}_1}{\sigma_y\cdot\sigma_{x_1}}=\frac{66,4-6,8\cdot 9,4}{1,83\cdot 1,56}=0,869$$

$$r_{yx_2}=\frac{\overline{y\cdot x_2}-\bar{y}\cdot\bar{x}_2}{\sigma_y\cdot\sigma_{x_2}}=\frac{44,5-6,8\cdot 6,3}{1,83\cdot 1,42}=0,639$$

$$r_{x_1x_2}=\frac{\overline{x_1\cdot x_2}-\bar{x}_1\cdot\bar{x}_2}{\sigma_{x_1}\cdot\sigma_{x_2}}=\frac{60,3-9,4\cdot 6,3}{1,56\cdot 1,42}=0,488$$

Значения парных коэффициентов корреляции указывают на достаточно тесную связь сменной добычи нефти на одного рабочего y с мощностью буровой x_1 и на умеренную связь с уровнем автоматизации работ x_2. В то же время межфакторная связь r_{x1x2} не очень сильная ($r_{x1x2}=0,49<0,7$), что говорит о том, что оба фактора являются информативными, т.е. и x_1, и x_2 необходимо включить в модель.

Теперь рассчитаем совокупный коэффициент корреляции R_{yx1x2}. Для этого сначала найдем определитель матрицы парных коэффициентов корреляции:

$$\Delta r=\begin{vmatrix}1 & 0,87 & 0,64\\ 0,87 & 1 & 0,49\\ 0,64 & 0,49 & 1\end{vmatrix}=0,139064$$

и определитель матрицы межфакторной корреляции:

$$\Delta r_{11}=\begin{vmatrix}1 & 0,49\\ 0,49 & 1\end{vmatrix}=0,7599$$

$$R_{yx_1x_2}=\sqrt{1-\frac{\Delta r}{\Delta r_{11}}}=\sqrt{1-\frac{0,139064}{0,7599}}=0,904$$

Тогда коэффициент множественной корреляции по формуле:

$$R_{yx_1x_2}=\sqrt{1-\frac{\Delta r}{\Delta r_{11}}}=\sqrt{1-\frac{0,139064}{0,7599}}=0,904$$

Т.е. можно сказать, что 81,7% (коэффициент детерминации $R^2_{yx1x2}=0,817$) вариации результата объясняется вариацией представленных в уравнении признаков, что указывает на весьма тесную связь признаков с результатом.

$$R_{yx_1x_2}=\sqrt{\sum\beta_i\cdot r_{yx_i}}=\sqrt{0,728\cdot 0,87+0,285\cdot 0,64}=0,903$$

Скорректированный коэффициент множественной детерминации

$$R=1-\left(1-R^2\right)\cdot\frac{n-1}{n-m-1}=1-\left(1-0,817\right)\cdot\frac{10-1}{10-2-1}=0,765$$

указывает на умеренную связь между результатом и признаками. Это связано с малым количеством наблюдений.

Теперь найдем частные коэффициенты корреляции по соответствующим формулам, рассмотренным ранее:

$$r_{yx_1\cdot x_2}=\sqrt{1-\frac{1-R^2_{yx_1x_2}}{1-r^2_{yx_2}}}=\sqrt{1-\frac{1-0,817}{1-0,408}}=0,831;$$

$$r_{yx_2\cdot x_1}=\sqrt{1-\frac{1-R^2_{yx_1x_2}}{1-r^2_{yx_1}}}=\sqrt{1-\frac{1-0,817}{1-0,755}}=0,503.$$

$$r_{yx_1\cdot x_2}=\frac{r_{yx_1}-r_{yx_2}\cdot r_{x_1x_2}}{\sqrt{\left(1-r^2_{yx_2}\right)\cdot\left(1-r^2_{x_1x_2}\right)}}=\frac{0,869-0,639\cdot 0,488}{\sqrt{\left(1-0,489^2\right)\left(1-0,639^2\right)}}=0,830;$$

$$r_{yx_2\cdot x_1}=\frac{r_{yx_2}-r_{yx_1}\cdot r_{x_1x_2}}{\sqrt{\left(1-r^2_{yx_1}\right)\cdot\left(1-r^2_{x_1x_2}\right)}}=\frac{0,639-0,869\cdot 0,488}{\sqrt{\left(1-0,488^2\right)\left(1-0,869^2\right)}}=0,498.$$

Т.е. можно сделать вывод, что фактор x_1 оказывает более сильное влияние на результат, чем признак x_2.

Оценим надежность уравнения регрессии в целом и показателя связи с помощью F-критерия Фишера. Фактическое значение F-критерия

$$F_{\text{факт}}=\frac{R^2}{1-R^2}\cdot\frac{n-m-1}{m}=\frac{0,817}{1-0,817}\cdot\frac{10-2-1}{2}=15,63$$

Табличное значение F-критерия при пятипроцентном уровне значимости (α=0,05, k_1=2, k_2=10-2-1=7): $F_{табл}$=4,74. Так как $F_{факт}$=15,63> $F_{табл}$=4,10, то уравнение признается статистически значимым.

Оценим целесообразность включения фактора x_1 после фактора x_2 и x_3 после x_1 с помощью частного F-критерия Фишера

$$F_{x_1} = \frac{R^2_{yx_1x_2} - r^2_{yx_2}}{1 - R^2_{yx_1x_2}} \cdot (n-3) = \frac{0,817 - 0,408}{1 - 0,817} \cdot 7 = 15,65;$$

$$F_{x_2} = \frac{R^2_{yx_1x_2} - r^2_{yx_1}}{1 - R^2_{yx_1x_2}} \cdot (n-3) = \frac{0,817 - 0,755}{1 - 0,817} \cdot 7 = 2,37.$$

Табличное значение частного F-критерия при пятипроцентном уровне значимости (α=0,05, k_1=1, k_2=10-2-1=7): $F_{табл}$=5,59. Так как F_{x1}=15,65> $F_{табл}$=5,59, а F_{x2}=2,37<$F_{табл}$=5,59, то включение фактора x_1 в модель статистически оправдано и коэффициент чистой регрессии b_1 статистически значим, а дополнительное включение фактора x_2, после того, как уже введен фактор x_1, нецелесообразно.

Уравнение регрессии, включающее только один значимый аргумент x_2:

$$y=-2,754+1,016x_1$$

Подведем небольшие итоги по рассмотренной теме, мы рассмотрели, как можно использовать коэффициент множественной корреляции при оценке уравнения множественной регрессии. Введен также скоректированный коэффициент детерминации. Помимо стандартных коэффициентов корреляции существую еще и частные коэффициента корреляции. Значимость уравнения множественной регрессии в целом, так же как и в парной регрессии, оценивается с помощью F-критерия Фишера, оценка параметров модели осуществляется при помощи критерия Стьюдента и целесообразность включения факторов осуществляется при помощи с помощью частного критерия Фишера. Рассчитав все необходимы показатели можно охарактеризовать уравнение множественной регрессии. Насколько эффективно оно описывает ту область в котором проводится исследование.

Для закрепления материала, в котором изложены все проблемы исследования уравнения парной и множественной регрессии ответим на следующие тестовые вопросы.