# ЭКОНОМИКО-МАТЕМАТИЧЕСКИЕ МЕТОДЫ И МОДЕЛИРОВАНИЕ

## МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

***Методические указания к лабораторным работам для студентов бакалавриата направления 21.03.02***

**САНКТ-ПЕТЕРБУРГ**
**2020**

Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное
учреждение высшего образования
Санкт-Петербургский горный университет

Кафедра информатики и компьютерных технологий

# ЭКОНОМИКО-МАТЕМАТИЧЕСКИЕ МЕТОДЫ И МОДЕЛИРОВАНИЕ

## МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

*Методические указания к лабораторным работам*
*для студентов бакалавриата направления 21.03.02*

САНКТ-ПЕТЕРБУРГ
2020

УДК 519.86:622.3.012 (073)

**ЭКОНОМИКО-МАТЕМАТИЧЕСКИЕ МЕТОДЫ И МОДЕЛИРОВАНИЕ. МНОЖЕСТВЕННАЯ РЕГРЕССИЯ.** Методические указания к лабораторным работам / Санкт-Петербургский горный университет. Сост.: *В.В. Беляев, Т.Р. Косовцева*. СПб, 2020. 42 с.

Методические указания содержат сведения, необходимые для лабораторных работ по проведению корреляционно-регрессионного анализа. Приведены необходимые теоретические сведения и примеры выполнения заданий по исследованию корреляционных и регрессионных связей между характеристиками экономических процессов, которые являются теоретической основой построения эконометрических моделей. Все решения выполнены с использованием электронных таблиц MS Excel, в том числе с применением надстройки «Пакет анализа».

Предназначены для студентов бакалавриата направления 21.03.02 «Землеустройство и кадастры» дневной формы обучения.

Научный редактор доц. *А.Б. Маховиков*

Рецензент канд. техн. наук *К.В. Столяров* (Корпорация «Телум Инж»)

# ВВЕДЕНИЕ

Как правило, реальные экономические явления достаточно сложны и выявление характера связи между различными свойствами (параметрами) таких явлений является сложной задачей. Парная регрессия, рассмотренная в предыдущих лабораторных работах, описывает исследуемую характеристику экономического явления (отклик) в зависимости от одной объясняющей характеристики (фактора) в предположении, что влиянием других факторов можно пренебречь. Адекватное уравнение в этом случае удается построить далеко не всегда, поскольку причиной изменения отклика является одновременное воздействие множества факторов. Для того, чтобы учесть это воздействие необходимо использовать *модель множественной регрессии.*

Построение модели множественной регрессии включает несколько этапов:

- выбор формы связи (уравнения регрессии);
- отбор факторных признаков.

Выбор формы связи затрудняется тем, что, теоретическая зависимость между признаками может быть выражена большим числом различных функций. Поскольку уравнение регрессии строится главным образом для объяснения и количественного отображения взаимосвязей, оно должно хорошо отражать сложившиеся между откликом и исследуемыми факторами фактические связи.

В данной работе описан математический аппарат для построения линейного уравнения множественной регрессии.

# ЛАБОРАТОРНАЯ РАБОТА. МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

*Цель: освоить на практике нахождение с помощью табличного процессора MS Excel числовых характеристик множественной регрессии, а также изучить основные свойства теории корреляции.*

## 1. ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ. БАЗОВЫЕ ПОНЯТИЯ

Будем предполагать, что несколько переменных $\boldsymbol{X_1, X_2, ..., X_p}$ (объясняющих переменных, предикторов, факторных признаков, регрессоров) оказывают воздействие на значения зависимой переменной $\boldsymbol{Y}$ (отклик, результативный признак.

В этом случае целесообразно строить уравнение множественной регрессии.

Множественная регрессия – уравнение связи зависимой переменной $Y$ с независимыми переменными $X_1, X_2, ..., X_p$:

$$Y = f\left(X_1, X_2, ..., X_p\right)$$

Наиболее простой и самой употребляемой является *модель множественной линейной регрессии,* которая имеет вид

$$Y = b_0 + b_1 X_1 + b_2 X_2 + ... + b_p X_p \qquad (1)$$

где $b_0, b_1, ..., b_p$ - параметры уравнения.

Пусть имеется *n*-наблюдений, тогда исходные данные представимы в виде матрицы размерности *n* на *p* и вектора размерности *n* :

$$\begin{bmatrix} X_1^1 & X_2^1 & ... & X_p^1 \\ X_1^2 & X_2^2 & ... & X_p^2 \\ ... & ... & ... & ... \\ X_1^n & X_2^n & ... & X_p^n \end{bmatrix}, \begin{bmatrix} Y_1 \\ . \\ . \\ . \\ Y_n \end{bmatrix}. \qquad (2)$$

Все элементы *i*-ой строки $\left(X_1^i \quad X_2^i \quad ... \quad X_p^i\right)$ и *i*-ого элемента вектора $Y_i$ - результаты *i*-ого наблюдения. Будем предполагать, что все наблюдения независимы и получены примерно в одинаковых условиях. В этом случае набор данных, определяемый соотношениями (2) называют *пространственной выборкой или пространственными данными (cross section data)*.На практике эти значения часто получаются как результаты некоторого эксперимента, поэтому их часто называют наблюдаемыми или экспериментальными или эмпирическими значениями.

Для оценки параметров уравнения множественной регрессии применяют метод наименьших квадратов (МНК). Идея этого метода была подробно рассмотрена в лабораторной работе «Линейная парная регрессия»[1]. Все соображения и выводы применимы и в случае множественной линейной регрессии с поправкой на количество факторов.

Рассчитаем $\hat{Y}_i$ - теоретические значения отклика, подставив в уравнение (1) значений факторных переменных *i*-го наблюдения. В результате получим величину

$$\hat{Y}_i = b_0 + b_1 X_1^i + b_2 X_2^i + ... + b_p X_p^i , \qquad (3)$$

Значения $\hat{Y}_i$ практически никогда не будут совпадать с наблюдаемыми значениями $Y_i$ .

Разность между наблюдаемыми значениями $Y_i$ и значениями $\hat{Y}_i$ , рассчитанным по уравнению регрессии, называется *регрессионным остатком* в наблюдении *i* и обозначается $\varepsilon_i$ :

$$\varepsilon_i = Y_i - \hat{Y}_i . \quad i = \overline{1,n} , \qquad (4)$$

Отметим, что $\varepsilon_i$ , $i = \overline{1,n}$ являются случайными величинами, которые также называют *случайными компонентами, случайными членами, возмущениями или остатками*.

С учетом соотношения (4), справедливо соотношение

$$Y_i = \hat{Y}_i + \varepsilon_i = b_0 + b_1 X_1^i + b_2 X_2^i + ... + b_p X_p^i + \varepsilon_i . \qquad (5)$$

Присутствие в этом соотношении случайной компоненты $\varepsilon_i$, обусловлено следующими причинами:

- ошибками спецификации, то есть отбора факторов, и выбора связи между явлениями;
- ошибками измерения.

Будем полагать, что относительно $\varepsilon$ выполняется ряд утверждений, известных как *условия Гаусса-Маркова:*

1. Равенство нулю математического ожидания регрессионных остатков:
   $M(\varepsilon_i) = 0,\ i = 1,...,n$; (6)
2. Постоянство дисперсии регрессионных остатков (гомоскедастичность остатков):
   $M(\varepsilon_i^{\,2}) = D(\varepsilon_i) = \sigma^2$; (7)
3. Отсутствие систематической связи (корреляции) между значениями регрессионных остатков в любых двух наблюдениях: $M(\varepsilon_i \cdot \varepsilon_j) = 0 \quad (i \neq j)$; (8)
4. $X_1, X_2,..., X_p$ - неслучайные величины.

Для определения параметров $b_0, b_1,..., b_p$ уравнения множественной линейной регрессии по МНК составляется сумма $S_{\text{ост}}$ - *остаточная сумма квадратов*

$$S_{\text{ост}} = \sum_{i=1}^{n} (Y_i - \hat{Y}_{i\,x_1 x_2 \ldots x_p})^2 . \quad (9)$$

Она равна сумме квадратов отклонений (остатков) наблюдаемых (эмпирических) значений отклика $Y_i$ от теоретических значений $\hat{Y}_i$ в точке $X_i$ . Чтобы подчеркнуть её зависимость от параметров уравнения регрессии $b_0, b_1,..., b_p$ , обозначим её как функцию от этих параметров через $S(b_0, b_1,..., b_p)$.

$$S(b_0, b_1,..., b_p) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 . \quad (10)$$

*Цель метода наименьших квадратов (МНК) заключается в выборе таких оценок* $b_0, b_1, ..., b_p$ *, для которых сумма квадратов отклонений (остатков) будет минимальной.*

Для того чтобы найти набор коэффициентов $b_0, b_1, ..., b_p$ , которые доставляют минимум функции $S(b_0, b_1, ..., b_p)$ , используем необходимое условие экстремума функции нескольких переменных - равенство нулю частных производных

$$\frac{\partial S(b_0, b_1, ..., b_p)}{\partial b_0} = 0; \quad \frac{\partial S(b_0, b_1, ..., b_p)}{\partial b_1} = 0; \quad ... \frac{\partial S(b_0, b_1, ..., b_p)}{\partial b_p} = 0$$

В результате преобразований получаем следующую систему нормальных уравнений:

$$\begin{cases} \sum Y = n \cdot b_0 + b_1 \cdot \sum X_1 + b_2 \cdot \sum X_2 + ... + b_p \cdot \sum X_p \\ \sum YX_1 = b_0 \sum X_1 + b_1 \cdot \sum X_1^2 + b_2 \cdot \sum X_1 X_2 + ... + b_p \cdot \sum X_p X_1 \\ ..... \\ \sum YX_p = b_0 \sum X_p + b_1 \cdot \sum X_1 X_p + b_2 \cdot \sum X_p X_2 + ... + b_p \cdot \sum X_p^2 \end{cases} \qquad (11)$$

Для ее решения может быть применен любой известный метод решения системы линейных уравнений.

Коэффициенты $\{b_j\}_{j=1}^{p}$ в уравнении (3) называются *коэффициентами множественной регрессии*. Величина коэффициента $b_j$ показывает среднее изменение отклика *Y* при изменении фактора $\boldsymbol{X_j}$ на единицу.

Другой вид уравнения множественной регрессии - уравнение регрессии в стандартизованном масштабе:

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + ... + \beta_p t_{x_p}, \qquad (12)$$

где: $t_y = \frac{Y - \overline{Y}}{\sigma_y}, t_{x_i} = \frac{X_i - \overline{X}_i}{\sigma_{x_i}}$ - стандартизованные переменные;

$i = 1,.., p$ , $\boldsymbol{p}$ - число неизвестных;

$\overline{Y}$ , $\overline{X}_i$ - средние значения;

$\sigma_y, \sigma_{x_i}$ - средние квадратические отклонения;

$\beta_i$ -стандартизованные коэффициенты регрессии.

В силу того, что стандартизованные переменные заданы как центрированные (средние значения $\bar{t}_y = \bar{t}_x = 0$ ) и нормированные (средние квадратические отклонения $\sigma_{t_y} = \sigma_{t_x} = 1$ ), стандартизованные коэффициенты регрессии сравнимы между собой, и с их помощью можно ранжировать факторы по силе их воздействия на результат.

Для определения коэффициентов уравнения множественной регрессии в стандартизованном масштабе так же применим МНК. Коэффициенты $\{\beta_j\}_{j=1}^{p}$ можно получить, решая систему, аналогичную системе (7). Эту систему можно преобразовать, и тогда, стандартизованные коэффициенты регрессии определяются из следующей системы уравнений:

$$\begin{cases} r_{yx_1} = \beta_1 + \beta_2 r_{x_2x_1} + \beta_3 r_{x_3x_1} + ... + \beta_p r_{x_Px_1} \\ r_{yx_2} = \beta_1 r_{x_2x_1} + \beta_2 + \beta_3 r_{x_3x_2} + ... + \beta_p r_{x_px_2} \\ ..... \\ r_{yx_p} = \beta_1 r_{x_px_1} + \beta_2 r_{x_px_2} + \beta_3 r_{x_3x_p} + ... + \beta_p \end{cases}, \qquad (13)$$

где

- $r_{x_ix_j}$ - коэффициент парной корреляции между факторами $X_i$ и $X_j$,
- $r_{yx_j}$ - коэффициент парной корреляции между откликом $Y$ и фактором $X_j$ .

Отметим, что связь коэффициентов множественной регрессии $b_j$ со стандартизованными коэффициентами $\beta_j$ описывается соотношением

$$b_j = \beta_j \frac{\sigma_y}{\sigma_{x_j}}. \qquad (14)$$

Стандартизованный коэффициент регрессии $\beta_j$ показывает, на сколько величин $\sigma_y$ в среднем изменится отклик при увеличении *j-го* фактора на одну величину $\sigma_{x_j}$ .

*Средние коэффициенты эластичности* для линейной регрессии рассчитываются по формуле :

$$\overline{Э}_{yx_j} = b_j \frac{\overline{X}_j}{\overline{Y}}. \qquad (15)$$

Средний коэффициент эластичности $\overline{Э}_{yx_j}$ показывает на сколько процентов в среднем изменится отклик при изменении его среднего значения фактора $X_j$ на один процент, при неизменном значении остальных факторов.

## 2. КОЭФФИЦИЕНТ ДЕТЕРМИНАЦИИ

Тесноту совместного влияния факторов на результат показывает *множественной детерминации.*

Качество построенной модели в целом оценивается коэффициентом множественной детерминации, который определяется формулой:

$$R^2_{yx_1x_2\ldots x_p} = 1 - \frac{S_{\text{ост}}}{S_{\text{полн}}}, \qquad (16)$$

где $S_{\text{ост}} = \sum_{i=1}^{n} (Y_i - \hat{Y}_{i\,x_1x_2\ldots x_p})^2$ - *остаточная сумма квадратов отклонений*,

$S_{\text{полн}} = \sum_{i=1}^{n} (Y_i - \overline{Y})^2$ - общая сумма квадратов отклонений значений $Y_i$ от среднего арифметического значения отклика $Y$ .

Для линейной регрессии справедливо следующее равенство:

$S_{\text{полн}} = S_{\text{ост}} + S_{\text{регр}}$ ,

где $S_{\text{регр}} = \sum_{i=1}^{n} (\hat{Y}_{i x_1 x_2 \ldots x_p} - \overline{Y})^2$ *регрессионная сумма квадратов отклонений*.

Остаточная сумма квадратов отклонений $S_{\text{ост}}$ характеризует суммарное отклонение наблюдаемых (эмпирических) данных от теоретических значений, найденных по уравнению регрессии. Регрессионная или факторная сумма квадратов отклонений $S_{\text{регр}}$ характеризует разброс теоретических значений относительно среднего арифметического значения наблюдаемого значения (отклика).

Все свойства коэффициента детерминации $R^2_{yx_1x_2\ldots x_p}$ указаны в лабораторной работе [1]. Так, значение этого коэффициента лежит в пределах от нуля до единицы. Это значение показывает долю объясненной вариации результативного признака (отклика) за счет включенных в уравнение $p$ факторов, т.е. насколько хорошо уравнение, полученное с помощью регрессионного анализа, объясняет взаимосвязь между откликом и факторами. Доля необъясненной вариации отклика других, не учтенных в модели факторов, равна $1 - R^2$ . Коэффициент детерминированности служит показателем тесноты связи между независимой переменной и факторами. Показателю тесноты связи можно дать качественную оценку (шкала Чеддока)

*Таблица 1*

**Шкала Чеддока**

| ***Количественная мера тесноты связи*** | ***Качественная характеристика силы связи*** |
|---|---|
| 0,1-0,3 | Слабая |
| 0,3-0,5 | Умеренная |
| 0,5-0,7 | Заметная |
| 0,7-0,9 | Высокая |
| 0,9-0,99 | Весьма высокая |

Величину $R^2$ для уравнения множественной регрессии в стандартизованном масштабе можно определить по формуле

$$R^2 = \sum \beta_i \cdot r_{yx_i} \qquad (17)$$

## 3. ОЦЕНКА НАДЕЖНОСТИ УРАВНЕНИЯ МНОЖЕСТВЕННОЙ РЕГРЕССИИ

*Проверить значимость уравнения регрессии* - значит установить насколько хорошо математическая модель, выражающая зависимость отклика от факторов, согласуется с экспериментальными данными, с учетом количества наблюдений и количества факторов в уравнении. Оценка значимости уравнения регрессии в целом сводится к проверке того, что величина $R^2$ не случайно отлична от нуля. Для оценки значимости уравнения множественной регрессии в целом используется ***F***-критерий Фишера.

Выдвигаем нулевую гипотезу $H_0 : R^2 = 0$ . Это возможно, когда уравнение регрессии незначимо, т.е. связь между откликом и факторами отсутствует. Альтернативная гипотеза $H_1 : R^2 > 0$ , в этом случае уравнение регрессии адекватно описывает связь между откликом и факторами.

Схема проведения дисперсионного анализа приведена в табл.2. Схемы применения ***F***-критерия Фишера для оценки значимости уравнения множественной регрессии и уравнения парной регрессии одинаковы. Различие состоит только в одном – в определении числа степеней свободы $df_{\text{регр}}$ и $df_{\text{ост}}$ .

Существует соотношение между числом *степеней свободы* $df$ (числом свободы независимого варьирования признака) для общей, факторной и остаточной сумм квадратов:

$$df_{\text{полн}} = df_{\text{ост}} + df_{\text{регр}} .$$

Для множественной линейной регрессии:

$$df_{\text{полн}} = n - 1;\ df_{\text{регр}} = p;\ df_{\text{ост}} = n - p - 1 ,$$

где $n$ - число единиц совокупности, $p$ - число факторов, включенных в уравнение регрессии.

$$D_{\text{полн}} = \frac{S_{\text{полн}}}{df_{полн}} = \frac{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}{n-1}; \qquad (14)$$

$$D_{\text{регр}} = \frac{S_{\text{регр}}}{df_{регр}} = \frac{\sum_{i=1}^{n}\left(\hat{Y}_i - \overline{Y}\right)^2}{p}; \qquad (15)$$

$$D_{\text{ост}} = \frac{S_{\text{ост}}}{df_{\text{ост}}} = \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{n-p-1}. \qquad (16)$$

*Таблица 2*

***Схема проведения дисперсионного анализа***

| ***Источники вариации:*** | ***Вариация, объясненная за счет регрессии*** | ***Остаточная вариация*** | ***Общая вариация*** |
|---|---|---|---|
| Число степеней свободы | $df_{\text{регр}} = p$ ($p$ - количество факторов) | $df_{\text{ост}} = n - p - 1$ | $df_{\text{полн}} = n - 1$ |
| Сумма квадратов отклонений | $S_{\text{регр}} = \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2$ | $S_{\text{ост}} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ | $S_{\text{полн}} = \sum_{i=1}^{n}(y_i - \overline{y})^2$ |
| Дисперсия на одну степень свободы | $D_{\text{регр}} = \frac{S_{\text{регр}}}{df_{\text{регр}}}$ | $D_{\text{ост}} = \frac{S_{\text{ост}}}{df_{\text{ост}}}$ | $D_{\text{полн}} = \frac{S_{\text{ост}}}{df_{\text{полн}}}$ |
| Фактическое значение критерия Фишера | $F_{\text{набл}} = \frac{D_{\text{регр}}}{D_{\text{ост}}}$ | | |
| Табличное значение критерия Фишера | $F_{\text{крит}}$ определяется по уровню значимости и числу степеней свободы числителя $df_{регр}$ и знаменателя $df_{ост}$ | | |

Критерий Фишера определяется следующим соотношением:

$$F_{\text{набл}} = \frac{D_{\text{регр}}}{D_{\text{ост}}} \qquad (17)$$

Использование критерия Фишера предполагает вычисление $F_{\text{набл}}$ и его сравнение с табличным значением $F_{\text{крит}}$, которое зависит от уровня значимости $\boldsymbol{\alpha}$ и числа степеней свободы для регрессионной и остаточной сумм. $F_{\text{крит}}$ определяется либо с помощью таблиц, либо с использованием специализированных пакетов программ, например, в MS Excel для этого может быть использована функция **FРАСПРОБР().**

Если $F_{\text{набл}} > F_{\text{крит}}$, нулевая гипотеза $H_0$ об отсутствии связи признаков отклоняется и делается вывод о справедливости гипотезы $H_1$ (о существенности этой связи, значимости уравнения регрессии). Если же величина $F_{\text{набл}}$ окажется меньше табличной, то есть $F_{\text{набл}} < F_{\text{крит}}$, то вероятность нулевой гипотезы $H_0$ выше заданного уровня значимости (например, **0.05**) и гипотеза $H_0$ не может быть отклонена без серьезного риска сделать неправильный вывод о наличии линейной связи между факторами $X_1$ ,$X_2$, ... ,$X_p$ и откликом ***Y*** . В этом случае уравнение регрессии считается статистически незначимым, и это уравнение нельзя использовать для анализа и прогноза.

Значение $F_{\text{набл}}$ может быть вычислено как по формуле (17), так и с помощью коэффициента детерминированности по формуле (18).

$$F_{\text{набл}} = \frac{D_{\text{регр}}}{D_{\text{ост}}} = \frac{R^2}{1-R^2} \cdot \frac{n-p-1}{p}, \qquad (18)$$

Частный ***F***-критерий оценивает статистическую значимость каждого из факторов в уравнении. В общем виде для фактора $\boldsymbol{X_i}$ частный ***F***-критерий определится как

$$F_{x_i} = \frac{R^2_{yx_1x_2\ldots x_i\ldots x_p} - R^2_{yx_1x_2\ldots x_{i-1}x_{i+1}\ldots x_p}}{1 - R^2_{yx_1x_2\ldots x_i\ldots x_p}} \cdot \frac{n-p-1}{1}. \qquad (19)$$

Необходимость вычисления такой оценки обусловлена тем, что не каждый внесенный в модель фактор будет увеличивать долю объясненной вариации результативного признака. Также 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 различной в зависимости от последовательности введения его в модель.

Частный *F*-критерий, вычисленный по формуле (19) построен на сравнении прироста факторной дисперсии, обусловленного влиянием дополнительного включенного фактора (числитель) с остаточной дисперсией на одну степень свободы по регрессионной модели в целом (знаменатель).

Вычисленный частный *F*-критерий $F_{x_i}$ сравнивается с табличным значением $F_{\text{крит}}$, который зависит от уровня значимости $\boldsymbol{\alpha}$ и числа степеней свободы: 1 и $(n-p-1)$. $F_{\text{крит}}$ определяется либо с помощью таблиц, либо с использованием функций, например, функции MS Excel **FРАСПРОБР().**

Если $F_{x_i} > F_{\text{крит}}$, то дополнительное включение фактора $\boldsymbol{X_i}$ в модель статистически оправдано и коэффициент регрессии $\boldsymbol{b_i}$ при $\boldsymbol{X_i}$ статически значим. Если же величина $\boldsymbol{F_{x_i}}$ окажется меньше табличной, то есть $F_{x_i} < F_{\text{крит}}$, то дополнительное включение в модель фактора $\boldsymbol{X_i}$ статистически не оправдано, поскольку существенно не увеличивает долю объясненной вариации отклика. При этом коэффициент регрессии $\boldsymbol{b_i}$ при $\boldsymbol{X_i}$ статистически незначим, что еще раз подтверждает нецелесообразность включения этого фактора в модель.

С частным *F*-критерием тесно связан $\boldsymbol{t}$ -критерий Стьюдента для проверки значимости коэффициентов регрессии.

Оценка значимости коэффициентов множественной регрессии с помощью ***t*** -критерия Стьюдента производится с помощью величины $t_{b_i}$ , вычисляемой по формулам

$$t_{b_i} = \frac{b_i}{m_{b_i}}, \qquad (20)$$

или

$$\left|t_{b_i}\right| = \sqrt{F_{x_i}} \qquad (21)$$

где $m_{b_i}$ – средняя квадратическая ошибка коэффициента регрессии $\boldsymbol{b_i}$, которая может быть определена по следующей формуле:

$$m_{b_i} = \frac{\sigma_y \cdot \sqrt{1 - R^2_{yx_{1\ldots}x_p}}}{\sigma_{x_i} \cdot \sqrt{1 - R^2_{x_i x_{1\ldots} x_p}}} \cdot \frac{1}{\sqrt{n - p - 1}} \quad , \qquad (22)$$

где $R^2_{x_i x_{1\ldots} x_p}$ - коэффициент множественной детерминации для зависимости фактора $\boldsymbol{X_i}$ от всех других факторов уравнения множественной детерминации, все остальные обозначения очевидны.

Заметим, что формула (21) полностью аналогична формуле (5.18)[1]. Сама процедура проверки значимости соответствующего коэффициента полностью аналогична процедуре проверки в лабораторной работе [1].

При представлении результатов множественной регрессии наряду с уравнением и скорректированным коэффициентом множественной детерминации принято приводить значения $t_{b_i}$ . На практике если наблюдаемые значения $t_{b_i} > 3$, то это означает, что значение этого коэффициента статистически достоверно, а уравнение может быть использовано для прогнозирования.

При эконометрическом исследовании необходимо стремиться к увеличению числа наблюдений, так как большой объем наблюдений является одной из предпосылок признания значимым как уравнения регрессии, так и его коэффициентов. Значимость этих

величин является необходимым условием построения адекватных статистических моделей.

Как показывает практика, для того, чтобы уравнение было адекватным, необходимо, чтобы количество наблюдений $n$ превышало количество определяемых коэффициентов регрессии $p$ в 6-7 раз.

## 4. СКОРРЕКТИРОВАННЫЙ ИНДЕКС МНОЖЕСТВЕННОЙ ДЕТЕРМИНАЦИИ

Скорректированный (исправленный, adjustable) коэффициент множественной детерминации $R^2_{adj}$ содержит поправку на число степеней свободы и рассчитывается по формуле :

$$R^2_{adj} = 1 - (1 - R^2) \cdot \frac{(n-1)}{(n-p-1)} \qquad (23)$$

Скорректированный коэффициент множественной детерминации $R^2_{adj}$ используется для сопоставления моделей содержащих различное количество факторов.

Чем больше ***p***, тем больше различие между $R^2_{adj}$ и $R^2$. Чем больше объем выборки ***n***, тем меньше это различие.

Существенно различным может быть изменение $R^2$ и $R^2_{adj}$ при включении дополнительного фактора в уравнение регрессии. Если этот фактор существенно влияет на отклик, то увеличатся значения как $R^2$ так и $R^2_{adj}$. Если вновь добавленный фактор несущественно влияет на отклик, то значение $R^2$, как правило, увеличивается (может быть незначительно), а значение $R^2_{adj}$ - уменьшается. Очевидно, что в этом случае такой фактор в уравнение включать не целесообразно.

## 5. ЧАСТНАЯ КОРРЕЛЯЦИЯ

*Частные коэффициенты (или индексы) корреляции*, измеряющие влияние на $Y$ фактора $X_i$ при устранении влияния

других факторов, включенных в уравнение регрессии, можно определить по формуле:

$$r_{yx_1x_2\ldots x_{i-1}x_{i+1}\ldots x_p} = \sqrt{1-\frac{1-R^2_{yx_1x_2\ldots x_i\ldots x_p}}{1-R^2_{yx_1x_2\ldots x_{i-1}x_{i+1}\ldots x_p}}} \quad , \qquad (24)$$

где $R^2_{yx_1x_2\ldots x_i\ldots x_p}$ - коэффициент детерминированности для уравнения регрессии, в которое включены факторы $X_1, X_2, \ldots, X_p$ ; $R^2_{yx_1x_2\ldots x_{i-1}x_{i+1}\ldots x_p}$ - коэффициент детерминированности для уравнения регрессии, в которое включены факторы $X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots X_p$, т.е. фактор $X_i$ исключен из уравнения.

Частные коэффициенты корреляции изменяются в пределах от **0** до **1**.

Нетрудно показать, что величина, стоящая под радикалом в правой части равенства (24) может быть преобразована к следующему виду:

$$1-\frac{1-R^2_{yx_1x_2\ldots x_i\ldots x_p}}{1-R^2_{yx_1x_2\ldots x_{i-1}x_{i+1}\ldots x_p}} = \frac{R^2_{yx_1x_2\ldots x_i\ldots x_p}-R^2_{yx_1x_2\ldots x_{i-1}x_{i+1}\ldots x_p}}{1-R^2_{yx_1x_2\ldots x_{i-1}x_{i+1}\ldots x_p}}$$

Правая часть последнего равенства представляет собой отношение приращения объясненной часть вариации отклика за счет включения фактора $X_i$ в уравнение регрессии к необъясненной доле вариации отклика, имевшей место до введения фактора $X_i$ в уравнение регрессии.

Таким образом, величина $r_{yx_i\cdot x_1x_2\ldots x_p}$ характеризует возрастание коэффициента детерминации за счет введения в уравнение регрессии фактора $X_i$ . Благодаря этому частные коэффициенты корреляции могут быть использованы для ранжирования влияния факторов на результат.

Так, при двух факторах и *i*=1 частный коэффициент корреляции $r_{yx1\cdot x2}$ может быть вычислен по формуле

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1 x_2}}{\sqrt{(1 - r_{yx_2}{}^2) \cdot (1 - r_{x_1 x_2}{}^2)}} \qquad (25)$$

Коэффициент $r_{yx_1 \cdot x_2}$ показывает тесноту связи между $Y$ и $X_1$ при неизменном уровне фактора $X_2$, включенного в уравнение регрессии.

Аналогично $r_{yx_2 \cdot x_1}$ можно определить по формуле

$$r_{yx_2 \cdot x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1 x_2}}{\sqrt{(1 - r_{yx_1}{}^2) \cdot (1 - r_{x_1 x_2}{}^2)}} \qquad (26)$$

Коэффициенты частной корреляции используют для оценки целесообразности включения фактора в уравнение регрессии.

## 6. МАТРИЧНАЯ ФОРМА ЗАПИСИ

Матричная форма записи для определения коэффициентов множественной линейной регрессии полностью аналогична таковой для парной регрессии (5.28) [1] , т.е.

$$B = (X^{\mathrm{T}} \cdot X)^{-1} \cdot X^{\mathrm{T}} \cdot y \qquad (27)$$

где $\boldsymbol{X}$ матрица размерности $n \cdot (p+1)$, В – вектор коэффициентов размерности $(p+1)$

$$X = \begin{bmatrix} 1 & X_{11} & . & X_{1p} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ 1 & X_{n1} & . & X_{np} \end{bmatrix}, \; y = \begin{bmatrix} Y_1 \\ . \\ . \\ . \\ Y_n \end{bmatrix}, \; B = \begin{bmatrix} b_0 \\ b_1 \\ . \\ b_{\mathrm{p}} \end{bmatrix}$$

## 7. МУЛЬТИКОЛЛИНЕАРНОСТЬ ФАКТОРОВ

При построении уравнения множественной регрессии может возникнуть проблема *мультиколлинеарности* факторов.

Одним из условий построения корректной регрессионной модели является условие линейной независимости факторов. Если это условие нарушается, т.е. *если один из факторов может быть выражен через несколько других, то говорят что, существует полная коллинеарность*.

Это порождает множество проблем. Например, применение формулы (18) невозможно, поскольку матрица $(X^{\mathrm{T}} \cdot X)^{-1}$ не может быть вычислена (определитель $\det(X^{\mathrm{T}} \cdot X) = 0$).

На практике полная коллинеарность встречается редко, гораздо чаще встречается ситуация, когда между факторами наблюдается высокая степень корреляции, и тогда говорят о наличии *мультиколлинеарности* факторов.

В этом случае применение формулы (18) формально возможно, поскольку матрица $(X^{\mathrm{T}} \cdot X)^{-1}$ может быть вычислена (определитель $\det(X^{\mathrm{T}} \cdot X) \neq 0$ , но близок к нулю), поэтому полученные значения найденных коэффициентов будут обладать «плохими свойствами».

**Основные отрицательные проявления мультиколлинеарности заключаются в следующем:**

- Значения найденных коэффициентов модели имеют неправильные с точки зрения теории знаки или неоправданно большие (маленькие) значения.
- Небольшие изменения исходных данных приводит к существенному изменению найденных коэффициентов модели
- Оценки имеют большие стандартные ошибки, малую значимость (хотя вся модель в целом является значимой).
- Невозможно оценить воздействие на отклик каждого фактора в отдельности.

Когда два фактора сильно коррелированы, говорят о коллинеарности факторов.

Считается, что *два фактора* явно *коллинеарны*, т.е. находятся между собой в линейной зависимости, если

$$\left| r_{x_i x_j} \right| \geq 0{,}7. \qquad (28)$$

Для решения проблемы мультиколлинеарности зависимые факторы исключают из модели.

В случае двух явно коллинеарных факторов уравнения регрессии рекомендуется один исключить. Предпочтение при этом отдается тому фактору, который при достаточно тесной связи с результатом имеет наименьшую тесноту связи с другими факторами.

## 8. ПОСТРОЕНИЕ ПРОГНОЗА С ПОМОЩЬЮ УРАВНЕНИЯ РЕГРЕССИИ

Рассмотрим применение уравнения регрессии для построения точечного и интервального прогноза. Точечный прогноз $\hat{Y}$ может быть получен путем подстановки в уравнение регрессии (1) значений факторов.

Результат точечного прогноза маловероятен. Поэтому находят интервальную оценку прогноза.

Для получения интервальной оценки необходимо воспользоваться формулой аналогичной формуле парной регрессии, которая для множественной регрессии имеет следующий вид.

$$\hat{Y} - \varepsilon \leq \widetilde{Y} \leq \hat{Y} + \varepsilon \qquad (29)$$

где $\varepsilon$ - полуширина доверительного интервала.

Точечное значение $\hat{Y}$ является серединой доверительного интервала, $\left(\hat{Y} - \varepsilon\right)$ - левой границей, $\left(\hat{Y} + \varepsilon\right)$ - правой границей.

Величина равна $\varepsilon$ половине ширины доверительного интервала и может быть вычислена по формуле

$$\varepsilon = t_{\text{крит}} \cdot s_{\hat{y}} \qquad (30)$$

где $t_{\text{крит}}$ - критическое значение распределения Стьюдента с числом степеней свободы равным $n$-$p$-1;

$s_{\hat{y}}$ - стандартная ошибка групповой средней

$$s_{\hat{y}} = s \cdot \sqrt{(X0)^T \left(X^T X\right) X0} \qquad (31)$$

$$X0 = \begin{pmatrix} 1 \\ x_{1,0} \\ x_{2,0} \\ \dots \\ x_{p,0} \end{pmatrix}, \; X = \begin{bmatrix} 1 & X_{11} . & X_{1\,p} \\ . & . \; . & . \\ . & . \; . & . \\ . & . \; . & . \\ 1 & X_{n1} . & X_{n\,p} \end{bmatrix}, \qquad (32)$$

$X0$ – вектор значений факторов, определяющий точку в $p$+1 пространстве, в которой строим прогноз;

$X$ – матрица, по которой было построено уравнение.

$S$ – стандартное отклонение остаточной дисперсии или стандартная ошибка уравнения регрессии:

$$s = \sqrt{D_{\text{ост}}} \qquad (33)$$

Стандартная ошибка уравнения регрессии $s$ может быть вычислена с помощью инструмента “Регрессия” надстройки «Пакет Анализа» MS Excel .

**ПРИМЕР**

Исследовать зависимость между стоимостью грузовой автомобильной перевозки $Y$ (тыс. руб), весом груза $X_1$ (тонн) и расстоянием $X_2$ (тыс. км) по 20 транспортным компаниям (табл.3).

*Таблица 3*

| ***№n/n*** | ***Y*** | ***$X_1$*** | ***$X_2$*** |
|---|---|---|---|
| 1 | 51 | 35 | 2 |
| 2 | 16 | 16 | 1,1 |
| 3 | 74 | 18 | 2,55 |
| 4 | 7,5 | 2 | 1,7 |
| 5 | 33 | 14 | 2,4 |
| 6 | 26 | 33 | 1,55 |

*Продолжение таблицы 3*

| ***№n/n*** | ***Y*** | $X_1$ | $X_2$ |
|---|---|---|---|
| 7 | *11,5* | 20 | 0,6 |
| 8 | *52* | 25 | 2,3 |
| 9 | *15,8* | 13 | 1,4 |
| 10 | *8* | 2 | 2,1 |
| 11 | *26* | 21 | 1,3 |
| 12 | *6* | 11 | 0,35 |
| 13 | *5,8* | 3 | 1,65 |
| 14 | *13,8* | 3,5 | 2,9 |
| 15 | *6,2* | 2,8 | 0,75 |
| 16 | *7,9* | 17 | 0,6 |
| 17 | *5,4* | 3,4 | 0,9 |
| 18 | *56* | 24 | 2,5 |
| 19 | *25,5* | 9 | 2,2 |
| 20 | *7,1* | 4,5 | 0,95 |

**Требуется построить и оценить линейную модель множественной регрессии по следующему плану**:

1. Вычислить описательные статистики для отклика и всех факторов.

2. Оценить визуально, построив соответствующие облака рассеяния величины $Y$ в зависимости от $X_1$ и $X_2$, целесообразность использования линейного уравнения регрессии.

3. Вычислить и проанализировать:

- линейные коэффициенты парной и частной межфакторной корреляции;
- линейные коэффициенты парной и частной корреляции между каждым фактором и откликом.

4. Написать уравнение множественной регрессии $Y = b_0 + b_1 X_1 + b_2 X_2$, оценить значимость его параметров, пояснить их экономический смысл. Коэффициенты уравнения вычислить двумя способами, используя:

- функцию ЛИНЕЙН();
- надстройку «Анализ Данных».

5. Написать уравнение множественной регрессии в стандартизованном масштабе $t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2}$, пояснить экономический смысл его параметров.

6. Вычислить средние частные коэффициенты эластичности $\overline{Э}_{yx_1}$ и $\overline{Э}_{yx_2}$. Пояснить их экономический смысл.

6. Вычислить коэффициентом множественной детерминации $R^2_{yx_1x_2}$ двумя способами:

- по определению, по формуле (8);
- с использованием матрицы парных коэффициентов корреляции по формуле (9).

7. Полученный результат сравнить с результатом, полученным с помощью надстройки «Анализ Данных» $\Rightarrow$ «Регрессия».

8. С помощью *F*-критерия Фишера дать оценку надежности уравнения регрессии в целом и показателя тесноты связи $R^2$, используя результат, полученный с помощью надстройки «Анализ Данных»;

9. Оценить значимость коэффициентов множественной регрессии с помощью *t*-критерия Стьюдента, используя результаты работы надстройки «Анализ Данных» $\Rightarrow$ «Регрессия».

10. С помощью частных *F*-критериев Фишера оценить целесообразность включения в уравнение множественной регрессии фактора $X_1$ после фактора $X_2$ и фактора $X_2$ после фактора $X_1$.

11. Найти точечный и интервальный прогноз значений отклика при условии, что значение каждого фактора меньше максимального значения на 10% величины размаха исходных данных.

## Пояснения по выполнению отдельных пунктов задания

Решение проведем с использованием электронных таблиц MS Excel.

***К пункту 1***

Исходные данные представлены на рис.1.*а*, и содержатся в интервале B3:D22, на рис.1.*б* приведены вычисленные средние значения, дисперсии и стандартные отклонения факторов и отклика.

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | **Множественная регрессия.** | | | | | | |
| 2 | N п.п. | y | x1 | x2 | yтеор | u-остатки | u^2 |
| 3 | 1 | 51 | 35 | 2 | 53,35668 | -2,356683199 | 5,553956 |
| 4 | 2 | 16 | 16 | 1,1 | 17,798 | -1,797996563 | 3,232792 |
| | | | | | | | |
| 21 | 19 | 25,5 | 9 | 2,2 | 26,32001 | -0,820010329 | 0,672417 |
| 22 | 20 | 7,1 | 4,5 | 0,95 | 2,237743 | 4,862257249 | 23,64155 |
| 23 | Суммы | 454,5 | 277,2 | 31,8 | | 1,59872E-14 | |
| 24 | Средн. | 22,725 | 13,86 | 1,59 | | | |
| 25 | Ст.откл | 19,8473896 | 10,04716 | 0,738004 | | | |

Рис.1.*а*. Исходные данные задачи в MS Excel

Описательные статистики для отклика и всех факторов $X_1$ и $X_2$, могут быть вычислены с помощью с помощью надстройки MS Excel «Пакет Анализа $\Rightarrow$ Описательные статистики» и представлены на рис. 1.*б*.

***К пункту 2***

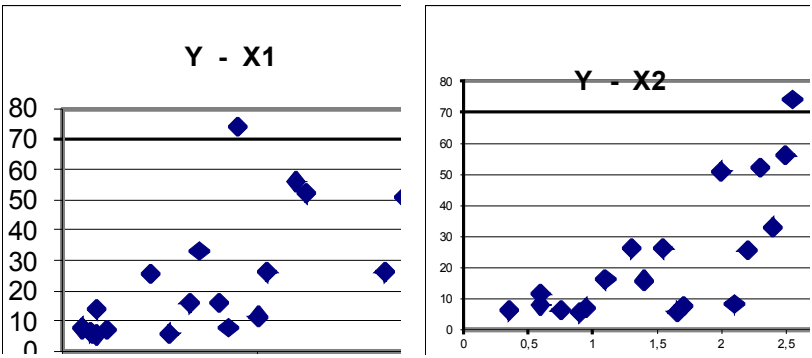
Вытянутость облака точек на диаграмме рассеяния (рис.2. *а*) вдоль наклонной прямой позволяет сделать предположение о том, что существует линейная связь между значениями переменных $X_1$ - весом груза и $Y$ - стоимостью грузовой автомобильной перевозки.

Анализируя рис.2.*б*, можно заметить наличие прямой линейной связи между значениями переменных $X_2$ - расстоянием и $Y$ - стоимостью грузовой автомобильной перевозки.

| | I | J | L | N |
|---|---|---|---|---|
| 3 | **Описательная статистика** | | | |
| 4 | | | | |
| 5 | | *y* | *x1* | *x2* |
| 6 | | | | |
| 7 | Среднее | 22,725 | 13,86 | 1,59 |
| 8 | Стандартная ошибка | 4,5533 | 2,3050 | 0,1693 |
| 9 | Медиана | 14,80 | 13,50 | 1,60 |
| 10 | Мода | 26,00 | 2,00 | 0,60 |
| 11 | Стандартное отклонение | 20,3630 | 10,3082 | 0,7572 |
| 12 | Дисперсия выборки | 414,6514 | 106,2583 | 0,5733 |
| 13 | Эксцесс | 0,6974 | -0,5669 | -1,2001 |
| 14 | Асимметричность | 1,2794 | 0,5556 | 0,0035 |
| 15 | Интервал | 68,60 | 33,00 | 2,55 |
| 16 | Минимум | 5,40 | 2,00 | 0,35 |
| 17 | Максимум | 74,00 | 35,00 | 2,90 |
| 18 | Сумма | 454,50 | 277,20 | 31,80 |
| 19 | Счет | 20 | 20 | 20 |
| 20 | **Прогноз** | | 31,70 | 2,65 |
| 21 | | | | |
| 22 | | | | |

=L17-L15*0,1

Рис.1 *б*. Описательная статистика для исходных данных задачи с помощью с помощью надстройки MS Excel «Пакет Анализа – Описательные статистики».




а) б)

Рис. 2 Облака рассеяния:

*а*) $Y$ - $X_1$ и *б*) $Y$ - $X_2$

***К пункту 3***

Значения **линейных коэффициентов парной корреляции** определяют тесноту попарно связанных переменных, использованных в данном уравнении множественной регрессии. **Линейные коэффициенты частной корреляции** оценивают тесноту связи значений двух переменных, исключая влияние всех других переменных, представленных в уравнении множественной регрессии. Матрицу парных коэффициентов корреляции переменных можно рассчитать, используя инструмент «Анализ данных» ⇒«Корреляция». Для этого:

1). В главном меню последовательно выберите пункты **Данные ⇒Анализ данных⇒Корреляция**. Щелкните по кнопке ***ОК***;

2). Заполните диалоговое окно ввода данных и параметров вывода (рис.3).

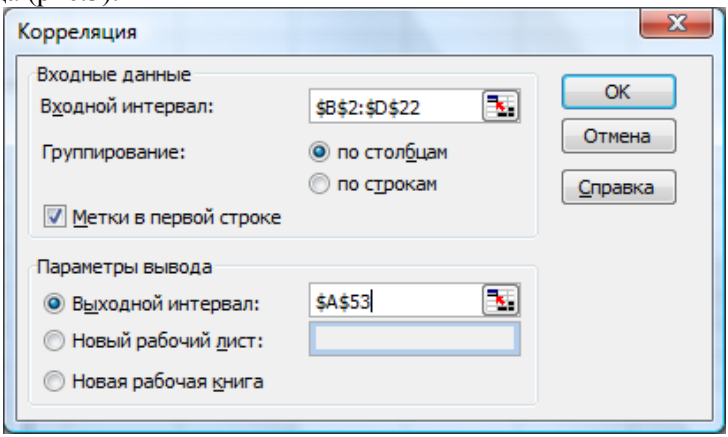



Рис.3 Диалоговое окно ввода данных и параметров вывода для вычисления коэффициентов парной корреляции

Значения **коэффициентов парной корреляции** указывают на заметную связь стоимости перевозок $Y$ как с весом груза – $X_1$, так и расстоянием – $X_2$ ($r_{yx1}$=0,66 и $r_{yx2}$=0,63). В то же время межфакторная связь $r_{x1x2}$ =0,12 довольно слабая, т.е. явной мультиколлинеарности нет. В связи с вышеизложенным, можно

сделать предварительный вывод, что нет оснований исключать факторы $X_1$ или $X_2$ из данной модели.

**Коэффициенты частной корреляции** дают более точную характеристику тесноты связи двух признаков, чем коэффициенты парной корреляции, так как очищают парную зависимость от взаимодействия данной пары признаков с другими признаками, представленными в модели (рис.4).

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 51 | **Надстройка-"Анализ данных"-" Корреляция"** | | | | | |
| 52 | **Значение коэффициентов парной корреляции** | | | | | |
| 53 | | *y* | *x1* | *x2* | | |
| 54 | y | 1 | | | | |
| 55 | x1 | 0,6552333 | 1 | | | |
| 56 | x2 | 0,6345813 | 0,124662 | 1 | | |
| 57 | **Коэффициенты частной корреляции** | | | | | |
| 58 | $r_{yx\,1\cdot x\,2} = \frac{r_{yx\,1} - r_{yx\,2}\cdot r_{x\,1\,x\,2}}{\sqrt{(1-r_{yx\,2}^{\,2})\cdot(1-r_{x\,1\,x\,2}^{\,2})}} =$ | | | | | |
| 59 | | | | | | **0,7513101** |
| 60 | | | | | | |
| 61 | $r_{yx\,2\cdot x\,1} = \frac{r_{yx\,2} - r_{yx\,1}\cdot r_{x\,1\,x\,2}}{\sqrt{(1-r_{yx\,1}^{\,2})\cdot(1-r_{x\,1\,x\,2}^{\,2})}} =$ | | | | | |
| 62 | | | | | | **0,7376567** |
| 63 | | | | | | |
| 64 | $r_{x\,1\,x\,2\cdot y} = \frac{r_{x\,1\,x\,2} - r_{yx\,2}\cdot r_{yx\,1}}{\sqrt{(1-r_{yx\,2}^{\,2})\cdot(1-r_{yx\,1}^{\,2})}} =$ | | | | | |
| 65 | | | | | | **-0,498662** |
| 66 | | | | | | |
| 67 | | | | | | |

Рис.4 Результаты вычисления коэффициентов корреляции (интервал A54:D56) и коэффициентов частной корреляции

Вычислим коэффициенты частной корреляции по рекуррентным формулам:

$$r_{x1x2\cdot y} = \frac{r_{x1x2} - r_{yx2}\cdot r_{yx1}}{\sqrt{(1-r_{yx2}^{\,2})\cdot(1-r_{yx1}^{\,2})}} = -0.49$$

$$r_{yx2\cdot x1} = \frac{r_{yx2} - r_{yx1} \cdot r_{x1x2}}{\sqrt{(1 - r_{yx1}^{\ 2}) \cdot (1 - r_{x1x2}^{\ 2})}} = 0.73$$

$$r_{yx1\cdot x2} = \frac{r_{yx1} - r_{yx2} \cdot r_{x1x2}}{\sqrt{(1 - r_{yx2}^{\ 2}) \cdot (1 - r_{x1x2}^{\ 2})}} = 0.75$$

Наиболее тесно связаны $Y$ и $X_1$ ($r_{yx1\ x2}$ =0,7513), связь $Y$ и $X_2$ чуть слабее*:* $r_{yx2\ x1}$=0,7376.

Между факторная зависимость $X_1$ и $X_2$ не очень сильная $|r_{x1x2\ \ y}|$ = 0,4987 , что подтверждает отсутствие коллинеарности между факторами.

Если сравнить коэффициенты парной и частной корреляции, то можно увидеть, что из-за наличия между факторной зависимости они отличаются друг от друга:

$r_{yx1}$= 0,6552; $r_{yx1\ x2}$= 0,7513; $r_{yx2}$=0,6346; $r_{yx2\ x1}$=0, 7376.

Частные коэффициенты корреляции между $Y$ и $X_1$, $Y$ и $X_2$ свидетельствуют о более сильных взаимосвязях переменных, чем это показывают значения парных коэффициентов корреляции. Это произошло потому, что парный коэффициент корреляции $r_{x1x2}$ занизил тесноту связи между $Y$ и $X_1$ , $Y$ и $X_2$

***К пункту 4***

*Вычисление параметров линейного уравнения множественной регрессии*.

Нахождение коэффициентов регрессии можно выполнить, используя функцию ЛИНЕЙН() (рис.5).

Нахождение коэффициентов регрессии можно провести с помощью инструмента «Анализ Данных» ⇒ «Регрессия» (рис.6).

Следует помнить, что в отличие от парной регрессии в диалоговом окне при заполнении параметра «входной интервал X» следует указать не один столбец, а все столбцы, содержащие значения факторных признаков.

Результат приведен на рис.7.

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 99 | | | | | | |
| 100 | **Уравнение множественной регрессии.** | | | | | |
| 101 | Лин | 15,10401 | 1,156057 | -17,31332 | | |
| 102 | | 3,352977 | 0,24629 | 6,44711 | | |
| 103 | | 0,739853 | 10,98002 | #Н/Д | | |
| 104 | | 24,17386 | 17 | #Н/Д | | |
| 105 | | 5828,843 | 2049,535 | #Н/Д | | |
| 106 | | | | | | |
| 107 | | X2 | X1 | 1 | | |
| 108 | Ура | Y=15,10*X2+1,16*X1-17,31 | | | | |
| 109 | | | | | | |

Рис.5. Результаты применения функции ЛИНЕЙН().

По результатам всех вычислений уравнение множественной регрессии имеет вида

$$Y = -17.3 + 1.16 \cdot X_1 + 15.10 \cdot X_2$$

Величины $b_1$ и $b_2$ указывают, что с увеличением значений $X_1$ и $X_2$ на единицу отклик увеличивается соответственно на 1,16 и на 15,10 тыс.руб.

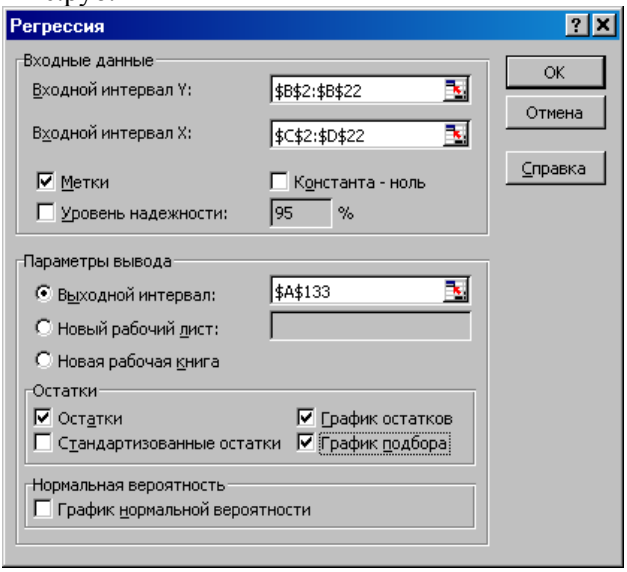



Рис..6 Диалоговое окно «Регрессия» инструмента «Анализ Данных»

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 133 | **ВЫВОД ИТОГОВ** | | | | | | |
| 134 | | | | | | | |
| 135 | ***Регрессионная статистика*** | | | | | | |
| 136 | **Множественный R** | **0,860147** | | | | | |
| 137 | **R-квадрат** | **0,739853** | | | | | |
| 138 | **Нормированный R-квадрат** | **0,709248** | | | | | |
| 139 | **Стандартная ошибка** | **10,98002** | | | | | |
| 140 | **Наблюдения** | **20** | | | | | |
| 141 | | | | | | | |
| 142 | **Дисперсионный анализ** | | | | | | |
| 143 | | ***df*** | ***SS*** | ***MS*** | ***F*** | ***Значимость F*** | |
| 144 | **Регрессия** | **2** | **5828,843** | **2914,421** | **24,17386** | **1,06993E-05** | |
| 145 | **Остаток** | **17** | **2049,535** | **120,5609** | | | |
| 146 | **Итого** | **19** | **7878,378** | | | | |
| 147 | | | | | | | |
| 148 | | ***Коэффициенты*** | ***Стандартная ошибка*** | ***t-статистика*** | ***P-Значение*** | ***Нижние 95%*** | ***Верхние 95%*** |
| 149 | **Y-пересечение** | **-17,31332** | **6,44711** | **-2,685439** | **0,015643** | **-30,91555429** | **-3,711089** |
| 150 | **x1** | **1,156057** | **0,24629** | **4,693892** | **0,000209** | **0,636430441** | **1,675683** |
| 151 | **x2** | **15,10401** | **3,352977** | **4,504657** | **0,000313** | **8,029837723** | **22,17818** |
| 152 | | | | | | | |

Рис.7 Результаты применения инструмента «Анализ Данных» ⇒ «Регрессия»

## ***К пункту 5***

Для вычисления коэффициентов уравнения регрессии в стандартизованном масштабе используем формулы (6).

$$\beta_1 = b_1 \frac{\sigma_{X_1}}{\sigma_Y} = 1{,}16 \cdot \frac{10{,}05}{19{,}85} = 0{,}585,\ \beta_2 = b_2 \frac{\sigma_{2_1}}{\sigma_Y} = 15{,}10 \cdot \frac{0{,}74}{19{,}85} = 0{,}56$$

С учетом этого, уравнение регрессии в стандартном масштабе будет иметь вид:

$$t_Y = 0{,}58 t_{X_1} + 0{,}56 t_{X_2}$$

То есть, с ростом груза на одну сигму при неизменном расстоянии стоимость грузовых автомобильных перевозок увеличивается в среднем на 0,58 сигмы.

Поскольку значения коэффициентов отличаются друг от друга незначительно, то влияние на стоимость грузовых автомобильных обоих факторов приблизительно одинаково.

***К пункту 6***

Рассчитаем средние коэффициенты эластичности

$$\overline{Э}_{YX_1} = f'(\overline{X}_1)\frac{\overline{X}_1}{\overline{Y}} = b_1 \frac{\overline{X}_1}{b_0 + b_1\overline{X}_1 + b_2\overline{X}_2} = 1{,}16 \cdot 13{,}86 /$$

$$\left(-17{,}31 + 1{,}16 \cdot 13{,}86 + 15{,}10 \cdot 1{,}59\right) = 0{,}71$$

$$\overline{Э}_{YX_1} = f'(\overline{X}_2)\frac{\overline{X}_2}{\overline{Y}} = b_2 \frac{\overline{X}_2}{b_0 + b_1\overline{X}_1 + b_2\overline{X}_2} = 1{,}05 \text{ ё}$$

С увеличением среднего веса груза на 1% от его среднего уровня средняя стоимость перевозок возрастет на 0,71% от своего среднего уровня; при увеличении среднего расстояния перевозок на 1% - средняя стоимость доставки груза увеличится на 1,05%. Различия в силе влияния факторов на результат, полученные при сравнении уравнения регрессии в стандартизованном масштабе и коэффициентов эластичности, объясняются тем, что при вычислении коэффициентов эластичности учитывают поведение уравнения регрессии в окрестности средних значений.

***К пункту 7***

Величина коэффициента множественной детерминации $R^2$, рассчитанная по определению (по формуле (8) и с использованием коэффициентов уравнения множественной регрессии в стандартизованном масштабе (по формуле (9) оказалась одинаковой и равной 0.74. Расчеты приведены на рис.9.

Полученный результат совпадает с результатом, полученным с помощью надстройки «Анализ Данных» $\Rightarrow$«Регрессия», содержащимся в ячейке B136 («Множественный R») на рис.7, лист MS Excel в режиме отображения формул приводится на рис.8.

Поскольку коэффициент множественной детерминации оценивает долю вариации результата за счет представленных в уравнении факторов в общей вариации результата и $R^2_{yx_1x_2} = 0{,}7398$, то эта доля составляет 74 % и указывает на весьма высокую степень обусловленности вариации результата вариацией

факторов, иными словами – на весьма тесную связь факторов с результатом. О шкале Чеддока сила связи оценивается как высокая.

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 181 | **Вычисление коэффициента множественной детерминации** | | | | | | |
| 182 | | | | | | | |
| 183 | **R^2=1-Sост/Sполн.** | | | | | **=** | **0,7399** |
| 184 | | | | | | | |
| 185 | **R^2=beta1*ryx1+beta2*ryx2** | | | | | **=** | **0,7399** |
| 186 | | | | | | | |
| 187 | **Вычисление исправленного коэффициента** | | | | | | |
| 188 | **множественной детерминации** | | | | | | |
| 189 | | | | | | | |
| 190 | **(R^2)adj=1-(1-R^2)*(n-1)/(n-p-1)** | | | | | **=** | **0,7092** |

Рис.8 Вычисление коэффициента множественной детерминации в MS Excel в режиме отображения данных.

| | A | F | G |
|---|---|---|---|
| 181 | **Вычисление коэффициента мн** | | |
| 182 | | | |
| 183 | **R^2=1-Sост/Sполн.** | **=** | **=1-H23/E23** |
| 184 | | | |
| 185 | **R^2=beta1*ryx1+beta2*ryx2** | **=** | **=C156*C168+B156*D168** |
| 186 | | | |
| 187 | **Вычисление исправленного ко** | | |
| 188 | **множественной детерминации** | | |
| 189 | | | |
| 190 | **(R^2)adj=1-(1-R^2)*(n-1)/(n-p-1)** | **=** | **=1-(1-B137)*B146/B145** |

Рис.9 Вычисление коэффициента множественной детерминации в MS Excel в режиме отображения формул.

## ***К пункту 8***

Оценку надежности уравнения регрессии в целом и показателя тесноты связи $R^2$ дает $F$- критерий Фишера.

По данным дисперсионного анализа, представленным в интервале ячеек A142:F146 на рис.10 , $F_{набл} = 24{,}16$.

Вероятность случайно получить такое значение $F$ – критерия составляет $1{,}07*10^{-5}$ , ячейка F144 («Значимость F»), что не превышает допустимый уровень значимости 5%. Следовательно, полученное значение не случайно, оно сформировалось под

влиянием существенных факторов, т.е. подтверждается статистическая значимость всего уравнения и показателя тесноты связи $R^2_{yx_1x_2}$.

***К пункту 9***

Оценку значимости коэффициентов множественной регрессии произведем с помощью ***t*** -критерия Стьюдента с использованием формулу (20).

Значения случайных ошибок параметров $b_0$, $b_1$ и $b_2$ с учетом округления:

$m_{b0}$=6,4471; $m_{b1}$=0,2463; $m_{b2}$=3,3530

Эти значения содержатся в диапазоне ячеек B102:D102 (рис.8), который является частью результата применения функции ЛИНЕЙН.

Эти значения используются для расчета $t$ - критерия Стьюдента по формуле (20):

$$t_{b_0} = \frac{-17{,}3133}{6{,}4471} = -2{,}68;\ t_{b_1} = \frac{1{,}1560}{0{,}2463} = 4{,}69;\ t_{b_2} = 4{,}50.$$

Модули вычисленных величин следует сравнить с $t_{крит.}$, определяемым с заданным уровнем значимости и числом степеней свободы равным $n$-$p$-1.

Значение $t_{крит}$ равно 2.10. Поскольку модули значений наблюдаемых значений больше критического, гипотеза о равенстве нулю коэффициентов отвергается.

Это позволяет сделать вывод о существенности данного параметра, который формируется под воздействием неслучайных причин. Здесь статистически значимыми являются все коэффициенты $b_0$ , $b_1$ и $b_2$.

Процедура проверки значимости коэффициентов уравнения с помощью инструмента «Анализ Данных» $\Rightarrow$ «Регрессия» существенно проще, поскольку все промежуточные операции выполняются автоматически. На рис.7 в интервале B149:E151 приведены значения коэффициентов регрессии, стандартных

ошибок, $t$ – статистики ($t$-наблюдаемые) и $P$ – значения соответственно.

Для анализа значимости коэффициентов регрессии столбец «*P*-значение» в интервале E86:E88: если он меньше принятого нами уровня значимости (в настоящей работе уровень значимости принят равным 0,05), делают вывод о неслучайной природе данного значения коэффициента, т.е. о том, что он статистически значим и надежен. В противоположном случае принимается гипотеза о случайной природе значения этого коэффициента уравнения. Здесь все $P$<0,05, что позволяет подтвердить сделанный ранее вывод о статистической значимости всех параметров регрессии

Очень важно помнить, что все найденные коэффициенты $b_0$, $b_1$ и $b_2$.являются точечными оценками коэффициентов регрессии, при этом интервальными оценками с надежностью 95% являются интервалы, границы которых указаны в таблице 4.

*Таблица 4*

| Коэффициент | Интервал | |
|---|---|---|
| | Левая граница | Правая граница |
| $b_0$ | -30,92 | -3,71 |
| $b_1$ | 0,636 | 1,676 |
| $b_2$ | 8,030 | 22,178 |

Знание этих интервалов позволяет получить важные выводы о том, что с увеличением веса груза на одну тонну (при неизменном значении расстояния) с вероятностью 95% стоимость поездки в среднем возрастет на величину от 0,636 до 1,676 тыс.руб. Увеличение расстояния на одну тыс. км (при неизменном значении веса груза) с вероятностью 95% увеличивает в среднем стоимость поездки на величину от 8,030 до 22,178 тыс.руб.

***К пункту 10***

Для оценки целесообразности включения в модель фактора $X_1$ после фактора $X_2$ и фактора $X_2$ после фактора $X_1$ вычислим значения частных F-критериев Фишера:

$$F_{\text{частн } x_2} = \frac{R^2_{yx1x2} - R^2_{yx1}}{1 - R^2_{yx1x2}} \cdot \frac{n-p-1}{1} = \frac{0.7398 - 0.6552^2}{1 - 0.7398} \cdot \frac{20-2-1}{1} = 20.29$$

$$R^2_{yx1=} r^2_{yx1}$$

$$F_{\text{частн } x_1} = \frac{R^2_{yx1x2} - R^2_{yx2}}{1 - R^2_{yx1x2}} \cdot \frac{n-p-1}{1} = \frac{0.7398 - 0.6345^2}{1 - 0.7398} \cdot \frac{20-2-1}{1} = 22.03$$

$$R^2_{yx2=} r^2_{yx2}$$

Частный $F$-критерий – $F_{\text{частн } x_2}$ показывает статистическую зависимость включения фактора $X_2$ в модель после того, как в нее включен фактор $X_1$ . $F_{\text{частн } x_2}$ = 20,29. Найдем $F_{крит}$=4,45 при принятом уровне значимости $a$ =0,05 (5%) (число степеней свободы числителя и знаменателя равны 1 и 17 соответственно).

$F_{\text{частн } x_2}$ = 20,29 > $F_{\text{крит}}$ =4,45. Следовательно, включение в модель фактора $X_2$ – расстояния, после того, как уравнение включен фактор $X_1$ – вес груза, статистически целесообразно: прирост факторной дисперсии за счет дополнительного признака $X_2$ оказывается значительным, существенным; фактор $X_2$ следует включать в уравнение после фактора $X_1$ .

Поменяем первоначальный порядок включения факторов в модель и рассмотрим вариант включения $X_1$ после $X_2$ . Для этого вычислим $F_{\text{частн } x_1}$ , оно равно 22,03 при том же уровне значимости $a$ =0,05 (5%). $F_{\text{крит}}$ =4,45 и $F_{\text{частн } x_1} > F_{\text{крит}}$ . Следовательно, значение частного $F$ -критерия для дополнительно включенного фактора $X_1$ не случайно, является статистически значимым, надежным, достоверным: прирост факторной дисперсии за счет дополнительного фактора $X_1$ является существенным.

Фактор $X_1$ должен присутствовать в уравнении, в том числе в варианте, когда он дополнительно включается после фактора $X_2$ .

***К пункту 11***

Вычислим значения каждого фактора $X_1^{\text{прогн.}}$ (тонн) - вес груза и $X_2^{\text{прогн.}}$ (тыс. км) - расстояние, в которых будем строить прогноз. В качестве прогнозных значений возьмем величину равную $x_{\max} - 0.1 \cdot (x_{\max} - x_{\min})$, где $x_{\max}$ и $x_{\min}$ максимальное и минимальное значения факторов в таблице исходных данных. Вычислим прогнозные значения каждого фактора с учетом того, что максимальные значения и размахов $(x_{\max} - x_{\min})$ уже были вычислены и приведены на рис.1.*б*. $X_1^{\text{прогн.}} = 31{,}7$; $X_2^{\text{прогн.}} = 2{,}645$.

Все вычисления приведены на рис.10-16. Заметим, что перед вычислением все необходимые данные были скопированы на новую страницу MS Excel. Для расчета точечного прогноза $\hat{Y}$ подставим полученные результаты вычислений в уравнение множественной регрессии

$$\hat{Y} = -17.3 + 1.16 \cdot X_1^{прогн.} + 15.10 \cdot X_2^{прогн.} =$$

$$= -17.3 + 1.16 \cdot 31.7 + 15.10 \cdot 2.645 = 59.28$$

Полученное прогнозное значение будет серединой интервального прогноза. Для того, чтобы вычислить ширину доверительного интервала необходимо вычислить выражение (30). Вектор $X0$ расположен в интервале C2:C4 (рис.10) и содержит прогнозные значения факторов, при этом первый элемент всегда равен единице.

| | A | B | C |
|---|---|---|---|
| 1 | | Прогнозирование | |
| 2 | | | 1 |
| 3 | | X0= | 31,7 |
| 4 | | | 2,645 |
| 5 | | | |
| 6 | X | | |
| 7 | x0 | x1 | x2 |
| 8 | 1 | 35,00 | 2,00 |
| 9 | 1 | 16,00 | 1,10 |
| | | | |
| 26 | 1 | 9,00 | 2,20 |
| 27 | 1 | 4,50 | 0,95 |

Рис.10 Расчет интервального прогноза. Режим отображения данных. (начало)

Матричные операции для вычисления $\sqrt{(X0)^T\left(X^TX\right)X0}$ проведены на рис. 12.

| | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | | | | | | | | | | | | | | | | | | | | |
| 2 | X0__transp | 1 | 31,7 | 2,645 | | | | | | | | | | | | | | | | | | | |
| 3 | | | | | | | | | | | | | | | | | | | | | | | |
| 4 | | | | | | | | | | | | | | | | | | | | | | | |
| 5 | | | X_transp | | | | | | | | | | | | | | | | | | | | |
| 6 | | | x0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 7 | | | x1 | 35 | 16 | # | 2 | # | # | # | # | 2 | # | # | 3 | 4 | 3 | # | 3 | # | 9 | 4,5 | |
| 8 | | | x2 | 2 | 1,1 | 3 | 2 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 0 | 2 | 3 | 1 | 1 | 1 | 3 | 2,2 | 0,95 |

Рис.11 Расчет интервального прогноза. Режим отображения данных. Транспонирование матриц (продолжение)

| | F | G | H | I | J | K | L |
|---|---|---|---|---|---|---|---|
| 14 | X_transp*X | | | | (X_transp*X)^(-1) | | |
| 15 | 20,0 | 277,2 | 31,8 | | 0,3 | 0,0 | -0,1 |
| 16 | 277,2 | 5860,9 | 459,2 | | 0,0 | 0,0 | 0,0 |
| 17 | 31,8 | 459,2 | 61,5 | | -0,1 | 0,0 | 0,1 |
| 18 | | | | | | | |
| 19 | | | | | | | |
| 20 | | | | | | | |
| 21 | | | | | | | |
| 22 | X0_transp*(X_transp*X)^(-1) | | | | | | |
| 23 | -0,194 | 0,0081 | 0,083 | | | | |
| 24 | | | | | | | |
| 25 | X0_transp*(X_transp*X)^(-1)*X0 | | | | | | |
| 26 | 0,282 | | | | | | |
| 27 | КОРЕНЬ(X0_*(X_transp*X)^(-1)*X0) | | | | | | |
| 28 | 0,531 | | | | | | |

Рис.12 Расчет интервального прогноза. Режим отображения данных.

Вычисление $\sqrt{(X0)^T\left(X^TX\right)X0}$ (продолжение)

В ячейку G34 записано значение $s$ – стандартное отклонение остаточной дисперсии или стандартная ошибка уравнения регрессии, которая была получена с помощью инструмента «Анализ Данных» ⇒ «Регрессия» содержится в ячейке G139 (рис.7).

Искомая величина $\varepsilon$ – половина ширины доверительного интервала вычислена по формуле (30) и содержится в ячейке G36 рис.13.

| | F | G | H | I | J | K |
|---|---|---|---|---|---|---|
| 27 | КОРЕНЬ(X0_*(X_transp*X)^(-1)*X0) | | | | | |
| 28 | 0,531 | | | | | |
| 29 | Gamma | 0,95 | | | | |
| 30 | n= | 20 | | | | |
| 31 | p= | 2 | | | | |
| 32 | t_крит= | 2,1098 | | | | |
| 33 | | | | | | |
| 34 | s= | 10,98 | | | | |
| 35 | | | | | | |
| 36 | eps= | 12,297 | | | | |
| 37 | | | | | | |
| 38 | | | | | | |
| 39 | | Коэффициент | | Точечный прогноз | | |
| 40 | Y-пере | -17,313 | | Y_прогн= | 59,3 | |
| 41 | X1 | 1,1561 | | | | |
| 42 | X2 | 15,104 | | Доверительный интервал | | |
| 43 | | | | Y_нижн | | Y_верхн |
| 44 | | | | 46,987 | | 71,581 |

Рис.13 Расчет точечного и интервального прогноза. Окончание. Режим отображения данных

Границы доверительного интервала вычислены с использованием формулы (29) и помещены в ячейки I44 и K44.

Все вычисления в режиме отображения формул приведены на рис.14 -16.

| | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|
| 14 | X_transp*X | | | | | | (X_transp*X)^(-1) | |
| 15 | =МУМНОЖ(A8:C27;G6:Z8) | =МУМНОЖ(A | =МУМНОЖ(A8:C27;G6:Z8) | | | | =МОБР(F15:H17) | =M( |
| 16 | =МУМНОЖ(A8:C27;G6:Z8) | =МУМНОЖ(A | =МУМНОЖ(A8:C27;G6:Z8) | | | | =МОБР(F15:H17) | =M( |
| 17 | =МУМНОЖ(A8:C27;G6:Z8) | =МУМНОЖ(A | =МУМНОЖ(A8:C27;G6:Z8) | | | | =МОБР(F15:H17) | =M( |
| 22 | X0_transp*(X_transp*X)^(-1) | | | | | | | |
| 23 | =МУМНОЖ(E2:G2;L15:N17) | =МУМНОЖ(E | =МУМНОЖ(E2:G2;L15:N17) | | | | | |
| 24 | | | | | | | | |
| 25 | X0_transp*(X_transp*X)^(-1)*X0 | | | | | | | |
| 26 | =МУМНОЖ(F23:H23;C2:C4) | | | | | | | |
| 27 | КОРЕНЬ(X0_*(X_transp*X)^(-1)*X0) | | | | | | | |
| 28 | =КОРЕНЬ(F26) | | | | | | | |

Рис.14 Расчет интервального прогноза. Режим отображения формул. Операции с матрицами и вычисление $\sqrt{(X0)^T\left(X^TX\right)X0}$ (продолжение)

| | F | G |
|---|---|---|
| 28 | =КОРЕНЬ(F26) | |
| 29 | Gamma= | 0,95 |
| 30 | n= | 20 |
| 31 | p= | 2 |
| 32 | t_крит= | =СТЬЮДРАСПОБР(1-G29;G30-G31-1) |
| 33 | | |
| 34 | s= | 10,98002099 |
| 35 | | |
| 36 | eps= | =F28*G32*G34 |
| 37 | | |
| 38 | | |
| 39 | | Коэффициент |
| 40 | Y-пересечение | -17,31332187 |
| 41 | X1 | 1,156056725 |
| 42 | X2 | 15,10400985 |

Рис.15 Расчет интервального прогноза. Режим отображения формул. Вычисление ширины интервала (продолжение)

| | I | J | K | L |
|---|---|---|---|---|
| 39 | Точечный прогноз | | | |
| 40 | Y_прогн= | =МУМНОЖ(E2:G2;G40:G42) | | |
| 41 | | | | |
| 42 | Доверительный интервал | | | |
| 43 | Y_нижн | | Y_верхн | |
| 44 | =J40-G36 | | =J40+G36 | |

Рис.16. Расчет точечного и интервального прогноза. Окончание. Режим отображения формул.

Интервальной оценкой прогноза с указанными значениями факторов является доверительный интервал с надежностью 95 % [46,99 ; 71,59 ] тыс.руб.

Так, стоимость перевозки груза весом 13,86 тонн на расстояние 1,59 тыс. км с вероятностью 95 % будет лежать в пределах [46,99 ; 71,59 ].

**Общий вывод** состоит в том, что множественная линейная модель

$$\hat{Y} = -17{,}31 + 1{,}16 \cdot X_1 + 15{,}10 \cdot X_2$$

с факторами $X_1$ и $X_2$ имеет коэффициент детерминированности $R^2_{yx_1x_2}$ =0,73. Она содержит информативные факторы $X_1$ и $X_2$.

Уравнение парной регрессии является простым, хорошо детерминированным, пригодным для анализа и для прогноза

.

## ЗАДАНИЕ

Для ряда регионов представлена информация об объёмах $Y$ **(у.е.)** продаж фирмы «Галактика» и ее затратах на рекламу в этих регионах – $X_1$, а также индекс потребительских доходов в этих регионах – $X_2$. Построить и оценить линейную модель множественной регрессии по плану, приведенному в примере, изложенном выше.

Исходные данные взять из файла «*LabRab_6.xls*».

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК


1. Эконометрика. Парная линейная регрессия. Методические указания к лабораторным работам для студентов направлений подготовки бакалавриата 21.03.02 и 38.03.01 / Сост.: *В.В. Беляев, Т.Р. Косовцева.* СПб, 2016. - 50 с.
2. *Магнус Я.Р.* Эконометрика. Начальный курс. Учебник для вузов. - / Я.Р. Магнус, П.К. Катышев, А.А. Пересецкий; М., «Дело» 2000. - 400 с.
3. Эконометрика./ Учебник для бакалавров. Под ред. Елисеевой И.И., М., «Проспект», 2014. - 288 с..
4. Практикум по эконометрике./ Под редакцией Елисеевой И.И., М., «Финансы и статистика», 2004. - 192 с.
5. *Тихомиров Н.* Эконометрика. Учебник / Н. Тихомиров, Е.М. Дорохина: М.: «Экзамен», 2006 . - 512 с.
6. *Кремер Н. Ш.* Эконометрика. Учебник для вузов, / Н.Ш. Кремер, Б.А. Путко М.: М.: Юнити, 2005. - 311 с.
7. *Арженовский С.В*. Эконометрика: учебное пособие/ С.В. Арженовский, О.Н. Федосова. Рост.гос.экон.университет – Ростов н/Д., 2002. - 102 с.

# СОДЕРЖАНИЕ

ВВЕДЕНИЕ ..... 3

ЛАБОРАТОРНАЯ РАБОТА. МНОЖЕСТВЕННАЯ РЕГРЕССИЯ ..... 4

1. Теоретические сведения. Базовые понятия ..... 4
2. коэффициент детерминации ..... 9
3. Оценка надежности уравнения множественной регрессии ..... 11
4. Скорректированный индекс множественной детерминации ..... 16
5. Частная корреляция ..... 16
6. матричная форма записи ..... 18
7. Мультиколлинеарность факторов ..... 18
8. Построение прогноза с помощью уравнения регрессии ..... 20

ПРИМЕР ..... 21

ЗАДАНИЕ ..... 40

БИБЛИОГРАФИЧЕСКИЙ СПИСОК ..... 41

**ЭКОНОМИКО-МАТЕМАТИЧЕСКИЕ МЕТОДЫ И МОДЕЛИРОВАНИЕ**

**МНОЖЕСТВЕННАЯ РЕГРЕССИЯ**

***Методические указания к лабораторным работам для студентов бакалавриата направления 21.03.02***

Сост.: *В.В. Беляев, Т.Р. Косовцева*

Печатается с оригинал-макета, подготовленного кафедрой информатики и компьютерных технологий

Ответственный за выпуск *Т.Р. Косовцева*

Лицензия ИД № 06517 от 09.01.2002

Подписано к печати 22.01.2020. Формат 60×84/16.
Усл. печ. л. 2,4. Усл.кр.-отт. 2,4. Уч.-изд.л. 1,8. Тираж 75 экз. Заказ 18. С 1.

Санкт-Петербургский горный университет
РИЦ Санкт-Петербургского горного университета
Адрес университета и РИЦ: 199106 Санкт-Петербург, 21-я линия, 2