Эконометрика – это наука, изучающая конкретные количественные и качественные взаимосвязи экономических объектов и процессов с помощью математических и статистических методов и моделей. (Большой Энциклопедический Словарь – М., БРЭ, 1977)

Эконометрические методы – это, прежде всего, методы статистического анализа конкретных экономических данных.

Оценка результатов эконометрического моделирования достигается посредством решения качественной и количественной проблемы. Качественная составляющая заключается в установлении соответствия между построенной моделью и основополагающей экономической концепцией, а количественная – в точности аппроксимации имеющейся информации данными расчётов.

С практической точки зрения к основным задачам эконометрики можно отнести:

- построение эконометрических моделей – представление экономических моделей в математической форме, удобной для проведения эмпирического анализа. Данную проблему называют проблемой спецификации, которую можно решить несколькими способами;
- оценку параметров построенной модели, позволяющую характеризовать адекватность модели реальными данными. Указанная задача решается на этапе параметризации;
- проверку качества полученной модели в целом. Данная задача реализуется на этапе верификации;
- использование построенной модели для прогнозирования.

Модель парной линейной регрессии является частным случаем модели многомерной регрессии. Её исследование представляет самостоятельный интерес, так как она имеет многие характерные свойства общих многомерных моделей, но более наглядна и проста для изучения.

Линейный парный корреляционно-регрессионный анализ
Однофакторный дисперсионный анализ

1. На основе данных, приведённых в Приложении 1 и соответствующих Вашему варианту, построить в MS Excel поле корреляции и сформулировать гипотезу о форме связи. Один из признаков, соответствующих Вашему варианту, будет играть роль факторного X, другой – результативного Y.
2. Рассчитать параметры выборочного уравнения линейной регрессии с помощью МНК. Пояснить экономический смысл параметров выборочного уравнения регрессии.
3. Рассчитать и пояснить экономический смысл косвенных признаков качества линейного уравнения регрессии: средней ошибки аппроксимации и среднего коэффициента эластичности.
4. Рассчитать дисперсии на одну степень свободы и построить таблицу дисперсионного анализа. Оценить статистическую значимость уравнения регрессии в целом с помощью F-критерия Фишера.
5. Рассчитать коэффициент детерминации, скорректированный коэффициент детерминации, коэффициент линейной парной корреляции. Сделать вывод о характере тесноты связи.
6. Оценить статистическую значимость параметров регрессии и линейного коэффициента корреляции с помощью t-критерия Стьюдента. Построить доверительные интервалы для параметров уравнения регрессии и линейного коэффициента корреляции для однопроцентного и пятипроцентного уровня значимости.
7. Выполнить прогноз ожидаемого значения признака-результата Y при прогнозном значении признака-фактора X, составляющего $(100+N+1)\%$ от среднего уровня $\bar{X}$, т.е. $X=\left(1+\dfrac{N+1}{100}\right)\cdot\bar{X}$, где N – номер варианта. Оценить точность прогноза, рассчитав ошибку прогноза и его доверительный интервал с вероятностью 95% и 99%.
8. На поле корреляции X, Y построить линию регрессии с указанием на диаграмме её уравнения и коэффициента детерминации. На промежутке изменения факторного признака X построить гиперболы – левую и правую границы интервалов для прогнозных значений.
9. Результаты дисперсионного и корреляционно-регрессионного анализа проверить с помощью надстройки MS Excel Анализ данных Регрессия.

Внимание: как выбирать данные из приложения 1
Например, у Вас вариант 12 (последние 2 цифры зачётки).
Из Таблицы 1 Приложения 1 выбираются соответствующие данные:

Вариант	**Номер начального наблюдения из Таблицы 2**	**Номер конечного наблюдения из Таблицы 2**	**Номер признаков из Таблицы 2**
12	**6**	**55**	**4,5**

Далее, по Таблице 2 выделяются соответствующие строки и столбцы для анализа

[illegible]	Собственные оборотные средства, млн. руб.	Балансовая прибыль, млн. руб.	Дебиторская задолженность, млн. руб.	Дивиденды, начисленные по результатам деятельности X	Курсовая цена акции, руб. Y
А	1	2	3	4	5
1	1011	107	37	20,33	92
2	799	102	64	20,04	83
3	995	107	71	19,87	95
4	1243	122	36	20,48	124
5	1507	108	51	20,13	96
6	947	108	41	20,26	106
7	1015	97	78	19,89	70
8	1169	109	42	19,92	97
9	1051	101	68	19,78	76
10	1372	116	34	20,23	112
11	1463	113	49	20,46	113
12	684	112	40	20,07	109
13	1251	106	56	20,23	91
14	1376	111	45	20,36	95
15	1193	113	44	20,28	115
16	1386	122	40	20,52	114
17	1631	118	47	20,28	123
18	1735	119	47	19,97	116
19	1181	102	49	19,97	85
20	922	100	65	19,57	91
21	1281	103	54	19,94	82
22	1333	113	59	20,29	105
23	1632	124	36	20,83	124
24	635	95	70	19,59	70
25	949	102	64	19,76	84
26	788	112	48	20,19	106
27	1728	124	30	20,66	128
28	1773	116	58	19,95	105
29	1679	118	48	20,61	121
30	1085	100	69	20,03	79
31	1214	99	58	19,78	82
32	1422	107	49	20,23	80
33	523	87	76	19,78	37
34	1025	109	59	20,09	101
35	1083	106	74	20,13	98
36	1466	113	54	20,56	98
37	1643	123	36	20,51	124
38	287	82	75	19,71	39
39	704	104	51	20,1	88
40	1177	112	35	20,32	108
41	1792	116	47	20,37	112
42	2072	106	33	20,03	80
43	1178	120	38	20,65	120
44	1204	105	58	20,19	88
45	1208	114	32	20,34	104
46	1416	107	58	20,27	94
47	1185	115	44	20,89	107
48	1220	96	68	19,85	82
49	1211	104	64	19,87	84
50	1288	108	35	20,2	101
51	918	102	54	20,33	98
52	809	102	70	20,2	89
53	1188	120	19	20,46	118
54	1294	106	28	20,17	90
55	1425	114	54	20,63	123
56	1514	112	48	19,79	107
57	1577	112	44	20,34	97

Регрессионный анализ

Модель линейной парной регрессии

Имеется два ряда эмпирических данных $X=(x_i)$ и $Y=(y_i)$, где y_i – результаты наблюдений в точках x_i, здесь и далее $i=\overline{1,n}$. Множество точек (x_i, y_i) на координатной плоскости называется *полем корреляции*.

Пусть по расположению эмпирических точек можно предположить наличие линейной корреляционной зависимости между величинами X и Y. В общем виде *теоретическую линейную парную регрессионную модель* можно представить так:

$$Y = M\left(Y\middle|X\right) = a + b\cdot X + \varepsilon \text{, или } y_i = M\left(Y\middle|X = x_i\right) = a + b\cdot x_i + \varepsilon_i,$$

где

Y – объясняемая (результирующая, зависимая, эндогенная) переменная,
X – объясняющая (факторная, независимая, экзогенная) переменная или регрессор;
a, b – теоретические параметры (числовые коэффициенты) регрессии, подлежащие оцениванию;
$\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ – теоретические случайные отклонения (возмущения, ошибки).

Задача линейного регрессионного анализа состоит в том, чтобы по имеющимся статистическим данным (x_i, y_i) построить так называемое *эмпирическое* (*выборочное*) *уравнение регрессии,* получив наилучшие оценки неизвестных параметров a и b,

$$\hat{Y} = \hat{a} + \hat{b}\cdot X \text{ или } \hat{y}_i = \hat{a} + \hat{b}\cdot x_i,$$

где

$\hat{Y} = (\hat{y}_i)$ – предсказанные (прогнозируемые значения объясняемой переменной), $\hat{y}_i$ – оценки условного математического ожидания $M\left(Y\middle|X = x_i\right)$;
$\hat{a}$ и $\hat{b}$ – оценки неизвестных параметров a, b, называемые *эмпирическими коэффициентами регрессии*.

Если по данной выборке найдены эмпирические коэффициенты регрессии $\hat{a}$ и $\hat{b}$, то уравнение теоретической регрессионной модели можно выписать с их использованием:

$$Y = \hat{a} + \hat{b}\cdot X + \hat{\varepsilon} \text{ или } y_i = \hat{a} + \hat{b}\cdot x_i + \hat{\varepsilon}_i,$$

где

$\hat{\varepsilon} = (\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n)$ – оценки теоретических случайных отклонений ε, ошибки (остатки) модели.

Оценка коэффициентов линейной парной регрессии. Метод наименьших квадратов. Теорема Гаусса-Маркова

Классический подход к оцениванию параметров линейной регрессии основан на методе наименьших квадратов (МНК). В МНК оценки параметров модели $\hat{a}$ и $\hat{b}$ строятся так, чтобы минимизировать сумму квадратов ошибок модели по всем наблюдениям:

$$S\left(\hat{a},\hat{b}\right)=\sum_{i=1}^{n}\hat{\varepsilon}_i^2=\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2=\sum_{i=1}^{n}\left(y_i-\hat{a}-\hat{b}\cdot x_i\right)^2\to\min.$$

Необходимым условием существования минимума функции двух переменных $S\left(\hat{a},\hat{b}\right)$ является равенство нулю её частных производных:

$$\begin{cases}\dfrac{\partial S}{\partial\hat{a}}=-2\sum\left(y_i-\hat{a}-\hat{b}\cdot x_i\right)=0,\\ \dfrac{\partial S}{\partial\hat{b}}=-2\sum\left(y_i-\hat{a}-\hat{b}\cdot x_i\right)\cdot x_i=0,\end{cases}\qquad \sum=\sum_{i=1}^{n}.$$

В результате для определения эмпирических коэффициентов регрессии $\hat{a}$, $\hat{b}$ получается *линейная система нормальных уравнений регрессии*

$$\begin{cases}\hat{b}\cdot\sum x_i^2 & + & \hat{a}\cdot\sum x_i & = & \sum y_i\cdot x_i\\ \hat{b}\cdot\sum x_i & + & \hat{a}\cdot n & = & \sum y_i\end{cases}.$$

Из системы нормальных уравнений регрессии находятся *оценки параметров регрессии*:

$$\hat{b}=\frac{\overline{X\cdot Y}-\overline{X}\cdot\overline{Y}}{\overline{X^2}-\overline{X}^2},\ \hat{a}=\overline{Y}-\hat{b}\cdot\overline{X},$$

где $\overline{X}=\dfrac{\sum x}{n}$, $\overline{Y}=\dfrac{\sum y}{n}$ - средние значения факторов X и Y, $\overline{X\cdot Y}=\dfrac{\sum xy}{n}$ - среднее значение $X\cdot Y$, $\overline{X^2}=\dfrac{\sum x^2}{n}$ - среднее значение X^2.

Учитывая, что $\operatorname{cov}(X,Y)=\overline{X\cdot Y}-\overline{X}\cdot\overline{Y}$ и $\mathrm{MSX}=\sigma^2\left(X\right)=\overline{X^2}-\overline{X}^2$ (*the Mean of the Sum of the Squared deviations from the Mean X*), оценка $\hat{b}$ может быть вычислена по формуле

$$\hat{b}=\frac{\operatorname{cov}(X,Y)}{\mathrm{MSX}}=\frac{\operatorname{cov}(X,Y)}{\sigma^2\left(X\right)}.$$

С геометрической точки зрения минимизация суммы квадратов ошибок модели означает выбор единственной прямой (из всех прямых с параметрами $\hat{a}$ и $\hat{b}$), которая ближе всего «прилегает» по ординатам к системе выборочных точек $\left(x_i,y_i\right)$.

Для правильного построения и анализа линейной регрессионной модели необходимо выполнение ряда требований, сформулированных в *теореме Гаусса-Маркова*.

Предположения (*условия Гаусса-Маркова*) *линейной парной регрессионной модели* ($i, j = \overline{1,n}$):

1. Спецификация модели: $Y = a + b \cdot X + \varepsilon$.
2. X – детерминированная (неслучайная) величина, при этом предполагается, что среди значений $\forall i \neq j$: $x_i \neq x_j$, (x_i не все одинаковые).
3.

3a. $M\left(Y\right) = \hat{Y} \Rightarrow M\left(\varepsilon\right) = 0$.

3b. Гомоскедастичность: $M\left(\varepsilon_i^2\right) = D\left(\varepsilon_i\right) = D\left(y_i\right) = \sigma^2$.

Условие независимости дисперсии ошибки от номера наблюдения называется *гомоскедастичностью*; случай, когда условие гомоскедастичности не выполняется, называется *гетероскедастичностью.*

3c. Отсутствие автокорреляции: $\operatorname{cov}\left(\varepsilon_{i-1}, \varepsilon_i\right) = M\left(\varepsilon_{i-1}, \varepsilon_i\right) = \operatorname{cov}\left(y_{i-1}, y_i\right) = 0$;

Отсутствие автокорреляции означает некоррелированность ошибок для разных наблюдений, т.е. результаты наблюдений не коррелированны, так что ковариации равны нулю.

Автокорреляция — статистическая взаимосвязь между последовательностями величин одного ряда, взятыми со сдвигом.

4. Результаты наблюдений и возмущения являются нормально распределенными случайными величинами: $Y \sim N\left(\hat{Y}, \sigma^2\right)$, $\varepsilon \sim N\left(0, \sigma^2\right)$.

Теорема Гаусса-Маркова. В предположениях линейной парной регрессионной модели 1-4 оценки параметров $\hat{a}$ и $\hat{b}$, полученные МНК, имеют наименьшую дисперсию в классе всех линейных несмещённых оценок.

Определив с помощью МНК коэффициенты выборочного уравнения регрессии, можно пояснить их экономический смысл таким образом:

1. можно сказать, что увеличение X на одну единицу (в единицах измерения X) приведёт к увеличению значения Y в среднем на $\hat{b}$ единиц (в единицах измерения Y);
2. постоянная $\hat{a}$ дает прогнозируемое значение Y (в единицах Y), если $X = 0$. Это может иметь или не иметь ясного смысла в зависимости от конкретной ситуации.

Косвенные признаки качества уравнения регрессии: средняя ошибка аппроксимации и средний коэффициент эластичности

Качество уравнения регрессии оценивается с помощью *средней ошибки аппроксимации*

$$A = \frac{1}{n}\sum\left|\frac{Y-\hat{Y}}{Y}\right|\cdot 100\% .$$

Качество уравнения регрессии считается хорошим, если $A < 8-10\%$. Средняя ошибка аппроксимации показывает, на сколько процентов в среднем фактические значения Y отличаются от расчётных $\hat{Y}$.

Средний коэффициент эластичности показывает, на сколько изменился результат Y от своей средней величины $\bar{Y}$ при изменении фактора X на 1% от своего среднего значения $\bar{X}$ и определяется по формуле

$$Э = \frac{\partial\hat{Y}}{\partial x}\cdot\frac{\bar{X}}{\bar{Y}} .$$

Для линейной парной регрессии $\frac{\partial\hat{Y}}{\partial x} = \hat{b}$, поэтому *средний коэффициент эластичности* для линейной парной регрессии имеет вид

$$Э = \hat{b}\cdot\frac{\bar{X}}{\bar{Y}} .$$

Дисперсионный анализ

Проверка адекватности модели регрессии

Качество построенной модели регрессии исследуется методами дисперсионного анализа.

Рассматривается основное уравнение дисперсионного анализа:

$\mathrm{SST} = \mathrm{SSR} + \mathrm{SSE}$,

где

- $\mathrm{SST} = \sum\left(Y-\bar{Y}\right)^2$ – *Total Sum of Squares*, общая сумма квадратов отклонений наблюдений от общего среднего, «скорректированная»;
- $\mathrm{SSR} = \sum\left(\hat{Y}-\bar{Y}\right)^2$ – *Regression Sum of Squares*, сумма квадратов отклонений предсказанных значений от общего среднего, «факторная», «обусловленная регрессией» или «объяснённая регрессией»;
- $\mathrm{SSE} = \sum\left(Y-\hat{Y}\right)^2$ – *Error Sum of Squares*, сумма квадратов отклонений наблюдений от предсказанных значений, «остаточная» или «необъяснённая регрессией».

Каждая сумма квадратов связана с *числом степеней свободы* df (*degree of freedom*). Число степеней свободы показывает, сколько независимых элементов информации (из n чисел y_i) необходимо для образования данной суммы квадратов.

Например, для SST $df_T = n-1$. Действительно, из n разностей $\left(y_i - \overline{Y}\right)^2$ только $n-1$ независима, т.к. для образования SST достаточно $n-1$ значений $y_1, y_2, \ldots, y_{n-1}$. Значение y_n для $\left(y_n - \overline{Y}\right)^2$ можно определить, зная $\overline{Y}$: $y_n = n \cdot \overline{Y} - \sum_{i=1}^{n-1} y_i$.

Для SSR $df_R = m$, где число констант регрессии, независимо определяемых по эмпирическим данным. В случае линейной парной регрессии $df_R = m = 1$, т.к. только оценка $\hat{b}$ определяется исключительно по данным выборки, оценка $\hat{a}$ зависима от $\hat{b}$. В случае парной линейной регрессии, $\hat{Y} = \hat{a} + \hat{b} \cdot X$, $\hat{a} = \overline{Y} - \hat{b} \cdot \overline{X}$. Прогнозируемые значения объясняемой переменной можно выписать, используя лишь оценку $\hat{b}$: $\hat{Y} = \hat{a} + \hat{b} \cdot X = \overline{Y} - \hat{b} \cdot \overline{X} + \hat{b} \cdot X = \hat{b} \cdot \left(X - \overline{X}\right) + \overline{Y}$.

Существует равенство между числом степеней свободы SST, числом степеней свободы SSR и SSE. Аналогично основному уравнению дисперсионного анализа можно записать:

$$df_T = df_R + df_E,$$

где

$df_T = n-1$ – число степеней свободы SST;
$df_R = m$ – число степеней свободы SSR;
$df_E = n-m-1$ – число степеней свободы SSE.

В случае парной линейной регрессии $df_T = n-1$, $df_R = 1$, $df_E = n-2$.

Адекватность линии регрессии зависит от того, какая часть SST обусловлена SSR, а какая SSE. Для сравнения влияния сумм квадратов используется понятие *среднего квадрата отклонений* или *дисперсии на одну степень свободы*:

$\mathrm{MST} = \dfrac{\mathrm{SST}}{df_T} = \dfrac{\sum\left(Y - \overline{Y}\right)^2}{n-1}$ – *the Mean of Total Sum of Squares*, исправленная дисперсия фактора *Y*;

$\mathrm{MSR} = \dfrac{\mathrm{SSR}}{df_R} = \dfrac{\sum\left(\hat{Y} - \overline{Y}\right)^2}{m}$ – *the Mean of Regression Sum of Squares*, дисперсия, объяснённая регрессией;

$\mathrm{MSE} = \dfrac{\mathrm{SSE}}{df_E} = \dfrac{\sum\left(Y - \hat{Y}\right)^2}{n-m-1}$ – *the Mean of Error Sum of Squares*, необъяснённая дисперсия или *квадрат стандартной ошибки регрессии*.

В случае парной линейной регрессии:

$$\mathrm{MST} = \frac{\mathrm{SST}}{n-1} = \frac{\sum\left(Y-\bar{Y}\right)^2}{n-1},$$

$$\mathrm{MSR} = \frac{\mathrm{SSR}}{1} = \frac{\sum\left(\hat{Y}-\bar{Y}\right)^2}{1},$$

$$\mathrm{MSE} = \frac{\mathrm{SSE}}{n-2} = \frac{\sum\left(Y-\hat{Y}\right)^2}{n-2}.$$

В случае линейной регрессии с нормально распределёнными ошибками статистика $F = \frac{\mathrm{MSR}}{\mathrm{MSE}}$ имеет точное распределение Фишера $F\left(m;n-m-1\right)$ для выборок любого объёма.

Статистическая значимость уравнения регрессии в целом проверяется F-критерием Фишера. Для этого требуется сравнить дисперсии на одну степень свободы, соответствующие факторной сумме квадратов MSR (дисперсия, объяснённая регрессией) и остаточной сумме квадратов MSE (необъяснённая дисперсия).

При сравнении дисперсий используются статистические методы проверки гипотез. В качестве основной (нулевой) гипотезы выдвигается гипотеза $\mathbf{H}_0$ о равенстве факторной и остаточной дисперсий:

$\mathbf{H}_0$: MSR = MSE (уравнение регрессии статистически незначимо).

Наряду с основной (проверяемой) гипотезой выдвигается альтернативная (конкурирующая) гипотеза $\mathbf{H}_1$ о неравенстве факторной и остаточной дисперсий:

$\mathbf{H}_1$: MSR $\neq$ MSE (уравнение регрессии статистически значимо).

В случае если основная гипотеза окажется неверной, принимается альтернативная гипотеза.

Вычисляется наблюдаемое значение критерия Фишера:

$$F = \frac{\mathrm{MSR}}{\mathrm{MSE}}$$

Наблюдаемое значение критерия Фишера F сравнивается с табличным значением $F_{\text{табл}}\left(\alpha;k_1;k_2\right)$ (Таблица 1 Приложения 1) при заданном уровне значимости α и числе степеней свободы $k_1 = m$ – для большей дисперсии и $k_2 = n-m-1$ для меньшей дисперсии. В случае парной линейной регрессии $k_1 = 1$, $k_2 = n-2$, а наблюдаемое значение критерия Фишера можно вычислить через коэффициент детерминации:

$$F = \frac{\mathrm{MSR}}{\mathrm{MSE}} = \frac{\mathrm{SSR}}{\mathrm{SSE}} \cdot \frac{n-m-1}{m} = \frac{\dfrac{\mathrm{SSR}}{\mathrm{SST}}}{1-\dfrac{\mathrm{SSR}}{\mathrm{SST}}} \cdot \frac{n-m-1}{m} = \frac{R^2}{1-R^2} \cdot \frac{n-m-1}{m}.$$

$F = \frac{R^2}{1-R^2} \cdot (n-2)$.

Если $F > F_{\text{табл}}(\alpha; k_1; k_2)$, гипотеза $\mathbf{H}_0$ отклоняется, и с вероятностью $(1-\alpha) \cdot 100\%$ признаётся статистическая значимость уравнения регрессии. В противном случае, когда $F < F_{\text{табл}}(\alpha; k_1; k_2)$, нет оснований отвергать гипотезу $\mathbf{H}_0$, откуда делается вывод о статистической не значимости уравнения регрессии в целом.

Оценка значимости уравнения регрессии обычно дается в виде таблицы дисперсионного анализа:

Таблица дисперсионного анализа

Источник	**Число степеней свободы *df***	**Сумма квадратов**	**Среднеквадратичное значение**	***F***	
				наблюдаемое	**табличное (критическое)**
Регрессия	m	SSR	MSR = SSR/m	F = MSR/MSE	$F_{\text{табл}}(\alpha; m; n-m-1)$
Ошибка	$n - m - 1$	SSE	MSE = SSE/($n - m - 1$)		
Всего	$n - 1$	SST	MST = SST/($n - 1$)		

Корреляционный анализ

Оценка тесноты связи с помощью выборочного коэффициента корреляции

Уравнение регрессии всегда дополняется показателем тесноты связи. При использовании линейной регрессии в качестве такого показателя выступает *линейный коэффициент корреляции* r_{XY}. Существует несколько видов формулы линейного коэффициента корреляции, основные из них:

$$r_{XY} = \frac{\operatorname{cov}(X,Y)}{\sigma(X)\cdot\sigma(Y)} = \frac{\operatorname{cov}(X,Y)}{\sqrt{\mathrm{MSX}\cdot\mathrm{MSY}}}, \qquad r_{XY} = \hat{b}\cdot\frac{\sigma(X)}{\sigma(Y)} = \hat{b}\cdot\sqrt{\frac{\mathrm{MSX}}{\mathrm{MSY}}}.$$

Значения коэффициента корреляции по модулю меньше единицы: $|r_{XY}| \le 1$.

Корреляционная связь между переменными называется *прямой*, если $r_{XY} > 0$, и *обратной*, если $r_{XY} < 0$.

Если значение коэффициента корреляции r_{XY} равно ± 1, связь представлена линейной функциональной зависимостью. При этом все наблюдаемые значения располагаются на линии регрессии.

При $r_{XY} = 0$ корреляционная связь между признаками в линейной форме отсутствует и линия регрессии параллельна оси *OX*.

Характеристику тесноты связи факторов по значению r_{XY} определяет шкала *Чеддока*:

Показатель тесноты связи r_{XY}	0,1 – 0,3	0,3 – 0,5	0,5 – 0,7	0,7 – 0,9	0,9 – 0,99
Характеристика силы связи	Слабая	Умеренная	Заметная	Высокая	Весьма высокая

Оценка тесноты связи с помощью показателя детерминации

Из основного уравнения дисперсионного анализа значение суммы квадратов, обусловленной регрессией равно $\mathrm{SSR} = \mathrm{SST} - \mathrm{SSE}$. Рассмотрим долю этой суммы в общей сумме квадратов отклонений фактора *Y*.

Коэффициентом детерминации, или долей объяснённой дисперсии (объяснённой суммы квадратов отклонений) в общей дисперсии фактора *Y*, называется

$$R^2 = \frac{\mathrm{SSR}}{\mathrm{SST}} = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}}.$$

В случае парной линейной регрессии, $\hat{Y} = \hat{a} + \hat{b} \cdot X$, $\hat{a} = \bar{Y} - \hat{b} \cdot \bar{X}$. Прогнозируемые значения объясняемой переменной можно выписать, используя лишь оценку $\hat{b}$:

$$\hat{Y} = \hat{a} + \hat{b} \cdot X = \bar{Y} - \hat{b} \cdot \bar{X} + \hat{b} \cdot X = \hat{b} \cdot \left(X - \bar{X} \right) + \bar{Y}.$$

Тогда объяснённая регрессией сумма квадратов для линейной парной регрессии явно выражается через оценку $\hat{b}$ и дисперсию экзогенного фактора *X*:

$$\mathrm{SSR} = \sum \left(\hat{Y} - \bar{Y} \right)^2 = \sum \left(\hat{b} \cdot \left(X - \bar{X} \right) + \bar{Y} - \bar{Y} \right)^2 = \hat{b}^2 \cdot \sum \left(X - \bar{X} \right)^2 = \hat{b}^2 \cdot n \cdot \mathrm{MSX}.$$

Общая сумма квадратов связана с дисперсией эндогенного фактора *Y*:

$$\mathrm{SST} = \sum \left(Y - \bar{Y} \right)^2 = n \cdot \mathrm{MSY}.$$

Отсюда:

$$R^2 = \frac{\mathrm{SSR}}{\mathrm{SST}} = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}} \hat{b}^2 \cdot \frac{\mathrm{MSX}}{\mathrm{MSY}} = r_{xy}^2$$

Итак, для линейной парной регрессии коэффициент детерминации равен квадрату коэффициента линейной корреляции.

Коэффициент детерминации для модели с $\hat{a} \neq 0$ принимает значения от 0 до 1. Чем ближе значение коэффициента к 1, тем сильнее зависимость. При оценке регрессионных моделей это интерпретируется как соответствие модели данным. Для приемлемых моделей предполагается, что коэффициент детерминации должен быть хотя бы не меньше 50%. Модели с коэффициентом детерминации выше 80% можно признать достаточно хорошими (коэффициент корреляции превышает 90%). Значение коэффициента детерминации 1 означает функциональную зависимость между переменными.

Для учёта соотношения количества наблюдений и количества оцениваемых параметров при оценке качества модели (чтобы была возможность сравнивать модели с разным числом факторов так, чтобы число регрессоров (факторов) не влияло на статистику) применяется *скорректированный коэффициент детерминации*, в котором используются несмещённые оценки дисперсий:

$$\tilde{R}^2 = 1 - \frac{\mathrm{MSE}}{\mathrm{MST}}.$$

Поскольку $\mathrm{SSE} = (n-m-1)\cdot \mathrm{MSE}$, $\mathrm{SST} = (n-1)\cdot \mathrm{MST}$, $R^2 = 1 - \frac{\mathrm{SSE}}{\mathrm{SST}} = 1 - \frac{\mathrm{MSE}}{\mathrm{MST}} \cdot \frac{n-m-1}{n-1}$, откуда $\frac{\mathrm{MSE}}{\mathrm{MST}} = \left(1-R^2\right)\cdot \frac{n-1}{n-m-1}$, формула для скорректированного коэффициента детерминации примет вид:

$$\tilde{R}^2 = 1 - \left(1-R^2\right)\cdot \frac{n-1}{n-m-1} \le R^2.$$

Скорректированный коэффициент детерминации применяется для решения двух типов задач:

– оценка тесноты связи между объясняемой и объясняющей переменной. Необходимо обратить внимание на близость к нескорректированному коэффициенту детерминации. Модель считается качественной, если показатели велики и несильно отличаются друг от друга.
– сравнение моделей с различным числом экзогенных факторов. При прочих равных условиях, предпочтение отдается той модели, у которой скорректированный коэффициент детерминации больше.

Следует отметить, что скорректированный коэффициент детерминации нельзя использовать в формулах, где применяется обычный коэффициент детерминации, поскольку скорректированный коэффициент детерминации нельзя интерпретировать как долю вариации объясняемой переменной, обусловленную вариацией факторов, включенных в модель.

Проверка статистической значимости коэффициентов регрессии и корреляции

Эмпирическое уравнение регрессии определяется на основе конечного числа статистических данных. Очевидно, что коэффициенты $\hat{a}$ и $\hat{b}$ эмпирического уравнения регрессии являются случайными величинами, изменяющимися от выборки к выборке. Возникает необходимость определения их статистической значимости.

Для сравнения оценок коэффициентов регрессии с их предполагаемыми теоретическими значениями, используются статистические методы проверки гипотез.

Для каждого из параметров определяется стандартные ошибки $m_b = \sqrt{\frac{\mathrm{MSE}}{\mathrm{MSX}} \cdot \frac{1}{n}}$, $m_a = m_b \cdot \sqrt{\overline{x^2}}$.

Для проверки гипотезы $\mathbf{H}_0$: $\hat{b} = b$ против двусторонней альтернативной гипотезы $\mathbf{H}_1$: $\hat{b} \neq b$ используется *t*-статистика $t_b = \dfrac{\hat{b} - b}{m_b}$, которая имеет распределение Стьюдента с $n-2$ степенями свободы. На доверительном уровне α нулевая гипотеза отвергается при $\left|t_b\right| > t_{\text{табл}}\left(\alpha/2; k = n-2\right)$. Табличные значения распределения Стьюдента для двусторонней критической области приводятся в Таблица 2 Приложения 2.

Наиболее просто *t*-статистика выглядит при гипотезе $\mathbf{H}_0$: $b = 0$, против двусторонней конкурирующей $\mathbf{H}_1$: $b \neq 0$, а именно, $t_b = \dfrac{\hat{b}}{m_b}$. Гипотеза в такой постановке обычно называется *гипотезой о статистической значимости коэффициента регрессии.* Значение $\left|t_b\right| > t_{\text{табл}}\left(\alpha/2; k = n-2\right)$ позволяет сделать вывод об отличии от нуля (на соответствующем уровне значимости α) коэффициента регрессии b (статистической значимости коэффициента регрессии b) и, следовательно, о наличии влияния (связи) *X* на *Y*.

Из неравенства $P\left(\left|t_b\right| < t_{\text{табл}}\left(\alpha/2; n-2\right)\right) = 1-\alpha$, получается двусторонний $100\cdot(1-\alpha)\%$-ный доверительный интервал для параметра регрессии b: $b \in \left[\hat{b} - \Delta_b; \hat{b} + \Delta_b\right]$, где $\Delta_b = t_{\text{табл}}\left(\alpha/2; n-2\right)\cdot m_b$ - предельная ошибка.

По аналогичной схеме на основе *t*-статистики $t_a = \dfrac{\hat{a}}{m_a}$ проверяется гипотеза о статистической значимости коэффициента a. Из неравенства $P\left(\left|t_a\right| < t_{\text{табл}}\left(\alpha/2; n-2\right)\right) = 1-\alpha$, получается двусторонний $100\cdot(1-\alpha)\%$-ный доверительный интервал для параметра регрессии a: $a \in \left[\hat{a} - \Delta_a; \hat{a} + \Delta_a\right]$, где $\Delta_a = t_{\text{табл}}\left(\alpha/2; n-2\right)\cdot m_a$ - предельная ошибка.

Используя *t*-статистику $t_R = \dfrac{\sqrt{R^2}}{m_R}$, где $m_R = \sqrt{\dfrac{\text{MSE}}{\text{SST}}} = \sqrt{\dfrac{\text{SSE}}{\text{SST}}\cdot\dfrac{1}{n-m-1}} = \sqrt{\dfrac{1-R^2}{n-m-1}}$ - стандартная ошибка, можно оценить и значимость коэффициента детерминации. Для парной линейной регрессии стандартная ошибка $m_r = \sqrt{\dfrac{1-R^2}{n-2}}$, откуда следует соотношение *t*- и *F*-критериев для парной линейной регрессии:

$$t_r^2 = \frac{R^2}{1-R^2}\cdot(n-2) = F\ .$$

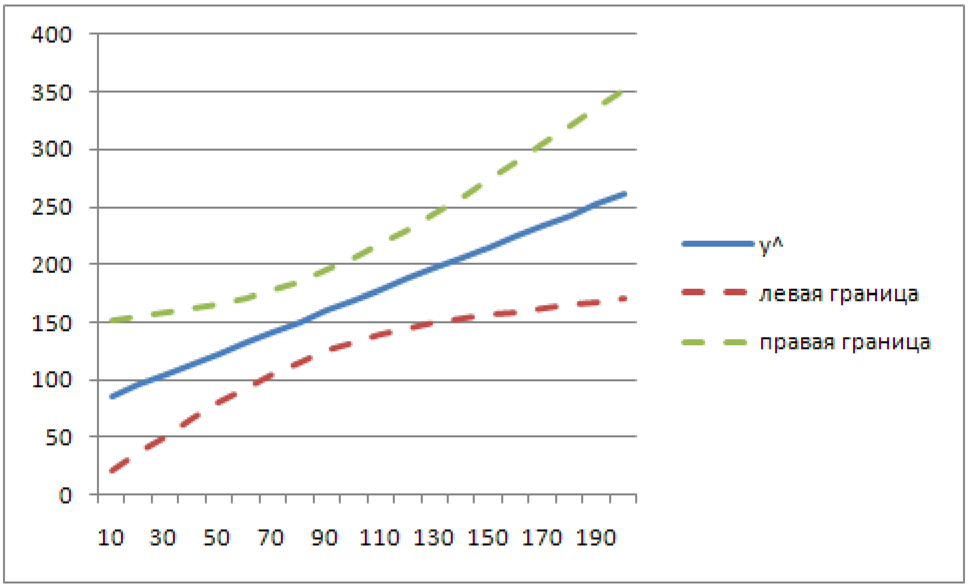
Из неравенства $P\left(\left|t_r\right| < t_{\text{табл}}\left(\alpha/2; n-2\right)\right) = 1-\alpha$, получается двусторонний $100\cdot(1-\alpha)\%$-ный доверительный интервал для коэффициента линейной корреляции r_{XY}: $\left[r_{XY}-\Delta_r; r_{XY}+\Delta_r\right]$, где $\Delta_r = t_{\text{табл}}\left(\alpha/2; n-2\right)\cdot m_r$ - предельная ошибка.

Если в границы доверительных интервалов попадает ноль, т.е. нижняя граница отрицательна, а верхняя положительна, то оцениваемый параметр признается статистически незначимым.

Прогноз по уравнению регрессии

В прогнозных расчётах по уравнению регрессии определяется предсказываемое значение y^* как *точечный прогноз* $\hat{Y}(x^*)$, т.е. путём подстановки в уравнение регрессии $\hat{Y} = \hat{a} + \hat{b}\cdot X$ соответствующего значения $X = x^*$. Вероятность реализации точечного прогноза практически равна нулю. Поэтому рассчитывается доверительный интервал прогноза с большой надёжностью $100\cdot(1-\alpha)\%$: $y^* \in \left[\hat{Y}(x^*) - \Delta_{\hat{Y}(x^*)}; \hat{Y}(x^*) + \Delta_{\hat{Y}(x^*)}\right]$, где $m_{\hat{Y}(x^*)} = \sqrt{\dfrac{\text{MSE}}{n}\cdot\left(n+1+\dfrac{\left(x^*-\bar{X}\right)^2}{\text{MSX}}\right)}$ - стандартная ошибка прогноза, $\Delta_{\hat{Y}(x^*)} = t_{\text{табл}}\left(\alpha/2; n-2\right)\cdot m_{\hat{Y}(x^*)}$ - предельная ошибка прогноза. Доверительный интервал прогноза определяет границы, за пределами которых могут оказаться не более $100\cdot\alpha\%$ точек наблюдений при $X = x^*$.

На промежутке изменения факторного признака строятся гиперболы – левая и правая границы интервала прогнозных значений y^*:



Визуализация и расчёт показателей статистического анализа с помощью MS Excel

Построение поля корреляции и линии регрессии

При построении поля корреляции и линии регрессии используется Мастер диаграмм MS Excel.

Построение поля корреляции:

1) выделить диапазон X, Y вместе с названиями
2) выбрать Точечную диаграмму;
3) указать место размещения диаграммы.

Добавление линии регрессии (линии тренда):

1) щелчком левой кнопки мыши выделить поле корреляции;
2) щелчком правой кнопки мыши войти в контекстное меню и выбрать Добавить линию тренда;
3) в поле Линия тренда на вкладке Тип выбрать вид линии тренда и задать соответствующие параметры;
4) отобразить на диаграмме уравнение регрессии и коэффициент детерминации R^2, установив соответствующие флажки на закладке Параметры.

Определение параметров выборочного уравнения линейной регрессии с использованием встроенной статистической функции ЛИНЕЙН

Порядок выполнения:

1) ввести исходные данные;
2) выделить область пустых ячеек 5 × 2 (5 строк, 2 столбца) для вывода результатов регрессионной статистики, которая будет выводиться в порядке, указанном в следующей таблице:

Значение коэффициента $\hat{b}$	Значение коэффициента $\hat{a}$
Стандартная ошибка m_b	Стандартная ошибка m_a
Коэффициент детерминации R^2	Стандартная ошибка регрессии $\sqrt{\text{MSE}}$
Наблюдаемое значение F-статистики	Число степеней свободы SSE $df_E = n - m - 1$
Регрессионная сумма квадратов SSR	Остаточная сумма квадратов SSE

3) вызвать Мастер функций (Shift+*F3*);
4) в окне Категория Мастера функций выбрать Статистические, в окне Функция – ЛИНЕЙН;
5) заполнить аргументы функции:
- *Известные_значения_у* – диапазон, содержащий данные результативного признака;
- *Известные_значения_х* – диапазон, содержащий данные факторного признака;
- *Константа* – логическое значение, которое указывает на наличие или отсутствие свободного члена в уравнении;

  если *Константа* = 0, то свободный член рассчитывается обычным образом;
  если *Константа* = 1, то свободный член равен 0.

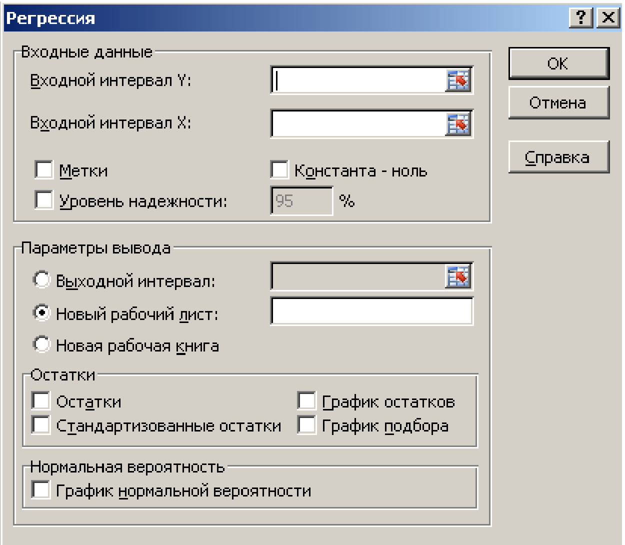
- *Статистика* – логическое значение, которое указывает, выводить дополнительную информацию по регрессионному анализу (*Статистика* =1) или нет (*Статистика* =0).

6) чтобы раскрыть таблицу 5 × 2, нажать на клавишу <*F2*>, а затем – на комбинацию клавиш <Ctrl>+<Shift>+<Enter>.

Использование инструмента анализа данных Регрессия для вычисления регрессионной статистики и дисперсионного анализа

Порядок действий:

1) настроить параметры MS Excel, подключив в надстройках Пакет анализа;
2) на вкладке Данные выбрать Анализ данных/Регрессия;
3) заполнить диалоговое окно ввода данных:



Диалоговое окно Регрессия

- Входные данные

Входной интервал Y. Ввести ссылку на диапазон зависимых данных. Диапазон должен состоять из одного столбца.

Входной интервал X. Ввести ссылку на диапазон независимых данных. Эти данные будут расположены слева направо в порядке возрастания. Максимальное число независимых переменных равно 16.

Метки. Установить флажок, если первая строка или первый столбец входного диапазона содержит заголовки. Снять этот флажок, если заголовки отсутствуют. В этом случае подходящие заголовки для данных выходной таблицы будут созданы автоматически.

Уровень надежности. Установить флажок, чтобы включить в выходную таблицу итогов дополнительный уровень. В соответствующее поле введите уровень надежности $100\cdot(1-\alpha)\%$, который следует применить, дополнительно к уровню 95%, применяемому по умолчанию.

Константа – ноль. Установить флажок, чтобы линия регрессии прошла через начало координат.

- Параметры вывода

Выходной интервал. Ввести ссылку на левую верхнюю ячейку выходного диапазона. Отвести как минимум семь столбцов для выходной таблицы итогов, которая будет включать в себя: результаты дисперсионного анализа, коэффициенты, стандартную погрешность вычисления Y, среднеквадратичные отклонения, число наблюдений, стандартные погрешности для коэффициентов.

Новый рабочий лист. Установить переключатель в это положение, чтобы открыть новый лист в книге и вставить результаты анализа, начиная с ячейки A1. При необходимости введите имя для нового листа в поле, расположенном напротив соответствующего положения переключателя.

Новая рабочая книга. Установить переключатель в это положение для создания новой книги, в которой результаты будут добавлены в новый лист.

- Остатки

Остатки. Установить флажок для включения остатков в выходную таблицу.

Стандартизированные остатки. Установить флажок для включения стандартизированных остатков в выходную таблицу.

График остатков. Установить флажок для построения графика остатков для каждой независимой переменной.

График подбора. Установить флажок для построения графика зависимости предсказанных значений от наблюдаемых.

- Нормальная вероятность

График нормальной вероятности. Установить флажок для построения графика нормальной вероятности.

По окончанию расчета на рабочий лист выводится три группы результатов.

Первая группа результатов, – Регрессионная статистика, включает в свой состав:

- *Множественный R* – коэффициент множественной корреляции;
- *R- квадрат* – множественный коэффициент детерминации;
- *Нормированный R- квадрат* – скорректированный коэффициент детерминации;
- *Стандартная ошибка* – стандартная ошибка регрессии;
- *Наблюдения* – количество наблюдений.

Вторая группа результатов, – Дисперсионный анализ, включает в свой состав:

- *df* – степени свободы (*degree of freedom*);
- *SS* – сумма квадратов отклонений (*Sum of squares*);

- *MS* – средний квадрат отклонения (*Mean square*);
- *F* – отношение дисперсий;
- *Значимость F* – критическое значение квантиля распределения Фишера, на котором отвергается нулевая гипотеза отсутствия влияния фактора.

Построчно в таблице выводятся показатели, характеризующие изменчивости: присущую модели и случайную.

Третья группа результатов включает в свой состав значения коэффициентов регрессии, а также статистики, на основании которых проверяется значимость влияния фактора для каждого коэффициента, включенного в модель:

- *Коэффициенты* – значения коэффициентов;
- *Стандартная ошибка* – стандартная ошибка коэффициентов;
- *t-статистика* – значение статистики критерия;
- *P- значение* – уровень значимости отклонения гипотезы равенства коэффициентов нулю;
- *Нижние 95%* – нижняя граница доверительного интервала, в котором находится значение коэффициента генеральной совокупности;
- *Верхние 95%* – верхняя граница доверительного интервала, в котором находится значение коэффициента генеральной совокупности.

При необходимости есть возможность вывести таблицу стандартных и простых остатков, где для каждого значения ряда выводится предсказанное значение, с которым сопоставляется остаток, представляющий разность между прогнозным и реальным значением ряда.

Кроме вывода табличной информации, есть возможность просмотреть графики остатков, что позволяет визуально проконтролировать качество подбора модели и отсутствие закономерности в остатках.

Прогнозирование

С помощью Мастера Диаграмм: в Параметрах линии тренда ввести Прогноз на заданное количество периодов

С помощью Мастера функций: использование функции Статистические/ПРЕДСКАЗ и Статистические/ТЕНДЕНЦИЯ для прогнозирования линейного тренда.

ПРИЛОЖЕНИЕ 1

Таблица 1

Вариант	Номер начального наблюдения из Таблицы 2	Номер конечного наблюдения из Таблицы 2	Номер признаков из Таблицы 2	Вариант	Номер начального наблюдения из Таблицы 2	Номер конечного наблюдения из Таблицы 2	Номер признаков из Таблицы 2
1	***2***	***3***	***4***	***5***	***6***	***7***	***8***
1	1	50	1,2	**51**	26	75	1,3
2	1	50	3,4	**52**	26	75	4,5
3	2	51	1,3	**53**	27	76	1,4
4	2	51	4,5	**54**	27	76	2,5
5	3	52	1,4	**55**	28	77	1,5
6	3	52	2,5	**56**	28	77	2,3
7	4	53	1,5	**57**	29	78	1,2
8	4	53	2,3	**58**	29	78	3,4
9	5	54	1,2	**59**	30	79	1,3
10	5	54	3,4	**60**	30	79	4,5
11	6	55	1,3	**61**	31	80	1,4
12	6	55	4,5	**62**	31	80	2,5
13	7	56	1,4	**63**	32	81	1,5
14	7	56	2,5	**64**	32	81	2,3
15	8	57	1,5	**65**	33	82	1,2
16	8	57	2,3	**66**	33	82	3,4
17	9	58	1,2	**67**	34	83	1,3
18	9	58	3,4	**68**	34	83	4,5
19	10	59	1,3	**69**	35	84	1,4
20	10	59	4,5	**70**	35	84	2,5
21	11	60	1,4	**71**	36	85	1,5
22	11	60	2,5	**72**	36	85	2,3
23	12	61	1,5	**73**	37	86	1,2
24	12	61	2,3	**74**	37	86	3,4
25	13	62	1,2	**75**	38	87	1,3
26	13	62	3,4	**76**	38	87	4,5
27	14	63	1,3	**77**	39	88	1,4
28	14	63	4,5	**78**	39	88	2,5
29	15	64	1,4	**79**	40	89	1,5
30	15	64	2,5	**80**	40	89	2,3
31	16	65	1,5	**81**	41	90	1,2
32	16	65	2,3	**82**	41	90	3,4
33	17	66	1,2	**83**	42	91	1,3
34	17	66	3,4	**84**	42	91	4,5
35	18	67	1,3	**85**	43	92	1,4
36	18	67	4,5	**86**	43	92	2,5

Продолжение Таблицы 1

1	***2***	***3***	***4***	***5***	***6***	***7***	***8***
37	19	68	1,4	**87**	44	93	1,5
38	19	68	2,5	**88**	44	93	2,3
39	20	69	1,5	**89**	45	94	1,2
40	20	69	2,3	**90**	45	94	3,4
41	21	70	1,2	**91**	46	95	1,3
42	21	70	3,4	**92**	46	95	4,5
43	22	71	1,3	**93**	47	96	1,4
44	22	71	4,5	**94**	47	96	2,5
45	23	72	1,4	**95**	48	97	1,5
46	23	72	2,5	**96**	48	97	2,3
47	24	73	1,5	**97**	49	98	1,2
48	24	73	2,3	**98**	49	98	3,4
49	25	74	1,2	**99**	50	99	1,3
50	25	74	3,4	**0**	50	99	4,5

ПРИЛОЖЕНИЕ 1

Таблица 2

№ наблюдений	Собственные оборотные средства, млн.руб.	Балансовая прибыль, млн.руб.	Дебиторская задолженность, млн.руб.	Дивиденды, начисленные по результатам деятельности, млн.руб.	Курсовая цена акции, руб.
A	***1***	***2***	***3***	***4***	***5***
1	1011	107	37	20,33	92
2	799	102	64	20,04	83
3	995	107	71	19,87	95
4	1243	122	26	20,48	124
5	1507	108	51	20,13	96
6	947	108	41	20,26	106
7	1015	97	78	19,89	70
8	1169	109	43	19,92	97
9	1051	101	68	19,78	76
10	1372	116	34	20,23	112
11	1463	113	49	20,46	113
12	684	112	40	20,07	109
13	1251	106	56	20,23	91
14	1376	111	45	20,26	95
15	1193	113	44	20,28	115
16	1386	122	40	20,52	114
17	1631	118	47	20,28	133
18	1735	119	47	19,97	116
19	1181	102	49	19,97	85
20	922	100	65	19,57	91
21	1281	103	54	19,94	82
22	1333	113	59	20,29	105
23	1632	124	36	20,83	124
24	635	95	70	19,59	70
25	949	102	64	19,76	84
26	788	112	48	20,19	106
27	1728	124	30	20,66	128
28	1773	116	58	19,95	105
29	1679	118	48	20,61	121
30	1085	100	69	20,03	79
31	1214	99	58	19,78	82
32	1422	107	49	20,22	80
33	523	87	76	19,78	37
34	1025	109	59	20,09	101

Продолжение Таблицы 2

А	*1*	*2*	*3*	*4*	*5*
35	1083	106	74	20,13	98
36	1466	113	54	20,56	98
37	1642	123	36	20,51	134
38	387	82	75	19,71	39
39	704	104	51	20,1	88
40	1177	112	35	20,32	108
41	1792	116	47	20,37	112
42	2072	106	33	20,03	80
43	1178	120	28	20,65	120
44	1304	105	58	20,19	88
45	1308	114	32	20,24	104
46	1416	107	58	20,27	94
47	1185	115	44	20,69	107
48	1220	96	68	19,85	82
49	1311	104	64	19,87	84
50	1288	108	25	20,2	101
51	918	102	54	20,33	98
52	809	102	70	20,2	89
53	1188	120	19	20,46	118
54	1394	106	28	20,17	90
55	1435	114	54	20,62	123
56	1514	112	48	19,79	107
57	1577	112	44	20,34	97
58	1579	122	39	20,51	126
59	1210	122	26	20,04	147
60	1448	108	58	20,39	88
61	1468	114	28	20,27	111
62	1661	113	47	20,06	121
63	989	108	58	20,39	104
64	1007	102	62	19,94	63
65	1030	112	62	19,95	99
66	1099	113	42	20,23	114
67	1197	110	67	20,49	99
68	1386	107	72	20,61	94
69	1498	117	45	20,56	124
70	1672	120	35	20,42	117
71	484	93	69	19,73	64
72	1060	89	62	19,42	52
73	1612	118	36	20,17	114
74	1120	103	42	19,87	78
75	947	98	52	20,26	85

Продолжение Таблицы 2

A	*1*	*2*	*3*	*4*	*5*
76	1102	95	56	20,04	57
77	1302	106	66	20,34	98
78	1477	123	32	20,63	119
79	820	110	68	20,32	94
80	1231	104	47	20,06	94
81	1311	103	59	20,04	83
82	1843	122	29	20,62	118
83	1215	114	36	20,53	116
84	1284	112	57	20,18	96
85	1336	115	54	20,4	117
86	1412	109	60	20,26	93
87	1447	108	45	19,79	81
88	1593	114	54	20,33	103
89	1663	107	49	20,24	86
90	1114	98	81	19,83	79
91	863	104	61	19,97	92
92	932	107	49	20,1	85
93	978	105	68	20,01	89
94	1621	123	53	20,21	121
95	1199	119	39	20,4	125
96	999	95	49	19,66	69
97	935	93	76	19,37	61
98	1494	120	48	20,25	116
99	817	98	72	19,82	82

ПРИЛОЖЕНИЕ 2. СТАТИСТИКО-МАТЕМАТИЧЕСКИЕ ТАБЛИЦЫ

1. Таблица значений *F – критерия Фишера* при уровне значимости $\alpha = 0,05$

k_2 \ k_1	1	2	3	4	5	6	8	12	24	∞
1	161,45	199,50	215,72	224,57	230,17	233,97	238,89	243,91	249,04	254,32
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3.02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	2,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,64	1,28
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,85	1,63	1,26
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,80	1,57	1,14
300	3,87	3,03	2,64	2,41	2,25	2,13	1,97	1,79	1,55	1,10
400	3,86	3,02	2,63	2,40	2,24	2,12	1,96	1,78	1,54	1,07
500	3,86	3,01	2,62	2,39	2,23	2,11	1,96	1,77	1,54	1,06
1000	**3,85**	**3,00**	**2,61**	**2,38**	**2,22**	**2,10**	**1,95**	**1,76**	**1,53**	1,03

2. Критические значения *t – критерия Стьюдента* при уровне значимости 0,10, 0,05, 0,01 (двухсторонний)

Число степеней свободы	Уровень значимости α =0,10	Уровень значимости α =0,05	Уровень значимости α =0,01
1	6,3138	12,706	63,657
2	2,9200	4,3027	9,9248
3	2,3534	3,1825	5,8409
4	2,1318	2,7764	4,6041
5	2,0150	2,5706	4,0321
6	1,9432	2,4469	3,7074
7	1,8946	2,3646	3,4995
8	1,8595	2,3060	3,3554
9	1,8331	2,2622	3,2498
10	1,8125	2,2281	3,1693
11	1,7959	2,2010	3,1058
12	1,7823	2,1788	3,0545
13	1,7709	2,1604	3,0123
14	1,7613	2,1448	2,9768
15	1,7530	2,1315	2,9467
16	1,7459	2,1199	2,9208
17	1,7396	2,1098	2,8982
18	1,7341	2,1009	2,8784
19	1,7291	2,0930	2,8609
20	1,7247	2,0860	2,8453
21	1,7207	2,0796	2,8314
22	1,7171	2,0739	2,8188
23	1,7139	2,0687	2,8073
24	1,7109	2,0639	2,7969
25	1,7081	2,0595	2,7874
26	1,7056	2,0555	2,7787
27	1,7033	2,0518	2,7707
28	1,7011	2,0484	2,7633
29	1,6991	2,0452	2,7564
30	1,6973	2,0423	2,7500
40	1,6839	2,0211	2,7045
60	1,6707	2,0003	2,6603
120	1,6577	1,9799	2,6174
∞	1,6449	1,9600	2,5758