

В.А. Балаш, О.С. Балаш, А.И. Землянухин

Эконометрика

Учебное пособие

Рекомендовано
Учебно-методическим объединением
по образованию в области статистики
в качестве учебного пособия
для студентов высших учебных заведений,
обучающихся по экономическим специальностям

Саратов
2005

УДК 311(075.8)

ББК 60.6я73

П69

АВТОРЫ:

В.А. Балаш, О.С. Балаш, А.И. Землянухин

РЕЦЕНЗЕНТЫ:

кафедра статистики Мордовского государственного университета (заведующий
кафедрой доктор экономических наук,
профессор **Ю.В. Сажин**);

Эконометрика: Учеб. пособие – Саратов: 2005. – с.

П69

ISBN 5-279-.....

Учебное пособие подготовлено по дисциплине “Эконометрика”. Оно содержит краткий обзор основных понятий эконометрики, по всем главам представлены решения типовых примеров в программе Excel. Учебное пособие включает задачи по изучаемой дисциплине. В приложениях даны математико-статистические таблицы.

Учебное пособие предназначено для студентов, аспирантов и преподавателей экономических вузов. Он может быть полезен широкому кругу практиков-аналитиков.

0702000000 – 012

П – 2004

010(01) – 2005

ISBN5-279-.....

УДК 311(075.8)

ББК 60.6я73

© Коллектив авторов, 2005

Оглавление

Введение	4
1. Парная линейная регрессия.....	7
2. Множественное линейное уравнение регрессии	23
3. Эконометрические модели с переменной структурой. Фиктивные переменные.....	32
4. Тест Чоу	36
5. Сравнение «длинной» и «короткой» регрессии	39
6. Гетероскедастичность	43
6.1. Тесты на гетероскедастичность	45
6.1.1. Тест Голдфелда-Квандта (Goldfeld-Quandt).....	45
6.1.2. Тест Бреуша-Пагана(Breus-Pagan).....	48
6.1.3. Тест Вайта (White).....	50
6.2. Коррекция на гетероскедастичность	52
Задачи для самостоятельного решения	59
Приложения.....	64

Введение

Введение

Регрессионное управление имеет вид:

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \dots + \beta_m x_m + \varepsilon_i, i = 1, 2, \dots, n,$$

где индекс i обозначает номер наблюдения, y_i – эндогенные или зависимые (объясняемые) переменные, x_1, \dots, x_m – экзогенные или независимые (объясняющие) переменные, ε_i – регрессионные остатки.

В простейшем случае парной линейной регрессии исследуется модель вида:

$$y_i = \alpha + \beta x_i + \varepsilon_i.$$

Решение задачи регрессионного анализа состоит из оценки коэффициентов α и β по их фактическим значениям (эти оценки обозначают $\hat{\alpha}, \hat{\beta}$), вычисление ожидаемого значения y по формуле $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$

(при этом $y_i - \hat{y}_i = \varepsilon_i$) и статистической проверки взаимосвязи между зависимыми и независимыми переменными.

В настоящее время уровень развития информационных технологий позволяет существенно упростить процесс эконометрического моделирования с использованием специализированных программных продуктов, таких как EViews, RATS, STATA, SPSS и других. Эти программы сочетают в себе удобство графического интерфейса и гибкости в выборе задач, основанную на использовании командного языка. Однако для их эксплуатации необходим достаточно высокий уровень общей компьютерной грамотности. Поэтому для большинства «средних пользователей» оптимальным является использование табличного процессора MS Excel, интегрированного в пакете MS Office, начиная с Excel 7.0 for Windows 95.

Табличный процессор MS Excel включает в себя программную надстройку «пакет анализа» и библиотеку из 78 статистических функций. Такой набор инструментов, как правила, вполне достаточен для проведения всестороннего статистического анализа информации.

Данное учебное пособие призвано помочь студенту, аспиранту или офисному служащему в освоении основ эконометрики с использованием MS Excel.

Отношение к эконометрике в нашей стране имеет характерный диапазон от полного неприятия, при отсутствии информации о предмете до неоправданного энтузиазма. Приведем две типичные цитаты. Первая датируется 1978 годом и взята из учебника А.Г. Гранберга «Математические модели социалистической экономики»: «Эконометрика

расширила арсенал приемов апологетики капиталистического строя. На основе математических моделей оптимизации, равновесия, теории игр модернизируются старые и создаются новые экономические теории. Буржуазные экономисты используют авторитет математики, наивную веру многих людей в абсолютную истинность математических формул для того, чтобы внушить доверие к своим теоретическим выводам, а возможности гармоничного и эффективного развития капиталистической экономики при взаимной заинтересованности всех социальных слоев общества».

Вторая цитат открывает учебник по эконометрике 2003 под реакцией И.И. Елисеевой: «эконометрика является одновременно нашим телескопом и нашим микроскопом для изучения окружающего экономического мира».

Сегодня эконометрика относится к числу базовых экономических дисциплин, и авторы будут считать свою цель достигнутой. Если это учебное пособие поможет в ее освоении и использовании.

1.1. Модели.

Эконометрика – это наука, связанная с эмпирическим выводом экономических законов, т.е. используются данные для того, чтобы получить количественные зависимости для экономических соотношений.

Данные, как правило, не являются экспериментальными, т.к. мы не можем проводить многократные эксперименты. Но это только малая часть работы эконометриста, он также моделирует экономические модели, основанные на экономической теории и эмпирических данных, оценивает неизвестные величины в этих моделях, делает прогнозы и оценивает их точность, дает рекомендации по экономической политике. Во всей этой деятельности существенным является использование моделей. Эйнштейн говорил: «Модели должны быть настолько простыми, насколько возможно, но не проще». В большинстве случаев экономические законы выражаются в относительно простой математической форме.

Рассмотрим функцию потребления $\ln C = \beta_0 + \beta_1 \ln Y + \beta_2 \ln P$, где $\beta_0, \beta_1, \beta_2$ – константы; C – потребление некоторого продукта на душу населения в некотором году; Y – реальный доход на душу населения в этом году; P – индекс цен на данный продукт, скорректированный (дефлированный) на общий индекс стоимости жизни. Это уравнение поведения, которое описывает потребителя по отношению к покупке данного пищевого продукта в зависимости от относительного уровня цен на продукты и реального душевого дохода.

Закон будет определен, как только мы найдем значения $\beta_0, \beta_1, \beta_2$. Соответственно задача эконометрики – определить и оценить эти коэффициенты из подходящего набора наблюдений. Но эта задача не является единственной.

1.2. Типы моделей.

Математические модели широко применяются в бизнесе, экономике, общественных науках, в исследовании экономической активности и даже в исследовании политических процессов. Можно выделить 3 основных класса моделей, которые используются при анализе и прогнозах:

1. . Модели временных рядов:

а) модель тренда: $y(t) = T(t) + E_t$, где $T(t)$ – временной тренд заданного параметрического вида; E_t – случайный (стохастический) компонент.

б) модель сезонности: $y(t) = S(t) + E_t$, где $S(t)$ – периодический сезонный компонент.

В) модели тренда и сезонности одновременно: $T(t)+S(t)+E_t$ (аддитивная модель); $T(t)S(t)+E_t$ (мультипликативная модель).

К моделям временных рядов относятся множества более сложных моделей, таких, как: модели адаптивного прогноза, модели авторегрессии, модели скользящей средней и др. Их общей чертой является то, что они объясняют поведение временного ряда исходя только из предыдущих значений. Такие модели могут применяться, например, для изучения и прогнозирования объемов продаж авиабилетов, спроса на мороженое, краткосрочного прогноза процентных ставок и т.д.

2. Регрессионные модели с одним уравнением: в таких моделях зависимая переменная y представляется в виде следующей функции: $f(x, \beta) = f(x_1, \dots, x_k; \beta_1, \dots, \beta_p)$, где x_1, \dots, x_k – независимые объясняющие переменные, β_1, \dots, β_p – параметры. В зависимости от вида функции $f(x, \beta)$ модели делятся на линейные и нелинейные.

Пример. Можно исследовать спрос на мороженое, как функцию от времени, температуры воздуха, среднего уровня дохода.

Область применения таких моделей, даже линейных, значительно шире, чем применение моделей временных рядов. Проблема оценки, отбора значимых параметров и т.п. посвящен огромный объем литературы. Тема является стержневой в эконометрике и в данном курсе.

3. Системы одновременных уравнений: эти модели описываются системами уравнений. Они могут состоять из тождеств и регрессионных уравнений, каждое из которых кроме объясняющих переменных, включает в себя системы переменных из других уравнений системы.

Пример. Модель спроса и предложения (S-D) (система одновременных уравнений требует использовать более сложный математический аппарат) Пусть Q_t^D – спрос на товар в момент времени t ; Q_t^S – предложение товара в момент времени t ; P_t – цена товара в момент времени t ; Y_t – реальный доход на душу населения в момент времени t .

Сопоставим систему уравнений «спрос-предложение».

Предложение: $Q_t^S = \alpha_1 + \alpha_2 P_t + \alpha_3 P_{t-1} + E_t$; $Q_t = Q_t^S = Q_t^D$.

Цена товара P_t и спрос на товар $Q_t^D = Q_t^S = Q_t$ определяются из уравнений моделей, т.е. являются эндогенными переменными. Предопределенными переменными являются доход Y^D и значение цены товара в предыдущий момент времени P_{t-1} .

1.3. Типы данных.

При моделировании экономических процессов мы имеем дело с двумя типами данных (пространственные данные и временные ряды).

Примером **пространственных данных** являются набор сведений: объем производства, количество работников, доход и др. по разным фирмам в один и тот же момент времени. Это называется пространственным срезом. Другим примером могут служить данные по покупке-продаже наличной валюты в обменных пунктах Москвы. Примером **временных данных** могут быть ежеквартальные данные по инфляции, заработной плате, ежедневный курс \$ на ММВБ.

Отличительной чертой всех данных является то, что они упорядочены по времени. Кроме того, в близкие моменты времени часто бывают зависимыми.

1. Парная линейная регрессия

Уравнением регрессии называется функция, описывающая зависимость условного среднего значения результативного признака y от заданных значений аргументов.

Предположим, что из двумерной генеральной совокупности (x, y) взята выборка объемом n , где (x_i, y_i) результат i -го наблюдения $i=1, \dots, n$. В этом случае *модель парной линейной регрессии* имеет вид:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

где y_i – зависимая переменная; x_i – независимая переменная; ε_i – ошибка. Допустим, что ε_i – это независимые, нормально распределенные случайные величины с нулевым математическим ожиданием $M(\varepsilon_i) = 0$, дисперсией ошибки, независящей от номера наблюдения $D\varepsilon_i = \sigma^2$.

Значения параметров β_0 и β_1 неизвестны. Их оценки, рассчитанные по выборочным данным, обозначают b_0 и b_1 . Подставляя оценки в уравнение получим выборочное уравнение регрессии:

$$\hat{y} = b_0 + b_1 x.$$

Чаще всего для расчета оценок коэффициентов регрессии применяют **метод наименьших квадратов** (МНК). В качестве оценок неизвестных параметров β_0 и β_1 берут такие значения выборочных характеристик b_0 и b_1 , которые минимизируют сумму квадратов отклонений y_i от \hat{y}_i :

$$Q = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2 \rightarrow \min.$$

Для отыскания минимума найдем частные производные Q и приравняем их к нулю:

$$\begin{cases} \frac{\partial Q}{\partial b_0} = -2 \sum_{i=1}^n (y_i - b_0 - b_1 x_i) = 0; \\ \frac{\partial Q}{\partial b_1} = -2 \sum_{i=1}^n (y_i - b_0 - b_1 x_i) x_i = 0. \end{cases}$$

Получим систему уравнений:

$$\begin{cases} nb_0 + b_1 \sum x_i = \sum y_i; \\ b_0 \sum x_i + b_1 \sum x_i^2 = \sum x_i y_i. \end{cases} \quad (1.1)$$

Данная система называется **системой нормальных уравнений**.

Решая эту систему относительно b_0 и b_1 , получим оценки коэффициентов регрессии:

$$b_0 = \bar{y} - b_1 \bar{x}, \quad (1.2)$$

$$b_1 = \frac{\sum x_i y_i - \frac{1}{n} \sum x_i \sum y_i}{\sum x_i^2 - \frac{1}{n} (\sum x_i)^2}, \quad (1.3)$$

где средние значения \bar{x} и \bar{y} находятся по формулам:

$$\bar{x} = \frac{\sum x_i}{n}; \quad \bar{y} = \frac{\sum y_i}{n}. \quad (1.4)$$

По полученному уравнению регрессии

$$\hat{y}_i = b_0 + b_1 x_i \quad (1.5)$$

получают расчетные (прогнозные) значения переменной y для каждого i наблюдения, то есть $\hat{y}_i(x_i)$. Для этого в уравнение (1.5) подставляют известные значения независимой переменной x_i ($i = 1, 2, \dots, n$).

Определяют характеристики качества построенной регрессионной модели (1.5). Для этого проводят анализ остатков модели $e_i = y_i - \hat{y}_i$ (для всех $i = 1, 2, \dots, n$).

Находят **несмещенную оценку остаточной дисперсии**:

$$\hat{S}^2 = \frac{1}{n-2} \sum (y_i - \hat{y}_i)^2 = \frac{\sum e_i^2}{n-2} = \frac{Q_{ост}}{n-2}. \quad (1.6)$$

Величину $S = \sqrt{\hat{S}^2}$ называют **стандартной ошибкой остатков**.

Наряду с получением точечных оценок коэффициентов регрессии важно знать, насколько точны эти оценки. Для этого вычисляют стандартные ошибки коэффициентов и строят доверительные интервалы для коэффициентов.

Стандартные ошибки коэффициентов b_0 и b_1 вычисляют по формулам:

$$\hat{S}_{b_0} = \sqrt{\frac{\hat{S}^2 \sum x_i^2}{n \sum x_i^2 - (\sum x_i)^2}}; \quad (1.7)$$

$$\hat{S}_{b_1} = \sqrt{\frac{\hat{S}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}. \quad (1.8)$$

Доверительным интервалом называется такой интервал, относительно которого можно с заранее выбранной вероятностью утверждать, что он содержит значения прогнозируемого показателя.

Интервальная оценка для параметра β_0 :

$$b_0 - t_\gamma \hat{S}_{b_0} \leq \beta_0 \leq b_0 + t_\gamma \hat{S}_{b_0}, \quad (1.9)$$

где t_γ определяется из таблицы распределения Стьюдента для **уровня значимости** $\alpha=1-\gamma$ и числа степеней свободы $\nu=n-2$. Обычно уровень значимости берут равным 0,05 или 0,01. γ называют **доверительной**

вероятностью или надежностью, \hat{S}_{b_0} - стандартной ошибкой коэффициента b_0 .

Аналогично определяется интервальная оценка для коэффициента β_1 :

$$b_1 - t_{\gamma} \hat{S}_{b_1} \leq \beta_1 \leq b_1 + t_{\gamma} \hat{S}_{b_1}. \quad (1.10)$$

Если переменная x не оказывает влияния на y , то прогноз будет одинаковым для всех значений независимых переменных x . В этом случае использовать уравнение регрессии для прогноза не имеет смысла. Значимость уравнения проверяют с помощью F критерия Фишера. Рассчитывают следующие суммы квадратов:

1. Общая сумма квадратов:

$$Q_{общ} = \sum (y_i - \bar{y})^2. \quad (1.11)$$

2. Сумма квадратов остатков:

$$Q_{ост} = \sum (y_i - \hat{y}_i)^2. \quad (1.12)$$

3. Сумма квадратов, обусловленная регрессией:

$$Q_{регр} = \sum (\hat{y}_i - \bar{y})^2. \quad (1.13)$$

Выполняется условие:

$$\sum (y_i - \bar{y})^2 = \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2, \text{ то есть} \\ Q_{общ} = Q_{регр} + Q_{ост}. \quad (1.14)$$

Для оценки значимости уравнения регрессии, то есть проверки гипотезы $H_0: \beta_1=0$ рассчитывают статистику $F_{набл}$:

$$F = \frac{\sum (\hat{y}_i - \bar{y})^2 / 1}{\sum (y_i - \hat{y}_i)^2 / (n-2)} = \frac{Q_{регр} / 1}{Q_{ост} / (n-2)}. \quad (1.15)$$

По таблице F -распределения находят критическое значение $F_{кр}(\alpha; \nu_1, \nu_2)$, где α - уровень значимости; ν_1, ν_2 - число степеней свободы, причем для модели парной регрессии $\nu_1=1, \nu_2=n-2$. Если выполняется $F_{набл} > F_{кр}$, то гипотеза отвергается и уравнение считается значимым.

Коэффициентом детерминации или долей объясненной дисперсии называется коэффициент:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{Q_{ост}}{Q_{общ}}. \quad (1.16)$$

Если $R^2 = 1$, то линия регрессии точно соответствует всем наблюдениям, если $R^2 = 0$, то регрессия не объясняет вариацию признаков, то есть отсутствует связь между x и y .

Коэффициент детерминации R^2 также может применяться для проверки гипотезы о значимости уравнения регрессии, то есть гипотезы $H_0: \beta_1=0$.

Для этого рассчитывают статистику:

$$F = \frac{R^2}{1 - R^2} (n - 2), \quad (1.17)$$

имеющую распределение Фишера с числом степеней свободы $\nu_1=1$, $\nu_2=n-2$.

Если подставить в (1.17) значение коэффициента детерминации (1.16), то получим статистику, рассчитываемую по формуле (1.15), то есть статистики (1.15) и (1.17) одинаковы. Аналогично проводится проверка гипотезы H_0 . Если выполняется $F_{набл} > F_{крит}$, то гипотеза отвергается и уравнение считается значимым.

Если в целом уравнение регрессии значимо, проверяют значимость каждого коэффициента.

Установление значимости коэффициента регрессии β_j сводится к проверке гипотезы $H_0: \beta_j=0$.

Для проверки гипотезы рассчитывают:

$$t_{набл_j} = \frac{b_j}{\hat{S}_{b_j}}. \quad (1.18)$$

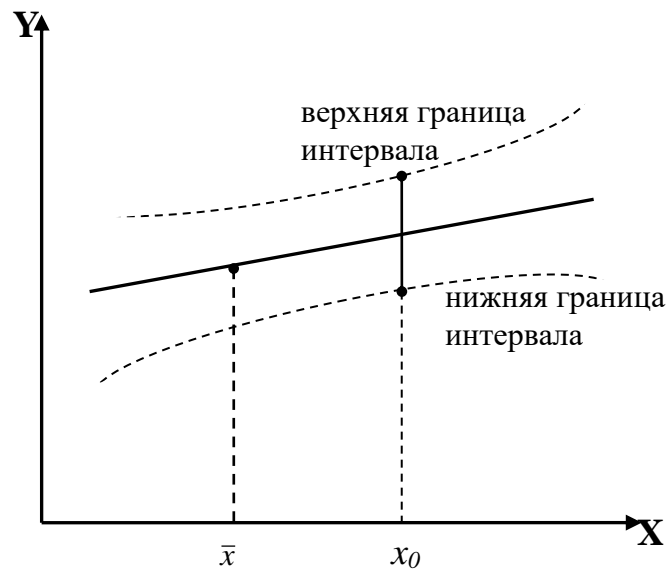
Из таблиц распределения Стьюдента (t -распределения) находят критическое значение $t_{кр}(\alpha; \nu=n-2)$, где α - уровень значимости, ν - число степеней свободы.

Если $|t_{набл}| > t_{кр}$, то гипотеза H_0 отвергается и коэффициент считается значимым. Если $|t_{набл}| \leq t_{кр}$, то гипотеза H_0 не отвергается.

Интервальная оценка для прогноза $y_{прогноз}$ при $x=x_0$ находится следующим образом:

$$y_{прогн} \in \left[(b_0 + b_1 x_0) \pm t_\gamma \hat{S} \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \right], \quad (1.19)$$

где t_γ определяется по таблицам распределения Стьюдента для заданного уровня значимости α и числа степеней свободы $\nu=n-2$.



Доверительный интервал имеет наименьшую величину, когда $x_0 = \bar{x}$, а по мере удаления x_0 от \bar{x} ширина доверительного интервала увеличивается, и точность оценки y снижается.

Задача 1.1. На основании данных об издержках обращения и изменения товарооборота, приведенных в таблице:

Издержки обращения, тыс. руб. (y)	1,1	1,3	1,4	1,1	1,9	1,7	1,4	1,2	1,9	1,9
Товарооборот, тыс. руб. (x)	4	5	6	3	8	9	6	4	6	9

и предположения, что генеральное уравнение регрессии имеет вид

$$y = \beta_0 + \beta_1 x + \varepsilon, \text{ требуется:}$$

- Найти оценку и проверить на 5% уровне значимости уравнения регрессии, то есть гипотезу $H_0: \beta_1 = 0$;
- Построить таблицу дисперсионного анализа для расчета F -критерия Фишера;
- Найти коэффициент детерминации R^2 ;
- Найти интервальную оценку для прогноза издержек обращения y при товарообороте $x = 10,0$ тыс. руб.

Решение

Расчеты проведем в программе Excel.

В столбец А занесем данные задачи по переменной x (товарооборот). В столбец В – данные переменной y (издержки обращения) (см. рис.1.1).

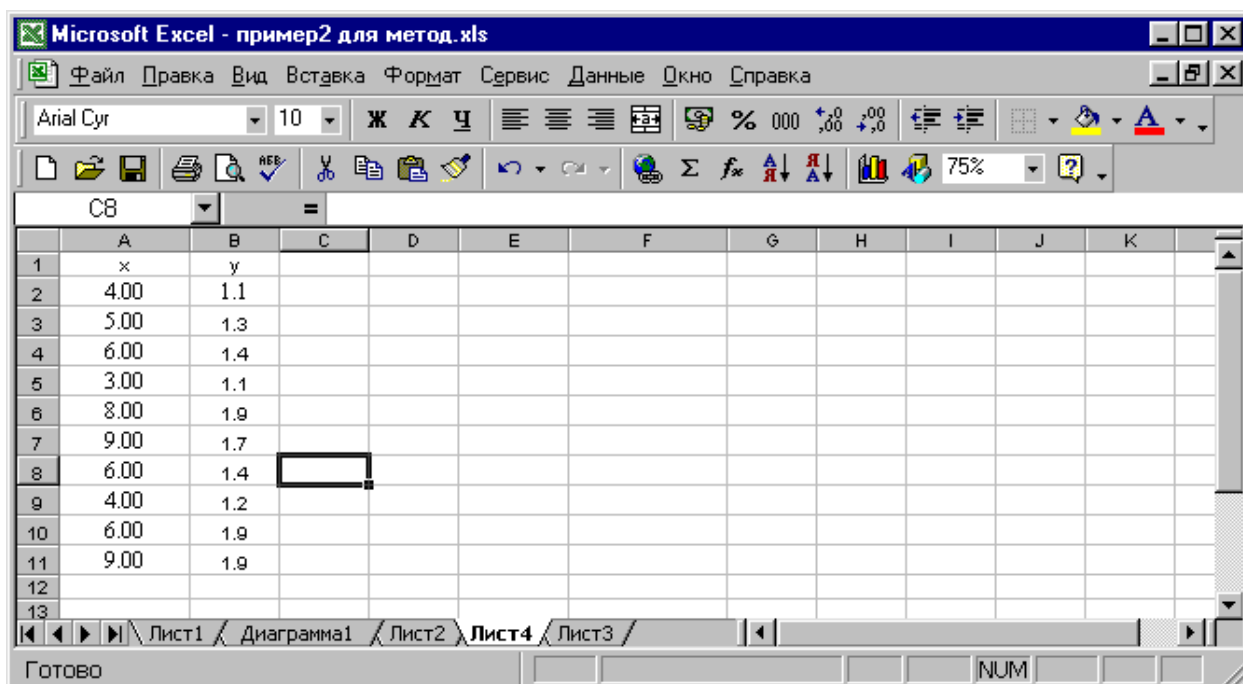


Рис. 1.1. Данные задачи

Для того, чтобы проверить, существует ли зависимость между признаками, построим диаграмму рассеивания. Для этого выделите столбцы x и y (ячейки A1:B11). На панели инструментов нажмите значок «Диаграммы» и при появлении окна (Мастер диаграмм) выберите точечную диаграмму (см. рис.1.2).

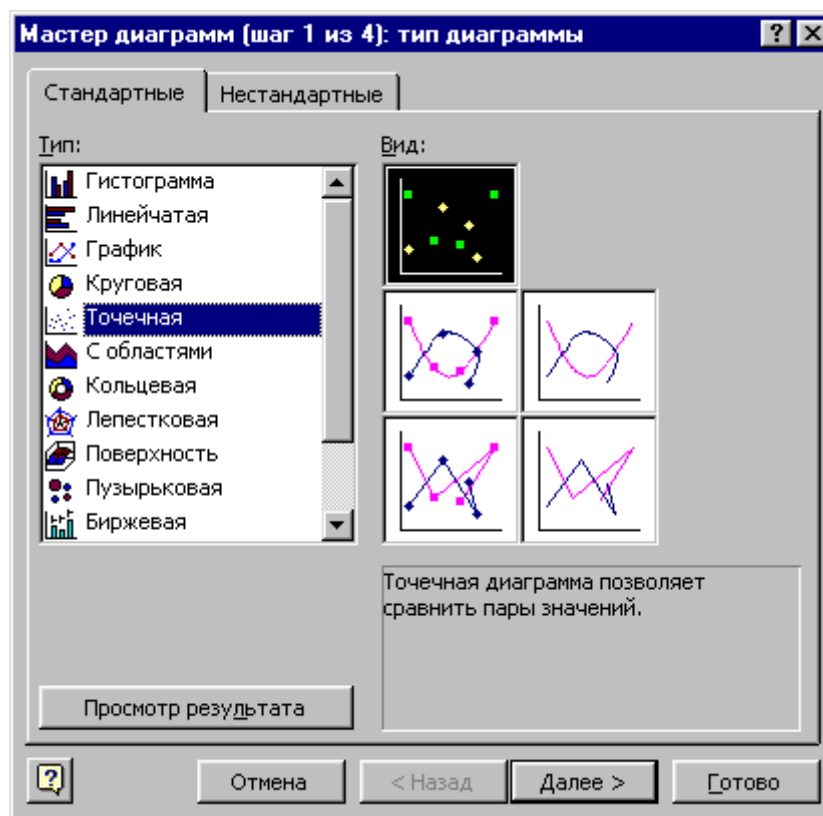


Рис. 1.2. Выбор точечной диаграммы

Нажимая клавишу «Далее» несколько раз, постройте точечную диаграмму (поле корреляции) на отдельном листе.

На рис. 1.3 видно, что между признаками x и y действительно наблюдается линейная связь.

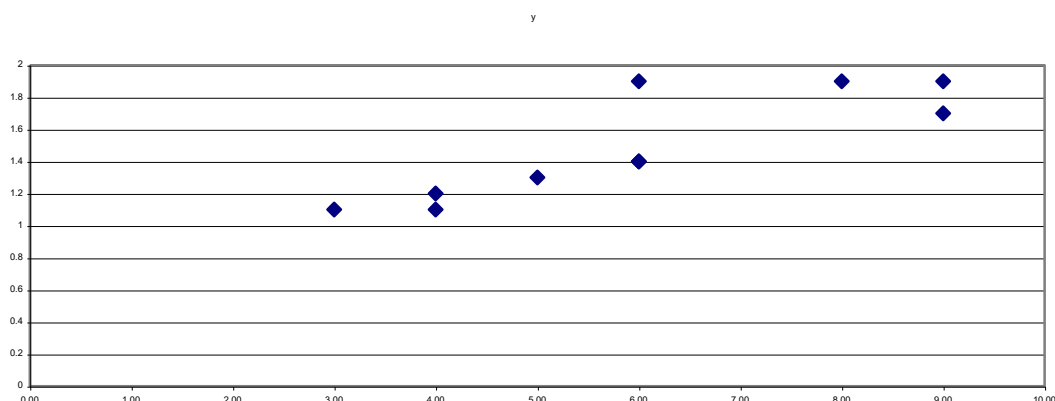


Рис. 1.3. Поле корреляции

Для нахождения коэффициентов регрессии необходимо решить систему (1.1) и вычислить столбцы x^2 и xy . Для этого в ячейку C2 вводится формула ($=A2^2$) (см. на рис.1.4 строку формул), чтобы формула ввелась, нажмите ОК. Затем курсор поместите в нижний правый угол ячейки и при появлении знака +, растяните ячейку вниз. Во всех ячейках столбца значения x возведутся в квадрат.

Аналогично рассчитывается столбец произведений xy : в ячейке D2 введена формула: ($=A2*B2$).

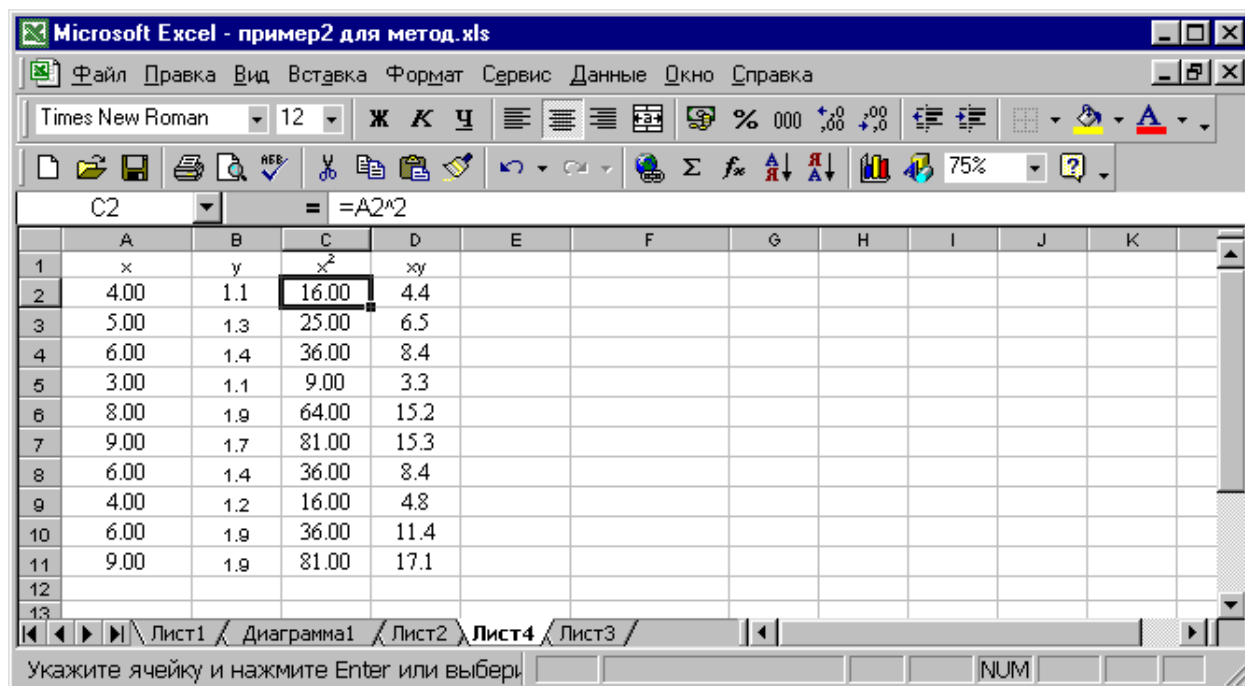


Рис.1.4. Вычисление столбцов

В ячейках A12, B12, C12, D12 стоят суммы по столбцам x , y , x^2 , xy . Для их нахождения необходимо в ячейке A12 выбрать на панели инструментов знак Σ , на экране выделится столбец, который необходимо суммировать. Затем следует нажать клавишу Enter – и в ячейке A12 появится сумма столбца x (60,00).

Аналогично рассчитываются суммы по остальным столбцам (рис. 1.5).

	A	B	C	D	E	F	G	H	I	J	K
1	x	y	x^2	xy							
2	4.00	1.1	16.00	4.4							
3	5.00	1.3	25.00	6.5							
4	6.00	1.4	36.00	8.4							
5	3.00	1.1	9.00	3.3							
6	8.00	1.9	64.00	15.2							
7	9.00	1.7	81.00	15.3							
8	6.00	1.4	36.00	8.4							
9	4.00	1.2	16.00	4.8							
10	6.00	1.9	36.00	11.4							
11	9.00	1.9	81.00	17.1							
12	60.00	14.90	400.00	94.80							
13											
14											
15											
16											
17											
18											
19											

Рис. 1.5. Вычисление сумм по столбцам таблицы

После того, как найдены суммы по столбцам, найдем оценки коэффициентов регрессии. Для решения системы уравнений методом наименьших квадратов (МНК) воспользуемся формулами (1.2) и (1.3).

Предварительно рассчитаем средние значения x и y по формулам (1.4). Для этого в ячейку B14 введем формулу ($=A12/10$), где 10 – это объем выборки. Аналогично находим в ячейке B15 среднее значение y .

Расчет коэффициентов регрессии по формулам (1.2) и (1.3) приведен на рис. 1.6. Причем, сначала рассчитывается коэффициент b_1 в ячейке B17 (см. строку формул на рис.1.6), а затем коэффициент b_0 .

Для расчета b_0 в ячейку B16 введена формула: ($=B15-B17*B14$).

Microsoft Excel - пример2 для метод.xls

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

Arial Cyr 10 Ж К Ц

В17 = =(D12-(1/10)*A12*B12)/(C12-(1/10)*(A12^2))

	A	B	C	D	E	F	G	H	I	J	K	L
1	x	y	x ²	xy								
2	4.00	1.1	16.00	4.4								
3	5.00	1.3	25.00	6.5								
4	6.00	1.4	36.00	8.4								
5	3.00	1.1	9.00	3.3								
6	8.00	1.9	64.00	15.2								
7	9.00	1.7	81.00	15.3								
8	6.00	1.4	36.00	8.4								
9	4.00	1.2	16.00	4.8								
10	6.00	1.9	36.00	11.4								
11	9.00	1.9	81.00	17.1								
12	60.00	14.90	400.00	94.80								
13												
14	x _{ср}	6										
15	y _{ср}	1.49										
16	b ₀	0.68										
17	b ₁	0.135										
18												

Лист1 / Диаграмма1 / Лист2 / Лист3 /

Готово

Рис. 1.6. Расчет коэффициентов регрессии

Таким образом, получили оценку уравнения регрессии:

$$\hat{y} = 0,68 + 0,135x. \quad (1.20)$$

Коэффициент $b_1 = 0,135$ показывает, что при изменении товарооборота на 1 тыс. руб., издержки обращения увеличивается на 0,135 тыс. руб.

Найдем прогнозные значения y_{np} . Для этого подставляют в уравнение регрессии (1.20) все значения x . Для этого в ячейку E2 вводят формулу (см. на рис. 1.7 строку формул). Следует учесть, что для всех данных x значения коэффициентов $b_0 = 0,68$ и $b_1 = 0,135$ не меняются, поэтому для ячеек, в которых находятся значения коэффициентов регрессии при наборе формулы, то есть для ячеек B16 и B17, следует нажать клавишу F4 на клавиатуре – в результате в строке формул появится знак \$.

Сумма y_{np} должна быть равна сумме y (14,9).

	A	B	C	D	E	F	G	H	I	J	K	L
1	x	y	x^2	xy	$y_{пр}$							
2	4.00	1.1	16.00	4.4	1.22							
3	5.00	1.3	25.00	6.5	1.355							
4	6.00	1.4	36.00	8.4	1.49							
5	3.00	1.1	9.00	3.3	1.085							
6	8.00	1.9	64.00	15.2	1.76							
7	9.00	1.7	81.00	15.3	1.895							
8	6.00	1.4	36.00	8.4	1.49							
9	4.00	1.2	16.00	4.8	1.22							
10	6.00	1.9	36.00	11.4	1.49							
11	9.00	1.9	81.00	17.1	1.895							
12	60.00	14.90	400.00	94.80	14.90							
13												
14	$x_{ср}$	6										
15	$y_{ср}$	1.49										
16	b_0	0.68										
17	b_1	0.135										
18												

Рис. 1.7. Нахождение прогнозных значений y

Проведем анализ полученного уравнения.

Найдем остатки $y_i - \hat{y}_i$, оценку остаточной дисперсии \hat{s}^2 , стандартных ошибок коэффициентов.

Для оценки значимости уравнения регрессии и для нахождения значения F критерия, рассчитаем по формулам (1.11), (1.12) и (1.13) $Q_{общ}$, $Q_{ост}$ и $Q_{регр}$, то есть рассчитаем таблицу дисперсионного анализа (рис. 1.8).

Обратите внимание на расчет $Q_{общ}$ и $Q_{регр}$ с использованием среднего значения y (см. строку формул на рис.1.8).

	A	B	C	D	E	F	G	H	I	J	K	L
	x	y	x ²	xy	Y _{гр}	Y-Y _{гр}	(Y-Y _{гр}) ²	(Y-Y _{ср}) ²	(Y _{ср} -Y _{гр}) ²			
1	4.00	1.1	16.00	4.4	1.22	-0.12	0.0144	0.1521	0.0729			
2	5.00	1.3	25.00	6.5	1.355	-0.055	0.00303	0.0361	0.01822			
3	6.00	1.4	36.00	8.4	1.49	-0.09	0.0081	0.0081	0			
4	3.00	1.1	9.00	3.3	1.085	0.015	0.00022	0.1521	0.16402			
5	8.00	1.9	64.00	15.2	1.76	0.14	0.0196	0.1681	0.0729			
6	9.00	1.7	81.00	15.3	1.895	-0.195	0.03802	0.0441	0.16402			
7	6.00	1.4	36.00	8.4	1.49	-0.09	0.0081	0.0081	0			
8	4.00	1.2	16.00	4.8	1.22	-0.02	0.0004	0.0841	0.0729			
9	6.00	1.9	36.00	11.4	1.49	0.41	0.1681	0.1681	0			
10	9.00	1.9	81.00	17.1	1.895	0.005	2.5E-05	0.1681	0.16402			
11	60.00	14.90	400.00	94.80	14.90	0.00		0.26	0.99	0.73		
12							Qост	Qобщ	Qрег			
13	x _{ср}	6										
14	y _{ср}	1.49										
15	b ₀	0.68										
16	b ₁	0.135										
17												
18												

Рис. 1.8. Расчет таблицы однофакторного дисперсионного анализа

Как видно из рис. 1.8 выполняется соотношение (1.14):

$$Q_{\text{общ}} = Q_{\text{рег}} + Q_{\text{ост}} = 0,73 + 0,26 = 0,99.$$

Для оценки значимости уравнения регрессии проверим гипотезу $H_0: \beta_1 = 0$.

Рассчитаем F -статистику по формуле (1.15) в ячейке E14 (рис. 1.9):

$$F_{\text{набл}} = \frac{\sum (y_i - \hat{y}_i)^2 / 1}{\sum (y_i - \hat{y}_i)^2 / (n - 2)} = \frac{Q_{\text{рег}} / 1}{Q_{\text{ост}} / (n - 2)} = \frac{0,73 / 1}{0,26 / 8} = 22,43.$$

По таблице F -распределения находят $F_{\text{кр}}$ с числом степеней свободы $\nu_1 = 1$, $\nu_2 = n - 2 = 10 - 2 = 8$. Найдем его с помощью математических функций: определим критическое значение распределения Фишера $F_{\text{крит}} = \text{FRASПОБР}(0,05; 1; 8) = 5,3176$ (ячейка G14 на рис. 1.9). Для нахождения этого числа необходимо на панели инструментов выбрать клавишу f_x - появится окно «Мастер функций», нажать курсором на строку «Статистические функции» и в правом окне найти FRASПОБР (см. строку формул на рис. 1.9). Вероятность – это уровень значимости $\alpha = 0,05$.

Так как $F_{\text{набл}} = 22,43 > F_{\text{кр}} = 5,3176$, то гипотеза отвергается и уравнение считается значимым.

Проверим значимость каждого коэффициента регрессии.

Рассчитаем несмещенную оценку остаточной дисперсии по формуле (1.6) (ячейка E15):

$$\hat{S}^2 = \frac{1}{n-2} \sum (y_i - \hat{y}_i)^2 = \frac{Q_{ост}}{n-2} = \frac{0,26}{8} = 0,0325,$$

и стандартную ошибку коэффициентов по формулам (1.7) и (1.8).

Для расчета стандартной ошибки коэффициента b_1 нашли сумму в ячейке J12:

$$\sum_{i=1}^n (x_i - \bar{x})^2 = 40.$$

В ячейках E15, E16 и E17 найдены оценки соответствующие остаточной дисперсии и дисперсии коэффициентов регрессии, то есть $\hat{S}_{b_0}^2 = 0,0325$ и $\hat{S}_{b_1}^2 = 0,00081$ В ячейках – G15, G16, G17 - стандартные ошибки или квадратные корни из дисперсий коэффициентов:

$$\hat{S}_{b_0} = 0,18;$$

$$\hat{S}_{b_1} = 0,0285.$$

Корень вычислили с помощью математической функции КОРЕНЬ (число), используя «Мастер функции».

Для проверки гипотезы $H_0: \beta_0=0$ рассчитали статистику в ячейке I16:

$$t_{набл} = \frac{b_0}{\hat{S}_{b_0}} = 3,77.$$

Находим критическое значение распределения Стьюдента с помощью статистической функции СТЬЮДРАСПОБР(0,05;8) = 2,3 (ячейка K16), где вероятность (уровень значимости) равна 0,05 и число степеней свободы $n-2=10-2=8$.

Так как $t_{набл}=3,77 > t_{кр}=2,3$ то гипотеза H_0 отвергается и коэффициент b_0 считается значимым.

Проверим гипотезу $H_0: \beta_1=0$. Рассчитаем в ячейке I17:

$$t_{набл} = \frac{b_1}{\hat{S}_{b_1}} = 4,736.$$

Так как $t_{набл}=4,736 > t_{кр}=2,3$ то гипотеза H_0 отвергается и коэффициент b_1 значим, то есть товарооборот значимо влияет на издержки обращения.

Найдем коэффициент детерминации:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{Q_{ост}}{Q_{общ}} = 0,74.$$

Таким образом, полученная модель регрессии адекватно описывает данные.

При выполнении всех расчетов получили рис.1.9.

	A	B	C	D	E	F	G	H	I	J	K
1	x	y	x ²	xy	y _{нр}	y-y _{нр}	(y-y _{нр}) ²	(y-y _{ср}) ²	(y _{ср} -y _{нр}) ²	(x-x _{ср}) ²	
2	4.00	1.1	16.00	4.4	1.22	-0.12	0.0144	0.1521	0.0729	4	
3	5.00	1.3	25.00	6.5	1.355	-0.055	0.00303	0.0361	0.01822	1	
4	6.00	1.4	36.00	8.4	1.49	-0.09	0.0081	0.0081	0	0	
5	3.00	1.1	9.00	3.3	1.085	0.015	0.00022	0.1521	0.16402	9	
6	8.00	1.9	64.00	15.2	1.76	0.14	0.0196	0.1681	0.0729	4	
7	9.00	1.7	81.00	15.3	1.895	-0.195	0.03802	0.0441	0.16402	9	
8	6.00	1.4	36.00	8.4	1.49	-0.09	0.0081	0.0081	0	0	
9	4.00	1.2	16.00	4.8	1.22	-0.02	0.0004	0.0841	0.0729	4	
10	6.00	1.9	36.00	11.4	1.49	0.41	0.1681	0.1681	0	0	
11	9.00	1.9	81.00	17.1	1.895	0.005	2.5E-05	0.1681	0.16402	9	
12	60.00	14.90	400.00	94.80	14.90	0.00	0.26	0.99	0.73	40.00	
13							Qост	Qобщ	Qперп		
14	x _{ср}	6		F набл	22.4307692	F крит	5.31764				
15	y _{ср}	1.49		S ²	0.0325	S	0.18028				
16	b ₀	0.68		S ² _{b0}	0.0325	S _{b0}	0.18028	t _{b0}	3.77196	t _{крит}	2.30601
17	b ₁	0.135		S ² _{b1}	0.0008125	S _{b1}	0.0285	t _{b1}	4.73611		

Рис. 1.9. Анализ уравнения регрессии

Найдем 95% доверительные интервалы для каждого коэффициента регрессии, а затем для прогнозного значения $x=2$.

Интервальная оценка для параметра β_0 :

$$b_0 \in [b_0 \pm t_{\gamma} \hat{S}_{b_0}] = [0,68 \pm 2,3 \cdot 0,18028] = [0,68 \pm 0,4146]$$

или $0,68-0,4141 \leq b_0 \leq 0,68+0,4146$.

В ячейке D22 найдена верхняя граница доверительного интервала для коэффициента b_0 , то есть $0,68+0,4146=1,096$, в ячейке D22 ($0,68-0,4146=0,264$) – нижняя граница. Следовательно, коэффициент b_0 изменяется в интервале от 0,264 до 1,096 (рис. 1.10).

Аналогично интервальная оценка для коэффициента β_1 :

$$\beta_1 \in [b_1 \pm t_{\gamma} \hat{S}_{b_1}] = [0,135 \pm 2,3 \cdot 0,0285] = [0,135 \pm 0,0657]$$

или $0,135-0,0657 \leq b_0 \leq 0,135+0,0657$.

В ячейках D24, D25 приведены расчеты для нижней и верхней границ коэффициента b_1 .

	A	B	C	D	E	F	G	H	I	J	K
1	x	y	x ²	xy	y _{пр}	y-y _{пр}	(y-y _{пр}) ²	(y-y _{пр}) ²	(y _{пр} -y _{ср}) ²	(x-x _{ср}) ²	
2	4.00	1.1	16.00	4.4	1.22	-0.12	0.0144	0.1521	0.0729	4	
3	5.00	1.3	25.00	6.5	1.355	-0.055	0.00303	0.0361	0.01822	1	
4	6.00	1.4	36.00	8.4	1.49	-0.09	0.0081	0.0081	0	0	
5	3.00	1.1	9.00	3.3	1.085	0.015	0.00022	0.1521	0.16402	9	
6	8.00	1.9	64.00	15.2	1.76	0.14	0.0196	0.1681	0.0729	4	
7	9.00	1.7	81.00	15.3	1.895	-0.195	0.03802	0.0441	0.16402	9	
8	6.00	1.4	36.00	8.4	1.49	-0.09	0.0081	0.0081	0	0	
9	4.00	1.2	16.00	4.8	1.22	-0.02	0.0004	0.0841	0.0729	4	
10	6.00	1.9	36.00	11.4	1.49	0.41	0.1681	0.1681	0	0	
11	9.00	1.9	81.00	17.1	1.895	0.005	2.5E-05	0.1681	0.16402	9	
12	60.00	14.90	400.00	94.80	14.90	0.00	0.26	0.99	0.73	40.00	
13							Qост	Qобщ	Qперп		
14	x _{ср}	6		F набл	22.4307692	F крит	5.31764				
15	y _{ср}	1.49		S ²	0.0325	S	0.18028				
16	b ₀	0.68		S ² _{b0}	0.0325	S _{b0}	0.18028	t _{b0}	3.77196	t _{крит}	2.30601
17	b ₁	0.135		S ² _{b1}	0.0008125	S _{b1}	0.0285	t _{b1}	4.73611		
18											
19		R ²	0.73711								
20											
21	Доверительный интервал										
22	b ₀	Верхняя граница	1.09572								
23		Нижняя граница	0.26428								
24	b ₁	Верхняя граница	0.20073								
25		Нижняя граница	0.06927								

Рис. 1.10. Расчет доверительных интервалов

Интервальная оценка для уравнения регрессии y при $x_0=10$:

$$y \in \left[(b_0 + b_1 x) \pm t_{\gamma} \hat{S} \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \right].$$

Для расчета этой величины необходимо подсчитать при $x_0=10$ $(x_0 - \bar{x})^2 = (10 - 1.49)^2$.

Расчет прогнозного значения для границ приведен в ячейках D24 и D25, формула для расчета верхней границы приведена в строке формул на рис.1.11. Для расчета нижней границы следует поставить знак «минус» после первой скобки.

В ячейке H26 приведено расчетное значение издержек обращения при $x=10$ тыс. руб. Для его нахождения в уравнение регрессии (1.20) подставлено значение $x=10$ и получено значение в ячейке H26 (2,03).

Microsoft Excel - пример2 для метод.xls

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

Arial Cyr 10 Ж К Ч

D26 =(\$B\$16+\$B\$17*10)+\$K\$16*\$G\$15*КОРЕНЬ((1/10+((10-\$B\$14)^2)/\$J\$12))

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	x	y	x ²	xy	Упр	У-Упр	(У-Упр) ²	(У-Упр) ²	(Уср-Упр) ²	(x-xср) ²			
2	4.00	1.1	16.00	4.4	1.22	-0.12	0.0144	0.1521	0.0729	4			
3	5.00	1.3	25.00	6.5	1.355	-0.055	0.00303	0.0361	0.01822	1			
4	6.00	1.4	36.00	8.4	1.49	-0.09	0.0081	0.0081	0	0			
5	3.00	1.1	9.00	3.3	1.085	0.015	0.00022	0.1521	0.16402	9			
6	8.00	1.9	64.00	15.2	1.76	0.14	0.0196	0.1681	0.0729	4			
7	9.00	1.7	81.00	15.3	1.895	-0.195	0.03802	0.0441	0.16402	9			
8	6.00	1.4	36.00	8.4	1.49	-0.09	0.0081	0.0081	0	0			
9	4.00	1.2	16.00	4.8	1.22	-0.02	0.0004	0.0841	0.0729	4			
10	6.00	1.9	36.00	11.4	1.49	0.41	0.1681	0.1681	0	0			
11	9.00	1.9	81.00	17.1	1.895	0.005	2.5E-05	0.1681	0.16402	9			
12	60.00	14.90	400.00	94.80	14.90	0.00	0.26	0.99	0.73	40.00			
13							Qост	Qобщ	Qрег				
14	xср	6		F набл	22.4307692	F крит	5.31764						
15	Уср	1.49		S ²	0.0325	S	0.18028						
16	b ₀	0.68		S ² _{b0}	0.0325	S _{b0}	0.18028	t _{b0}	3.77196	tkрит	2.30601		
17	b ₁	0.135		S ² _{b1}	0.0008125	S _{b1}	0.0285	t _{b1}	4.73611				
18													
19		R ²	0.73711										
20													
21	Доверительный интервал												
22	b ₀	Верхняя граница		1.09572									
23		Нижняя граница		0.26428									
24	b ₁	Верхняя граница		0.20073									
25		Нижняя граница		0.06927									
26	для x=10	Верхняя граница		2.32396	издержки обращения при x=10					2.03			
27		Нижняя граница		1.73604									
28													

Лист1 Диаграмма1 Лист2 Лист4 Лист3

Готово NUM

Рис. 1.11. Расчет доверительного интервала для x=10

Получено, что доверительный интервал для прогноза издержек обращения при товарообороте $x = 10$ тыс. руб. находится в пределах от 2,354 тыс. руб. до 1,736 тыс. руб. с 95 % уровнем надежности.

Таким образом, по полученному уравнению регрессии

$$\hat{y} = 0,68 + 0,135x$$

можно сформулировать следующие **выводы**:

- при изменении товарооборота на 1 тыс. руб., издержки обращения увеличивается на 0,135 тыс. руб.;
- коэффициент детерминации ($R^2=0,737$) показывает, что полученная модель регрессии адекватна данным;
- уравнение регрессии значимо и может применяться при прогнозе;
- товарооборот значимо влияет на издержки обращения;
- при товарообороте 10 тыс. руб. издержки обращения составят 2,03 тыс. руб.;

- доверительный интервал для прогноза издержек обращения при товарообороте $x = 10$ тыс. руб. находится в пределах от 2,354 тыс. руб. до 1,736 тыс. руб. с 95 % уровнем надежности.

В поле корреляции возможно построить уравнение линейной регрессии. Для этого на рис. 1.3 для построения прямой выделите курсором точки, соответствующие данным. Нажмите правую кнопку мыши и выберите «Добавить линию тренда», ОК и из предложенных линий тренда отберите линейную. В результате получится прямая регрессии (рис. 1.12).

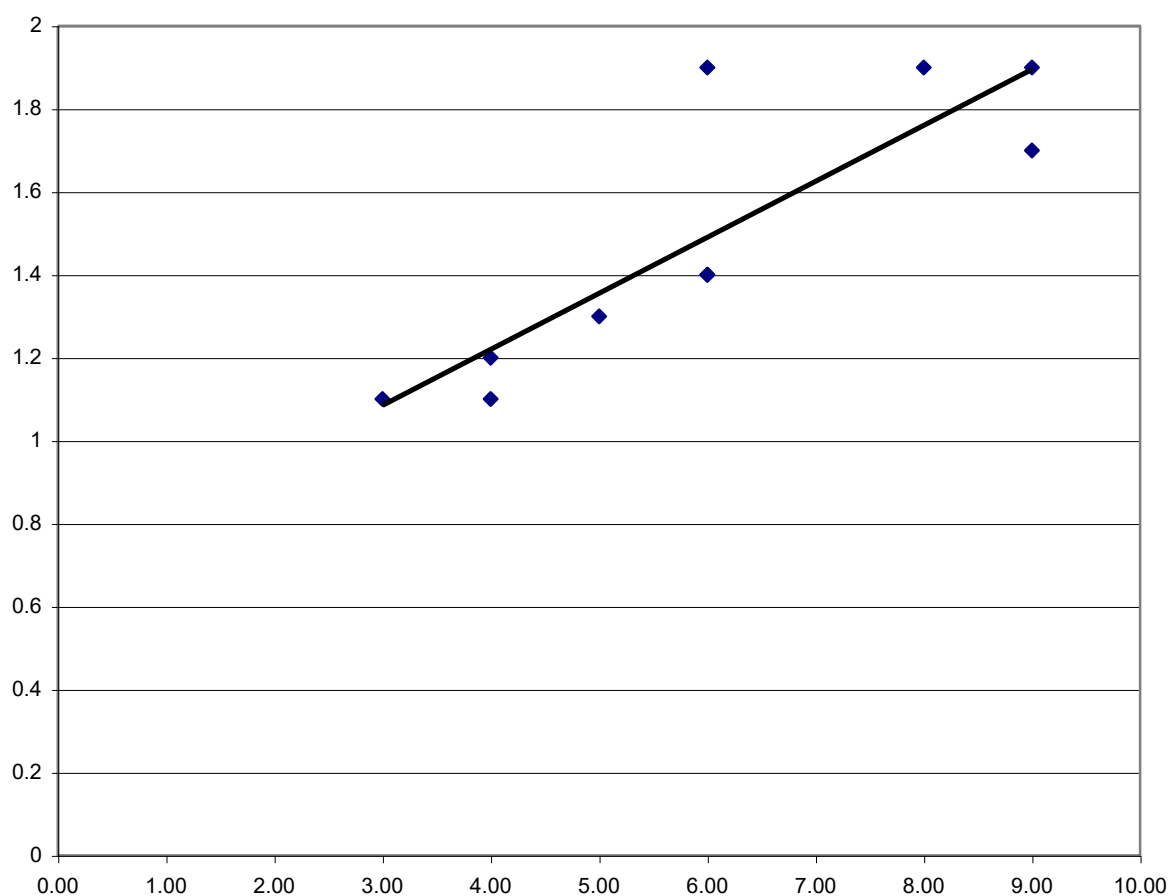


Рис. 1.12. Прямая регрессии

2. Множественное линейное уравнение регрессии

Пусть требуется построить линейную модель зависимости признака Y от набора независимых переменных X_1, \dots, X_k . Модель множественной линейной регрессии записывается в виде:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon, \quad (2.1)$$

где $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ – коэффициенты модели, ε – случайная величина, удовлетворяющая условиям: математическое ожидание ошибки равно нулю; дисперсия ошибки не зависит от номера наблюдения; ошибки разных наблюдений не зависят друг от друга.

Пусть для оценки неизвестных параметров β_j ($j = 1, \dots, k$) уравнения регрессии (2.1) взята случайная выборка объемом n и $y_i, x_{i1}, \dots, x_{ik}$ – результат i наблюдения, где $i=1, \dots, n$.

То есть имеется n наблюдений объясняемой переменной $y_i, i=1, \dots, n$, и n наблюдений k объясняющих переменных $x_{1i}, x_{2i}, \dots, x_{ki}, i=1, \dots, n$.

В матричной форме линейная модель имеет вид:

$$Y = X\beta + \varepsilon, \quad (2.2)$$

где $y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ – вектор-столбец наблюдений размерности n ;

$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n1} & \dots & x_{nk} \end{bmatrix}$ – матрица независимых переменных размерности $n \times (k+1)$;

$\beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix}$ – вектор-столбец неизвестных параметров, подлежащих оцениванию, размерности $(k+1)$;

$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$ – вектор-столбец случайных ошибок размерности n .

Применяя метод наименьших квадратов, получим вектор оценок коэффициентов модели:

$$b = (X^T X)^{-1} X^T Y. \quad (2.3)$$

Тогда выборочное уравнение регрессии имеет вид:

$$\hat{y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \dots + b_k x_{ki}. \quad (2.4)$$

Несмещенную оценку остаточной дисперсии найдем по формуле:

$$\hat{S}^2 = \frac{1}{n-k-1} \sum (y_i - \hat{y}_i)^2 = \frac{\sum e_i}{n-k-1} = \frac{Q_{ост}}{n-k-1}. \quad (2.5)$$

Для проведения регрессионного анализа рассчитывают следующие показатели:

- **коэффициент детерминации** имеет тот же самый смысл, что и R^2 для парной регрессии:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}; \quad (2.6)$$

- **множественный коэффициент корреляции**, являющийся мерой линейной зависимости между переменной y и факторами x_1, x_2, \dots, x_k :

$$R = \sqrt{R^2}; \quad (2.7)$$

- **скорректированный коэффициент детерминации** корректирует значение коэффициента R^2 с учетом числа независимых переменных и размером выборки, чтобы снизить влияние количества переменных:

$$R_{adj}^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2 / (n-k)}{\sum (y_i - \bar{y})^2 / (n-1)}; \quad (2.8)$$

- **стандартная ошибка** представляет собой стандартное отклонение для остатков. Измеряет рассеивание значений y относительно регрессии:

$$\hat{S} = \sqrt{\hat{S}^2} = \sqrt{\frac{1}{n-k-1} \sum (y_i - \hat{y}_i)^2}. \quad (2.9)$$

Для проверки значимости уравнения регрессии находят:

- **общую сумму квадратов**, которая объясняет общую изменчивость:

$$Q_{общ} = \sum (y_i - \bar{y})^2; \quad (2.10)$$

- **сумму квадратов остатков** – это часть дисперсии, не объясненная линейной зависимостью:

$$Q_{ост} = \sum (y_i - \hat{y}_i)^2; \quad (2.11)$$

- **сумму квадратов, обусловленных регрессией**, которая измеряет часть дисперсии, объясняемую регрессионной зависимостью:

$$Q_{рег} = \sum (\hat{y}_i - \bar{y})^2. \quad (2.12)$$

Выполняется условие:

$$\sum (y_i - \bar{y})^2 = \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2, \text{ то есть}$$

$$Q_{\text{общ}} = Q_{\text{регр}} + Q_{\text{ост.}} \quad (2.13)$$

Общая изменчивость у Изменчивость, объясненная
линейной зависимостью Остаточная или
необъясненная
изменчивость

Для оценки значимости уравнения регрессии, то есть проверки гипотезы $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ рассчитывают статистику:

$$F = \frac{\sum (\hat{y}_i - \bar{y})^2 / k}{\sum (y_i - \hat{y}_i)^2 / (n - k - 1)} = \frac{Q_{\text{регр}} / k}{Q_{\text{ост}} / (n - k - 1)}, \quad (2.14)$$

имеющую распределение Фишера-Снедекора (F -распределение).

По таблице F -распределения находят критическое значение $F_{кр}(\alpha; \nu_1, \nu_2)$, где α - уровень значимости; ν_1, ν_2 - число степеней свободы, $\nu_1 = k, \nu_2 = n - k - 1$. Если выполняется $F_{\text{набл}} > F_{кр}$, то гипотеза отвергается и уравнение считается значимым.

Значимость коэффициентов регрессии β_j проверяется с помощью t -критерия. Рассчитывают статистику:

$$t_{\text{набл}_j} = \frac{b_j}{\hat{S}_{b_j}}, \quad (2.15)$$

где $\hat{S}_{b_j}^2 = \hat{S}^2 [(X^T X)^{-1}]_{jj}$ - дисперсия коэффициента регрессии b_j ;

\hat{S}^2 - несмещенная оценка остаточной дисперсии (2.5);

$[(X^T X)^{-1}]_{jj}$ - элементы обратной матрицы, стоящие на главной диагонали;

\hat{S}_{b_j} - стандартная ошибка коэффициента b_j . Измеряет отклонение имеющихся данных y_i от их оценок \hat{y}_i .

Из таблиц распределения Стьюдента (t -распределения) находят критическое значение $t_{кр}(\alpha; \nu = n - k - 1)$, где ν - число степеней свободы.

Если $|t_{\text{набл}}| > t_{кр}$, то гипотеза $H_0: \beta_j = 0$ отвергается и коэффициент считается значимым. Если $|t_{\text{набл}}| \leq t_{кр}$, то гипотеза H_0 не отвергается.

Доверительные интервалы для значимых коэффициентов регрессии строятся по формуле:

$$\beta_j \in \{b_j \pm t_\gamma \hat{S} \sqrt{[(X^T X)^{-1}]_{jj}}\}, \quad (2.16)$$

где t_γ определяется из таблицы распределения Стьюдента для уровня значимости $\alpha = 1 - \gamma$ и числа степеней свободы $\nu = n - k - 1$.

Задача 2

Имеются данные о зависимости прибыли y (тыс. руб) от расходов на рекламу (тыс. руб) x_1 и стоимости основных фондов x_2 (тыс. руб) Данные приведены в таблице 2.1.

На основании данных таблицы:

- построить уравнение регрессии $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$,
- найти оценку и проверить на 5% уровне значимости уравнения регрессии, то есть гипотезу $H_0: \beta_1 = \beta_2 = 0$;
- проверить значимость каждого коэффициента регрессии;
- Найти коэффициент детерминации R^2 .

Таблица 2.1

x_1	12	14	15	16	20	21	24	25	26	23	25	26	19	27	25	29	28	30	31	33
x_2	6,2	5,2	6,5	6,9	10,2	15,5	17,2	16,2	20,5	16,9	18,2	19,5	26,2	35,2	26,5	26,9	30,2	35,5	40,0	35,0
y	10	12	15	16	17	20	22	35	41	25	27	23	26	45	39	38	41	42	42	45

Решение

Решим задачу в программе Excel с помощью пакета «Анализ данных».

Для проведения регрессионного анализа на панели инструментов выберите «Сервис», затем «Надстройки» и при появлении окна поставить флажок на «Пакет анализа», нажать ОК.

Для решения задачи введем данные задачи в столбцы А, В и С (рис.2.1).

	A	B	C	D	E	F	G	H
1	x1	x2	y					
2	12	6,2	10					
3	14	5,2	12					
4	15	6,5	15					
5	16	6,9	16					
6	20	10,2	17					
7	21	15,5	20					
8	24	17,2	22					
9	25	16,2	35					
10	26	20,5	41					
11	23	16,9	25					
12	25	18,2	27					
13	26	19,5	23					
14	19	26,2	26					
15	27	35,2	45					
16	25	26,5	39					
17	29	26,9	38					
18	28	30,2	41					
19	30	35,5	42					
20	31	40	42					
21	33	35	45					
22								
23								
24								

Рис. 2.1. Входные данные

Выберите еще раз «Сервис» и «Анализ данных». При появлении окна «Анализ данных» - нажмите курсором на инструмент анализа: «Регрессия» (см. рис. 2.2).

После выбора «Регрессии» появляется окно «Регрессия» (см. рис. 2.3).

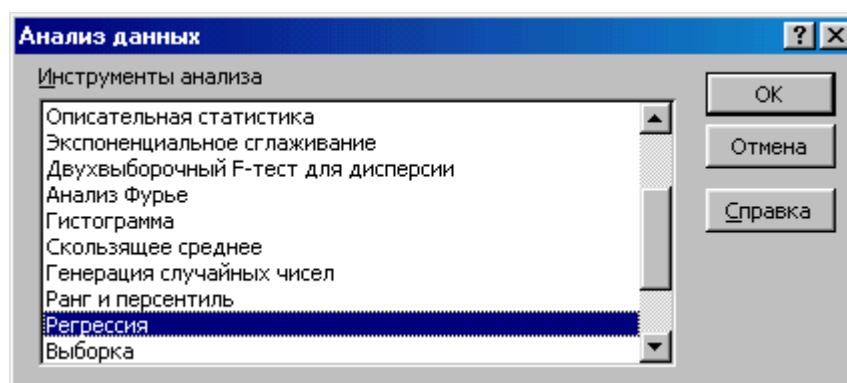


Рис. 2.2. Окно «Анализ данных»

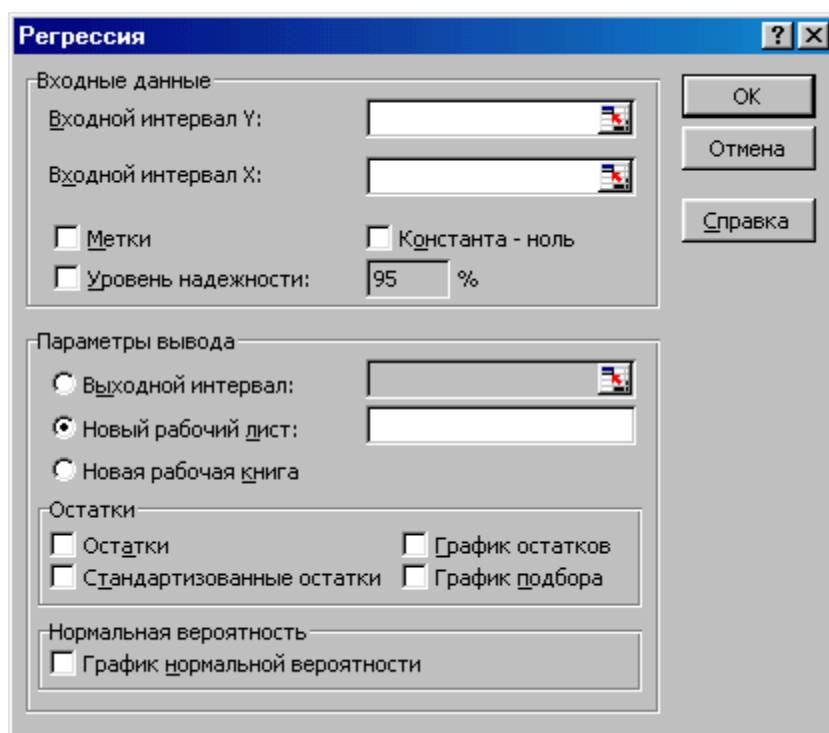


Рис. 2.3. Окно «Регрессия»

В категории **Входные данные** окна «Регрессия» необходимо указать:

- **Входной интервал Y** – диапазон зависимых переменных, состоящих из одного столбца;
- **Входной интервал X** - диапазон независимых переменных, подлежащих анализу. Может состоять из одного или более столбцов. Максимальное количество столбцов равно 16;
- **Константа-ноль** – для построения регрессии без свободного члена (b_0), то поставьте флажок в этом окне;
- **Уровень надежности** – для включения дополнительного уровня надежности, установите флажок в соответствующем окне. По умолчанию уровень надежности равен 95%;
- **Остатки** (остатки, стандартизованные остатки, график остатков и график подбора) – используются как дополнительный анализ параметров вывода.

Укажем курсором столбцы входных данных (см. рис. 2.4). Выбор столбцов проводится курсором.

Входной интервал Y – это столбец C2 – C45.

Входной интервал X – это два столбца A2 до B21.

Необходимо заполнить **Параметры вывода**: можно установить **Выходной интервал**, то есть указать ячейку, в которой появится результат. Если выходной интервал не устанавливается, то решение появится на новом листе Excel.

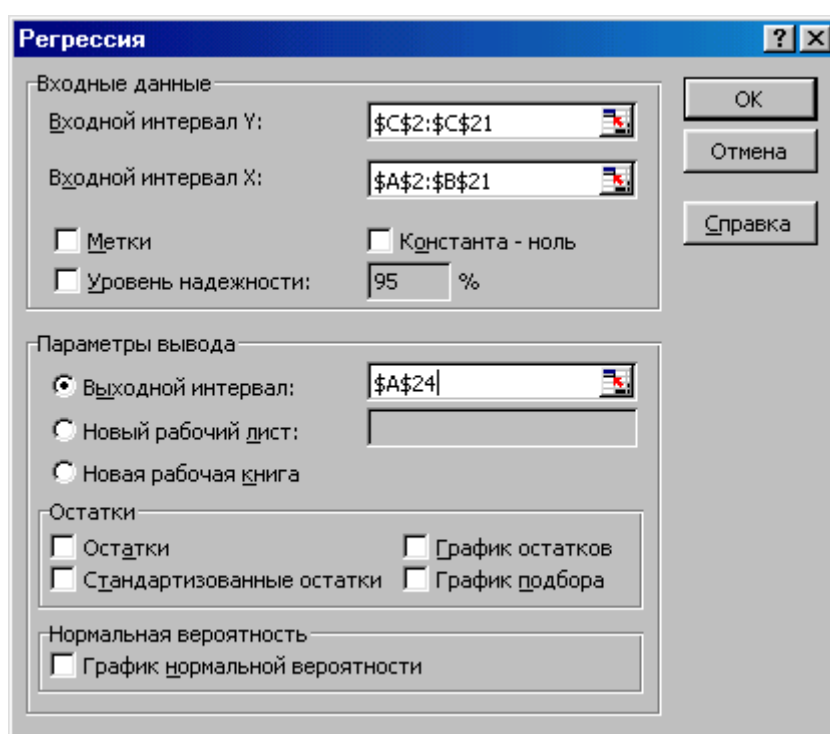


Рис. 2.4. Выбор диапазона регрессии

Нажав ОК, получим результаты решения задачи (см. рис. 2.5).

	A	B	C	D	E	F	G	H
24	ВЫВОД ИТОГОВ							
25								
26	Регрессионная статистика							
27	Множественный R	0,932052367						
28	R-квадрат	0,868721614						
29	Нормированный R-квадрат	0,853277098						
30	Стандартная ошибка	4,573752373						
31	Наблюдения	20						
32								
33	Дисперсионный анализ							
34		df	SS	MS	F	Значимость F		
35	Регрессия	2	2353,323417	1176,661708	56,2479016	3,19626E-08		
36	Остаток	17	355,6265831	20,91921077				
37	Итого	19	2708,95					
38								
39		Коэффициенты	стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%	Нижние 95,0%
40	Y-пересечение	-4,09490343	5,469608398	-0,74866483	0,46429244	-15,63478465	7,444977689	-15,63478465
41	Переменная X 1	0,917617916	0,366777547	2,501837757	0,02286165	0,143783853	1,691451979	0,143783853
42	Переменная X 2	0,561001848	0,199528978	2,811630946	0,01200773	0,140031915	0,981971781	0,140031915
43								
44								
45								
46								
47								
48								
49								
50								
51								
52								
53								
54								

Рис. 2.5. Вывод итогов построения регрессии

В столбце «Коэффициенты» получены коэффициенты уравнения регрессии.

Коэффициент $b_0 = -4,09$ в Таблице анализа – это Y -пересечение.

Таким образом, получили уравнение регрессии:

$$\hat{y} = -4,09 + 0,92x_1 + 0,56x_2.$$

Коэффициент $b_1 = 0,92$ показывает, что при увеличении расходов на рекламу на 1 тыс. руб. прибыль увеличивается в среднем на 0,92 тыс. руб., а увеличение стоимости основных фондов на 1 тыс. руб. приводит к увеличению прибыли в среднем на 0,56 тыс. руб.

Стандартные ошибки коэффициентов:

$$\hat{S}_{b_j} = \hat{S} \sqrt{[(X^T X)^{-1}]_{jj}},$$

где \hat{S}^2 - стандартная ошибка (2.9);

$[(X^T X)^{-1}]_{jj}$ - элементы обратной матрицы, стоящие на главной диагонали.

Они составляют: $\hat{S}_{b_0} = 5,46$; $\hat{S}_{b_1} = 0,37$; $\hat{S}_{b_2} = 0,1995$.

Для проверки значимости коэффициентов регрессии рассчитываются t -статистики по формуле (2.15).

Получены следующие t -статистики (рис. 2.5):

$$t_{b_0} = -0,75; t_{b_1} = 2,5; t_{b_2} = 2,81.$$

Находим критическое значение распределения Стьюдента для вероятности (уровня значимости) 0,05 и число степеней свободы $\nu = n - k - 1 = 20 - 2 - 1 = 17$.

Критическое значение находим из таблиц распределения Стьюдента или с помощью статистической функции $\text{СТЮДРАСПОБР}(0,05;17) = 2,11$.

Для проверки гипотезы $H_0: \beta_j = 0$ сравниваем полученные значения для всех коэффициентов $t_{\text{набл}}$ с $t_{\text{кр}} = 2,11$. Получим, что все коэффициенты значимы, кроме b_0 . То есть на прибыль значимо оказывают влияние расходы на рекламу и стоимость основных фондов.

Для проверки значимости коэффициентов можно использовать P -значения (ячейки E40-E42).

По величине P -значения возможно определять значимость коэффициентов, не находя критическое значение t -статистики. Если значение t -статистики велико, то соответствующее значение вероятности значимости мало – меньше 0,05, и можно считать, что коэффициент регрессии значим. И наоборот, если значение t -статистики мало, соответственно вероятность значимости больше 0,05 – коэффициент считается незначимым.

Для коэффициентов b_1 значения вероятности близко к нулю (0,02), следовательно, коэффициент b_1 можно считать значимым. Аналогично определяется значимость коэффициента b_2 .

Далее на рис. 2.5 представлены доверительные интервалы (нижняя и верхняя границы), рассчитываемые по формуле (2.15).

В разделе **Регрессионная статистика** получили:

- коэффициент детерминации $R^2 = 0,98$, рассчитывается по формуле (2.6), и показывает, что модель достаточно хорошо описывает данные, так как R^2 близок к 1;
- скорректированный коэффициент детерминации имеет тот же смысл, что и R^2 , но считается, что он точнее отражает степень адекватности модели;
- стандартная ошибка, рассчитываемая по формуле (2.9).

В **Дисперсионном анализе** вычисляются:

- df – число степеней свободы;
- SS – суммы квадратов разностей;
- MS – оценки дисперсий;
- F – вычисленное значение критерия Фишера;
- *Значимость F .*

Сумма квадратов регрессии вычисляется по формуле (2.12):

$$Q_{\text{рег}} = \sum (\hat{y}_i - \bar{y})^2 = SS_1;$$

- сумма квадратов остатков:

$$Q_{\text{ост}} = \sum (y_i - \hat{y}_i)^2 = SS_2;$$

- общая сумма квадратов:

$$Q_{общ} = \sum (y_i - \bar{y})^2 = SS.$$

Выполняется условие (2.13):

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2 = SS_1 + SS_2 = SS.$$

То есть $2353,32 + 355,62 = 2706,95$.

Число степеней свободы df для SS_1 равно $df_1 = 2$ (k - число независимых переменных или факторов), для SS_2 : $df_2 = n - k - 1 = 20 - 2 - 1 = 17$, для SS : $df = n - 1 = 20 - 1 = 19$.

Получены оценки средних квадратов:

$$MS_1 = \frac{SS_1}{df_1} = \frac{2353,35}{2} = 1176,66;$$

$$MS_2 = \frac{SS_2}{df_2} = \frac{355,62}{17} = 20,919;$$

наблюдаемое значение F -критерия (2.14):

$$F_{набл} = \frac{SS_1 / df_1}{SS_2 / df_2} = \frac{1176,66}{20,919} = 56,25.$$

Сравним полученное значение $F_{набл}$ с критическим. Так как $F_{крит} = 3,59 < F_{набл} = 56,25$, то гипотеза $H_0: \beta_1 = \beta_2 = 0$ отвергается и уравнение считается значимым.

Значимость F – это вероятность значимости для F критерия. В нашем случае она фактически равна нулю ($3,19 \cdot 10^{-8} = 0$), то есть гипотеза $H_0: \beta_1 = \beta_2 = 0$ отвергается и уравнение считается значимым.

3. Эконометрические модели с переменной структурой. Фиктивные переменные

Ранее рассматривались регрессионные модели с постоянными значениями коэффициентов на множестве исходных данных. Их также называют *эконометрическими моделями с постоянной структурой*. При построении таких моделей предполагается, что взаимосвязи между зависимой и независимыми переменными постоянны и не подвержены изменениям ни во времени, ни в пространстве.

Однако, кроме переменных, включенных в модель, на уровень анализируемого показателя обычно влияет большое число сопутствующих факторов. Обеспечить постоянство внешних условий для всех объектов, вошедших в выборку, удастся достаточно редко.

При анализе реальных социально-экономических процессов часто приходится констатировать, что со временем под влиянием вновь появившихся условий, факторов и масштаба, взаимосвязи между качественными и количественными, зависимыми и независимыми факторами меняются. В таких случаях модели с постоянной структурой

являются недостаточно точными для объяснения закономерностей меняющихся явлений и для их анализа прибегают к построению так называемых *моделей с переменной структурой*. Они позволяют учесть изменения значений коэффициентов в рамках одной эконометрической модели.

Если в ходе сбора статистических данных имеет место косвенное воздействие на них во времени или в пространстве некоторых качественных факторов - *сопутствующих переменных*, то линейные модели с переменной структурой учитывают скачкообразные сдвиги в значениях коэффициентов регрессии.

Использование моделей с переменной структурой позволяет решить ряд проблем, вызванных неполной спецификацией эконометрической модели: невключение по разным причинам в уравнение регрессии существенных факторов (из-за отсутствия исходных данных, слабости содержательного обоснования и т.д.), неверной функциональной формы модели.

Чаще всего предполагают, что сопутствующие переменные носят качественный характер. Например, на потребление населением газированной воды оказывает влияние качественная «сопутствующая переменная» – сезонность. За этой качественной переменной скрывается целый комплекс факторов: изменение длины светового дня, среднемесячной температуры, распределение праздничных дней по месяцам и т.д. При анализе расходов на продукты питания, очевидно, что существуют различия в образе жизни городского и сельского населения, которые можно попытаться учесть с помощью переменной «тип поселения»: областной центр, город, село.

Для того, чтобы учесть влияние сопутствующих переменных на структуру эконометрической модели, в регрессионную модель вводят *фиктивные переменные* или *дамми-переменные* (*dummy-variables*).

Под *фиктивной переменной* или *дамми-переменной* понимают переменную, которая равна единице для конкретной части выборочной совокупности, и нулю - для оставшейся части

Задача 3

Допустим, что исследователю необходимо изучить тест, который предсказывает будущую производительность труда работников. Восемь женщин и семь мужчин выполнили предусмотренные тестом задания. После чего была оценена производительность индексом, принимающим значение от 0 до 10. Данные приведены в таблице 3.1.

Таблица 3.1

Данные по производительности работников

Оценка производительности	Данные теста способностей	Пол 0 – женский; 1 - мужской
6	65	0
3	50	0
5	42	0
12	99	0
3	31	0
8	85	0
5	67	0
10	91	0
8	98	1
1	40	1
9	99	1
6	90	1
8	93	1
4	72	1
7	80	1

Уравнение регрессии будем искать в виде:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2, \quad (3.1)$$

где x_1 – переменная теста способностей;

x_2 – фиктивная переменная «пол»:

$$x_{2i} = \begin{cases} 0, & \text{если } i \text{ индивид женщина;} \\ 1, & \text{если } i \text{ индивид мужчина.} \end{cases}$$

Уравнение (3.1) эквивалентно следующим уравнениям:

$$\hat{y} = b_0 + b_1 x_1 + 0 \cdot x_2 = b_0 + b_1 x_1 \quad - \text{ для женщин;}$$

$$\hat{y} = b_0 + b_1 x_1 + b_2 \cdot 1 = (b_0 + b_2) + b_1 x_1 \quad - \text{ для мужчин.}$$

Коэффициент b_2 представляет влияние мужского пола на производительность труда, b_1 - влияние разницы в результатах теста способностей. Переменную x_2 можно рассматривать как переключающую переменную, которая включена, когда рассматриваются данные для мужчины и выключена, когда данные относятся к женщине.

Введем данные в таблицу Excel и построим уравнение с помощью Пакета анализа. Результаты представлены на рис. 3.1.

Регрессионная статистика						
Множественный R	0.93635					
R-квадрат	0.87674					
Нормированный R-квадрат	0.8562					
Стандартная ошибка	1.11636					
Наблюдения	15					

Дисперсионный анализ						
	df	SS	MS	F	значимость F	
Регрессия	2	106.378	53.1891	42.678698	3.506E-06	
Остаток	12	14.9552	1.24627			
Итого	14	121.333				

	Коэффициент	стандартная ошибка	t-статистика	P-Значение	верхние 95%	нижние 95%	верхние 95.0%	нижние 95.0%
Y-пересечение	-1.81	0.98409	-1.8392	0.0907398	-3.954144	0.33417	-3.9541	0.33417
Переменная X 1	0.12543	0.01361	9.2182	8.567E-07	0.0957863	0.15508	0.09579	0.15508
Переменная X 2	-2.2969	0.6149	-3.7354	0.0028458	-3.636636	-0.9571	-3.6366	-0.9571

Рис. 3.1. Вывод итогов

Как видно из рис. 3.1 получено уравнение регрессии:

$$\hat{y} = -1,81 + 0,125x_1 - 2,3x_2. \quad (3.2)$$

Коэффициент при x_1 показывает, что при увеличении оценки теста на 1 балл, производительность труда увеличивается в среднем на 0,125 оценки.

Коэффициент при переменной x_2 применим только для мужчин. Он свидетельствует, что оценка производительности труда для мужчин на 2,3 балла ниже по сравнению с женщинами, при условии, что значение теста способностей у них одинаково.

Коэффициенты уравнения при переменных x_1 и x_2 значимы. Следовательно, на оценку производительности значения теста и пол опрашиваемого оказывают значимое влияние.

Можно записать уравнение отдельно для мужчин и женщин:

$$\hat{y} = -1,81 + 0,125x_1 \quad \text{для женщин};$$

$$\hat{y} = -1,81 + 0,125x_1 - 2,3 = -4,11 + 0,125x_1 \quad \text{для мужчин}.$$

Эти уравнения можно использовать для прогноза.

Рассчитаем производительность труда при прохождении теста на 70 баллов:

$$\hat{y} = -1,81 + 0,125 \cdot 70 = 6,94 \quad \text{для женщин};$$

$$\hat{y} = -4,11 + 0,125 \cdot 70 = 4,64 - \text{для мужчин.}$$

4. Тест Чоу

Пусть совокупность состоит из двух подвыборок. Допустим, что число наблюдений в первой подвыборке равно n_A , во второй n_B . У нас есть альтернатива: объединить подвыборки и оценивать одну объединенную регрессию или строить отдельные регрессии для каждой подвыборки.

Запишем уравнения регрессии для каждой из частей совокупности:

$$y_i = \beta_0^A + \beta_1^A x_{i1} + \beta_2^A x_{i2} + \dots + \beta_k^A x_{ik} + \varepsilon_i^A, \quad i = 1, \dots, n_A,$$

$$y_i = \beta_0^B + \beta_1^B x_{i1} + \beta_2^B x_{i2} + \dots + \beta_k^B x_{ik} + \varepsilon_i^B, \quad i = n_A + 1, \dots, n_A + n_B.$$

Если коэффициенты регрессии в обеих частях достаточно близки, то их можно считать регрессионно однородными и рассматривать не два отдельных уравнения, а одно общее уравнение, рассчитанное по совокупности в целом.

Проверяемая гипотеза имеет вид:

$$H_0 : \beta_0^A = \beta_0^B, \beta_1^A = \beta_1^B, \dots, \beta_k^A = \beta_k^B, D(\varepsilon_i^A) = D(\varepsilon_i^B) = \sigma^2.$$

Г. Чоу (Chow) предложил тест для проверки гипотезы H_0 . Рассчитываются суммы квадратов остатков для регрессий подвыборок $Q_{ост}^A, Q_{ост}^B$ и по объединенной выборке $Q_{ост}^P$.

Равенство между $Q_{ост}^P = Q_{ост}^A + Q_{ост}^B$ будет иметь место только при совпадении коэффициентов регрессии для объединенной регрессии и регрессий подвыборок. В общем случае при разделении выборки будет наблюдаться улучшение качества уравнения, что можно представить, как $Q_{ост}^P - Q_{ост}^A - Q_{ост}^B$. Это имеет свою цену: используются $(k+1)$ дополнительных степеней свободы, так как вместо $(k+1)$ параметров для одной объединенной регрессии мы теперь должны оценить в сумме $(2k+2)$ параметров (k — число объясняющих переменных, единица соответствует постоянному члену). После разделения выборки, однако, остается необъясненная сумма квадратов остатков $Q_{ост}^A + Q_{ост}^B$ и, кроме того $(n - 2k - 2)$ степеней свободы.

Для того, чтобы определить, является ли значимым улучшение качества уравнения после разделения выборки, используется F -статистика:

$$F = \frac{(Q_{ост}^P - Q_{ост}^A - Q_{ост}^B)/(k+1)}{(Q_{ост}^A + Q_{ост}^B)/(n - 2k - 2)}, \quad (4.1)$$

которая имеет распределение Фишера с $(k+1)$ и $(n - 2k - 2)$ степенями свободы.

Задача 4

По данным задачи 3.1 с помощью теста Чоу проверьте, что коэффициенты регрессионных уравнений для мужчин и женщин одинаковы.

Решение

Построим уравнение регрессии зависимости производительности труда от результатов теста отдельно для мужчин и женщин (рис. 4.1 и 4.2).

Microsoft Excel - пример2 для метод. xls

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

Arial Cyr 10 Ж К Ч

C54 = 10.8281117696867

	A	B	C	D	E	F	G	H	I	J
42	ВЫВОД ИТОГОВ									
43										
44	Регрессионная статистика									
45	Множественный R	0.92394								
46	R-квадрат	0.85367								
47	Нормированный R-квадрат	0.82929								
48	Стандартная ошибка	1.34339								
49	Наблюдения	8								
50										
51	Дисперсионный анализ									
52		df	SS	MS	F	значимость F				
53	Регрессия	1	63.1719	63.1719	35.004379	0.001038				
54	Остаток	6	10.8281	1.80469						
55	Итого	7	74							
56										
57		Коэффициент	стандартная ошибка	статистика t	P-Значение	Верхние 95%	нижние 95%	Верхние 95%	нижние 95%	Верхние 95.0%
58	Y-пересечение	-1.6901	1.46351	-1.1548	0.2920736	-5.271166	1.89098	-5.2712	1.89098	
59	Переменная X 1	0.12362	0.02089	5.91645	0.001038	0.0724959	0.17475	0.0725	0.17475	
60										
61										
62	ВЫВОД ИТОГОВ									

Лист1 / Диаграмма1 / Лист2 / Лист4 / Лист3 /

Действия Автофигуры

Готово NUM

Рис. 4.1. Вывод уравнения зависимости производительности труда от результатов теста для женщин

Как видно из рис. 4.1 значение Q остаточное для этого уравнения равно 10,8281.

Microsoft Excel - пример2 для метод.xls

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

Arial Cyr 10 Ж К Ч

C74 = 4.0920140798592

63										
64	Регрессионная статистика									
65	Множественный R	0.95534								
66	R-квадрат	0.91267								
67	Нормированный R-квадр	0.8952								
68	Стандартная ошибка	0.90466								
69	Наблюдения	7								
70										
71	Дисперсионный анализ									
72		df	SS	MS	F	начисность F				
73	Регрессия	1	42.7651	42.7651	52.254376	0.0007904				
74	Остаток	5	4.09201	0.8184						
75	Итого	6	46.8571							
76										
77	Коэффициент артная статистика P-Значение Нижние 95% Верхние 95% Нижние 95% Верхние 95%									
78	Y-пересечение	-4.3422	1.49023	-2.9138	0.0332532	-8.17296	-0.5115	-8.173	-0.5115	
79	Переменная X 1	0.12831	0.01775	7.22872	0.0007904	0.0826846	0.17394	0.08268	0.17394	
80										
81										
82										
83										

Лист1 / Диаграмма1 / Лист2 / Лист4 / Лист3 /

Действия Автофигуры

Готово NUM

Рис. 4.2. Вывод уравнения зависимости производительности труда от результатов теста для мужчин

Для мужчин значение Q остаточное равно 4,09.

Построим уравнение регрессии для всей выборки и мужчин и женщин (рис. 4.3).

Microsoft Excel - пример2 для метод.xls

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

Arial Cyr 10 Ж К Ч

C54 = 32.344546853685

43										
44	Регрессионная статистика									
45	Множественный R	0.8564								
46	R-квадрат	0.73342								
47	Нормированный R-квадр	0.71292								
48	Стандартная ошибка	1.57735								
49	Наблюдения	15								
50										
51	Дисперсионный анализ									
52		df	SS	MS	F	начисность F				
53	Регрессия	1	88.9888	88.9888	35.766592	4.589E-05				
54	Остаток	13	32.3445	2.48804						
55	Итого	14	121.333							
56										
57	Коэффициент артная статистика P-Значение Нижние 95% Верхние 95% Нижние 95% Верхние 95%									
58	Y-пересечение	-1.604	1.38828	-1.1554	0.2687295	-4.603187	1.3952	-4.6032	1.3952	
59	Переменная X 1	0.10804	0.01807	5.98052	4.589E-05	0.0690121	0.14707	0.06901	0.14707	
60										

Лист1 / Диаграмма1 / Лист2 / Лист4 / Лист3 /

Действия Автофигуры

Готово NUM

Рис. 4.2. Вывод уравнения зависимости производительности труда от результатов теста для всей выборки

Как видно Q остаточное для всей выборки равно 32,345.

Проверим гипотезу $H_0: \beta_0^A = \beta_0^B, \beta_1^A = \beta_1^B$, то есть коэффициенты при переменной «Результаты теста» одинаковы для мужчин и женщин.

Воспользуемся формулой (4.1).

$$F = \frac{(Q_{ост}^P - Q_{ост}^A - Q_{ост}^B)/(k+1)}{(Q_{ост}^A + Q_{ост}^B)/(n-2k-2)} = \frac{(32,345-10,83-4,092)/(1+1)}{(10,83+4,092)/(15-2-2)} = \frac{8,7115}{1,36} = 6,4.$$

Найдем $F_{крит}$ для числа степеней свободы $v_1=k+1=1+1=2$ и $v_2=n-2k-2=15-2-2=11$ по таблице распределения Фишера или используя функцию Excel ФРАСПОБР(0,05; 2; 11) = 3,98.

Так как $F_{набл}=6,4 > F_{крит}=3,98$, то гипотеза о равенстве коэффициентов регрессии отвергается.

5. Сравнение «длинной» и «короткой» регрессии

Иногда из большого числа независимых переменных требуется отобрать наиболее существенные объясняющие переменные, влияющие на зависимую переменную. Для этого используют тест проверки «длинной» и «короткой» регрессий.

Рассмотрим два уравнения регрессии:

$$(короткое) \quad y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + u_i;$$

$$(длинное) \quad y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \dots + \beta_{k+q} x_{ik+q} + v_i.$$

Какое из двух уравнений выбрать?

Проверим гипотезу $H_0: \beta_{k+1} = \beta_{k+2} = \dots = \beta_{k+q} = 0$.

Найдем сумму квадратов остатков для «длинной» (unrestricted) модели $Q_{ост}^{UR}$.

Найдем сумму квадратов остатков для «короткой» (restricted) модели $Q_{ост}^R$.

Рассчитаем значение $F_{набл}$:

$$F = \frac{(Q_{ост}^R - Q_{ост}^{UR})/q}{Q_{ост}^{UR}/(n-k-q-1)}. \quad (5.1)$$

Если $F_{набл} > F_{кр}(v_1=q, v_2=n-k-q-1)$, гипотеза отвергается (выбираем длинную регрессию), в противном случае – выбираем короткую регрессию.

Задача 5

По данным таблицы 4.1 изучается зависимость индекса человеческого развития Y от переменных:

$X1$ – ВВП 1997г., % к 1990г.;

$X2$ – суточная калорийность питания населения, ккал на душу населения;

$X3$ – ожидаемая продолжительность жизни при рождении 1997г., число лет;

$X4$ – расходы на конечное потребление в текущих ценах, % к ВВП;

$X5$ – расходы домашних хозяйств, % к ВВП;

$X6$ – валовое накопление, % к ВВП.

Таблица 5.1

Данные по странам

Страна	Y	$X1$	$X2$	$X3$	$X4$	$X5$	$X6$
Австрия	0.904	115	3343	77	75.5	56.1	25.2
Австралия	0.922	123	3001	78.2	78.5	61.8	21.8
Белоруссия	0.763	74	3101	68	78.5	59.1	25.7
Бельгия	0.923	111	3543	77.2	77.7	63.3	17.8
Великобритания	0.918	113	3237	77.2	84.4	64.1	15.9
Германия	0.906	110	3330	77.2	75.9	57	22.4
Дания	0.905	119	3808	75.7	76	50.7	20.6
Индия	0.545	146	2415	62.6	67.5	57.1	25.2
Испания	0.894	113	3295	78	78.2	62	20.7
Италия	0.9	108	3504	78.2	78.1	61.8	17.5
Канада	0.932	113	3056	79	78.6	58.6	19.7
Казахстан	0.74	71	3007	67.6	84	71.7	18.5
Китай	0.701	210	2844	69.8	59.2	48	42.4
Латвия	0.744	94	2861	68.4	90.2	63.9	23
Нидерланды	0.921	118	3259	77.9	72.8	59.1	20.2
Норвегия	0.927	130	3350	78.1	67.7	47.5	25.2
Польша	0.802	127	3344	72.5	82.6	65.3	22.4
Россия	0.747	61	2704	66.6	74.4	53.2	22.7
США	0.927	117	3642	76.7	83.3	67.9	18.1
Украина	0.721	46	2753	68.8	83.7	61.7	20.1
Финляндия	0.913	107	2916	76.8	73.8	52.9	17.36
Франция	0.918	110	3551	78.1	79.2	59.9	16.8
Чехия	0.833	99.2	3177	73.9	71.5	51.5	29
Швейцария	0.914	101	3280	78.6	75.3	61.2	20.3
Швеция	0.923	105	3160	78.5	79	53.1	14.1

Построить регрессию зависимости Y от всех независимых переменных.

Исключая незначимые переменные, определить, какую регрессию возможно применить для анализа и прогноза.

Решение

Построим длинную регрессию зависимости Y от всех переменных X . Результаты приведены на рис. 5.1.

Регрессионная статистика						
Множественный R	0.982					
R-квадрат	0.96433					
Нормированный R-квадрат	0.95244					
Стандартная ошибка	0.02209					
Наблюдения	25					

Дисперсионный анализ						
	df	SS	MS	F	Значимость F	
Регрессия	6	0.2374788	0.0395798	81.103888	4.81695E-12	
Остаток	18	0.0087842	0.00048801			
Итого	24	0.246263				

	Коэффициент	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%	Нижние 95%
Y-пересечение	-0.6423	0.1555882	-4.12799567	0.0006316	-0.96914631	-0.3153885	-0.96914631
Переменная X 1	-0.0005	0.0002297	-2.08850177	0.0512303	-0.00096214	2.853E-06	-0.00096214
Переменная X 2	4.3E-05	2.043E-05	2.11029345	0.0490895	1.91439E-07	8.604E-05	1.91439E-07
Переменная X 3	0.01858	0.0015965	11.6387351	8.264E-10	0.015223661	0.0219318	0.015223661
Переменная X 4	0.00133	0.0014612	0.91134356	0.3741584	-0.0017382	0.0044015	-0.0017382
Переменная X 5	-0.0014	0.0012336	-1.13754504	0.2702268	-0.00399486	0.0011884	-0.00399486
Переменная X 6	0.00022	0.0014205	0.15342937	0.8797667	-0.00276643	0.0032023	-0.00276643

Рис. 5.1. Результаты «длинной» регрессии

Получили следующее уравнение регрессии:

$$\hat{y} = -0,6423 - 0,0005x_1 + 0,000043x_2 + 0,01858x_3 + 0,00133x_4 - 0,0014x_5 + 0,00022x_6.$$

По значениям t -статистики или P -значениям видно, что значимое влияние на зависимую переменную Y оказывают переменные $X1$, $X2$, $X3$ и незначимое – переменные $X4$, $X5$, $X6$.

Исключим незначимую переменную, для которой значение вероятности значимости наибольшая, то есть переменную $X6$.

Результат представлен на рис. 5.2.

	A	B	C	D	E	F	G	H	I	J
1	ВЫВОД ИТОГОВ									
2										
3	<i>Регрессионная статистика</i>									
4	Множественный R	0.981979216								
5	R-квадрат	0.96428318								
6	Нормированный R-кв	0.954884017								
7	Стандартная ошибка	0.021515885								
8	Наблюдения	25								
9										
10	<i>Дисперсионный анализ</i>									
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>значимость F</i>				
12	Регрессия	5	0.23747	0.04749	102.592	4.5E-13				
13	Остаток	19	0.0088	0.00046						
14	Итого	24	0.24626							
15										
16		<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95.0%</i>	<i>Верхние 95.0%</i>	
17	Y-пересечение	-0.62430192	0.09979	-6.2563	5.2E-06	-0.8332	-0.4154	-0.8332	-0.4154	
18	Переменная X 1	-0.00046413	0.0002	-2.3112	0.03221	-0.0009	-4E-05	-0.0009	-4E-05	
19	Переменная X 2	4.33835E-05	2E-05	2.18808	0.04136	1.9E-06	8.5E-05	1.9E-06	8.5E-05	
20	Переменная X 3	0.01844748	0.00132	14.0087	1.8E-11	0.01569	0.0212	0.01569	0.0212	
21	Переменная X 4	0.001278985	0.00138	0.92457	0.36678	-0.0016	0.00417	-0.0016	0.00417	
22	Переменная X 5	-0.00143789	0.00118	-1.2174	0.23836	-0.0039	0.00103	-0.0039	0.00103	
23										

Рис. 5.2. Регрессия без переменной X6

Как видно из рисунка 5.2 по сравнению с рисунком 5.1 значения коэффициентов детерминации практически не изменились. Значения коэффициентов регрессии изменилось незначительно, также незначимыми переменными остались X4, X5. Последовательно исключим эти незначимые переменные (проделайте это самостоятельно).

Получили «короткую» регрессию, включающую только переменные X1, X2, X3 (рис. 5.3).

Microsoft Excel - решение в exl.xls

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

Arial Cyr 10 Ж К Ч

B74 = ((C65-C40)/3)/(C40/(25-6-1))

	A	B	C	D	E	F	G
53	ВЫВОД ИТОГОВ						
54							
55	Регрессионная статистика						
56	Множественный R	0.980558547					
57	R-квадрат	0.961495065					
58	Нормированный R-квадрат	0.95599436					
59	Стандартная ошибка	0.021249473					
60	Наблюдения	25					
61							
62	Дисперсионный анализ						
63		df	SS	MS	F	Значимость F	
64	Регрессия	3	0.2367807	0.0789269	174.79488	5.22679E-15	
65	Остаток	21	0.0094823	0.00045154			
66	Итого	24	0.246263				
67							
68		Коэффициент	Стандарт	t-	P-	Верхние	Верхние
		ы	ная	статистика	Значение	95%	95%
69	Y-пересечение	-0.621146936	0.067386	-9.21773821	7.908E-09	-0.76128392	-0.48101
70	Переменная X 1	-0.000535121	0.0001447	-3.69690608	0.0013379	-0.00083614	-0.0002341
71	Переменная X 2	4.26372E-05	1.902E-05	2.2422267	0.035871	3.09216E-06	8.218E-05
72	Переменная X 3	0.018730357	0.0012796	14.6373317	1.723E-12	0.016069222	0.0213915
73							
74	F набл	0.476829531					
75	Fкрит	3.159911444					
76							

Бреуш-паган / множ. регр / Чоу / Гольфельд-ке

Готово NUM

Рис. 5.3. Результаты «короткой» регрессии

Полученная регрессия имеет вид:

$$\hat{y} = -0,6211 - 0,0005x_1 + 0,000043x_2 + 0,0187x_3.$$

Как видно по P -значениям, все коэффициенты значимы.

Определим, какая из регрессий предпочтительнее для анализа.

Рассчитаем F -статистику по формуле (5.1):

$$F = \frac{(Q_{оцн}^R - Q_{оцн}^{UR})/q}{Q_{оцн}^{UR}/(n-k-q-1)} = \frac{(0,0094823 - 0,0087842)/3}{0,0087842/(25-6-3-1)} = 0,47.$$

Так как $F_{набл} = 0,47 < F_{крит} = 3,16$, то гипотеза $H_0: \beta_4 = \beta_5 = \beta_6 = 0$ о не влиянии на индекс человеческого развития Y переменных X_4 (расходы на конечное потребление в текущих ценах), X_5 (расходы домашних хозяйств) и X_6 (валовое накопление) принимается и для дальнейшего анализа можно использовать «короткую» регрессию.

6. Гетероскедастичность

В рассмотренной ранее регрессионной модели, которую часто называют *классической*, предполагается, что случайные составляющие e_i имеют постоянную дисперсию и не коррелируют друг с другом, то есть

ковариационная матрица случайного вектора имеет вид $v(e) = s^2 I_n$. Это условие известно как *гомоскедастичность*, что означает «одинаковый разброс» (рис. 6.1).

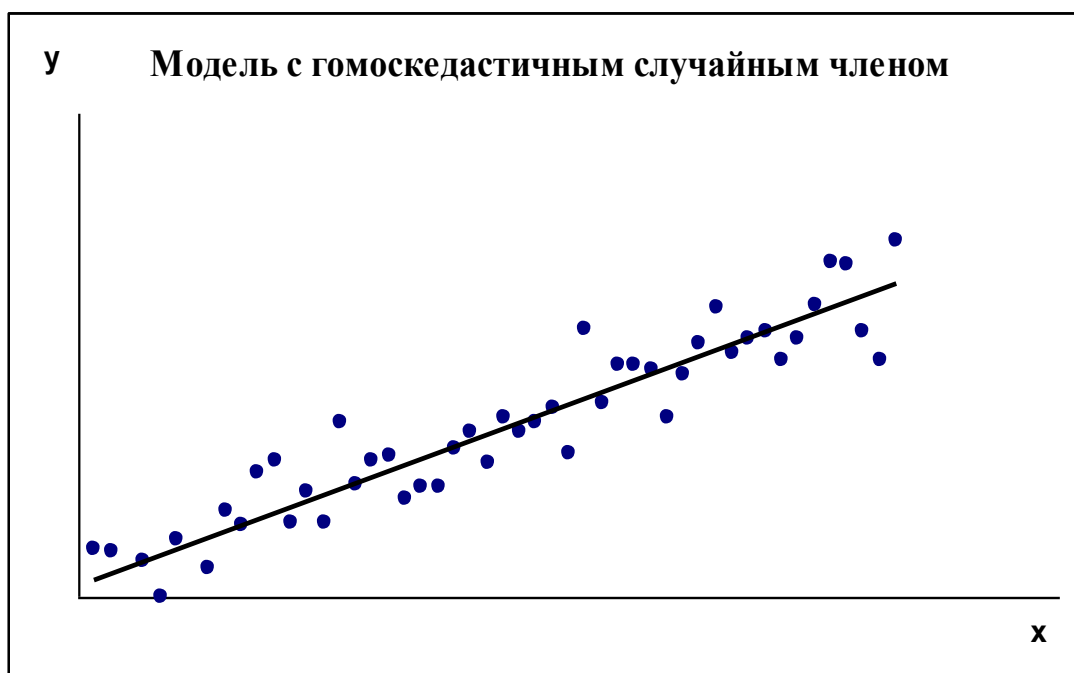


Рис. 6.1. Модель с гомоскедастичным случайным членом

Если ошибки не являются гомоскедастичными, то имеет место *гетероскедастичность*, что означает «неодинаковый разброс».

Гетероскедастичность может принимать разные формы, например, *непостоянство дисперсии или автокорреляция*.

На рисунке 6.2 представлен пример, когда ошибки не коррелированы, но имеют неодинаковую дисперсию.



Рис. 6.2. Модель с гетероскедастичным случайным членом

При гомоскедастичности коэффициенты регрессии, полученные методом наименьших квадратов (МНК), имеют наименьшую дисперсию среди всех несмещенных оценок, являющихся линейными функциями от наблюдений y . Если имеет место гетероскедастичность, то:

1. оценки МНК, которые мы использовали до сих пор, неэффективны. Можно найти другие оценки, которые имеют меньшую дисперсию и являются несмещенными;
2. стандартные ошибки, рассчитанные по обычной формуле (МНК), будут не верны. Они вычисляются на основе предположения, что распределение случайного члена гомоскедастично. Вполне вероятно, что стандартные ошибки будут занижены, а, следовательно, t -статистики – завышены и будет получено неправильное представление о точности уравнения регрессии.

6.1. Тесты на гетероскедастичность

Тесты на гетероскедастичность предназначены для ситуации, когда ошибки не коррелированы, но дисперсия ошибок не постоянна.

Проверяется основная гипотеза $H_0: \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2$ (модель гомоскедастична) против альтернативной гипотезы $H_1: \sigma_1^2 \neq \sigma_2^2 \neq \dots \neq \sigma_n^2$ (модель гетероскедастична).

Наиболее часто используют:

1. Тест Голдфелда-Квандта (Goldfeld-Quandt).
2. Тест Бреуша-Пагана (Breus-Pagan).
3. Тест Вайта (White).

6.1.1. Тест Голдфелда-Квандта (Goldfeld-Quandt)

Предполагается, что стандартные отклонения ошибки пропорционально значениям одной из независимых переменных.

Этапы тестирования:

1. Упорядочивают наблюдения по величине x .
2. Выборку разбивают на три части.
3. Рассчитывают регрессию для первой трети выборки, находят $Q_{ост}^I$.
4. Рассчитывают регрессию для последней трети выборки, находят $Q_{ост}^{II}$.
5. Находят

$$F_{набл} = Q_{ост}^{II} / Q_{ост}^I. \quad (6.1)$$

6. Если $F_{набл} > F_{крит}(\alpha; \nu_1 = \frac{n-d}{2} - k - 1; \nu_2 = \frac{n-d}{2} - k - 1)$, то имеет место гетероскедастичность.

Необходимо учитывать, что если в модели более одной объясняющей переменной, то число наблюдений $\frac{n-d}{2}$ должно быть больше, чем $k+1$, где k - число объясняющих переменных.

Пример 6.1

По данным таблицы 5.1 проверить гипотезу о гомоскедастичности, используя тест Гольфрейда-Квандта.

Решение

Упорядочим выборку по той переменной, по которой есть подозрение на гетероскедастичность, например, по x_1 . Для этого необходимо выделить весь массив переменной, в командной строке курсором выбрать «Данные», затем «Сортировка». Появится окно «Сортировка диапазона» (рис.6.3). Необходимо отметить «Сортировать по возрастанию», нажать ОК. В результате ваши данные отсортируются по возрастанию данных переменной x_1 .

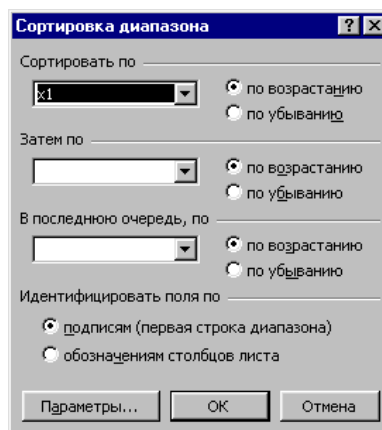


Рис. 6.3. Окно «Сортировка диапазона»

Разобьем 25 наблюдений приблизительно на 3 части.

Построим регрессию для первых 9 наблюдений (рис. 6.4) и для последних 9 переменных (рис.6.5). Для каждой регрессии найдем $Q_{ост}$.

Microsoft Excel - решение в exl.xls

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

Arial Cyr 10 Ж К Ч

C41 = 0.000392718099408069

	A	B	C	D	E	F	G
31	Регрессионная статистика						
32	Множественный R	0.996612896					
33	R-квадрат	0.993237264					
34	Нормированный R-	0.972949055					
35	Стандартная ошибка	0.014012817					
36	Наблюдения	9					
37							
38	Дисперсионный анализ						
39		df	SS	MS	F	Значимость F	
40	Регрессия	6	0.057678	0.009613028	48.95638108	0.020151314	
41	Остаток	2	0.000393	0.000196359			
42	Итого	8	0.058071				
43							
44		Кoeffициент	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	верхние 95%
45	Y-пересечение	0.426410144	0.237428	1.795952252	0.21434267	-0.595162698	1.447983
46	Переменная X 1	0.001111489	0.00043	2.584151343	0.122772634	-0.000739159	0.002962
47	Переменная X 2	3.0579E-05	4.95E-05	0.617847477	0.599654905	-0.000182371	0.000244
48	Переменная X 3	0.008109291	0.002896	2.799852729	0.107402145	-0.00435261	0.020571
49	Переменная X 4	-0.003180229	0.001479	-2.150082152	0.164526868	-0.009544373	0.003184
50	Переменная X 5	-0.000703988	0.001457	-0.483021037	0.676784926	-0.00697497	0.005567
51	Переменная X 6	-0.004319471	0.001829	-2.36201111	0.142027796	-0.01218785	0.003549

Готово

Рис. 6.4. Регрессия по первым 9 наблюдениям

Microsoft Excel - решение в exl.xls

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

Arial Cyr 10 Ж К Ч

C65 = 0.000121165282029202

	A	B	C	D	E	F	G	H
52	ВЫВОД ИТОГОВ							
53								
54								
55	Регрессионная статистика							
56	Множественный R	0.999580742						
57	R-квадрат	0.99916166						
58	Нормированный R-	0.99664664						
59	Стандартная ошибка	0.007783485						
60	Наблюдения	9						
61								
62	Дисперсионный анализ							
63		df	SS	MS	F	Значимость F		
64	Регрессия	6	0.144409	0.024068139	397.2778129	0.002512912		
65	Остаток	2	0.000121	6.05826E-05				
66	Итого	8	0.14453					
67								
68		Кoeffициент	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	верхние 95%	нижние 95%
69	Y-пересечение	-0.93944209	0.089253	-10.52555418	0.008905905	-1.323468984	-0.55542	-1.32
70	Переменная X 1	0.000175573	0.000421	0.418973254	0.717191965	-0.001636127	0.001987	-0.00
71	Переменная X 2	4.47173E-05	1.56E-05	2.867127501	0.103164827	-2.23893E-05	0.000112	-2.2E
72	Переменная X 3	0.022148257	0.000918	24.12039807	0.001714404	0.018197397	0.026099	0.018
73	Переменная X 4	0.000353631	0.001521	0.232532271	0.837753446	-0.006189778	0.006897	-0.00
74	Переменная X 5	-0.000252503	0.001193	-0.211571085	0.852043211	-0.005387589	0.004883	-0.00
75	Переменная X 6	-0.00185708	0.001623	-1.143875866	0.37112189	-0.008842434	0.005128	-0.00

Готово

Рис. 6.5. Регрессия по последним 9 наблюдениям

Найдем статистику (6.1):

$$F_{набл} = \frac{Q_{ост}^{II}}{Q_{ост}^I} = \frac{0.000121}{0.0004} = 0.3.$$

Так как

$$\begin{aligned} F_{набл} = 0,3 < F_{крит}(\alpha; \nu_1 = \frac{n-d}{2} - k - 1; \nu_2 = \frac{n-d}{2} - k - 1) = \\ = F_{крит}(0.05; \nu_1 = \frac{25-7}{2} - 6 - 1; \nu_2 = \frac{25-7}{2} - 6 - 1) = 19, \end{aligned}$$

то гипотеза о гомоскедастичности не отвергается.

Следует заметить, что переменная X_I гомоскедастична, но это не значит, что по всем остальным переменным модель может быть гетероскедастичной. Поэтому необходима дальнейшая проверка по остальным переменным.

6.1.2. Тест Бреуша-Пагана (Breus-Pagan)

Предполагается, что дисперсия случайной ошибки зависит от нескольких независимых переменных.

$$s_i^2 = \gamma_0 + \gamma_1 z_{i1} + \gamma_2 z_{i2} + \dots + \gamma_m z_{im}.$$

Этапы тестирования:

1. Рассчитывают МНК-оценки коэффициентов регрессии.
2. Находят остатки e_i .
3. Находят квадраты остатков e_i^2 .
4. Рассчитывают коэффициент детерминации R^2 для регрессии

$$e_i^2 = \gamma_0 + \gamma_1 z_{i1} + \gamma_2 z_{i2} + \dots + \gamma_m z_{im}.$$
5. Вычисляют $X_{набл}^2 = n R^2$. (6.2)
6. Если $X_{набл}^2$ превосходит критическое значение статистики Хи-квадрат для m степеней свободы, гетероскедастичность присутствует.

Пример 6.2

По данным таблицы 5.1 проверить гипотезу о гетероскедастичности, используя тест Бреуша-Пагана.

Решение

Рассчитаем регрессию по всем шести переменным, в результате получим регрессию (рис. 5.2). При построении регрессии необходимо вывести остатки e_i , для этого следует поставить флажок «Остатки» в параметрах **Регрессии** (рис. 2.3).

В результате получим таблицу остатков. Найдем квадраты остатков (рис. 6.6).

The screenshot shows an Excel window titled "Microsoft Excel - решение в exl.xls". The active sheet is "Лист1". The table contains the following data:

	A	B	C	D	E	F	G
54	ВЫВОД ОСТАТКА						
55							
56	наблюдения	предсказанно	Остатки	e^2			
57	1	0.747249	-0.02625	0.000689			
58	2	0.69815	0.04885	0.002386			
59	3	0.716943	0.023057	0.000532			
60	4	0.733329	0.029671	0.00088			
61	5	0.754018	-0.01002	0.0001			
62	6	0.833932	-0.00093	8.69E-07			
63	7	0.918069	-0.00407	1.66E-05			
64	8	0.946146	-0.02315	0.000536			
65	9	0.889514	0.023486	0.000552			
66	10	0.929458	-0.02946	0.000868			
67	11	0.904401	0.001599	2.56E-06			
68	12	0.937583	-0.01958	0.000383			
69	13	0.908875	0.014125	0.0002			
70	14	0.913761	-0.01976	0.00039			
71	15	0.91457	0.00343	1.18E-05			
72	16	0.931167	0.000833	6.95E-07			
73	17	0.899336	0.004664	2.17E-05			
74	18	0.908598	0.018402	0.000339			
75	19	0.903113	0.017887	0.00032			
76	20	0.91409	-0.00909	8.26E-05			
77	21	0.905275	0.016725	0.00028			
78	22	0.822441	-0.02044	0.000418			
79	23	0.917123	0.009877	9.76E-05			
80	24	0.586709	-0.04171	0.00174			

Рис. 6.6. Вывод остатка регрессии

Затем строим регрессию, в которой за зависимую переменную берется столбец квадратов остатков e_i^2 , а за независимые переменные – переменные $X_1, X_2, X_3, X_4, X_5, X_6$.

Результат представлен на рис. 6.7.

Microsoft Excel - решение в exl.xls

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

Arial Cyr 10 Ж К Ч

E88 =ХИ2ОБР(0.05;6)

	A	B	C	D	E	F
82						
83	ВЫВОД ИТОГОВ					
84						
85	Регрессионная статистика					
86	Множеств	0.828273				
87	R-квадрат	0.686036		хи-квадрат набл	17.1509	
88	Нормиров	0.603414		Хи-квадрат крит	12.59158	
89	Стандартн	0.000356				
90	Наблюден	25				
91						
92	Дисперсионный анализ					
93		df	SS	MS	F	значимость F
94	Регрессия	5	5.27E-06	1.05307E-06	8.303301	0.000267
95	Остаток	19	2.41E-06	1.26826E-07		

Бреуш-паган / множ. регр / Лист1 / Ч

Готово NUM

Рис. 6.7. Вывод остатка регрессии

Найдена статистика (6.2): $X^2_{набл} = nR^2 = 25 \cdot 0,68 = 17,15$.

Так как $X^2_{набл} = 17,15 > X^2_{крит} = 12,59$, то гипотеза о гомоскедастичности отвергается и модель считается гетероскедастичной.

Критическое значение распределения хи-квадрат найдено с помощью действий: $f_x \rightarrow$ Статистические \rightarrow ХИ2ОБР(m), где m – число переменных, входящих в уравнение регрессии (в данном случае 6).

6.1.3. Тест Вайта (White)

Этот тест аналогичен тесту Бреуша-Пагана. В качестве независимых переменных используются все регрессоры, их квадраты и попарные произведения.

$$s_i^2 = \gamma_0 + \gamma_1 x_{i1} + \gamma_2 x_{i2} + \dots + \gamma_k x_{ik} + \gamma_{k+1} x_{i1} x_{i2} + \gamma_{k+2} x_{i1} x_{i3} + \dots + \gamma_m x_{ik}^2.$$

Этапы тестирования:

1. Рассчитывают МНК-оценки коэффициентов регрессии.
2. Находят остатки e_i .
3. Находят квадраты остатков e_i^2 .
4. Находят оценку остаточной дисперсии \hat{s}^2 .
5. Рассчитывают R^2 для регрессии

$$e_i^2 = \gamma_0 + \gamma_1 x_{i1} + \gamma_2 x_{i2} + \dots + \gamma_k x_{ik} + \gamma_{k+1} x_{i1} x_{i2} + \gamma_{k+2} x_{i1} x_{i3} + \dots + \gamma_m x_{ik}^2.$$

6. Вычисляют $X^2_{набл} = nR^2$. (6.3)

7. Если $X^2_{набл}$ превосходит критическое значение статистики Хи-квадрат для m степеней свободы, то гетероскедастичность присутствует.

Пример 6.3

По данным таблицы 5.1 проверить гипотезу о гетероскедастичности, используя тест Вайта.

Решение

Так как число переменных, входящих в уравнение не может быть больше 16, оставим только первые четыре переменных. В таблице построим данные, соответствующие квадратам переменных и их перекрестным произведениям (рис. 6.8).

	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	x1	x2	x3	x4	x1x2	x1x3	x1x4	x2x3	x2x4	x3x4	x1 ²	x2 ²	x3 ²	x4 ²
1														
2	115	3343	77	75.5	384445	8855	8682.5	257411	252396.5	5813.5	13225	11175649	5929	5700.2
3	123	3001	78.2	78.5	369123	9618.6	9655.5	234678.2	235578.5	6138.7	15129	9006001	6115.24	6162.2
4	74	3101	68	78.5	229474	5032	5809	210868	243428.5	5338	5476	9616201	4624	6162.2
5	111	3543	77.2	77.7	393273	8569.2	8624.7	273519.6	275291.1	5998.44	12321	12552849	5959.84	6037.2
6	113	3237	77.2	84.4	365781	8723.6	9537.2	249896.4	273202.8	6515.68	12769	10478169	5959.84	7123.3
7	110	3330	77.2	75.9	366300	8492	8349	257076	252747	5859.48	12100	11088900	5959.84	5760.8
8	119	3808	75.7	76	453152	9008.3	9044	288265.6	289408	5753.2	14161	14500864	5730.49	577
9	146	2415	62.6	67.5	352590	9139.6	9855	151179	163012.5	4225.5	21316	5832225	3918.76	4556.2
10	113	3295	78	78.2	372335	8814	8836.6	257010	257669	6099.6	12769	10857025	6084	6115.2
11	108	3504	78.2	78.1	378432	8445.6	8434.8	274012.8	273662.4	6107.42	11664	12278016	6115.24	6099.6
12	113	3056	79	78.6	345328	8927	8881.8	241424	240201.6	6209.4	12769	9339136	6241	6177.9
13	71	3007	67.6	84	213497	4799.6	5964	203273.2	252588	5678.4	5041	9042049	4569.76	705
14	210	2844	69.8	59.2	597240	14658	12432	198511.2	168364.8	4132.16	44100	8088336	4872.04	3504.6
15	94	2861	68.4	90.2	268934	6429.6	8478.8	195692.4	258062.2	6169.68	8836	8185321	4678.56	8136.0
16	118	3259	77.9	72.8	384562	9192.2	8590.4	253876.1	237255.2	5671.12	13924	10621081	6068.41	5299.8
17	130	3350	78.1	67.7	435500	10153	8801	261635	226795	5287.37	16900	11222500	6099.61	4583.2
18	127	3344	72.5	82.6	424688	9207.5	10490.2	242440	276214.4	5988.5	16129	11182336	5256.25	6822.7
19	61	2704	66.6	74.4	164944	4062.6	4538.4	180086.4	201177.6	4955.04	3721	7311616	4435.56	5535.3
20	117	3642	76.7	83.3	426114	8973.9	9746.1	279341.4	303378.6	6389.11	13689	13264164	5882.89	6938.8
21	46	2753	68.8	83.7	126638	3164.8	3850.2	189406.4	230426.1	5758.56	2116	7579009	4733.44	7005.6
22	107	2916	76.8	73.8	312012	8217.6	7896.6	223948.8	215200.8	5667.84	11449	8503056	5898.24	5446.4
23	110	3551	78.1	79.2	390610	8591	8712	277333.1	281239.2	6185.52	12100	12609601	6099.61	6272.6
24	99.2	3177	73.9	71.5	315158.4	7330.88	7092.8	234780.3	227155.5	5283.85	9840.64	10093329	5461.21	5112.2
25	101	3280	78.6	75.3	331280	7938.6	7605.3	257808	246984	5918.58	10201	10758400	6177.96	5670.0
26	105	3160	78.5	79	331800	8242.5	8295	248060	249640	6201.5	11025	9985600	6162.25	624
27														

Рис. 6.8. Данные для построения регрессии

Построим уравнение регрессии для этих переменных. Получим коэффициент детерминации, равный 0,98 (рис. 6.9).

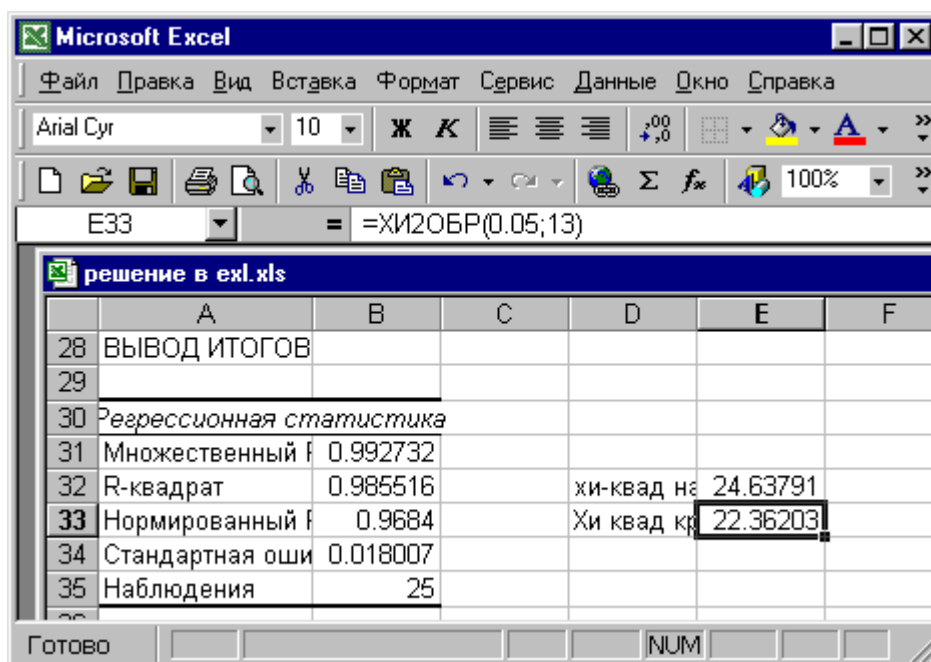


Рис. 6.9. Расчет теста Вайта

Рассчитаем статистику по формуле (6.3).

Так как $X^2_{набл} = nR^2 = 24,64 > X^2_{крит} = 22,36$ для числа степеней свободы, равного 13, то гипотеза о гомоскедастичности отвергается и модель можно считать гетероскедастичной.

6.2. Коррекция на гетероскедастичность

Рассмотрим модель:

$$y_i = \beta_0 + \beta_1 x_i + u_i.$$

Пусть ошибки u_i не коррелированы, но характеризуются разной дисперсией:

$$D(u_i) = \sigma_i^2,$$

то есть модель гетероскедастична.

Матрица ковариаций W вектора ошибок u имеет вид:

$$\Omega = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix}.$$

Формула обобщенного метода наименьших квадратов

$$b = (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} Y$$

сводится к **взвешенному методу наименьших квадратов**, так как матрица Ω^{-1} имеет простой вид:

$$\Omega^{-1} = \begin{pmatrix} 1/\sigma_1^2 & 0 & \dots & 0 \\ 0 & 1/\sigma_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1/\sigma_n^2 \end{pmatrix}.$$

Если s_i^2 — известны, то получить модель с гомоскедастичными остатками можно, используя в качестве весов наблюдений величины $1/\sigma_i$.

Разделим каждую строку матрицы данных на σ_i :

$$\frac{y_i}{\sigma_i} = \beta_0 \frac{1}{\sigma_i} + \beta_1 \frac{x_i}{\sigma_i} + \frac{u_i}{\sigma_i}. \quad (6.4)$$

В результате получим преобразованную модель:

$$y'_i = \beta_0 \frac{1}{\sigma_i} + \beta_1 x'_i + \varepsilon_i. \quad (6.5)$$

Найдем дисперсию $D(\varepsilon_i)$:

$$D(\varepsilon_i) = D\left(\frac{u_i}{\sigma_i}\right) = \frac{D(u_i)}{\sigma_i^2} = \frac{\sigma_i^2}{\sigma_i^2} = 1.$$

В итоге получим пересмотренную модель:

$$y'_i = \beta_0 h_i + \beta_1 x'_i + \varepsilon_i, \quad (6.6)$$

где $y'_i = \frac{y_i}{\sigma_i}$, $h_i = \frac{1}{\sigma_i}$, $x'_i = \frac{x_i}{\sigma_i}$, $\varepsilon_i = \frac{u_i}{\sigma_i}$.

Заметим, что модель (6.6) не включает свободный член, b_i — коэффициент регрессии при новой переменной h .

На практике дисперсии ошибок почти никогда не известны.

Однако, иногда можно предположить, что σ_i^2 пропорциональны некоторой переменной Z_i .

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

$$D u_i = \sigma_i^2$$

$$\sigma_i = \lambda Z_i.$$

Тогда в качестве весов наблюдений следует использовать величину $1/z_i$:

$$\frac{y_i}{z_i} = \beta_0 \frac{1}{z_i} + \beta_1 \frac{x_i}{z_i} + \frac{u_i}{z_i}.$$

В итоге получим пересмотренную модель:

$$y'_i = \beta_0 h_i + \beta_1 x'_i + \varepsilon_i, \quad (6.7)$$

где

$$y'_i = \frac{y_i}{z_i}, \quad h_i = \frac{1}{z_i}, \quad x'_i = \frac{x_i}{z_i}, \quad \varepsilon_i = \frac{u_i}{z_i}. \quad (6.8)$$

Подобрать простое преобразования для того, чтобы добиться гомоскедастичности удается не всегда. Часто гетероскедастичность является сигналом о неправильной функциональной форме модели.

В общем случае для коррекции гетероскедастичности используют следующую процедуру:

1. Рассчитывают МНК-оценки коэффициентов регрессии.
2. Находят остатки e_i .
3. Находят квадраты остатков e_i^2 .
4. Находят логарифмы квадратов остатков $\ln(e_i^2)$.
5. Рассчитывают регрессию $\ln(e_i^2) = \gamma_0 + \gamma_1 z_{i1} + \gamma_2 z_{i2} + \dots + \gamma_k z_{ik} + u_i$.
7. Получают прогноз $\ln(e_i^2)_{пр}$.
9. Находят веса наблюдений $w_i = \sqrt{\exp(\ln(e_i^2)_{прогноз})}$.
11. Полученные веса w_i используют во взвешенном методе наименьших квадратов.

Пример 6.4

Для регрессии, рассчитанной по таблице 5.1 сделать коррекцию на гетероскедастичность.

Решение

Регрессия, рассчитанная в примере 5 имеет вид:

$$\hat{y} = -0,6423 - 0,0005x_1 + 0,000043x_2 + 0,01858x_3 + 0,00133x_4 - 0,0014x_5 + 0,00022x_6.$$

Модель гетероскедастична согласно тесту Бреуша-Пагана (пример 6.2).

Возможно, что модель имеют неправильную форму.

Построим модель

$$\log Y = \beta_0 + \log \beta_1 X_1 + \beta_2 \log X_2 + \dots + \beta_k \log X_k + \varepsilon$$

и модель

$$\log Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon.$$

Проверим для каждой гипотезу о гетероскедастичности тестом Бреуша-Пагана. Прodelайте это самостоятельно, для нахождения логарифмов воспользуйтесь функцией: $f_x \rightarrow \text{Математические} \rightarrow \text{LOG}$.

Сделаем коррекцию на гетероскедастичность, используя взвешенный метод наименьших квадратов.

Рассчитываем МНК-оценки коэффициентов регрессии, найдем остатки e_i (не забудьте при нахождении регрессии поставить флажок «Остатки» в параметрах **Регрессии**).

Найдем квадраты остатков e_i^2 , логарифмы квадратов остатков $\ln(e_i^2)$ (рис.6.10).

Microsoft Excel - решение в exl.xls

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

Arial Cyr 10 Ж К Ч

E57 =LN(D57)

	A	B	C	D	E	F
52						
53						
54	ВЫВОД ОСТАТКА					
55						
56	Наблюдения	Предсказанное Y	Остатки	Квадрат остатков e^2	$\ln(e^2)$	
57	1	0.761775783	-0.040775783	0.001662664	-6.39933	
58	2	0.711707263	0.035292737	0.001245577	-6.68816	
59	3	0.724461374	0.015538626	0.000241449	-8.32885	
60	4	0.746432404	0.016567596	0.000274485	-8.20061	
61	5	0.742178867	0.001821133	3.31652E-06	-12.6166	
62	6	0.849293072	-0.016293072	0.000265464	-8.23403	
63	7	0.929738753	-0.015738753	0.000247708	-8.30326	
64	8	0.93573042	-0.01273042	0.000162064	-8.72752	
65	9	0.886735197	0.026264803	0.00068984	-7.27905	
66	10	0.930884627	-0.030884627	0.00095386	-6.95499	
67	11	0.908719199	-0.002719199	7.39404E-06	-11.8148	
68	12	0.934072422	-0.016072422	0.000258323	-8.2613	
69	13	0.909977379	0.013022621	0.000169589	-8.68213	
70	14	0.916309466	-0.022309466	0.000497712	-7.60549	
71	15	0.903209775	0.014790225	0.000218751	-8.42758	
72	16	0.92966807	0.00233193	5.4379E-06	-12.1221	
73	17	0.904506443	-0.000506443	2.56484E-07	-15.1762	
74	18	0.903146872	0.023853128	0.000568972	-7.47168	
75	19	0.907270807	0.013729193	0.000188491	-8.57646	
76	20	0.905726679	-0.000726679	5.28062E-07	-14.4541	
77	21	0.903472202	0.018527798	0.000343279	-7.97697	
78	22	0.811128691	-0.009128691	8.3333E-05	-9.39267	
79	23	0.919730067	0.007269933	5.28519E-05	-9.84802	
80	24	0.570049718	-0.025049718	0.000627488	-7.37379	
81	25	0.697074452	0.003925548	1.54099E-05	-11.0805	

Бреуш-паган / множ. регр / Лист1 / Чоу / Лист

Готово

Рис. 6.10. Нахождение логарифмов квадратов остатков

Построим регрессию от логарифмов остатков, независимые переменные – значения $X_1, X_2, X_3, X_4, X_5, X_6$.

$$\ln(e_i^2) = \gamma_0 + \gamma_0 z_{i1} + \gamma_0 z_{i2} + \dots + \gamma_0 z_{ik} + u_i.$$

Найдем прогноз $\ln(e_i^2)_{np}$ и веса наблюдений $w_i = \sqrt{\exp(\ln(e_i^2)_{прогноз})}$ (рис. 6.11).

Microsoft Excel - решение в exl.xls

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

Arial Cyr 10 Ж К Ц

E112 =КОРЕНЬ(D112)

	A	B	C	D	E
109	ВЫВОД ОСТАТКА				
110					
111	наблюдения	Предсказанное Y	Остатки	exp (предсказанное Y)	Веса w
112	1	-7.628626502	1.229292656	0.000486328	0.022053
113	2	-7.405688575	0.717532419	0.000607785	0.024653
114	3	-6.108910526	-2.219942167	0.002222971	0.047148
115	4	-9.206334399	1.00572129	0.000100401	0.01002
116	5	-11.68259698	-0.933996295	8.43942E-06	0.002905
117	6	-10.26619252	2.032161956	3.47896E-05	0.005898
118	7	-7.585866703	-0.717391801	0.000507575	0.022529
119	8	-9.673630452	0.946108692	6.2921E-05	0.007932
120	9	-8.026635015	0.747584	0.000326646	0.018073
121	10	-8.594304858	1.639311379	0.000185157	0.013607
122	11	-9.729736334	-2.085099821	5.9488E-05	0.007713
123	12	-9.520769365	1.259468519	7.33132E-05	0.008562
124	13	-8.29749617	-0.384638493	0.00024914	0.015784
125	14	-8.903354037	1.297865619	0.000135932	0.011659
126	15	-9.494076917	1.066499321	7.52965E-05	0.008677
127	16	-9.204813286	-2.917304885	0.000100554	0.010028
128	17	-10.56838574	-4.607812515	2.57163E-05	0.005071
129	18	-9.56771841	2.096038552	6.99508E-05	0.008364
130	19	-7.695637973	-0.880823658	0.000454807	0.021326
131	20	-12.42928284	-2.024768881	3.99973E-06	0.002
132	21	-8.928724912	0.95175877	0.000132527	0.011512
133	22	-10.52319821	1.130532352	2.6905E-05	0.005187
134	23	-10.50032343	0.652306991	2.75275E-05	0.005247
135	24	-6.352801901	-1.020983549	0.00174186	0.041736
136	25	-12.10107799	1.02057955	5.55352E-06	0.002357

Готово

Рис. 6.11. Нахождение весов w_i

Рассчитанные веса w_i используют во взвешенном методе наименьших квадратов.

Преобразуем имеющиеся по условию данные. Поделим каждое значение таблицы на соответствующие веса, при этом по строке деление на вес не меняется, а по столбцу – меняется. Не забудьте клавишу F4 - при необходимости нажмите на нее два или три раза.

Получим преобразованные данные рис. 6.12 (обратите внимание на строку формул). Дополнительно необходимо построить столбец 1/Вес.

	A	B	C	D	E	F	G	H	I
137									
138	Преобразование данных								
139	y	x1	x2	x3	x4	x5	x6	1/вес	
140	32.69418094	2085.9	124836.4496	3119.777599	3795.427	2797.824	911.4467	45.3456	
141	30.30019638	2474.31	109681.0322	2701.463292	3017.851	2157.926	920.769	40.56251	
142	15.69512525	1505.88	63777.35355	1433.770901	1781.609	1520.73	392.3781	21.20963	
143	76.1473251	7385.19	309479.4956	6786.393324	7834.292	5898.174	2564.857	99.7999	
144	256.1041623	32357.2	984830.6565	23545.06008	31049.19	21996.04	7917.199	344.226	
145	141.2278614	16818.5	538632.5518	12529.09839	12122.2	8731.374	4916.696	169.5413	
146	40.56917915	4483.03	145587.4263	3488.771861	3342.297	2716.448	901.0441	44.38641	
147	116.3600405	13237.1	398372.403	9896.276466	9959.31	6694.169	1777.548	126.0672	
148	50.51640369	5920.32	161342.6431	4249.353563	4083.363	2926.964	960.531	55.33012	
149	66.14120413	7936.94	257509.7547	5746.935737	5739.587	4541.696	1286.079	73.49023	
150	117.4663834	14261.9	431747.3032	10009.27682	9840.727	7390.269	2904.246	129.6538	
151	107.2139954	12847	414724.2893	9121.364966	9249.835	6995.772	1962.086	116.7908	
152	58.47632903	7032.36	224465.4754	4890.9779	4922.655	4010.348	1127.713	63.35464	
153	76.67897221	9692.08	282614.3327	6690.11167	6707.266	5317.781	1775.453	85.77066	
154	105.7925996	13022.4	373039.9183	8896.719707	9726.466	7387.043	1832.355	115.2425	
155	92.94279337	11268.8	304756.6272	7878.198151	7838.309	5843.828	1964.563	99.72403	
156	178.2642416	22677.4	659222.7429	15184.01173	14888.22	11062.64	4969.313	197.195	
157	110.8366495	13989.1	435455.3156	9170.626774	9959.755	8118.456	2164.124	119.5649	
158	43.1863185	5533.1	152816.734	3652.784159	3413.642	2771.239	947.1918	46.89068	
159	452.5150791	59502	1904063.449	37851.26131	38001.27	25350.84	10300.34	500.0167	
160	80.09011198	10684.5	260683.7593	6792.892361	6818.952	5368.296	1893.671	86.86563	
161	154.6172074	24484.3	644688.2063	13977.24131	15924.42	12589.16	4318.486	192.7895	
162	176.6835007	24777.6	638500.2454	14885.63259	12903.42	9053.362	4803.047	190.5971	
163	13.05839819	3498.21	57864.27822	1499.918765	1617.325	1368.137	603.8012	23.96036	
164	297.4635216	89111.8	1206827.754	29619.04965	25121.03	20368.4	17992.09	424.3417	
165									

Рис. 6.12. Преобразование данных

Построим регрессию без свободного члена. Для этого в окне «Константа-ноль» поставьте флажок.

Получим результаты регрессии (рис. 6.13). Это модель гомоскедастична, ее коэффициенты имеют ту же самую интерпретацию, а стандартные ошибки более точные:

$$y' = -0,618 \frac{1}{w_i} - 0,00036x'_1 + 0,0x'_2 + 0,018x'_3 + \\ + 0,0014x'_4 - 0,0014x'_5 - 0,00026x'_6.$$

Как видно из модели значения коэффициентов поменялось незначительно, однако переменная x_5 стала значимой, а x_6 – поменяла знак.

Microsoft Excel - решение в exl.xls

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

Arial Cyr 10 Ж К У

B183 = -0.000364950643992631

	A	B	C	D	E	F	G
167							
168	Регрессионная статистика						
169	Множественный R	0.99989					
170	R-квадрат	0.99978					
171	Нормированный R-квадрат	0.94415					
172	Стандартная ошибка	1.70505					
173	Наблюдения	25					
174							
175	Дисперсионный анализ						
176		df	SS	MS	F	Значимость F	
177	Регрессия	7	236614.932	33802.13314	11627.03507	5.26004E-30	
178	Остаток	18	52.32962598	2.907201444			
179	Итого	25	236667.2616				
180							
181							
182							
183	Переменная X 1	-0.00036	0.000145419	-2.509651217	0.021864698	-0.000670465	-5.9437E-05
184	Переменная X 2	4E-05	1.01145E-05	3.984864212	0.000868625	1.90552E-05	6.15549E-05
185	Переменная X 3	0.01826	0.001509101	12.09986273	4.41568E-10	0.015089408	0.021430419
186	Переменная X 4	0.00139	0.001150176	1.211319525	0.241436611	-0.001023201	0.003809661
187	Переменная X 5	-0.00141	0.000759776	-1.854228818	0.080165977	-0.003005031	0.000187433
188	Переменная X 6	-0.0002	0.001320395	-0.148108546	0.883903642	-0.002969611	0.002578487
189	Переменная X 7	-0.61799	0.206526064	-2.99228953	0.007815328	-1.051881272	-0.18409028
190							

Корр. коэф. корр. статистика P-Значение Нижние 95% Верхние 95%

Y-пересечение 0 #И/Д #И/Д #И/Д #И/Д #И/Д

Бреуш-паган / множ. регр / Лист1 / Чоу / Лист2 / коррекция /

Готово

Рис. 6.13. Регрессия, построенная взвешенным МНК

На значение коэффициента детерминации при взвешенном МНК обращать внимание не следует.

Задачи для самостоятельного решения

1-10. На основании данных о приросте курса акций за 10 месяцев и изменении валютного курса (%), приведенных в таблице в зависимости от варианта:

Изменение валютного курса x	Курс акций y в зависимости от варианта									
	1	2	3	4	5	6	7	8	9	10
4	7	1	1	5	9	11	1	5	5	1
5	8	5	2	4	8	12	5	6	6	2
8	5	2	5	7	10	15	9	8	9	5
7	4	3	4	8	11	15	8	7	8	6
9	9	6	8	9	9	17	7	9	5	9
5	6	4	7	8	8	19	9	9	3	8
2	8	8	9	9	2	20	11	6	2	7
6	5	7	8	10	7	22	12	11	4	6
4	2	9	10	11	6	21	13	12	7	11
7	1	5	2	12	5	26	15	15	9	12
8	5	4	5	15	3	25	14	16	8	15
9	4	6	10	11	2	28	16	20	10	16
5	4	9	8	12	1	30	20	21	11	20

и предположения, что генеральное уравнение регрессии имеет вид

$y = \beta_0 + \beta_1 x + \varepsilon$, требуется:

- а) Найти оценку и проверить на 5% уровне значимости уравнения регрессии, то есть гипотезу $H_0: \beta_1 = 0$;
- б) Построить таблицу дисперсионного анализа для расчета F -критерия Фишера;
- в) Найти коэффициент детерминации R^2 ;
- г) Найти интервальную оценку для прогноза при $x=11$;
- д) построить прямую линейной регрессии на диаграмме рассеивания.

12. По 30 наблюдениям проведено исследование зависимости результативного признака Y производительности труда от объясняющих факторов: X_1 – фондовооруженности и X_2 – объема заработной платы. Построено уравнение регрессии: $y = 23,55 + 7,14x_1 - 0,58x_2$ и найдены значения $S_{b0} = 2,58$, $S_{b1} = 7,85$, $S_{b2} = 0,74$. Можно ли утверждать на уровне значимости 0,05, что x_1 и x_2 оказывают влияние на Y .

13. Регрессия зависимой переменной Y на три независимые переменные на основе 18 наблюдений дала следующие результаты:

$$Y = 5,3 + 58 x_1 + 7,8 x_2 - 7 x_3$$

Стандартные ошибки	()	(1,4)	(1,0)	(1,8)
t- значения	(2,1)	()	()	()
95% доверительные границы	()	()	()	()

а) Заполните пропуски. б) Ход решения поясните.

в) Сделайте выводы о значимости коэффициентов регрессии

14. Построить регрессионную модель зависимости данных об объеме продаж Y в зависимости от:

X_1 – результат теста способностей к продаже;

X_2 – возраст;

X_3 – результат теста тревожности;

X_4 – опыт работы;

X_5 – средний балл школьного аттестата.

А) Постройте уравнение регрессии для составления способности кандидата стать хорошим продавцом.

Б) Сформулируйте выводы по полученному уравнению.

В) Оцените значимость уравнения регрессии и каждого коэффициента.

Д) спрогнозируйте результат объема продаж продавца без опыта работы, возрастом 22 года, имеющего средний балл аттестата 3,5 и с результатом теста тревожности 50.

Г) определите наилучшее уравнение для прогноза.

Объем продаж в месяц, в тыс. руб)	Результат теста способностей к продаже	Возраст	Результат теста тревожности	Опыт работы	Средний балл школьного аттестата
44	10	22,1	4,9	0	2,4
47	19	22,5	3,0	1	2,6
60	27	23,1	1,5	0	2,8
71	31	24,0	0,6	3	2,7
61	64	22,6	1,8	2	2,0
60	81	21,7	3,3	1	2,5
58	42	23,8	3,2	0	2,5
56	67	22,0	2,1	0	2,3
66	48	22,4	6,0	1	2,8
61	64	22,6	1,8	1	3,4
51	57	21,1	3,8	0	3,0
47	10	22,5	4,5	1	2,7
53	48	22,2	4,5	0	2,8
74	96	24,8	0,1	3	3,8
65	75	22,6	0,9	0	3,7
33	12	20,5	4,8	0	2,1
54	47	21,9	2,3	1	1,8
39	20	20,5	3,0	2	1,5
52	73	20,8	0,3	2	1,9
30	4	20,0	2,7	0	2,2
58	9	23,3	4,4	1	2,8
59	98	21,3	3,9	1	2,9

52	27	22,9	1,4	2	3,2
56	59	22,3	2,7	1	2,4
49	23	22,6	2,7	1	2,4
63	90	22,4	2,2	2	2,6
61	34	23,8	0,7	1	3,4
39	16	20,6	3,1	1	2,3
62	32	24,4	0,6	3	4,0
78	94	25,0	4,6	5	3,6

15. Постройте модель для прогноза годового объема продаж автомобилей в регионе по данным таблицы.

Регион	Годовой объем продаж автомобилей, млн. долл.	Количество пунктов обслуживания	Количество зарегистрированных автомобилей
1	52,3	2011	24,6
2	26,0	2850	22,1
3	20,2	650	7,9
4	16,0	480	12,5
5	30,0	1694	9,0
6	46,2	2302	11,5
7	35,0	2214	20,5
8	3,5	125	4,1
9	33,1	1840	8,9
10	25,2	1233	6,1
11	38,2	1699	9,5

Сформулируйте выводы по полученной модели. Проверьте значимость модели и полученных оценок коэффициентов регрессии.

Спрогнозируйте годовой объем продаж для 12 региона с 2500 пунктами обслуживания и 20,2 млн зарегистрированных автомобилей.

16. По результатам двух предварительных экзаменов X_1 и X_2 , среднего значения текущих оценок X_3 и результата окончательного экзамена Y для 20 студентов, получена следующая таблица.

X_1	X_2	X_3	y
87	85	2,7	91
100	84	3,3	90
91	82	3,5	83
85	60	3,7	93
56	64	2,8	43
81	48	3,1	75
77	67	3,1	63
86	73	3,0	78
79	90	3,8	98
96	69	3,7	99
93	60	3,2	54

92	69	3,1	63
100	86	3,6	96
80	87	3,5	89
100	96	3,8	97
69	51	2,8	50
80	75	3,6	74
74	70	3,1	58
79	66	2,9	87
95	83	3,3	57

Определите:

А) уравнение регрессии для прогноза окончательного экзамена на основе оценок двух предварительных экзаменов и текущего среднего балла. Является ли регрессия значимой. Поясните ответ.

Б) спрогнозируйте результат окончательного экзамена для студента с предварительными оценками 86 и 77 и средним баллом 3,4.

В) найдите наилучшую модель методом пошаговой регрессии. Сравните полученную регрессию. На основании F -критерия обоснуйте выбор модели

17. Приведена информация о 25 уже существующих горнолыжных лагерях в шт.

Вашингтон. Анализировались следующие переменные.

Y — стоимость одного дня пребывания в лагере;

X_1 — общая площадь лагеря в акрах;

X_2 — количество жилых помещений;

X_3 — наличие смывных туалетов;

X_4 — наличие плавательного бассейна;

X_5 — наличие канатных подъемников;

X_6 — количество дополнительных мест развлечения.

Лагерь	Y	X_1	X_2	X_3	X_4	X_5	X_6
1	7,00	40	32	0	0	0	2
2	8,50	20	47	1	0	1	2
3	9,00	45	18	1	1	1	1
4	8,00	110	32	1	0	1	3
5	8,00	30	54	1	0	1	2
6	7,00	50	30	1	0	1	3
7	7,75	35	30	1	0	1	2
8	8,00	18	40	1	0	1	1
9	8,50	23	60	1	1	1	1
10	8,50	9	60	1	0	1	3
11	9,00	52	50	1	1	1	2
12	7,00	25	21	0	0	1	1
13	9,00	250	30	1	0	1	2
14	8,50	140	70	1	1	1	2
15	9,00	120	80	1	1	1	1
16	7,50	60	50	1	1	1	2

17	8,50	120	35	1	. 0	1	2
18	9,00	173	25	1	1	1	2
19	8,00	100	75	1	0	1	2
20	9,50	134	35	1	1	1	1
21	7,50	114	120	0	1	1	2
22	7,50	2	17	0	0	1	2
23	7,50	32	15	0.	1	0	3
24	9,00	25	30	1	1	1	2
25	7,50	66	100	1	0	1	2

Построить уравнение регрессии зависимости стоимости проведенного дня. Построить уравнение регрессии, проверить его значимость и значимость каждого коэффициента. Сделать выводы по полученным коэффициентам.

18. Построить уравнение регрессии по данным таблицы 1 приложения зависимость объема продаж от численности работников, стоимости материальных и нематериальных ресурсов, издержек на производство, оплатой труда, расходов на рекламу и исследований на разработки и рекламу. Сформулировать выводы о значимости модели и значениях коэффициентов.

19. Построить модель зависимости логарифма заработной платы от возраста, образования, профессиональных данных, общего стажа, срока пребывания в должности, рыночной стоимости фирмы, процента собственности в фирме, доходов фирмы и объемами продаж по данным таблицы 2 приложения.

20. Построить модель зависимости заработной платы, премий и других выплат от возраста, образования, профессиональных данных, общего стажа, срока пребывания в должности, рыночной стоимости фирмы, процента собственности в фирме, доходов фирмы и объемами продаж по данным таблицы 2 приложения.

21. Построить модель зависимости общих выплат от возраста, образования, профессиональных данных, общего стажа, срока пребывания в должности, рыночной стоимости фирмы, процента собственности в фирме, доходов фирмы и объемами продаж по данным табл. 2 приложения.

Приложения

Таблица 1

Основные экономические показатели за 2000г. для 266 компаний США

N/N	Объем продаж (млн долл.)	Численность работников (тыс. чел.)	Стоимость материальных ресурсов (млн долл.)	Стоимость нематериальных ресурсов (млн долл.)	Издержки на производство (млн долл.)	Оплата труда (млн долл.)	Расходы на рекламу (млн долл.)	Расходы на исследования и разработки (млн долл.)
1	3 221,8008	42,0000	147,9000	30,6000	2 285,2007	599,7998	118,3000	28,0000
2	1 690,6001	20,9050	93,0000	29,1000	1 057,2002	343,2000	114,9000	8,9000
3	2 197,2764	39,0000	66,8670	55,8600	1 387,0679	661,3997	95,5680	11,1820
4	2 357,8206	23,3000	59,5560	69,6080	1 743,7952	25,6320	51,9170	8,5000
5	8 129,0000	35,0000	297,0000	29,0000	7 423,0000	1 178,0000	12,8000	9,2530
6	11 851,0000	23,0000	394,0000	20,0000	10 942,0000	2 556.0000	11,6530	14,6000
7	323,8606	3,9000	2,5900	4,2880	233,5300	22,8350	3,5290	30,7320
8	660,4856	8,3780	10,9840	3,3720	582,2649	25,6250	44,9990	64,8730
9	4351,1601	50,9120	102,7080	217,0920	4 156,8671	12,8360	66,2640	8,7790
10	985,8357	5,5000	16,6010	29,5900	874,1287	19,5000	112,3860	18,3650
11	3 802,5581	39,6000	206,1020	157,3520	2 997,2703	518,0000	139,7290	16,4130
12	2 576,0464	22,6000	50,6690	47,0790	1 885,9053	349,4910	48,8170	9,5000
13	106,0160	28,0000	1,3120	42,0000	84,6590	35,5550	22,9370	8,7330
14	5 669,8945	46,8810	103,0000	31,1000	4 424,3007	785,0000	141,3000	18,5000
15	319,6570	2,8940	4,5770	2,2090	246,6980	42,8370	87,0000	1,1000
16	511,7217	10,1000	19,5600	27,0000	286,2288	48,9990	1,8700	23,6520
17	884,6189	22,8010	58,0940	33,0000	467,4436	36,5000	16,0350	29,6320
18	166,3750	2,3000	3,9510	5,2890	111,0310	31,0000	4,0230	38,5420
19	59,1310	18,0000	1,1400	14,5000	43,7430	26,3210	90,3250	56,9820

N/N	Объем продаж (млн долл.)	Численность работников (тыс. чел.)	Стоимость материальных ресурсов (млн долл.)	Стоимость нематериальных ресурсов (млн долл.)	Издержки на производство (млн долл.)	Оплата труда (млн долл.)	Расходы на рекламу (млн долл.)	Расходы на исследования и разработки (млн долл.)
20	136,6970	3,1000	2,0090	18,4930	105,3300	15,8880	46,3000	8,6330
21	767,8799	8,1000	37,4250	18,0560	519,3948	112,1350	21,8470	2,7860
22	61,3280	1,1390	1,3880	26,3250	35,2020	17,3140	2,4270	88,5230
23	445,6387	5,8000	18,9780	12,6000	213,2880	12,1000	62,8060	1,4600
24	2259,6316	16,0270	228,7270	27,3350	1 696,3772	421,8057	116,5990	9,6000
25	624,8040	8,7000	86,4030	2,8080	408,4707	168,0200	33,4700	9,4440
26	329,9578	4,0000	14,9460	8,3710	225,0410	20,9850	12,9790	32,0000
27	308,7327	2,1070	14,8080	43,5920	239,1300	36,5000	18,1220	1,8510
28	598,9507	5,0000	39,7150	27,8920	481,9436	45,0000	39,8230	0,7500
29	172,7920	1,5760	1,6590	23,5420	118,7090	48,2000	7,9090	26,3330
30	910,8406	7,0000	14,4610	5,5880	677,2527	7,0000	58,2130	1,8000
31	142,1830	1,6000	5,5880	72,5190	126,9660	1,6000	2,7310	57,2710
32	425,0828	6,8330	72,5190	31,8030	256,2837	6,8330	12,1440	44,1550
33	4 337,9140	36,1000	306,0220	101,4290	2344,1631	36,1000	270,2576	16,1100
35	209,4520	2,9980	14,4690	14,6060	176,4890	2,0780	1,8970	714,9990
36	62,4180	3,8000	3,7390	7,6680	34,4700	2,9980	44,0500	121,3300
37	4 300,0000	95,5000	412,2886	157,6030	2 108,5503	5,1000	257,6807	11,6440
38	390,6829	5,1000	30,8480	10,8550	225,1080	6,3000	18,3780	33,4770
39	270,0127	6,3000	40,0340	22,4540	189,8000	2,0000	4,9080	43,7430
40	97,9660	2,0000	6,9940	5,2500	64,5920	31,9700	2,5900	18,9700
41	66,4090	12,5261	3,7570	1,0090	57,2310	33,2000	59,1300	14,9460
42	56,5550	3,9000	1,6240	6,9940	44,0550	53,5000	19,5600	1,6590
43	3267,9551	31,9790	502,0398	45,6140	2517,7566	754,8977	3,9510	57,7210
44	2 745,7439	43,9680	251,0340	16,1110	1 638,7969	45,0000	161,2000	108,1480
45	2 609,0000	33,2000	248,0001	10,0000	1 874,0000	564,0000	18,0000	83,0000

N/N	Объем продаж (млн долл.)	Численность работников (тыс. чел.)	Стоимость материальных ресурсов (млн долл.)	Стоимость нематериальных ресурсов (млн долл.)	Издержки на производство (млн долл.)	Оплата труда (млн долл.)	Расходы на рекламу (млн долл.)	Расходы на исследования и разработки (млн долл.)
46	1 677,6016	11,6440	284,6089	87,4830	1 185,9717	24,4530	6,4840	36,1310
47	6887,6210	53,5000	1 075,1719	84,0390	4721,9570	1 375,7996	44,0700	231,4690
48	10584,1990	132,1400	714,2002	22,6000	7 353,5000	3 204,2688	93,4000	377,1001
49	2912,7644	45,8540	195,2680	45,6430	2 189,5293	879,6548	14,9460	66,0560
50	309,5820	66,8000	275,3079	67,3120	2913,9036	993,3997	1,6590	40,5470
51	1 946,4766	24,4530	121,3300	6,2920	1 403,4976	546,0508	35,2020	40,0810
52	9254,1171	151,2000	1 431,0906	121,3300	6 187,7851	2 125,2012	95,9510	334,8057
53	5018,6914	62,8510	479,8997	1,6240	3 478,0989	1 318,0999	9,2530	144,3000
54	1 510,7798	15,3000	207,9320	63,5190	1 157,2117	13,9700	27,6660	39,7150
55	1 560,0750	22,7000	162,5190	61,9380	1 188,9126	18,4340	19,3190	24,7010
56	2 794,0000	37,4000	256,0999	7,3000	1 928,4988	780,7996	18,3650	70,1000
57	921,3689	13,9700	61,9380	18,4340	597,7000	45,1640	19,2020	22,6500
58	1 253,5430	13,0580	66,4310	13,9700	806,6758	236,5000	32,0000	48,6510
59	1 328,1138	13,1160	201,1960	31,2730	851,8938	1,1550	31,2730	33,5620
60	1 314,6299	27,3460	36,9330	43,0750	569,7327	6,4690	174,4610	42,1160
61	7 869,6914	113,3710	687,7998	90,2000	5 580,5976	1931,5005	76,5000	155,9000
62	73,0550	7,8240	26,5680	20,6650	38,9980	22,8990	43,0750	99,8430
63	108,5090	87,4350	5,6630	37,3860	77,1740	36,9990	90,2000	1,6500
64	1 422,4507	16,5000	100,4700	69,8820	1 060,5420	305,7000	6,3970	25,4520
65	87,4350	7,6550	8,5150	15,3750	51,3970	11,3940	69,8820	2,7200
66	7,8240	9,5280	26,6950	7,7640	6,7860	20,5720	4,2100	52,1780
67	868,7107	15,3400	42,4040	1,2120	686,0518	200,4850	10,4000	22,7240
68	137,3950	2,8750	14,1080	9,7470	112,2350	30,7620	83,1580	1,9000
69	753,8848	6,5480	24,2870	4,2120	596,5076	13,4000	88,8250	6,4200

N/N	Объем продаж (млн долл.)	Численность работников (тыс. чел.)	Стоимость материальных ресурсов (млн долл.)	Стоимость нематериальных ресурсов (млн долл.)	Издержки на производство (млн долл.)	Оплата труда (млн долл.)	Расходы на рекламу (млн долл.)	Расходы на исследования и разработки (млн долл.)
70	1 445,0166	27,0030	84,1490	99,9080	786,8777	1,9360	39,8650	76,1870
71	3062,6316	49,6190	67,6310	83,1580	1 446,5227	668,9910	243,0450	74,5240
72	2 450,4285	32,6000	81,9220	88,8250	906,9639	6,7120	423,2698	90,5730
73	141,2580	1,3040	4,5050	6,7300	95,1540	3,7000	9,9040	9,7580
74	6,8030	5,1000	9,5230	1,4590	2,3980	12,2490	0,7230	11,9490
75	1 852,0896	25,4000	89,5500	57,7900	672,7947	4,5070	28,4910	148,0770
76	365,7217	4,9030	17,0620	16,7160	217,5420	3,4720	6,7300	11,8950
77	1 981,4397	28,7000	155,8530	141,2700	668,7720	634,0596	55,2940	161,3500
78	2362,1326	40,7000	110,1000	99,8430	1 055,4187	11,3940	75,7000	113,1280
79	357,0696	5,5500	12,6430	52,1780	141,2700	2,1330	36,8860	18,9510
80	220,3790	3,7000	10,7860	9,7580	67,1220	20,5720	7,1610	6,2610
81	1 082,4927	17,9000	51,3360	52,1780	310,7820	315,8997	114,9660	65,6910
82	848,3799	17,1000	41,2990	11,9490	386,0066	16,0000	40,6150	61,6940
83	1 112,0386	16,5890	74,5790	44,6610	378,7710	7,3000	91,2150	77,3130
84	1 515,8816	37,0000	108,0460	52,3290	758,5320	469,9229	74,5950	61^8300
85	1 328,5508	19,9200	44,6810	6,2850	566,2200	323,7090	36,9560	115,5890
86	2 878,4956	58,0000	182,2670	348,1426	1 247,2339	1,1500	391,6277	85,3970
87	4312,0507	56,6000	169,2950	66,9970	2 672,3262	6,4600	260,3870	37,6540
88	54,3250	37,3860	1,0660	2,8130	26,5960	4,7670	0,7520	44,6610
89	122,9470	57,1720	13,7480	7,5620	94,6720	17,6580	1,4590	3,8670
90	2 014,7056	31,0000	74,7910		700,4778	503,6768	45,0900	21,1460
91	969,8328	18,5170	40,8340	54,2710	448,5286	9,4450	91,2690	8,5670
92	45,3670	8,3500	1,6430	7,0670	15,7310	2,1230	5,1820	52,3290
93	255,1320	3,3000	10,6420	20,2520	131,6750	12,2220	42,5670	6,2850
94	1 710,4700	31,7000	91,5640	54,7540	752,5889	530,2456	239,9010	42,0600

N/N	Объем продаж (млн долл.)	Численность работников (тыс. чел.)	Стоимость материальных ресурсов (млн долл.)	Стоимость нематериальных ресурсов (млн долл.)	Издержки на производство (млн долл.)	Оплата труда (млн долл.)	Расходы на рекламу (млн долл.)	Расходы на исследования и разработки (млн долл.)
95	365,8809	3,4800	20,0140	6,7300	177,5500	25,8740	16,7100	23,7910
96	33,2650	2,0870	1,5120	4,4840	19,7100	19,7100	1,1550	2,8890
97	53,7460	0,5250	2,0870	42,2810	16,1820	16,1800	7,6770	19,7100
98	52,8760	1,1420	2,4190	1,2160	27,1500	27,1500	6,4690	16,1820
99	9,6630	2,4190	12,7460	7,9670	5,6960	5,6950	0,4570	27,1500
100	1 451,6687	29,0000	86,6820	97,2690	505,8267	36,1200	137,7250	30,7620
101	321,3638	4,9110	13,1180	11,0840	268,0159	57,2600	1,1110	13,4000
102	156,4580	2,3500	4,5670	3,8620	114,1930	6,4800	4,7670	5,6960
103	52,1870	0,8650	1,5100	20,6490	36,5130	59,3250	18,0150	1,9360
104	447,2100	7,7670	12,7460	41,7940	280,3218	26,8120	9,4400	505,8267
105	86,8170	1,1000	1,2810	19,3850	57,2600	26,6950	2,1230	1,3420
106	1 132,3499	18,0150	16,8570	1,6970	785,0718	36,9240	25,8740	^т 6,7120
107	217,4120	3,2000	4,4840	10,5440	142,6020	57,5790	3,2520	3,7000
108	7,7640	86,6820	1,2810	7,2210	6,4800	9,5280	20,8580	268,0159
109	1 581,8760	20,8580	142,2810	5,8820	1 280,1670	359,0999	1,1000	12,2490
110	201,4650	1,1000	7,9670	1,3370	169,2630	57,5700	7,4000	114,1930
111	198,9010	0,9110	9,7470	0,4290	164,1940	73,9670	1,8400	36,5130
112	1 497,0076	7,4000	131,9400	6,0210	1 098,2969	99,4080	5,1000	280,3280
113	153,2290	1,8400	11,0840	3,4390	59,2350	9,2800	8,3500	4,5070
114	367,9246	5,1000	20,6490	11,2110	230,1690	73,9670	1,2110	11,3940
115	494,4136	8,3500	19,3850	3,1490	342,9849	6,4690	3,8030	2,1000
116	52,4550	1,2120	7,2210	7,0620	26,8120	4,7670	4,0510	57,2600
117	37,3860	0,8200	1,3370	44,3730	26,6950	9,4400	5,6000	785,0718
118	57,7120	13,1190	3,4390	0,7160	36,9240	2,1230	7,5620	0,9510

N/N	Объем продаж (млн долл.)	Численность работников (тыс. чел.)	Стоимость материальных ресурсов (млн долл.)	Стоимость нематериальных ресурсов (млн долл.)	Издержки на производство (млн долл.)	Оплата труда (млн долл.)	Расходы на рекламу (млн долл.)	Расходы на исследования и разработки (млн долл.)
119	586,4766	3.8030	44,3730	34,2780	391,3706	25,8740	2,8100	20,5720
120	476,2078	4,0510	34,2780	30,2360	244,7830	99,9080	5,8820	16,0000
121	15,3570	4,5671	16,8570	53,2830	9,5280	29,0000	6,0200	142,6020
122	393,6016	5,6000	30,2360	2,8890	265,3079	9,2800	11,2110	7,3000
123	4701,1210	7,5620	353,2830	48,6920	3 707,6846	4,9110	3,1490	6,4800
124	1 167,8340	2,8100	48,6920	8,4580	1 017,6038	2,3500	7,0620	59,2350
125	12298,3980	50,7000	1 221,8008	10,4000	9285,7109	1 016,5000	13,1160	64,6000
126	439,4727	1,9020	65,1100	39,8650	263,8108	51,1480	27,3460	31,2730
127	29127,0030	108,7000	1 897,0005	9,9040	20 032,0000	78,7700	16,5000	86,0000
128	1 993,6624	8.0000	43,4190	45,7820	1 755,5662	3,5730	31,1370	43,0750
129	4 660,8945	18,1000	636,1238	28,4900	3 675,6895	440,7996	3,4000	11,6000
130	976,4578	8,8280	14,8590	55,2940	879,3516	91,8000	15,3440	90,2000
131	3 834,9324	6,6610	316,7156	68,2690	3 557,4734	7,4050	2,8250	69,8820
132	9 535,7382	42,7800	1 107,3838	75,7000	7075,1875	971,0000	6,5480	29,7730
133	657,7776	1,2640	56,1460	36,8860	565,0176	14,4700	27,0030	4,2120
134	100,4570	43,0750	44,0680	7,1610	72,7830	22,0310	49,6110	83,1580
135	60334,5110	130,0000	4 186,9296	40,6150	45 999,0070	3 405,0000	32,6000	290,0000
136	2 150,0000	90,2110	311,7000	91,2150	1 460,7996	57,4030	1,3040	25,1000
137	18069,0000	58,3000	1 680,0000	74,5900	13442,0000	1 345,0000	25,4000	88,8250
138	109,7380	69,8870	32,2560	36,9560	97,0130	2,5200	4,9030	6,7300
139	592,7710	3,2520	123,7680	3,8770	420,3206	67,3300	28,7000	1,4590
140	4 642,3945	14,3280	353,5999	33,5620	4 085,0989	324,0000	40,7000	25,0000
141	2 072,4412	11,1480	270,1846	42,1160	1640,8118	1,2400	5,5500	4,9810
142	4 509,3828	13,3540	502,2720	1,6500	2 838,0845	236,4540	2,0370	12,8000

N/N	Объем продаж (млн долл.)	Численность работников (тыс. чел.)	Стоимость материальных ресурсов (млн долл.)	Стоимость нематериальных ресурсов (млн долл.)	Издержки на производство (млн долл.)	Оплата труда (млн долл.)	Расходы на рекламу (млн долл.)	Расходы на исследования и разработки (млн долл.)
143	34 736,0030	207,7000	1 760,7100	2,7200	26 053,9060	20,9400	3,7000	16,7160
144	1 191,0337	4,2070	255,6150	1,9000	865,6477	82,6730	0,2670	99,8430
145	312,7300	4,2120	76,5000	6,4200	452,4130	17,0050	17,9000	52,1780
146	1 553,1077	9,1500	343,9539	23,6410	988,8760	185,6600	12,5840	9,7580
147	6 997,7734	30,0080	956,1719	11,2330	4886,8125	720,5000	17,1000	58,2460
148	513,1880	5,1420	41,9800	41,9800	375,3599	25,0200	11,3330	11,9490
149	28 085,0030	94,8000	2913,0000	32,5600	20 632,0000	2 344,0000	89,0000	231,0000
150	11 062,8980	34,9740	1 774,3904	43,0250	8 259,7656	1 051,0000	16,5890	114,0000
151	23 232,4060	37,5750	1 049,6729	90,2110	19964,6050	994,0000	37,0000	89,7370
152	14961,5000	47,0110	1 744,0364	69,8870	10046,0000	1 126,7310	19,9200	80,3250
153	5 197,7070	24,1450	762,2510	4,2120	3 336,7566	431,9976	7,7130	15,0520
154	7 428,2343	33,7210	601,1216	6,7310	5 714,3085	9,7320	58,0000	21,0000
155	28 607,5030	67,8410	1 344,3777	10,4000	24 787,6050	1 572,7996	56,6000	52,0000
156	87,6100	6,7310	12,7120	39,8650	74,5510	31,5580	31,0000	44,6610
157	1 165,6736	3,5310	26,6780	9,9040	1 035,7129	6,6000	18,5170	2,4490
158	567,3650	1,5420	97,4910	28,9400	480,5110	23,5230	3,300	52,3290
159	5 954,9414	16,2970	732,0000	55,9240	4 540,4609	444,8997	31,7000	18,5000
160	368,0940	2,3150	15,0860	2,7160	319,4939	10,6050	3,4800	6,2850
161	751,7327	6,2550	51,1060	13,5380	606,8318	3,5230	6,8000	9,9000
162	895,4087	10,9000	145,5140	9,3840	681,9656	26,3250	39,0000	30,6000
163	1 063,2908	16,1790	51,1480	25,7670	746,2820	12,6000	16,6980	14,6320
164	1 306,0867	19,3970	78,7700	2,7490	1 021,4856	435,2998	23,3000	13,2830
165	140,4440	1,9190	3,5730	55,8600	122,3210	27,3350	35,0000	29,1000
166	4357,2812	52,1400	110,4470	12,0830	3 540,9612	1 235,0000	3,9000	55,8600

N/N	Объем продаж (млн долл.)	Численность работников (тыс. чел.)	Стоимость материальных ресурсов (млн долл.)	Стоимость нематериальных ресурсов (млн долл.)	Издержки на производство (млн долл.)	Оплата труда (млн долл.)	Расходы на рекламу (млн долл.)	Расходы на исследования и разработки (млн долл.)
19	263,9048	3,7000	7,4050	27,2080	203,3440	2,8080	8,3780	3,2500
168	6 184,8945	94,5000	398,2000	69,6080	5 224,0000	2 550,0000	50,9120	37,1000
169	257,6509	3,3640	14,4730	7,5700	190,4190	8,3710	5,5000	69,6080
170	50,5150	52,5350	29,1000	29,0000	18,0560	43,5920	39,6000	29,0000
171	419,6470	4,3020	22,0310	20,0000	341,5906	135,6000	22,6000	20,0000
172	1 227,4490	20,0000	57,4030	4,2880	999,7520	27,8920	28,0000	9,0000
173	779,3450	8,8000	22,0670	3,3700	678,4258	229,1270	46,8810	4,2880
174	72,1760	1,3000	2,5210	29,5900	50,9650	24,8290	2,8940	3,3700
175	3 248,0076	36,0620	263,6167	19,4460	2710,3455	974,3379	10,1000	29,5900
176	921,1270	12,6590	67,3340	10,5250	771,0059	23,5420	22,8010	157,3520
177	711,9827	12,5120	133,3850	45,0790	653,8069	351,4700	2,3000	45,0790
178	72,4110	1,0250	1,2400	42,0000	60,0820	93,0000	18,0000	42,0000
179	297,5686	4,1520	20,9420	0,8990	248,7160	123,1000	3,1000	31,1000
180	677,8489	6,0700	17,0050	31,1000	613,3047	169,2570	8,1000	2,2090
181	582,6238	1,4000	25,0290	2,2090	474,3450	66,8670	1,1390	27,0000
182	3 750,4109	38,1700	120,8280	27,0000	3 240,7886	1 132,6216	5,8000	33,0000
183	88,8070	2,3330	9,7320	7,2110	66,6540	59,5500	16,0270	5,2890
184	306,9397	2,8000	31,5880	33,0000	220,4980	2,5900	8,7000	14,5000
185	331,7366	5,2000	6,6000	11,0250	295,3848	10,9840	4,0000	18,4930
186	1 546,9500	8,9000	23,5230	5,2890	439,0479	16,0010	2,1000	18,0560
187	7,5910	30,6000	7,8900	14,5000	5,0480	50,6690	5,0000	26,3250
188	3 479,4573	41,3940	170,3720	18,4930	3 100,5391	1 177,5999	1,5760	15,0500
189	485,6138	6,6580	58,6750	3,5250	335,3318	42,0000	93,0000	1,4320
190	123,2280	2,0450	10,6050	5,2550	96,6630	20,9000	66,8670	12,6000

N/N	Объем продаж (млн долл.)	Численность работников (тыс. чел.)	Стоимость материальных ресурсов (млн долл.)	Стоимость нематериальных ресурсов (млн долл.)	Издержки на производство (млн долл.)	Оплата труда (млн долл.)	Расходы на рекламу (млн долл.)	Расходы на исследования и разработки (млн долл.)
191	488,2327	4,6500	20,4800	1,1111	402,8457	402,8400	77,0101	22,2426
192	100,7820	1,7030	2,4430	1,6800	88,7960	4,0000	21,0000	28,3032
193	165,7970	4,7660	3,2790	88,0003	120,1080	2,0000	4,0008	18,2022
194	274,8440	3,5500	21,7900	2,9530	213,1860	3,0000	3,7521	24,2628
195	11 049,5000	166,8480	667,7998	55,5000	9 955,3984	4485,1953	22,0007	52,5000
196	1 154,8477	14,4190	32,2360	4,0800	1 037,4727	424,4556	21,1234	30,3234
197	578,7107	11,4920	26,3000	8,0141	433,8230	1,0111	12,3456	5,3300
198	124,5440	1,8000	4,6280	1,9850	101,5300	23,6630	78,9101	36,3840
199	3711,2029	63,4000	303,3838	4,5720	2 729,9280	22,0222	91,0111	33,0000
200	124,8600	2,0000	5,2240	2,3200	79,7770	51,0000	21,3141	2,6500
201	2 466,0000	26,8650	161,7000	2,0202	2 028,7996	18,4021	3,2000	14,9000
202	2 829,2991	36,2000	156,8000	27,1000	2261,0000	930,2000	51,1617	25,2000
203	814,8196	14,8000	48,5520	16,0111	622,9507	204,9000	18,1920	1,4150
204	4051,7996	46,0000	349,5999	2,6000	3 036,5999	1 215,2996	21,2223	56,6000
205	67,0390	28,0005	3,5010	2,5170	54,9070	66,5620	24,2526	42,4446
206	240,5670	4,0000	5,5670	1,3220	184,1350	61,6900	2,5860	3,0470
207	45,2140	2,0000	1,4110	18,1010	38,0970	62,3201	27,2829	48,5052
208	69,9520	81,0000	33,3333	8,0033	65,4570	52,3302	30,3132	54,5658
209	54,5490	1,1270	1,7720	17,7200	42,5990	42,4444	33,3435	60,6264
210	317,4480	5,7840	12,6650	11,0330	254,1990	80,1010	36,3738	66,6870
211	847,9927	24,0000	85,0240	19,7930	664,9578	34,1021	39,4041	10,4000
212	467,9546	4,8450	13,1650	2,3810	400,5806	4,0999	42,4344	1,0011
213	126,6750	14,0007	7,7490	14,1441	109,6830	50,6410	45,4647	1,0022
214	85,7290	49,0000	2,1610	49,4949	72,8400	9,9901	48,4950	1,0033

N/N	Объем продаж (млн долл.)	Численность работников (тыс. чел.)	Стоимость материальных ресурсов (млн долл.)	Стоимость нематериальных ресурсов (млн долл.)	Издержки на производство (млн долл.)	Оплата труда (млн долл.)	Расходы на рекламу (млн долл.)	Расходы на исследования и разработки (млн долл.)
215	680,7666	8,2200	19,2340	77,7878	578,8528	9,8175	51,5253	1,3090
216	211,3230	1,5670	4,8350	15,6180	171,4130	65,0000	54,5556	1,8201
217	254,3030	3,1000	2,7620	2,3570	205,8410	42,4381	57,5859	2,0880
218	1 396,8108	29,4160	79,9820	28,2626	1 000,2886	3,8107	16,1580	3,4510
219	3981,0000	52,9000	188,3000	70,3000	3 120,5999	1 085,7996	75,8000	37,5000
220	3 943,0990	56,5320	259,5000	49,9000	3 352,3008	1 275,7002	60,6162	42,3000
221	1 260,2349	17,2880	103,0320	11,4810	1 055,9436	12,0000	63,6465	2,1133
222	973,2527	9,8850	25,4530	5,5580	848,7227	4,0877	66,6768	3,3210
223	19,9060	18,0002	5,6666	1,4100	16,5170	3,3876	69,7071	4,2242
224	66,8260	1,3200	6,1110	88,1388	48,9480	4,5222	72,7374	5,6210
225	178,7460	2,1980	5,5430	138,0000	138,5690	43,4350	75,7677	6,2155
226	26,7510	1,0560	8,8888	211,0113	17,9930	18,1111	78,7980	7,2102
227	20,5750	43,1111	7,7777	82,1003	13,9720	14,2222	81,8283	8,9712
228	51,5960	18,5216	1,6940	1,1620	38,8190	88,9922	81,0077	24,2601
229	106,1150	2,6000	4,6850	9,9210	64,0500	12,4243	77,0222	23,2810
230	8,5160	14,2421	12,0818	12,1402	5,9500	7,8246	22,4443	24,8588
231	308,8916	5,7000	15,8370	13,1402	144,7340	42,4444	47,8877	2,7060
232	753,8069	16,8750	37,4620	3,6210	491,1160	210,0050	16,4370	4,9340
233	41,2960	1,1080	2,5820	12,1213	28,1320	81,8118	12,5456	24,5293
234	145,6840	3,4190	13,3250	1,0087	105,1630	51,7100	51,8196	1,8480
235	51,3130	1,0000	1,5700	8,0025	35,9730	43,4400	21,4231	59,6085
236	21,4070	12,5358	18,7842	5,5554	12,9550	12,8888	37,8286	64,8821
237	585,6597	8,2000	56,0530	80,9960	359,8350	77,9999	13,6920	8,9610
238	516,7239	10,3000	17,9320	9,3610	376,4170	1,1007	5,6670	5,6000

N/N	Объем продаж (млн долл.)	Численность работников (тыс. чел.)	Стоимость материальных ресурсов (млн долл.)	Стоимость нематериальных ресурсов (млн долл.)	Издержки на производство (млн долл.)	Оплата труда (млн долл.)	Расходы на рекламу (млн долл.)	Расходы на исследования и разработки (млн долл.)
239	316,8147	7,0000	3,9360	12,1314	267,2456	2,0008	86,8686	76,7686
240	509,7000	10,0000	27,0360	15,1617	375,3457	179,9240	85,8686	3,6080
241	341,3887	7,1270	7,1570	8,1819	287,6907	9,0007	86,8888	86,7795
242	33,0660	1,0872	1,9540	9,2021	24,0720	12,7210	83,1111	95,9594
243	200,5920	4,0000	5,3320	20,0290	153,5480	7,6660	82,2222	94,9388
244	184,5810	4,0500	7,2780	10,3570	142,7160	8,7770	22,6665	1,0790
245	217,7520	4,0880	7,3840	10,1146	179,1020	78,3910	44,6621	89,9012
246	386,8118	7,4040	18,4880	47,1213	302,5586	2,9990	18,1716	3,8620
247	69,1530	12,1212	1,6190	48,1415	54,4310	11,3410	15,1413	13,8125
248	81,4650	1,6220	4,1890	16,4950	70,5080	4,4555	12,1110	47,8552
249	329,5518	6,0040	12,2520	8,0540	269,6377	12,1417	9,8765	51,9185
250	36,3870	133,0000	12,7246	51,5355	27,7690	21,8283	4,3210	54,3321
251	344,7937	7,5000	24,7400	57,5982	205,0610	92,9395	8,1234	4,8200
252	22,8030	84,1000	2,1060	83,4952	10,6830	96,9899	5,6788	43,8388
253	196,3030	5,4660	5,9730	-99,9242	142,1520	97,9294	12,4582	2,2710
254	31,5660	13,7124	8,1264	10,1115	22,3750	95,0092	14,5220	66,7274
237	585,6597	8,2000	56,0530	80,9960	359,8350	77,9999	13,6920	8,9610
238	516,7239	10,3000	17,9320	9,3610	376,4170	1,1007	5,6670	5,6000
239	316,8147	7,0000	3,9360	12,1314	267,2456	2,0008	86,8686	76,7686
240	509,7000	10,0000	27,0360	15,1617	375,3457	179,9240	85,8686	3,6080
241	341,3887	7,1270	7,1570	8,1819	287,6907	9,0007	86,8888	86,7795
242	33,0660	1,0872	1,9540	9,2021	24,0720	12,7210	83,1111	95,9594
243	200,5920	4,0000	5,3320	20,0290	153,5480	7,6660	82,2222	94,9388
244	184,5810	4,0500	7,2780	10,3570	142,7160	8,7770	22,6665	1,0790
245	217,7520	4,0880	7,3840	10,1146	179,1020	78,3910	44,6621	89,9012

N/N	Объем продаж (млн долл.)	Численность работников (тыс. чел.)	Стоимость материальных ресурсов (млн долл.)	Стоимость нематериальных ресурсов (млн долл.)	Издержки на производство (млн долл.)	Оплата труда (млн долл.)	Расходы на рекламу (млн долл.)	Расходы на исследования и разработки (млн долл.)
246	386,8118	7,4040	18,4880	47,1213	302,5586	2,9990	18,1716	3,8620
247	69,1530	12,1212	1,6190	48,1415	54,4310	11,3410	15,1413	13,8125
248	81,4650	1,6220	4,1890	16,4950	70,5080	4,4555	12,1110	47,8552
249	329,5518	6,0040	12,2520	8,0540	269,6377	12,1417	9,8765	51,9185
250	36,3870	133,0000	12,7246	51,5355	27,7690	21,8283	4,3210	54,3321
251	344,7937	7,5000	24,7400	57,5982	205,0610	92,9395	8,1234	4,8200
252	22,8030	84,1000	2,1060	83,4952	10,6830	96,9899	5,6788	43,8388
253	196,3030	5,4660	5,9730	-99,9242	142,1520	97,9294	12,4582	2,2710
254	31,5660	13,7124	8,1264	10,1115	22,3750	95,0092	14,5220	66,7274
255	108,8580	1,7000	1,2870	92,4445	45,9130	92,6666	1,4330	53,5422
256	83,6260	1,2320	4,1220	55,6677	45,0950	92,5555	13,5620	22,5673
257	390,8726	6,1660	17,3310	40,5880	296,8577	58,2130	18,0000	10,0000
258	363,9839	7,0160	11,2700	11,5610	234,6320	2,7310	6,4860	86,0000
259	52,2620	0,4420	5,1030	1,1500	43,5110	12,1440	44,0700	16,0000
260	228,6110	5,6500	1,8370	41,5600	161,4700	20,5400	14,9460	3,0730
261	60,8250	1,5000	1,4910	45,3100	41,6820	1,8970	1,6590	30,7300
262	16,6890	40,5000	57,6000	9,8540	9,8450	18,3700	35,2020	63,5300
263	39,8290	62,1000	3,9900	1,5610	32,6580	4,9080	9,2530	30,0000
264	28,9020	93,4000	1,1040	36,5000	23,1410	2,5900	18,3650	10,0000
265	8,7410	27,0000	55,6000	32,1400	6,3700	59,3100	27,6600	56,6660
266	61,9446	7,0000	35,0470	43,2000	432,3777	160,6660	19,2020	6,3530

блица

Заработная плата и другие данные о руководящих работниках 50 компаний США за 2000г.

N	Заработная плата (тыс.долл.)	Премия (тыс. доли.)	Другие выплаты (тыс. долл.)	Всего выплачено (тыс. долл.)	Возраст	Образование'	Профессиональные данные	Общий стаж	Срок пребывания в должности	Рыночная стоимость фирмы (млн долл.)	Процент собственности в фирме	Доходы фирмы (млн долл.)	Объемы продаж фирмы (млн долл.)
1	173	275	5	453	64	2	3	40	26	54,5	3,14	91	872
2	1441	459	78	1948	55	1	1	23	23	7,6	0,55	145	1227
3	1646	0	89	1735	47	2	7	5	5	21,7	0,52	-47	1712
4	294	325	24	643	65	1	3	29	23	8,9	0,89	44	1681
5	1254	105	102	1 461	63	1	6	23	8	3,6	0,05	201	5673
6	325	25	7	357	54	2	5	20	1	0,5	0,06	71	1 117
7	658	0	11	669	61	2	5	2	2	0,7	0,05	-187	1475
8	1723	289	82	2094	63	1	3	41	8	5,9	0,04	1 166	10818
9	504	69	24	597	57	2	5	27	13	1,4	0,03	377	2686
10	822	38	29	889	56	2	1	5	5	0,7	0,03	224	2201
11	374	129	11	514	57	2	4	3	3	4,1	0,17	79	661
12	447	11	8	466	48	2	1	17	1	0,2	0,01	189	1539
13	2781	0	52	2833	50	1	3	4	4	11,7	0,39	-332	11663
14	128	282	17	427	54	1	8	31	15	71,4	10,09	55	2366
15	1 782	0	74	1856	60	2	7	33	3	0,4	0,03	-507	4864
16	1 137	423	92	1 652	60	2	1	34	14	11,5	0,06	856	14949
17	761	20	1	782	49	2	9	18	9	1,3	0,07	14	5061
18	505	0	108	613	56	2	9	8	1	0,1	0,02	-29	1929
19	976	448	64	1488	58	1	3	9	8	9,4	0,56	126	2643

N	Заработная плата (тыс.долл.)	Премия (тыс. доли.)	Другие выплаты (тыс. долл.)	Всего выплачено (тыс. долл.)	Возраст	Образование'	Профессиональные данные	Общий стаж	Срок пребывания в должности	Рыночная стоимость фирмы (млн долл.)	Процент собственности в фирме	Доходы фирмы (млн долл.)	Объемы продаж фирмы (млн долл.)
20	434	12	1	447	50	2	9	5	1	0,3	0,03	54	1084
21	1 010	687	55	1 752	63	1	7	14	14	534,2	3,99	249	5137
22	956	1 452	89	2497	64	0	1	28	28	221,1	3,98	91	844
23	700	37	31	768	60	2	9	30	8	0,7	0,02	322	2097
24	1813	489	40	2342	71	1	4	46	34	9,6	0,83	99	835
25	3396	0	13	3409	64	0	9	30	30	29,4	0,31	-99	12021
26	2 108	38	98	2244	64	2	5	41	5	4,0	0,04	30	4451
27	597	0	4	601	59	2	4	35	5	0,1	0,12	-85	1 911
28	616	862	76	1 554	61	1	S	41	17	30,6	2,23	82	1 435
25	237	221	4	462	61	2	5	25	11	16,8	1,03	27	1 314
30	571	0	16	587	55	2	4	5	5	1,6	0,17	-76	2301
31	269	391	28	688	54	2	9	28	28	1 689,0	34,04	317	3277
32	721	101	71	893	60	2	5	36	15	2,0	0,04	417	4444
33	328	238	34	600	60	0	7	42	1	85,6	17,66	43	1 214
34	538	25	7	570	60	1	4	3	3	0,2	0,21	49	804
35	741	104	9	854	62	1	7	30	3	2,6	0,17	81	669
36	607	380	47	1 034	51	1	6	23	3	7,0	0,83	82	578
37	1 044	107	36	1 187	55	2	8	2	1	3,1	1,21	10	1214
38	2409	1487	143	4039	55	1	7	32	17	35,2	0,29	715	12651

N	Заработная плата (тыс.долл.)	Премия (тыс. доли.)	Другие выплаты (тыс. долл.)	Всего выплачено (тыс. долл.)	Возраст	Образование'	Профессиональные данные	Общий стаж	Срок пребывания в должности	Рыночная стоимость фирмы (млн долл.)	Процент собственности в фирме	Доходы фирмы (млн долл.)	Объемы продаж фирмы (млн долл.)
39	287	198	32	517	59	1	8	37	21	181,0	6,70	136	3180
40	567	15	34	616	51	2	1	16	9	0,3	0,01	237	2754
41	682	0	2	684	62	2	1	36	2	0,8	0,01	-1086	12794
42	1 226	174	56	1456	52	1	4	30	10	6,0	0,17	98	4439
43	952	80	56	1 088	45	2	7	11	11	2,6	0,17	48	415
44	432	0	12	444	50	2	7	25	3	0,1	0,01	-50	1 569
45	1085	440	97	1622	57	1	8	29	5	9,7	0,19	347	9886
46	1009	117	0	1 126	64	1	3	28	14	3,4	0,21	63	2545
47	1711	182	134	2027	62	2	2	25	5	3,4	0,04	806	8379
48	408	183	9	600	52	2	7	12	2	4,2	0,26	10	21351
49	543	13	2	558	62	2	5	34	12	0,2	0,01	265	2359
50	278	209	74	561	61	2	6	24	24	15,4	0,95	52	695

'0 — нет высшего образования; 1 — незаконченное высшее образование; 1 — законченное высшее образование.

Распределение Фишера-Снедекора

Значения $F_{табл}$, удовлетворяющие условию $P(F > F_{табл})$. Значение соответствует вероятности 0,05

Число степеней свободы	К1									
К2	1	2	3	4	5	6	7	8	9	10
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
31	4.16	3.30	2.91	2.68	2.52	2.41	2.32	2.25	2.20	2.15

Закон распределения Стьюдента (t- распределение)

Число степеней свободы	Уровень значимости		
	0,1	0,05	0,01
1	6,3137	12,7062	63,6559
2	2,9200	4,3027	9,9250
3	2,3534	3,1824	5,8408
4	2,1318	2,7765	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	1,7531	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	2,7970
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
50	1,6759	2,0086	2,6778
60	1,6706	2,0003	2,6603
100	1,6602	1,9840	2,6259