Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
**«Владимирский государственный университет
имени Александра Григорьевича и Николая Григорьевича Столетовых»
(ВлГУ)**

Методические рекомендации по выполнению лабораторных работ по дисциплине «Статистические методы в управлении инновациями»

для студентов направления
27.04.05«Инноватика»

Составитель:
доцент кафедры ТМС Иванченко А.Б.

Владимир, 2022

Методические рекомендации, содержащие основные указания по выполнению лабораторных работ по дисциплине «Статистические методы в управлении инновациями» для магистров, обучающихся по направлению 27.04.05 «Инноватика».

Методические рекомендации составлены на основе требований ФГОС ВО и ОПОП направления 27.04.05 «Инноватика». В качестве рекомендаций для организации эффективной работы студентов использованы методические пособия ведущих вузов России.

Рассмотрены и одобрены на заседании
УМК направления 27.04.05 «Инноватика»
и кафедры Технология машиностроения.
Протокол № 1 от 31.08.2022 г.
Рукописный фонд кафедры ТМС ВлГУ

Оглавление

Введение

Современная конкурентная среда продуцирует огромный информационный поток, лишая правильного восприятия действительности.Без современных технологий интеллектуального анализа данных долговременное управление реальными процессами и принятие правильныхрешений невозможно. Статистическое управление позволяет грамотнособрать данные, описать их структуру, понять и увидеть закономерности в массе вероятностных явлений. Статистические методы – это удивительно мощный инструмент управления. Даже простейшие методывизуального анализа позволяют прояснить сложную ситуацию, тщательно скрытую за нагромождением информации, исследовать её и принять доказательное решение.

Цель выполнения лабораторных работ - дать будущим магистрам направлениям «Инноватика» практические навыки по вопросам применения статистических методов в управлении инновациями, при контроле и управлении качеством.

Цель изучения данного курса – формирование у студентов целостного системного представления об управлении качеством как современной концепции управления, а также умений и навыков в области управления качеством продукции, услуг, работ, деятельности отечественных предприятий и организаций.

Данная дисциплина позволит студентам:

- усвоить теоретические знания, формирующие умения и навыки, обеспечивающие квалифицированную профессиональную деятельность;
- развить творческое мышление студентов, повысить их интеллектуальный уровень.

Теоретические и практические знания, получаемые при изучении данного курса, могутбыть использованы в дальнейшем освоении специальных дисциплин, при выполнениидипломной работы и в процессе профессиональной деятельности.

Приобретаемые компетенции.

В процессе освоения дисциплины у студентов развиваются следующие общекультурные, общепрофессиональные и профессиональныекомпетенции:

- способностью к абстрактному мышлению, анализу, синтезу (ОК-1);
- способность решать профессиональные задачи на основе математических методов и моделей для управления инновациями, компьютерных технологий в инновационной сфере (ОПК-3);
- способностью найти (выбрать) оптимальные решения при создании новой наукоемкой продукции с учетом требований качества, стоимости, сроков исполнения, конкурентоспособности и экологической безопасности (ПК-4);
- способностью выполнить анализ результатов научного эксперимента с использованием соответствующих методов и инструментов обработки (ПК-8).

Все лабораторные работы реализуются в процедурах пакета Statistica. Подробно с описанием пакета можно ознакомиться по специальной литературе [2–4].

Лабораторная работа № 1: «Освоение процедур программного пакета Statistica: элементы интерфейса, структура, режим работы».

Цель работы:ознакомиться со структурой и возможностями программного пакета Statistica**,** самостоятельно изучить основные средства описательной статистики в данном программном пакете (диаграмма рассеивания, трехмерный визуальный анализ данных, круговые диаграммы, произвести расчет согласно задания.

Порядок выполнения работы:

1. Осуществить запуск программного пакета Statisticaи ознакомиться с предназначением основных пунктов меню программного продукта.

2. Согласно данным Приложения 1 построить гистограммы (команда *Graphs/Histograms)* для двух переменных на одном графике в зависимости от номера вашего варианта N: VarN, VarN+1. Использовать опцию *Multiple* (несколько графиков на одной сетке) во вкладке *Quick.*

3. Для тех же переменных построить столбчатую диаграмму (Graphs/*2DGraphs/ BarColumnsPlots).*

4. Для переменной VarN построить круговую диаграмму (Pie *Chart -Counts).* Обратите внимание, как строится график *Pie Chart* при изменении переменной *Categories.*

5. Для переменных VarN, VarN+1, VarN+2 построить 3D график (Graphs/ *3DXYZGraphs/ SurfacePlots).*Обратить внимание, что трѐхмерный график можно разворачивать на любой угол в подменю «свойства графика / все свойства». Настроить графики, подписав переменные и оси.

6. Решить с помощью вероятностного калькулятора следующую задачу:

Известно, что в некоторой стране рост мужчин приближенно имеет нормальное распределение со средним 176 см и стандартным отклонением 7,63 см. Какова вероятность того, что рост случайно встреченного вами мужчины будет не менее 186 см?

7. Подготовить отчет о проделанной работе.

Краткая теория.Универсальная интегрированная система, предназначенная для статистического анализа, визуализации данных и разработки пользовательских приложений Statistica – это современный пакет, в котором реализованы все новейшие компьютерные и математические методыстатистического анализа данных. Программа снабжена подсказками, какие методы анализа существуют и какие из них лучше всего подходят для тех или иных задач.

Система избавляет пользователя от рутинных вычислений, наглядно отображает результаты полный набор классических и продвинутых методованализа данных помогает оптимально спланироватьбудущие эксперименты и созда□т высококачественные отч□ты, оставляяспециалисту удовольствие интерпретации результатов и формулировкивыводов.

Statistica позволяет:

—построить различные графики: гистограммы (Graphs/Histograms), графики рассеяния (Graphs/Scatterplots), круговые диаграммы (Graphs/ 2DGraphs/ PieCharts), построить 3D (3DXYZGraphs) и другие графики;

—вычислить вероятность, среднее значение и т. д., построить графики

различных распределений с помощью вероятностного калькулятора (Statistics/Probability Calculator);

—построитьдиаграммуПарето (Statistics/Industrial Statisticas&Six Sigma/ Quality Control Charts/ Pareto chart analysis);

—построитьдиаграммупричинирезультатов (Statistics/Industrial Statisticas&Six Sigma/ Process Analysis/ Cause-effect diagrams);

—Построитьконтрольныекарты (Statistics/Industrial Statisticas&Six Sigma/ Quality Control Charts);

—провестикластерныйанализ (Statistics/ Multivariable Exploratory Techniques/ Cluster Analysis);

—провестинелинейноеоценивание – регрессионныйанализ (Statistics/ Advanced Linear/ Nonlinear Models/Nonlinear Estimation);

—провестикорреляционныйанализ (Statistics/Basic Statistics/Correlation Matrices);

—рассчитатьстатистическиехарактеристикипеременных (Statistics/ Basic Statistics/ Descriptive Statistics);

—провестианализвременныхрядов (Statistics/ Advanced Linear/Nonlinear Model / Time Series Analysis/ Forecasting);

—организовать анализ с помощью других статистических методов, используемых в промышленности для обработки данных.

Набор данных в пакете Statistica – это прямоугольная таблица,столбцам которой соответствуют обрабатываемые переменные (Variables), а строкам отвечают наблюдения (Cases) значений переменных.В отличие от электронной таблицы Excel, где строки и столбцы могутбыть интерпретированы пользователем по собственному желанию.Для создания нового набора данных нужно, прежде всего, завестифайл с трафаретом таблицы нужных размеров. Для этого необходимоиспользовать модуль File/ New.

Визуальные методы анализа данных чрезвычайно важны для предварительного исследования. Многие скрытые явления становятся отч□тливыми, если для них найти подходящее графическое представление.Кроме того, многие сложные задачи решаются чрезвычайно простымиметодами описательной статистики.

График – это черт□ж, показывающий соотношение статистическихвеличин при помощи разнообразных геометрических и изобразительных средств. В пакете Statistica графический анализ проводится черезопцию Graphs.

Контрольные вопросы

1. Для чего необходим сбор данных?
2. Какие типы данных существуют?
3. Что такое «разведочный анализ данных»?
4. Как построить простейшие графики в программе Statistica?
5. Как формируются наборы данных в программе Statistica?
6. Для чего необходимы гистограммы и как их построить в программе Statistica?

Лабораторная работа № 2: «Первичная обработка данных, вычисление элементарных статистик в программном пакете Statistica».

Цель работы:ознакомиться сметодикой разведочного анализа данныхв программном пакете Statistica**,** изучить основные средства построения графиков и вычисление простейшихстатистик.

Порядок выполнения работы:

1. Составить таблицу исходных данных на основании данных ежедневного изменения индекса Московской межбанковской валютной биржи (в течении года) или, либо на основании ежедневного изменения стоимости чистых активов ПИФ ММВБ.
2. Построить гистограмму для выбранной переменной. Сравнить построение гистограммы для разных значений интервалов группировки, которые можно изменить в окне построения гистограммы.
3. Составить новую таблицу, разделив выбранную переменную по месяцам. При этом каждая переменная новой таблицы будет соответствовать одному месяцу. С помощью модуля *Statistics/ Basic Statistics/Tables* исследовать изменение среднего арифметического значения переменной и медианы. Усреднения проводить каждый месяц. Построить графики «ящик с усами» для всех средних значений и медиан (кнопка *Box&Whiskerplotforallvariables*окна *Descriptivestatistics*). На графике соединить средние значения прямыми линиями. Сделать выводы.

Краткая теория. Множество всех обследуемых объектов называется *генеральной совокупностью*. В большинстве случаев в силу того, что генеральнаясовокупность имеет очень много элементов, либо её элементы труднодоступны, обследуется некоторая часть генеральной совокупности –выборка.

Так как значения переменных не постоянны, нужно описывать ихизменчивость. Для этого придуманы описательные или дескриптивныестатистики: *минимум, максимум, среднее, дисперсия, стандартное отклонение, медиана, квартили, мода* и так далее. Идея этих статистикочень проста: вместо того чтобы рассматривать все значения переменной, а их может быть очень много, вначале стоит просмотреть описательные статистики.

Расч□т описательных статистик производится при помощи модуля*Statistics/ Basic Statistics/ Tables.* В этом модуле объединены наиболеечасто использующиеся на начальном этапе обработки данных процедуры. В стартовой панели модуля приводится перечень статистическихпроцедур этого модуля.

При вызове модуля *Descriptive statistics* (Описательные статистики) появляется диалоговое окно, в котором при помощи кнопки*Variables* следует выбрать переменные для анализа. Для построениягистограмм и таблиц частот используются кнопки *Frequency tables* и *Histograms* соответственно. Чтобы выбрать статистики, подлежащиевычислению, нужно воспользоваться вкладкой *Advanced* этого диалогового окна.

Возможен расчёт следующих описательных статистик:

1. Valid N – объем выборки.

2.**Mean** – среднее арифметическое.

3.**Median** – медиана.

3.**Sum** – сумма.

4.**Standard Deviation** – стандартное отклонение.

5. **Variance** – дисперсия.

6. **Standard error of mean** – стандартнаяошибкасреднего.

7. **95 % confidence limits of mean** – 95%-йдоверительныйинтервал для статистического анализа.

8.Minimum, maximum – минимальное и максимальное значения.

9.Lower, upper quartiles – нижняя и верхняя квартили.

10.Quartile range – интерквартильная широта.

11.Range – размах.

12.Skewness – асимметрия.

13.StandarderrorofSkewness– стандартнаяошибкаасимметрии.

14.Kurtosis– эксцесс.

15. StandarderrorofKurtosis– стандартнаяошибкаэксцесса.

Для визуализации описательных статистик можно построить график «ящики с усами». С помощью этого графика можно быстро оценитьданные на предмет структуры распределения, наличия неправдоподобных измерений, однородности наблюдений и так далее.Этолегкоможносделатьприпомощикнопки*Box & Whisker plot for all variables* окна*Descriptive statistics*. Предварительно необходимо обратиться к вкладке *Options* и установить одно из четырёхположений:

Median/Quart./Range – Медиана / Квартили / Размах;

Mean/SE/SD – Среднее / Ошибка среднего / Стандартное отклонение;

Mean/SD/1.96SD – Среднее / Стандартное отклонение / Интервал1,96*стандартного отклонения;

Mean/SE/1.96*SE – Среднее / Ошибка среднего / Интервал 1,96*ошибки среднего.

Контрольные вопросы

1. Дать определение основных описательных статистик.
2. Почему при оценке статистических переменных используетсявыборка из генеральной совокупности?
3. Как рассчитать основные статистики в программе Statistica?
4. Как построить простейшие графики «ящик с усами» в программеStatistica?
5. В чём заключается центральная предельная теорема?
6. Чем характеризуется нормальное распределение данных?
7. Как оценить настройку и наладку технологического процессав программе Statistica?

Лабораторная работа № 3: «Проверка статистических гипотез в программном пакете Statistica».

Цель работы:ознакомиться с методикой проверки статистических гипотезв программном пакете Statisticaна примерах контроля и управления качеством.

Порядок выполнения работы:

1. С помощью пакета Statistica показать утверждение теоремы Хинчина для заданного закона распределения случайных чисел X [0; 1].

Номер варианта	Распределение	Функция в программе
1	Равномерное	Rnd
2	Нормальное	Rndnormal
3	Пуассона	Poisson

2. Сформировать набор данных для последующего анализа в программе Statistica, состоящий из одной переменной и 100 наблюдений (см. Приложение 2).
3. По данной выборке объема n = 150 построить статистический ряд и сгруппированную выборку. Найти выборочное среднее, исправленную выборочную дисперсию, исправленное выборочное среднеквадратическое отклонение.
4. Проверить гипотезу о нормальном распределении случайной величины X с помощью критерия χ^2 критерия в программном пакете Statistica и вручную.
5. Сравнить графически наблюдаемые (Observed Frequency) и ожидаемые частоты (Expected Frequency): записать соответствующие столбцы в отдельную таблицу и построить график рассеяния (команды Graphs/ Scatterplots/ Variables/ OK). Является ли исследуемая переменная нормально распределённой?Почему?
6. Таблицу из восьми переменных (Var1...Var8) и 500 наблюдений заполнить случайными числами из интервала [0; 1]. Найти

Var9=Var1+Var2+...+Var8.

Для этого необходимо дважды щёлкнуть левой кнопкой мыши по Var9 и в появившемся окне «Long name (label orformula)» записать формулу. Построить гистограммы для Var1 и Var9 отдельно. Какие распределения вы получили? Объяснить полученный результат.

7. Для этих же данных проверить гипотезу о нормальном распределении случайных величин Var1 Var9 по критерию χ^2 автоматически, как это делалось в задании 3. Сравнить с результатами Задания 3 и объяснить полученный результат.

Краткая теория.Во многих случаях требуется решить, справедливо ли некоторое суждение. Если мы считаем, что исходные данные для таких суждений в той или иной мере носят случайный характер, то и ответы можно дать лишь с определённой степенью уверенности, и имеется некоторая вероятность ошибиться.

Весь статистический анализ основан на идее случайного выбора.Поскольку имеющиеся данные появились в результате случайного выбора из некоторой генеральной совокупности, то все суждения, основанные на этих данных, будут иметь вероятностный характер.

Рассмотрение вероятностных задач в строгой математической постановке приводит к понятию статистической гипотезы. Термин «гипотеза» означает предположение, которое вызывает сомнения и которое мы собираемся проверить. Проверка гипотез осуществляется с помощью критериев статистической оценки различий.

Статистический критерий – это правило, по которому принимается решение о принятии истинной и отклонении ложной гипотезыс высокой вероятностью. Критерии делятся на параметрические и непараметрические.

Параметрические критерии – это критерии, включающие в формулу расчёта параметры распределения, то есть средние и дисперсии(*t*-критерий Стьюдента, критерий *F* и др.).Непараметрические критерии – это критерии, не включающиев формулу расчёта параметров распределения и основанные на оперировании частотами или рангами (Q-критерий Розенбаума, критерийУилкоксона и др.).

При нормальном распределении признака параметрические критерии обладают большей мощностью, чем непараметрические критерии.

Схема проверки гипотез с помощью статистических критериев состоит из следующих трёх шагов.

1. Вычисляется эмпирическое (или фактическое, реальное) значение критерия $F_{эмп}$. Вычисляется число степеней свободы и уровень значимости.

2. По таблицам критических значений для выбранного критериянаходится так называемая критическая точка (или критическое значение) $F_{кр}$.

3. По соотношению эмпирического и критического значений критерия судят о том, подтверждается или опровергается нулевая гипотеза.

Например, если $F_{эмп}>F_{кр}$, гипотеза H_0 отвергается.В системе Statistica это делается автоматически.

В большинстве случаев для того, чтобы различия признавалисьзначимыми, необходимо, чтобы эмпирическое значение критерия превышало критическое, хотя есть критерии (например, Манна – Уитниили критерий знаков), в которых нужно придерживаться противоположного правила.

Число степеней свободы равно числу классов вариационного рядаминус число условий, при которых он был сформирован. К числу такихусловий относятся объём выборки, средние и дисперсии.

Уровень значимости – это вероятность отклонения нулевой гипотезы, в то время как она верна.Обычно при проверке статистических гипотез принимают триуровня значимости:

- 5%-й (вероятность ошибочной оценки $\alpha = 0,05$),
- 1%-й ($\alpha= 0,01$),
- 0,1%-й ($\alpha = 0,001$).

В промышленной статистике частосчитают достаточным 5%-й уровень значимости. При этом нулевую гипотезу не отвергают, если в результате исследования окажется, что вероятность ошибочности оценки относительно правильности принятойгипотезы превышает 5 %, то есть $\alpha> 0,05$. Если же $\alpha< 0,05$, то принятую гипотезу следует отвергнуть на взятом уровне значимости, Ошибкапри этом возможна не более чем в 5 % случаев, т. е. она маловероятна.

В пакете Statistica значение задаваемого уровня значимости не используется. Как правило, в выходных данных содержатся выборочные значения статистики критерия и вероятность того, что случайная величина превышает это выборочное значение при условии, что верна гипотеза H0. Эта вероятность называется *p*-значением (p-level).

Ошибки при принятии гипотез:

1. Ошибка, состоящая в том, что правильная гипотеза отклонена, в то время как она верна, называется ошибкой I рода.
2. Ошибка, состоящая в том, что правильная гипотеза принята, в то время как она неверна, называется ошибкой II рода.

При приёмочном контроле ошибка первого рода приводитк браковке партии с допустимой долей брака (риск производителя). Приконтроле производства – к вмешательству в налаженный процесс производства (ложная тревога). Ошибка второго рода приводит к принятиюпартии с недопустимой долей брака (риск потребителя). При

контролепроизводства – приводит к вмешательству в процесс производства, вышедший за допустимые границы.

*Гипотезы о виде распределения.*При проверке гипотез о параметрах генеральной совокупностиконтролируемого показателя предполагается, что закон распределенияизвестен. Однако на практике это не всегда имеет место. И тогда необходимо определить, какому закону распределения подчиняется исследуемая случайная величина.Для решения этой задачи используют статистические критерии, называемые *критериями согласия*.

Теория вероятностей позволяет пользоваться несколькими критериями согласия: критерий Пирсона (критерий χ^2), критерий Колмогорова, Смирнова и др.

Порядок действий при реализации критерия Пирсона в программном пакете Statistica:

1. *Statistics/ DistributionFitting*(подбор распределений)/*Continuous Distributions* (непрерывные распределения)/*OK*/*Normal* (нормальное распределение)/ *Variable/ Summary*. На экран выводится таблица для расчёта статистики критерия.
2. Во вкладке *Parameters* того же окна появятся оценки параметров число интервалов группировки (*Number of categories*) можно при необходимости изменить.
3. Для вычерчивания измеряемого и ожидаемого распределения нажимаем соответствующую кнопку (*Plotofobservedandexpected distribution*). Появится гистограмма, вверху которой написано рассчитанное значение χ2 (*Chi-Square test*), число степеней свободы (*df*) и уровень значимости (*p*). Именно *p*-уровень представляет собой вероятность ошибки, связанной с распространением наблюдаемого результата на всю выборку.

Гипотезы об однородности выборки.Пусть имеются выборки, извлечённые из различных совокупностей. Требуется проверить гипотезу о том, что исходные совокупностираспределены одинаково. Всистеме Statistica этагипотезапроверяетсявмодуле*Statistics/Advanced Linear/Nonlinear models/ Log-Linear Analysis ofFrequency Tables*.

Теорема Хинчина: Среднее арифметическое $\overline{X} = \left|\frac{X_1 + X_2 + \ldots + X_n}{n}\right|$ независимых случайных величин X_j, j=1, 2,..., n, имеющих одно и тожераспределение и конечное математическое ожидание m, сходится по вероятности при n–>∞ к m. Таким образом, при заданном □и достаточнобольшом nсобытие

$$\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - m\right| < \varepsilon$$

можно считать практически достоверным.

*Указание.*Постепенно увеличивая n, показать обоснованность теоремы Хинчина методом математической индукции. Равномерно распределенные случайные числа генерируются программно (Fill/StandardizeBlock / FillRandomValues). Среднее значение вычисляется тем же способом командойStatisticsofBlockData / BlockColumns / Means.

Контрольные вопросы

1. Что называется статистической гипотезой?
2. Какие критерии согласия вам известны?
3. Как проверить гипотезу о принадлежности к распределениюв программе Statistica?
4. Как проверить гипотезу об однородности выборок в программе Statistica?

Лабораторная работа № 4: «Корреляционный анализ в программе Statistica».

Цель работы: ознакомиться с назначением и ключевыми элементами корреляционного анализа, то есть установлению факта взаимной зависимости случайных величин.

Порядок выполнения работы:

1. Ознакомиться с краткой теорией – основами корреляционного анализа.
2. Спомощьюмодуля*Statistics/BasicStatistics/CorrelationMatrices*рассчитатькоэффициентыкорреляциидляпеременных, состоящихизстрокVarN, VarN+1 иVarN+2 (см. Приложение 3), гдеN – номервариантa. Выбиратьопции*OnevariablelistuSummary: Correlationmatrix* (таблицаскоэффициентамикорреляции); *Scatterplotmatrixforselectedvariables* (графическоеотображениезависимостей). Построитькорреляционную матрицу. Сделать выводы о взаимной зависимости переменных.
3. Изменить те же данные следующим образом. Первуюпеременную (пусть это Var1) оставить неизменной, а Var2 сделать равной 2*Var1; Var3 сделать равной 2*Var1+Var12. Рассчитать коэффициенты корреляции, построить корреляционнуюматрицу. Сделать выводы о взаимной зависимости переменных.

Краткая теория.Основная задача корреляционного анализа состоит в выявлении связи между случайными переменными. Например,на свободном рынке обычно наблюдается большая степень корреляциимежду размером урожая и рыночными ценами на соответствующуюпродукцию сельского хозяйства. Часто корреляция привлекает нашевнимание к причинно-следственным связям, существующим междуизучаемыми двумя рядами величин. В области естественных и общественных наук установление существенной корреляции часто заставляетнас искать возможные связи между явлениями, которые в противномслучае могли остаться незамеченными.

В экономике в большинстве случаев между переменными величинами существуют зависимости, когда каждому значению одной переменной соответствует не какое-то определенное, а множество возможных значений другой переменной. Иначе говоря, каждому значениюодной переменной соответствует определенное условное распределениедругой переменной. Такая зависимость получила название статистической.

Возникновение понятия статистической связи обусловливается тем,что зависимая переменная подвержена влиянию неконтролируемых илинеучтенных факторов, а также тем, что измерение значений переменныхнеизбежно сопровождается некоторыми случайными ошибками.

Статистическая зависимость между двумя переменными, при которой каждому значению одной переменной соответствует определенноеусловное математическое ожидание (среднее значение) другой, называется корреляционной.

Функциональная зависимость представляет собой частный случайкорреляционной. При функциональной зависимости с изменением значений некоторой переменной x однозначно изменяется определенноезначение переменной y, при корреляционной – определенное среднеезначение (математическое ожидание) y, а при статистической – определенное распределение переменной y. Каждая

корреляционная зависимость является статистической, но не каждая статистическая зависимость является корреляционной.

Статистические связи между переменными можно изучать методами корреляционного и регрессионного анализа. Основной задачей корреляционного анализа является выявление связи между случайными переменными и оценка ее степени.

Корреляция определяет степень, с которой значения двух переменных «пропорциональны» друг другу. Пропорциональность означаетпросто линейную зависимость. Корреляция высокая, если на графикезависимость «можно представить» прямой линией (с положительнымили отрицательным углом наклона).

В производственных условиях обычно информации, полученной издиаграмм рассеяния при условии их корректного построения, бываетдостаточно для того, чтобы оценить степень зависимости y от x. Но вряде случаев требуется дать количественную оценку степени связи между величинами x и y. Такой оценкой является коэффициент корреляции.

Коэффициент корреляцииr – это показатель, оценивающий тесноту линейной связи между признаками.

Отметим основные характеристики этого показателя:

1. Он может принимать значения от –1 до +1. Знак «+» означает, что связь прямая (когда значения одной переменной возрастают, значения другой переменной также возрастают), «–» означает, что связь обратная.
2. Чем ближе коэффициент к 1,величине коэффициента корреляции менее 0,3 связь оценивается какслабая, от 0,31 до 0,5 – умеренная, от 0,51 до 0,7 – значительная, от 0,71до 0,9 – тесная, 0,91 и выше – очень тесная.
3. Если все значения переменных увеличить (уменьшить) на однои то же число или в одно и то же число раз, то величина коэффициентакорреляции не изменится.
4. При $r = \pm 1$ корреляционная связь представляет линейнуюфункциональную зависимость. При этом все наблюдаемые значениярасполагаются на общей прямой.
5. При $r = 0$ линейная корреляционная связь отсутствует. Приэтом групповые средние переменных совпадают с их общими средними,а линии регрессии параллельны осям координат.
6. Равенство $r = 0$ говорит лишь об отсутствии линейной корреляционной зависимости (некоррелированности переменных), но не вообще об отсутствии корреляционной, а тем более, статистической зависимости.

Основываясь на коэффициентах корреляции, вы не можете строгодоказать причинной зависимости между переменными, однако можетеопределить ложные корреляции, т. е. корреляции, которые обусловленывлияниями «других», остающихся вне вашего поля зрения переменных.Основная проблема ложной корреляции состоит в том, что вы не знаете,кто является ее носителем. Тем не менее, если вы знаете, где искать, томожно воспользоваться частные корреляции, чтобы контролировать(частично исключенное) влияние определенных переменных.

Корреляция, совпадение или необычное явление сами по себе ничего не доказывают, но они могут привлечь внимание к отдельным вопросам и привести к дополнительному исследованию. Хотя корреляцияпрямо не указывает на причинную связь, она может служить ключом кразгадке причин. При благоприятных условиях на ее основе можносформулировать гипотезы, проверяемые экспериментально, когда возможен контроль других влияний, помимо тех немногочисленных, которые подлежат исследованию.

Очень важно установить логическую связь между двумя рядамиявлений или двумя совпадающими во времени явлениями, либо же датьим разумное объяснение.Иногда вывод об отсутствии корреляции важнее наличия сильнойкорреляции. Нулевая корреляция двух переменных может свидетельствовать о том, что никакого влияния одной переменной на другую несуществует, при условии, что мы доверяем результатам измерений.

Корреляционный анализ в программе Statistica проводят с помощьюмодуля*Statistics/ BasicStatistics/ CorrelationMatrices.*В стартовом окне для расчета квадратной матрицы используется кнопка *Onevariablelist.*Спомощьюкнопки*Twolists (rect. matrix)*можноограничиться выводом только необходимых переменных, если не требуются все возможные парные корреляции. Из списка выбирают переменные, между которыми будут рассчитаны парные коэффициенты корреляции. После нажатия на кнопку *Summary* или *Correlations* на экране появится корреляционная матрица.

Процедура *Correlation matrices*сразу же дает возможность проверить достоверность рассчитанных коэффициентов корреляции. Значение коэффициента корреляции может быть высоким, но не достоверным, случайным.

На практике часто изучают связи между порядковыми переменными, измеренными в так называемой порядковой шкале. В этой шкале можно установить лишь порядок, в котором объекты выстраиваютсяпо степени проявления признака (например, качество жилищных условий, тестовые баллы, экзаменационные оценки). Если, скажем, по некоторой дисциплине два студента имеют оценки «отлично» и «удовлетворительно», то можно лишь утверждать, что уровень подготовки поэтой дисциплине первого студента лучше, чем второго, но нельзя сказать, на сколько.

Оказалось, что в таких случаях проблема оценки тесноты связиразрешима, если упорядочить, или ранжировать объекты анализа постепени выраженности измеряемых признаков. При этом каждому объекту присваивается определенный номер, называемый рангом. Например, объекту с наименьшим проявлением (значением) признака присваивается ранг 1, следующему за ним – 2 и т. д. Объекты можнорасполагать и в порядке убывания проявления признака.

Ранжируя попарно связанные значения признаков, можно видеть,как они распределяются относительно друг друга. Если возрастающимзначениям одного признака соответствуют возрастающие значениядругого, то между ними существует положительная связь. Если же привозрастании значений одного признака значения другого последовательно уменьшаются, это указывает на наличие отрицательной связимежду ними. При отсутствии корреляции ранжированным значениямодного признака будут соответствовать самые различные значениядругого.

Определив ранги значений переменных, по их разностям можносудить о степени зависимости одного признака от изменений другого.Коэффициент ранговой корреляции Спирмена находится по формуле:

$$p = 1 - \frac{6\sum_{i=1}^{n}(r_i - s_i)^2}{n^3 - n},$$

где r_i и s_i – ранги i-го объекта по переменным x и y, n – число пар наблюдений (объем выборки). Если ранги всех объектов равны ($r_i = s_i, i = 1,2,\ldots,n$), то p = 1, то есть наблюдается полная прямая связь.

Коэффициент ранговой корреляцииКендаллатвычисляется по формуле:

$$\tau = 1 - \frac{4k}{n(n-1)},$$

где k – число инверсий (нарушений порядка) в ряду рангов второй переменной при условии, что ранги первой переменной упорядочены.

В пакете Statistica коэффициенты ранговой корреляции Кендаллаи Спирмена вычисляется в процедуре *Statistics/ Nonparametrics,* в появившемсястартовомокневыберитепункт*Correlations* (*SpearmanKendalltau,gamma)* / кнопка *OK.*

Контрольные вопросы

1. Какие виды зависимостей вам известны? Чем они характеризуются?
2. Что такое коэффициент корреляции, какие значения он можетпринимать?
3. Как коэффициент корреляции связан со степенью связи междупеременными?
4. Как провести корреляционный анализ в программе Statistica?

Лабораторная работа № 5: «Регрессионный анализ в программе Statistica».

Цель работы: ознакомиться с назначением и ключевыми элементами регрессионного анализа, то есть установлению вида взаимной зависимости случайных величин.

Порядок выполнения работы:

1. Ознакомиться с порядком проведения регрессионного анализа в программном пакете Statistica (см. краткая теория).
2. В таблице приведены данные о работе ипростое всего парка землеройной техники (в машино-часах).

Месяц	Простой	Работа
ноябрь	1130,01	4137,63
декабрь	734,42	3704
январь	265,4	1328,4
февраль	586,6	1961,6
март	666,7	1939,7
апрель	1232	3116
май	3888,35	8509,35
июнь	5465,39	12588,89
июль	7412,33	14875,5
август	7168,66	15388,08
сентябрь	7416,68	15450,67
октябрь	5001,41	11944,82

Необходимо:

2.1. Построить линейную зависимость времени простоя техники от времени работы и месяца.

2.2. Срогнозировать время простоя на следующий период и оценитьнедополученную прибыль в результате простоев.
3. Требуется выявить зависимость аварий на дорогах от количества автотранспорта для некоторого региона на основе результатов ежегодных наблюдений:

Год	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Количество аварий на дорогах	166	153	177	201	216	208	227	238	268	268	274
Количество зарегистрированных транспортных средств	352	373	411	441	462	490	529	577	641	692	743

4.Исходные данные приведены в Приложении 4.Считать первый столбец независимой переменной (Argument), а остальные (Var-N) – зависимой переменной. Номер зависимой переменной соответствует вашему варианту по спискугруппы. Построить график зависимости второй переменной Var-N отпервой: *Graphs / 2D Scatterplot;* во вкладке *Advanced* выбрать *Off,* в опции «свойства графика» соединить точки линиями. Выбирая в настройках *Advanced* аппроксимирующую функцию, определить, какая функцияявляется наилучшей аппроксимацией для предложенных данных. Объяснить, почему.

5. В модуле "Nonlinear Estimation" – «Нелинейное оценивание» собраны процедуры, позволяющие оценить нелинейные зависимости между данными. Вы можете выбрать различные модели зависимостей, задать собственную функцию, выбрать метод оцениваниянеизвестных параметров.

Провести нелинейный регрессионный анализ данных задания 4.Дляэтоговоспользоватьсямодулем *Statistics / Advanced Linear/ Nonlinear models / Nonlinear estimation / user-specified regression, least squares.*Кнопкой*Function to be estimated*ввестимодельнуюфункцию. Кнопкой *Variables* ввести зависимую переменную.
Найти коэффициенты нелинейной регрессии. Провести анализ остатков и показать, что найдено наилучшее решение.

Краткая теория.Регрессионный анализ является одним из наиболее распространенных методов обработки экспериментальных данных при изучениизависимостей в физике, биологии, экономике, технике и других областях. Он заключается в определении аналитического выражения, в котором изменение одной величины (называемой зависимой или результативным признаком) y обусловлено влиянием одной или несколькихнезависимых величин (факторов) x_1, x_2, …, x_n, а множество всех прочихфакторов, также оказывающих влияние на зависимую величину, принимается за постоянные и средние значения.

Регрессия может быть однофакторной (парной) и многофакторной(множественной). Если в качестве объясняющих факторов использоватьтолько три фактора x_1, x_2, x_3, то регрессионная модель может быть записана в виде:

$$y = f(x_1, x_2, x_3) + \varepsilon,$$

где f (x_1,x_2,x_3) – неслучайная составляющая отклика y, зависящая от x_1, x_2, x_3, а ε – остаток или случайная составляющая, обусловленная влиянием на отклик y множества

неучтённых и непредсказуемых факторов, а также ошибок измерений зависимой переменной.

Для простой (парной) регрессии в условиях, когда достаточно полно установлены причинно-следственные связи, можно использовать графическое изображение. При множественности причинных связей невозможно чётко разграничить одни причинные явления от других.

В этом случае наиболее приемлемым способом определения зависимости (уравнения регрессии) является метод перебора различных уравнений, реализуемый с помощью компьютера. Существуют различные регрессионные модели, определяемые выбором функции $f(x_1,x_2,x_3)$:

— простая линейная регрессия $y = b_0 + b_1 x + \varepsilon$;

— множественная регрессия $y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_{m-1} x_{m-1} + \varepsilon$;

— полиномиальная регрессия $y = b_0 + b_1 x + b_2 x^2 + \ldots + b_{m-1} x^{m-1} + \varepsilon$;

— регрессионная модель общего вида:

$$Y = b_0 + b_1 f_1(x_1,x_2,\ldots,x_n) + \ldots + b_{m-1} f_{m-1}(x_1,x_2,\ldots,x_n) + \varepsilon ,$$

где $f_i(x_1,x_2,\ldots,x_n)$, $i = 1,\ldots,m-1$ – заданные функции факторов.Параметры b_0, b_1, …, b_{m-1} называются коэффициентами регрессии.

В приведённые регрессионные модели коэффициенты b_0, b_1, …, b_{m-1}входят линейно. Такие модели называют линейными, а математическиеметоды анализа этих моделей – линейным регрессионным анализом.В некоторых случаях нелинейные модели с помощью специальныхлинеаризующих преобразований могут быть преобразованы в линейные.

После выбора вида регрессионной модели, используя результатынаблюдений зависимой переменной и факторов, нужно вычислитьоценки (приближённые значения) параметров регрессии, а затем проверить значимость и адекватность модели результатам наблюдений.

Порядок проведения регрессионного анализа следующий:

1. выбор модели регрессии, что заключает в себе предположениео зависимости функций регрессии от факторов;
2. оценка параметров регрессии в выбранной модели методом наименьших квадратов;
3. проверка статистических гипотез о регрессии.

В программном пакете *Statistica* линейный регрессионный анализ выполняется в модуле Statistics/Multiple Regression. В стартовом диалоговом окне этого модуля при помощи кнопки Variables указываются зависимая (*dependent)* и независимые (*independent*) переменные. В поле *Input file* указывается тип файлас данными:

Raw Data – данные в виде строчной таблицы (по умолчанию);

Correlation Matrix – данные в виде корреляционной матрицы.

Для вывода результатов и их анализа нажмите на кнопку ОК. Система произведет вычисления, и на экране появится окно результатов:

Dependent – имя зависимой переменной.

No. of cases – число наблюдений, по которым построена регрессия.

Multiple R – коэффициент множественной корреляции. Эта статистика полезна в множественной регрессии, когда вы хотите описать зависимости между переменными. Она может принимать значения от 0 до1 и характеризует тесноту линейной связи между зависимой и всеми независимыми переменными.

R? – квадрат коэффициента множественной корреляции (R^2), называемый коэффициентомдетерминации:

$$R^2 = \frac{SSR}{SST}.$$

Коэффициент детерминации является одной из основных статистикв данном окне, он показывает долю общего разброса (относительно выборочного среднего зависимой переменной), которая объясняется построенной регрессией. Чем ближе коэффициент детерминации к единице, тем качественнее найдена модель (объясняет поведение большегочисла точек).

Коэффициент детерминации обладает одним существенным недостатком. При равенстве числа независимых переменных q числу наблюдений n величина R^2 равна 1. По мередобавления переменных в уравнение значение R^2 неизбежно возрастает.

Это ведет к неоправданному предпочтению моделей с большим числомнезависимых переменных. Отсюда следует, что необходима поправкак R^2, которая бы учитывала число переменных и наблюдений. В результате получаем скорректированный коэффициент детерминации (*adjusted R?*) $\overline{R}^2$:

$$\overline{R}^2 = 1 - \frac{n-1}{n-q-1}(1-R^2).$$

Включение новой переменной в регрессионное уравнение увеличивает R^2 не всегда, а только в том случае, когда частный *F-критерий* при проверке гипотезы о значимости включаемой переменной больше или равен 1. В противном случае включение новой переменной уменьшает значение коэффициентов детерминации. Таким образом, скорректированный $\overline{R}^2$ можно с большим успехом (по сравнению с R^2) применять для выбора наилучшего подмножества независимых переменных в регрессионном уравнении.

F-критерий используется для оценки адекватности регрессионной модели, определяет отношение дисперсии оценки модели к дисперсии остатка и равен:

$$F = \frac{SSR/q}{SSE/(n-q-1)},$$

где *SSE* – сумма квадратов остатков.

Всякая сумма квадратов связана с числом степеней свободы. Эторазность между числом различных опытов и числом констант, найденных по этим опытам независимо друг от друга. Например, для *SSE* число степеней свободы равно числу опытов *n* минус (q + 1) коэффициентов регрессии.

Standard Error of estimate – стандартная ошибка оценки. Эта статистика является мерой рассеяния наблюдаемых значений относительнорегрессионной прямой.

Intercept – оценка свободного члена регрессии. Значение коэффициента b_0 в уравнении регрессии.

Std. Error – стандартная ошибка оценки свободного члена. Стандартная ошибка коэффициента b_0 в уравнении регрессии.

F – значения F-критерия для проверки гипотезы $b_1 = 0$.

df – число степеней свободы F-критерия.

p – уровень значимости.

t – *t-критерий* для проверки гипотезы о равенстве нулю свободногочлена уравнения.

Beta – коэффициенты *b* уравнения.

В информационной части прежде всего нужно смотреть на значение коэффициента детерминации. В нашем примере он равен 0,988...

При помощи кнопок диалогового окна *Multiple Regressions Results* результаты регрессионного анализа можно просмотреть болеедетально. Щелкните далее на кнопку *Summary:Regression rezults* (краткие результаты регрессии). Вы увидите таблицу с результатами анализа.

Во втором столбце таблицы (Beta) выводятся стандартизованныекоэффициенты регрессии, в третьем (Std.Err. of Beta) – их стандартныеотклонения. В случае множественной регрессии стандартизованные коэффициенты регрессии используются для сравнения влияния на зависимую переменную факторов, имеющих различную размерность.

Для оценки адекватности модели необходимо исследовать остатки. Остатки – это разность между исходными (наблюдаемыми) значениями зависимой переменной и предсказанными (модельными, Predicted values) значениями. Остатки должны быть нормально распределены, иметь нулевое среднее значение и постоянную дисперсию, независимо от величин зависимых и независимых переменных. Модель должна быть адекватна на всех отрезках интервала изменения зависимой переменной.

Вначале для оценки адекватности модели лучше всего использовать визуальные методы и затем, если потребуется, перейти к статистическим критериям.

Для исследования остатков в окне результатов регрессионного анализа необходимо выбрать вкладку Residuals/ assumptions/ prediction и нажать кнопку Performresidualanalysis. Для оценки адекватности модели построим нормальный вероятностный график остатков. В отобразившемся окне, перейдя к вкладке Quick, необходимо нажать кнопку Normalplotofresiduals.

Для выявления нестабильности дисперсии ошибки уравнения можно построить график зависимости регрессионных остатков от предсказанного значения зависимой переменной. Во вкладке Scatterplots нажмите кнопку Predictedvs. residuals.

Рассмотрим наиболее характерные формы этих графиков.

1. *Выделяющиеся точки графика*: некоторые из остатков могут поабсолютной величине сильно превосходить все остальные остатки. Еслимаксимальное значение остатка больше некоторого заранее выбранногочисла, то наблюдение, имеющее такой остаток, является аномальным(выброс). Величина остатка аномального наблюдения зависит от объемавыборки *n*.

Задача исключения аномальных данных не простая. С одной стороны, одно единственное такое наблюдение может обесценить все результаты регрессионного анализа. Тогда эти наблюдения нужно удалить.

С другой стороны, автоматическое удаление резко выделяющихся наблюдений без установления причин их возникновения оправдано тольков хорошо обкатанных регрессионных моделях, в которых основной интерес представляют только большинство данных.

2. Если остатки попадают в горизонтальную полосу с центром на оси абсцисс, модель можно рассматривать как адекватную.

3. Если эта полоса непрерывно расширяется, когда y возрастает, это указывает на отсутствие постоянства дисперсии. В этом случае нужно стабилизировать дисперсию путем преобразований или перейти к взвешенному метод наименьших квадратов.

4. Если эта полоса имеет вид линейного тренда (возрастает или убывает), то анализ ошибочен. Отрицательные остатки соответствуют малым значениям предсказанных значений y, положительные остатки – большим значениям. Этот

результат может получиться и при ошибочном пропуске свободного члена b_0. Другой способ исправить ситуацию – включить в регрессионную модель фактор, зависящий от номера наблюдения (или времени).

5. Если график имеет криволинейный вид, то модель неадекватна. В регрессионной модели не учтены факторы, оказывающие существенное влияние на зависимую переменную y. Необходимо вводить дополнительные члены (например, квадратичные и взаимодействия) или провести преобразование наблюдений, а затем повторить все вычисления и анализ остатков.

Очень удобным визуальным способом оценки адекватности регрессионной модели является анализ графика опытных и полученных по регрессионному уравнению значений зависимой переменной. Он строится при помощи кнопки Predicted vs. observed окна анализа остатков.

Важно просмотреть графики зависимости остатков от каждой из независимых переменных. Эти графики полезны для обнаружения нелинейной зависимости от переменных. Их легко просмотреть при помощи кнопки *Residualsvs. independentvar.* вкладки *Residuals*.

Остатки должны быть нормально распределены, т. е. на графике они должны представлять приблизительно горизонтальную полосу одинаковой ширины на всем ее протяжении. Коэффициент корреляции (*r*) между регрессионными остатками и переменными должен равняться нулю.

Присутствие нелинейного тренда в регрессионных остатках вызывает сомнение в адекватности модели и говорит о необходимости пересмотра модели – преобразования или ввода новых переменных, перехода к нелинейной модели.

Кнопка *Redundancy* предназначена для поиска выбросов. Выбросы – это остатки, которые значительно превосходят по абсолютной величине остальные. Выбросы дают данные, которые являются не типичными по отношению к остальным данным и требуют выяснения причин их возникновения. Выбросы должны исключаться из обработки, если они вызваны ошибками измерения. Для выделения выбросов, имеющихся в регрессионных остатках, предложены следующие метрики:

1. Расстояние Р.Д. Кука (*Cook's Distance*) показывает расстояние между коэффициентами уравнения регрессии после исключения из обработки каждой точки данных. Большое значение показателя Кука указывает на сильно влияющее наблюдение.

2. Расстояние Махаланобиса (*Mahalanobis Distance*) показывает,насколько каждое наблюдение отклоняется от центра статистическойсовокупности.

Просмотр величин остатков и специальных критериев для ихоценки осуществляется при помощи кнопки Summary окна исследования остатков.

Поиск наилучшей регрессионной модели – это громоздкий процесс. При помощи опции *Method* пользователь может отказаться от стандартного регрессионного анализа (Standard) и воспользоваться методами пошагового включения переменных в регрессионную модель (*Forward stepwise*) или пошагового исключения переменных (*Backward step wise*) из регрессионной модели. Эти методы можно использовать в сложных системах с большим числом переменных. Опция *Displaying results* вкладки *Stepwise* позволяет просматривать итоговые результаты регрессионного анализа (Summary *only)* или после каждогошага включения или исключения переменных (At *each step).*

*Гребневая регрессия*отличаетсяустойчивостью для случаев сильной коррелированности зависимых переменных друг с другом. В отличие от метода наименьших квадратов,дающего несмещенные оценки коэффициентов уравнения, в методегребневой регрессии оценки смещенные, но при этом они имеют меньшую

дисперсию. Поэтому такие оценки могут давать более точные иприемлемые для практического использования модели.

Для расчета коэффициентов уравнения гребневой регрессии следует отметить чекбокс в опции Ridge regression диалогового окна ModelDefinition.При практическом использовании метода гребневой регрессии одним из основных вопросов является выбор параметра λ (lambda). Существуют численные методы расчета этого параметра, но чаще используют простой опытный подход: начинают расчет при λ= 0, увеличиваютпараметр с малым шагом, например, 0,001 и следят за ошибкой регрессии и коэффициентами уравнения. Ошибка не должна увеличиваться, акоэффициенты должны стабилизироваться и при дальнейшем увеличении параметра мало изменяться. Значение принятого параметра является мерой смещения оценок от истинного значения, поэтому стараютсяне придавать слишком больших значений. Обычно выбирают меньше 0,5. Приλ=0 уравнение имеет коэффициенты классического метода*наименьших квадратов.*

Контрольные вопросы

1. Как провести линейный регрессионный анализ в программе Statistica?
2. Что такое фиксированная регрессия в программе Statistica?
3. Как интерпретировать результаты регрессионного анализа в программе Statistica?
4. Что такое остатки и как по ним оценить адекватность регрессионного анализа?
5. Что такое коэффициент детерминации? Как он связан с точно стью регрессионной модели?
6. Пояснить отличительные особенности гребневой регрессии.

Лабораторная работа № 6: «Статистические методы контроля качества».

Цель работы: ознакомиться с основными статистическими методами контроля качества предприятия продукции, с анализом процессов методами построения карт контроля качества.

Порядок выполнения работы:

1.В системе Statistica построить диаграмму причин и результатовспомощьюмодуля*Statistics / IndustrialStatistics&SixSigma/ProcessAnalysis/ Cause–Effect [Ishikawa, Fishbone] diagrams.*

2. Спомощьюмодуля *Statistics/ Industrial Statistics & Six Sigma/ Quality Control Charts/ Pareto chart analysis*построитьдиаграммуПарето. Причины и их число выбрать самостоятельно, можно любые, неотносящиеся к какому-либо процессу. Диаграммы построить:

2.1. только для переменной 1 «причина», без учета значимости причины для общего вклада в качество анализируемого процесса. Для этогона вкладке*Quick* выбрать*Codes (requirestabulationofdatabycodes)* –установлен по умолчанию, а в окне выбора переменных указать однупеременную с перечнем причин;

2.2. для переменных 1 «причина» и 2 «число», с учетом значимостипричины для общего вклада в качество анализируемого процесса. Дляэтогонавкладке *Quick*выбрать *Codes and counts (one variable with defect type, one variable with counts).*
В чем состоит разница между этими картами Парето?

3. С целью выяснения причин брака составлен контрольный листок в предположении, что причинами могут быть рабочий, станок или смена.

Распределение дефектов по рабочим, станкам и сменам

Рабочий	Станок	1 смена	2 смена	3 смена	Число дефектов на станках	Сумма дефектов рабочего
А	А 1			•	1	24
	А 2	• •	•		3	
	А 3	•	• • • • • • •	• • • • • • • • • • • •	20	
Б	Б 1	• •	• • • • • • •	• • • • • •	15	45
	Б 2	•	•	• • •	5	
	Б 3	• •	• • • • • • • •	• • • • • • • • • • • • • • •	25	
В	В 1	•	•	• •	4	52
	В 2	• •	• •	• • • •	8	
	В 3	• • •	• • • • • • • • • • • • • • • • •	• • • • • • • • • • • • • • • • •	40	

Определить, кто виноват, если это возможно.

4. Спомощьюмодуля *Statistics / Industrial Statistics & Six Sigma/Quality Control Charts/ Individuals & moving range*построитьконтрольнуюX-R карту индивидуальных значений для одной переменной своего варианта (VarN) согласно Приложения 5 и провести анализ качества производственного процесса. Проверить машинное построение карты расчетом средней линии.

5.Спомощьюмодуля *Statistics / Industrial Statistics & Six Sigma/Quality Control Charts/ X bar & R chart for variables*построитьконтрольнуюX-R карту производственного процесса для одной переменной своего варианта с группировкой по времени (VarN, Time) согласно Приложения 5 и провести анализ качества процесса. Сравнить построение карты с результатамипредыдущего задания. Чем отличается построение карты индивидуальных значений от карты процесса?

6.Спомощьюмодуля *Statistics / Industrial Statistics & SixSigma/ Quality Control Charts/ X-bar & R Chart for variables*построитьконтрольнуюX-R карту процесса для одной переменной своего варианта (VarN) и временных интервалов (Time, Day)

согласно Приложения 5 и провести анализ качества производственного процесса.

7. Построить контрольную С-карту индивидуальных значений для одной переменной своего варианта (VarN) и провести анализкачества процесса.

Краткая теория.Качество относится к числу важнейших критериев функционирования предприятия в условиях относительно насыщенного рынка и преобладающей неценовой конкуренции. Эффективно управлять процессами производства – значит активно использовать экономические и организационные рычаги воздействия на разработку, производство и эксплуатацию изделий. Качество продукции обеспечивается в первую очередь самим изготовителем на всех этапах жизненного цикла, начиная с проектирования и разработки, а также непрерывно в процессе производства. Для того чтобы выпускать продукцию высокого технического уровня и качества, необходимо эффективно управлять процессами формирования этих комплексных и обобщающих характеристик изделий.

При осуществлении контроля качества производится обязательный сбор данных, а затем их обработка. Но данные, касающиеся даже одного и того же параметра изделия, не могут быть многократно получены при идентичных условиях, так как в ходе процесса меняются отдельные процессы и обстоятельства. Поэтому при операциях, относящихся к контролю качества, приходится иметь дело с большим числом данных, характеризующих те или иные параметры изделия, условия процесса и т. д. Эти данные при повторных измерениях всегда оказываются несколько отличающимися от полученных в другое время и при других условиях, то есть всегда наблюдается разброс данных. Анализируя разброс данных, можно найти решение возникшей в процессе производствапроблемы.

При повторяющихся рабочих процессах, прежде всего при серийном и массовом производстве, определенные факторы снижения качества становятся типичными. Использование математико-статистическихметодов дает возможность исследовать протекание технологическогопроцесса.

В таком случае говорят, что процесс изготовления является статистически управляемым. Статистические методы позволяют обнаружить:где, когда, кем и при каких обстоятельствах вызваны те или иные помехи в производственном процессе. Это повышает чувство ответственности всех участников производственного процесса, способствует тесномусотрудничеству и рождает новое отношение к понятию «качество».

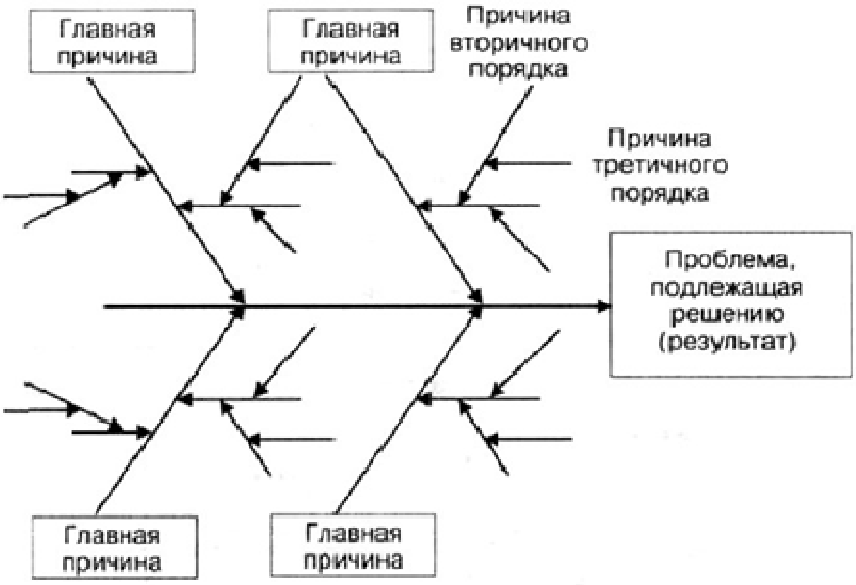
Диаграмма причин и результатов. Когда решается задача анализа возможных причин, ответственныхза тот или иной дефект или проблему, целесообразно эти причины определенным образом упорядочить, провести их классификацию, выявить максимально возможное их количество без риска упустить какую-нибудь из них. При этом очень важно обеспечить наглядность, т. е. ситуацию, при которой все причины и их отношение к результату постоянно находились бы в поле зрения.

Объектами исследования с помощью причинно-следственных диаграмм могут быть: появление дефектности изделий, увеличение расходов на устранение брака, падение спроса на продукцию на рынке,управление персоналом и т. д.

Диаграмму причин и результатов впервые внедрил в производственную практику профессор Токийского университета Каору Исикава(1953 г.).

Определение: Диаграмма причин и результатов — это диаграмма, котораяпоказывает отношение между показателями качества и воздействующими на него факторами.

Безусловно, это один из наиболее элегантных и широко используемых методов среди так называемых семи простых инструментов контроля качества. Иначе диаграмму Исикавы называют причинно-следственной диаграммой или «рыбий скелет».



Для построения причинно-следственной диаграммы данные заносятся в таблицу рабочего окна:

	Варка супа				
	1 Продукты	2 Технология варки	3 Условия варки	4 Повар	5 Оборудов. кухни
1	Вода	Закрытая крышка кастрюли	Время	Опыт работы	Исправная плита
2	Соль	Сначала варить бульон	Температура плиты	Квалификация	Большая кастрюля
3	Мясо	Снимать пену при варке бульона		Знание рецепта	Острый нож
4	Картофель	Нарезать продукты ножом			Доска для нарезки
5	Морковь	Не допускать разваривания			
6	Лук				
7	Лапша				
8	Приправы				

Затем в основном рабочем окне системы в выпадающем меню выберите команду *Statistics/ IndustrialStatistica&SixSigma/ ProcessAnalysis*. В появившемся окне выберите команду *Cause-effect (Ishikawa,Fishbone) diagrams* и нажмите OK. Появится окно, в котором с помощью кнопки *Variables* необходимо отметить,какие факторы будут находиться вверху «хребта рыбы», а какие внизу.

С помощью вкладок Arrows и Font sizes можно выбрать размер шрифтовдля надписей, толщину и угол наклона линий «костей». Все линии и надписи на диаграмме можноизменить и передвинуть.

Построенную диаграмму Исикавы необходимо постоянно совершенствовать. Это позволяет получить действительно ценный документ, который поможет в решении и других проблем, которые могут возникнуть в дальнейшем не только в связи с рассматриваемым показателем качества, но и при возникновении других дефектов или несоответствий.

Дальнейшая работа будет состоять в том, чтобы на основе наблюдений за реальным процессом, в результате которого потеря качества,установить действительную связь между исследуемым показателем качества и выбранными факторами (причинами), которые оказывают нанего наибольшее негативное воздействие.

Закон 80/20. Смысл закона, восходящего к работам социолога Вильфредо Парето, состоит в том, что за 80 % результата отвечает 20 % причин.

Поскольку подавляющую долю эффекта определяет лишь небольшая доля элементов, дающих максимальный вклад, их влияние оказывается непропорционально велико, поэтому этот закон также называютпринципом дисбаланса.

Под «результатом» процесса может пониматься, например, суммарный объем продаж многономенклатурного товара, благосостояниенаселения страны, объем товара на складе, количество жителей городови т. п.

Важным является то, чтобы число составляющих (количество ассортиментных позиций, население страны, количество городов и т. д.), было бы велико.

Популярность закона Парето определяется с одной стороны егочрезвычайной простотой и наглядностью, а с другой стороны – возможностью применения в анализе очень широкого круга процессов. Например:

- 80 % пыли подметается с 20 % пола, по которому чаще всего ходят;
- 80 % стирки уходит на 20 % одежды, которую чаще всего носят;
- 80 % покупок делают 20 % покупателей;
- 80 % телефонных звонков делают 20 % абонентов;
- 80 % продукции выпускают 20 % предприятий;
- 80 % работы делают 20 % людей;
- 80 % людей считают, что они входят в эти 20 %;
- 80 % пользования файлами осуществляется в пределах 20 % файлов;
- 80 % времени, отдаваемого чтению, тратится на 20 % газетных страниц;
- 80 % прибыли дают только 20 % клиентов;
- 80 % потерь на производстве дают только 20 % видов дефектов, а оставшиеся 80 % видов дефектов обуславливают остальные 20 % потерь.

Конечно, соотношение 80/20 не является абсолютным и универсальным, хотя, как правило, отклонения от этого соотношения не оченьвелики.

Анализ Парето. В большинстве случаев подавляющее число дефектов и связанныхс ними материальных потерь возникает из-за относительно небольшогочисла причин. Таким образом, выяснив причины появления основныхдефектов, можно устранить почти все потери, сосредоточив усилия наликвидации именно этих причин.

Диаграмма Парето – это инструмент, позволяющий распределить усилия для разрешения возникающих проблем и выявить основные причины, которые нужно проанализироватьв первую очередь.

С помощью анализа Парето можно выявить, какой из видов дефектов приносит наибольшие потери во времени или в материалах, какиедефекты встречаются наиболее часто. Можно анализировать экономические проблемы предприятия, социальные процессы в больших коллективах, психологические проблемы в группах и много других проблем, возникающих в производственной, экономической, социальной и других сферах деятельности. Диаграммы Парето применять целесообразнотолько в том случае, когда анализируется большое число видов дефектовили причин их появления и когда выявление группы существенных причин затруднено.

Диаграмма Парето по результатам деятельности предназначенадля выявления главной проблемы. Она отражает нежелательные результаты деятельности: дефекты, поломки, отказы, ремонты, возвратыпродукции, объем потерь, затраты, нехватку

запасов, ошибки в составлении счетов, срыв сроков поставок и прочее.

Диаграмма Парето по причинам отражает причины проблем, возникающих в ходе производства. Она используется для выявления главной из них: исполнитель работы, оборудование, сырье, метод работы,измерения.

Построение диаграммы Парето начинают с классификации возникающих проблем по отдельным факторам (например, проблемы, относящиеся к браку, к работе оборудования или исполнителей и т. д.). Затем производят сбор и анализ по каждому фактору, чтобы выяснить,какие из этих факторов являются превалирующими при решении проблем.

Всистеме Statistica диаграммуПаретоможнопостроитьспомощьюмодуля *Statistics/ Industrial Statistic & Six Sigma/ Quality Control Charts/ Pareto chart analysis*. В появившемся диалоговом окне необходимо выбрать формат для ввода данных и нажать *OK*. Если диаграмма строится только по причинам, используются настройки поумолчанию *–Codes (requirestabulationofdatacodes)*.Еслидиаграммастроитсяпопричинамистоимости, выбираемопцию*Codes and counts(one variable with defect type, one variable with counts)*.Диаграмму можно отредактировать с помощью панели рисования ис помощью настроек панели «свойства графиков».

Карты контроля качества. Изготовление продукции всегда связано с непостоянством условийпроизводства. Это приводит к изменениям качества изготовляемыхизделий. При хорошо спланированном и правильно осуществляемомпроцессе эти изменения незначительны. В таком случае говорят, чтопроцесс является *статистически подконтрольным.* Как правило, производственные процессы протекают в статистически регулируемом состоянии, однако случаются ситуации, когда под воздействием неслучайных причин процесс выходит из состояния статистическогоконтроля. В таких случаях необходимо как можно быстрее обнаружитьпричину этих вариаций, что без применения специальных методов сделать порой весьма трудно.

Для решения этой задачи используется механизм, разработанный в1924 году американским инженером Вальтером Шухартом, базирующийся на использовании контрольных карт, часто называемых картамиШухарта. Карты контроля качества, или контрольные карты служат дляпостоянного контроля за тем, чтобы производственный процесс оставался статистически подконтрольным. Основная цель применения контрольных карт – быстрое обнаружение характера изменений в производственных процессах *по результатам наблюдения за параметрамипродукции* с целью поиска их причин и корректировки процесса еще дотого, как начнет появляться бракованная продукция.

Все описанные ранее статистические методы дают возможностьзафиксировать состояние процесса в определенный момент времени.В отличие от них метод контрольных карт позволяет отслеживать состояние процесса во времени и более того – воздействовать на процессдо того, как он выйдет из-под контроля.

Контрольные карты – это линейные графики для оценки управляемости процесса по результатам сравнения отдельных измеренийс заданными контрольными границами.

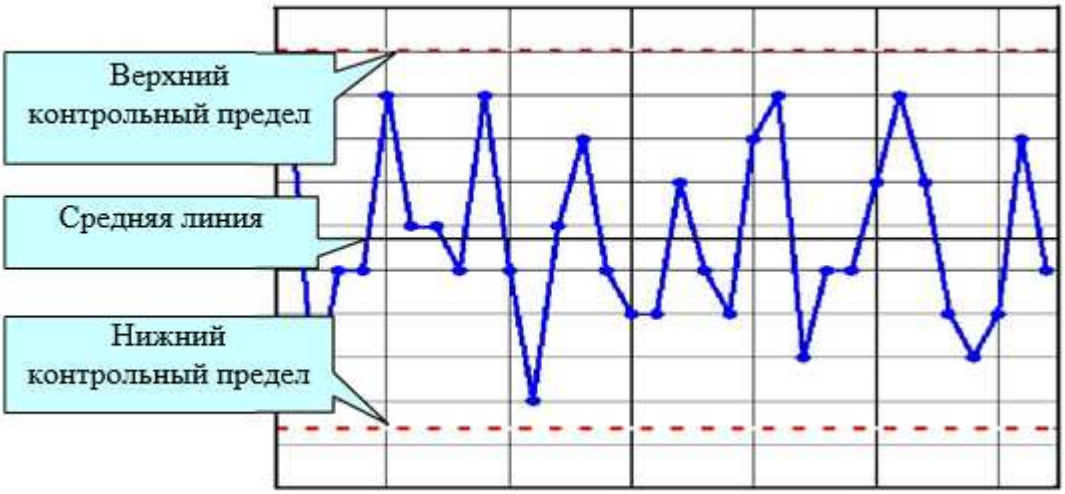


рис. Пример контрольной карты

Всякая контрольная карта состоит обычно из трех линий. Центральная (средняя) линия представляет собой требуемое среднее значение характеристики контролируемого параметра качества. Две другиелинии, одна из которых находится над центральной – верхний контрольный предел (UCL – Upper Control Level), а другая под ней – нижний контрольный предел (LCL – Lower Control Level), представляют собой максимально допустимые пределы изменения значенийконтролируемой характеристики (показателя качества), чтобы считатьпроцесс удовлетворяющим предъявляемым к нему требованиям.

Контрольные карты применяются как для анализа количественныхи данных, когда результаты измерений показателя качества непрерывны ивыражаются в числовой форме, так и в случае, когда информация обобъектах дискретна и ограничена выводом типа «годен»–«не годен».В первом случае применяются контрольные карты по количественномупризнаку, во втором – по альтернативному. Подробнее о видах картможно прочитать в работе [6].

Контрольная карта индивидуальных значений. Эта карта применяется, когда наблюдение производится над сравнительно небольшим числом объектов, и все они подвергаются контролю. Чаще всего это бывает при наладке и настройке процесса, когдапреследуется цель его предварительного исследования. Карта удобнатогда, когда процесс протекает в реальном времени и есть возможность оперативного вмешательства в него в случае выхода параметракачества за допустимые пределы.

Порядок построения карты следующий.

1. Запуститьмодуль *Statistics/ Industrial Statistics & Six Sigma/ Quality Control Charts.*На стартовой панели выбрать *Individuals &moving range (отдельные наблюдения и скользящий размах)* и нажатькнопку *ОК.* \

2. В появившемся диалоговом окне выбрать переменную с измерениями – *Measurements (observations)*инажать*ОК.*В результате будет построена контрольная карта. Имеется возможность группировки данных, если наблюдений слишкоммного. При этом у каждой выборки будет вычислено среднее значение,которое наносится на карту. Для группировки необходимо указать переменную *Part identifiers (code*

numbers), где должны быть номера выборок. Если объем каждой выборки постоянный, это можно указатьпрямо, отметив чекбокс *Constant number of samplesper part* и введя нужный объем выборки

Контрольная карта средних значений и размахов. Эта карта применяется при массовом производстве. Достоинство ее состоит, во-первых, в том, что она позволяет отслеживать во времениках настройку процесса, так и его наладку, а во-вторых, выводы относительно характеристик делаются на основе малых выборок из большогочисла рассматриваемых единиц продукции, что существенно удешевляет контроль текущих характеристик процесса [6].

Построение карты в системе Statistica:

1. Запуститьмодуль *Statistics/ Industrial Statistics & Six Sigma/ Quality Control Charts.*На стартовой панели выбрать *X-bar & Rchartforvariables* и нажать кнопку *OK.*

2. В появившемся диалоговом окне выбрать переменную с измерениями – *Measurements* и переменную – номера выборок *Sample Idents(opt.)* и нажать *OK.*

На *x-карте* скользящих средних все точки попадают внутрь контрольных границ. На контрольной карте скользящих размахов все точкитакже находятся внутри контрольных границ. Размахи служат оценкойизменчивости характеристик, поэтому можно сказать, что концентрациявещества подчиняется требованиям статистического контроля по уровню средних и изменчивости.

При построении *x-R* карты могут возникнуть следующие ситуации:

1. За границами регулирования находятся точки на *R-карте* и соответствующие им точки на *x-карте*. Это означает,что за счет обычных (внутренних) причин увеличилось технологическоерассеяние, т. е. возросла величина. В этом случае следует заняться поиском и устранением причин разладки процесса.

2. За границами регулирования находятся точки на *x-карте,* но приэтом соответствующие им точки на *R-карте* лежат в границах регулирования. Поскольку по *R-карте* выхода за границырегулирования нет, полное технологическое рассеяние остается прежним, т. е. наладка процесса не изменяется. Значит, есть все основанияпредполагать, что выход за границы регулирования по *x-карте* произошел потому, что распределение по *x* сместилось в сторону больших илименьших значений контролируемого признака. Это, как правило, является результатом воздействия на процесс какой-то особой внешней причины, изменяющей его настройку. Дальнейшие действия должны бытьсвязаны с поиском и устранением этой причины.

3. За границами регулирования наблюдаются точки на *R-карте,* атакже соответствующие им и не соответствующие точки на *x-карте.* Этоговорит о наличии как обычных, так и особых причин, ухудшающихпроцесс.

Часто встречается ситуация, когда влияние первой обнаруженнойособой причины настолько велико, что из-за нее не видно влияния других причин. В этом случае соответствующую точку исключают из набора данных и строят карту заново. Влияние других причин становитсявидимым. Таким образом, последовательно, шаг за шагом обнаруживаяособое поведение точек на контрольной карте и устанавливая их причины, делают процесс прозрачным, доступным нашему пониманию.

<u>Наблюдается серия точек</u>

Серия – это такое состояние процесса, при котором последовательные точки лежат по одну сторону от средней линии.Число таких точек называется длиной серии. Процесс нестабилен, если:

— серия состоит из 7 точек и более;

— 10 точек из 11 лежат по одну сторону от средней линии;
— не менее 12 точек из 14 лежат по одну сторону от средней линии;
— не менее 16 точек из 20 лежат по одну сторону от средней линии.

Причиной серии является внешнее воздействие на процесс, котороесдвигает центр рассеяния в ту или иную сторону от средней линии, изменяя настройку процесса.

Наблюдается дрейф

Дрейф – это не менее 7 поднимающихся или ниспадающих точек. Причинами появления дрейфа могут быть, например, такие факторы, как постепенный рост (падение) температурыокружающей среды, износ технологического оборудования, появление всредствах измерения прогрессирующих погрешностей, изменение физических и химических параметров процесса и другие неслучайныепричины.

Две и более близлежащих точки приближаются к границам регулирования (лежат за пределами 2-сигмовых границ). Точки считаются приблизившимися к границам регулирования, если они находятся за пределами плюс-минус 2 относительно среднейлинии, т. е. на расстоянии большем, чем 2/3 расстояния от средней линии до границы регулирования в так называемой *зоне внимания*.

Если выход точки за границы регулирования незначителен и вдальнейшем больше не повторяется, то вполне возможно, что этот фактне говорит о дестабилизации процесса. Но если выход за границы какой-либо одной точки составляет заметную величину, то вмешательствов процесс с целью его совершенствования в любом случае необходимо.

Наблюдается периодичность

Наличие подъемов и спадов с примерно одинаковыми интерваламитакже говорит о нестабильности процесса, причиной которойможет быть воздействие на процесс внешнего периодически изменяющегося фактора.

Точки приближаются к средней линии

Точки считаются приблизившимися к средней линии, если они лежат внутри полуторасигмовой зоны, то есть внутри линий, делящих пополам расстояние от средней линии до границ регулирования. В этомслучае следует изменить способ разбиения на выборки или группы, поскольку может оказаться, что смешаны данные из разных распределений.

Отмеченные в каждом рассмотренном случае выходы процессов изсостояния статистического регулирования несут в себе потенциальнуюугрозу получения брака в недалеком будущем, ибо однажды возникшиенестабильности в процессе всегда имеют тенденцию со временем нарастать.

Таким образом, контрольные карты и их грамотный анализ позволяют прогнозировать характер протекания производственных процессовв будущем и вовремя их останавливать для корректировки с целью предупреждения возможного появления бракованной продукции.

Контрольные карты накопленных сумм. Карты Шухарта нечувствительны к малым возмущениям процессов. При достаточно долговременном контроле информация о начальном этапе процесса теряется. В отличие от рассмотренных, контрольныекарты накопленных сумм – это карты с памятью. Они могут быть болеечувствительными к возмущениям, т. е. уже в самом начале сдвига уровня настройки процесса или изменения технологического рассеяния ониуказывают на необходимость вмешательства в процесс.

Таким образом, контрольные карты накопленных сумм следуетприменять в тех случаях, когда даже незначительные смещения уровнянастройки процесса недопустимы и подлежат скорейшему устранению.

Для построения контрольной карты накопленных сумм на стартовой панели необходимо выбрать вкладку Variables, а в ней –*CuSumchartforindividuals*инажатьОК. Дальнейшие действия аналогичны рассмотренному алгоритму для построения карт иднивидуальныхи средних значений.

Контрольные вопросы

1. Почему необходимо статистическое управление качеством?
2. Что такое диаграмма причин и результатов и как она строитсяв программе Statistica?
3. В чём состоит смысл закона 80/20?
4. Как строится в программе Statistica диаграмма Парето?
5. Как построить карты контроля качества в программе Statistica?
6. Почему важен совместный анализ карт средних значений и размахов?

Лабораторная работа № 7: «Кластерный анализ».

Цель работы: ознакомиться с основами кластерный анализ, методы которого применяют как для разведочного анализа, так и при исследованиях разнородных множеств.

Порядок выполнения работы:

1. Исходные данные – Приложение 6 - представляют собой результаты анкетного опроса работников предприятия по вопросам качества трудовой жизни. Для анализа выбрано 15 вопросов (переменные) и 41 наблюдение – ответы на вопросы по 10-бальной шкале (1– не удовлетворяет, 4 – скорее не удовлетворяет, 7 – скорее удовлетворяет, 10 – удовлетворяет). Ответило 15 руководителей (первые 15 наблюдений) и 26 рабочих.

Провести кластерный анализ с целью:

— выявить наиболее проблемные вопросы, волнующие всех работников;
— выявить вопросы, на которые руководители и рабочие отвечают диаметрально противоположно.
— По результатам анализа дать рекомендации для принятия управленческих решений: что необходимо сделать в первую очередь для улучшения качества трудовой жизни предприятия?

2. Исходные данные – Приложение 7 - представляют собой статистику отдела сбыта предприятия попроизводству стройматериалов. Для каждого вида продукции указанобъем продаж по месяцам и выручка.

Провести кластерный анализ с целью:

— сегментировать продукцию по месяцам года;
— повышения выручки путем рационализации выпуска нужной продукции по месяцам.

Предположить существование трех кластеров, на которые сегментируются выручка и объем продукции. По результатам анализа дать рекомендации для принятия управленческих решений: что необходимосделать для улучшения качества производственного процесса.

Краткая теория.Кластерный анализ включает в себя набор различных алгоритмовклассификации. Общий вопрос, задаваемый исследователями во многихобластях, состоит в том, как организовать наблюдаемые данные в наглядные структуры. Например, биологи ставят цель разбить животныхна различные виды, чтобы содержательно описать различия между ними. Задача кластерного анализа состоит в разбиении исходной совокупности объектов на группы схожих, близких между собой объектов. Этигруппы называют кластерами. Другими словами, кластерный анализ –это один из способов классификации объектов по их признакам [**5**]. Желательно, чтобы результаты классификации имели содержательную интерпретацию.

Результаты, полученные методами кластерного анализа, применяют в самых различных областях. В маркетинге – это сегментация конкурентов и потребителей. В психиатрии для успешной терапии являетсярешающей правильная диагностика симптомов, таких как паранойя,шизофрения и т. д. В менеджменте важна классификация поставщиков,выявление схожих производственных ситуаций, при которых возникаетбрак. В социологии – разбиение респондентов на однородные группы[16, 17]. В инвестировании важно сгруппировать ценные бумаги посходству в тенденции доходности, чтобы составить оптимальный инвестиционный портфель. В общем, всякий раз, когда необходимо классифицировать большое количество информации такого рода и представлять ее в виде, пригодном для дальнейшей обработки, кластерныйанализ оказывается весьма полезным и эффективным.

Достоинство кластерного анализа состоит в том, что он работает даже тогда, когда данных мало и не выполняются требованиянормальности распределений случайных величин и другие требования классических методов статистического анализа. Кластерный анализ позволяет провести объективную классификацию любых объектов, которые охарактеризованы рядом признаков. Изэтого можно извлечь ряд преимуществ:

1. Полученные кластеры можно интерпретировать, то есть описывать, какие же собственно группы существуют.

2. Отдельные кластеры можно выбраковывать. Это полезно в техслучаях, когда при наборе данных допущены определенные ошибки, врезультате которых значения показателей у отдельных объектов резкоотклоняются. При применении кластерного анализа такие объекты попадают в отдельный кластер.

3. Для дальнейшего анализа могут быть выбраны только те кластеры, которые обладают интересующими характеристиками.

Методы кластеризации. В пакете Statistica реализуются следующие методы кластеризации.

1. Иерархические алгоритмы – древовидная кластеризация. В основе иерархических алгоритмов лежит идея последовательной кластеризации. На начальном шаге каждый объект рассматривается как отдельный кластер. На следующем шаге некоторые из ближайших друг кдругу кластеров будут объединяться в отдельный кластер.

2. Метод К-средних. Этот метод используется наиболее часто. Онотносится к группе так называемых эталонных методов кластерногоанализа. Число кластеров К задается пользователем.

3. Двухвходовое объединение. При использовании этого методакластеризация проводится одновременно как по переменным (столбцам), так и по результатам наблюдений (строкам). Результатами процедуры являются описательные статистики по переменным и наблюдениям, а также двумерная цветная

диаграмма, на которой цветомотмечаются значения данных. По распределению цвета можно составить представление об однородных группах.

Нормирование переменных для кластеризации. Разбиение исходной совокупности объектов на кластеры связано свычислением расстояний между объектами и выбора объектов, расстояние между которыми наименьшее из всех возможных.

Наиболее часто используется привычное всем нам евклидово (геометрическое) расстояние. Эта метрика отвечает интуитивным представлениям о близости объектов в пространстве (как будто расстояния между объектами измерены рулеткой). Но для данной метрики нарасстояние между объектами могут сильно влиять изменения масштабов (единиц измерения). Например, если один из признаков измерен вмиллиметрах, а затем его значение переведены в сантиметры, евклидоворасстояние между объектами сильно изменится. Это приведет к тому,что результаты кластерного анализа могут значительно отличаться отпредыдущих.

Если переменные измерены в разных единицах измерения, то требуется их предварительная нормировка, то есть преобразование исходных данных, которое переводит их в безразмерные величины.

В пакете Statistica нормировка любой переменной *x* выполняется поформуле:

$$x_{норм} = \frac{x - \mu}{\sigma}.$$

Для этого нужно щелкнуть правой кнопкой мыши по имени переменной и в открывшемся меню выбрать последовательность команд:Fill/ StandardizeBlock/ StandardizeColumns.

Значения нормированной переменной станут равными нулю, а дисперсии – единице.

Метод К-средних в программе Statistica. Метод K-средних (K-means) разбивает множество объектов на заданное число K различных кластеров, расположенных на возможнобольших расстояниях друг от друга. Обычно, когда результаты кластерного анализа методом K-средних получены, можно рассчитать средниедля каждого кластера по каждому измерению, чтобы оценить, насколько кластеры различаются друг от друга. В идеале вы должны получитьсильно различающиеся средние для большинства измерений, используемых в анализе. Значения F-статистики, полученные для каждого измерения, являются другим индикатором того, насколько хорошо соответствующее измерение дискриминирует кластеры.

В программе Statistica кластерный анализ выполняется следующим образом.

1. Создать файл данных исходных данных.

2. Выбратьмодуль Statistics/ Multivariable Exploratory Techniques/Cluster Analysis. Нажать OK, в появившемся окне выбрать метод кластеризации *K-means clustering* и нажать OK.

3. В появившемся диалоговом окне необходимо установить следующие настройки.

Выбрать переменные кнопкой Variables.

Выбрать объекты кластеризации: это могут быть переменные –столбцы (Variables (columns)), либо наблюдения – строки (Cases(Rows)).

Выбрать число кластеров. Этот выбор делает пользователь, исходяиз собственных предположений о числе групп схожих объектов. Привыборе количества кластеров руководствуйтесь следующим: количествокластеров, по возможности, не должно быть слишком большим. Расстояние, на котором объединялись объекты

данного кластера, должнобыть, по возможности, гораздо меньше расстояния, на котором к этомукластеру присоединяется еще что-либо.

Далее необходимо выбрать начальное разбиение объектов по кластерам (Initial cluster centers). Пакет *Statistica* предлагает 1) выбрать наблюдения с максимальным расстоянием между центрами кластеров;2) рассортировать расстояния и выбрать наблюдения с постояннымиинтервалами (установка по умолчанию); 3) взять первые наблюдения зацентры и присоединять остальные объекты к ним.

Древовидная кластеризация – это пример иерархического алгоритма, принцип работы которого состоит в последовательном объединениив кластер сначала самых близких, а затем и все более отдаленных другот друга элементов. Большинство из этих алгоритмов исходит из матрицы сходства (расстояний), и каждый отдельный элемент рассматривается вначале как отдельный кластер.

После загрузки модуля кластерного анализа и выбора*Joining (tree clustering),* в окне ввода параметров кластеризации можноизменить следующие параметры:

Исходные данные (Input). Они могут быть в виде матрицы исследуемых данных (Raw *data)* и в виде матрицы расстояний (Distance*matrix).*

Кластеризацию (Cluster) наблюдений (Cases *(raw))* или переменных (Variable *(columns)),* описывающих состояние объекта.

Меры расстояния (Distance *measure).* Здесь возможен выбор следующих мер: евклидово расстояние (Euclidean *distances),* квадрат евклидова расстояния (Squared *Euclidean distances),* расстояние городскихкварталов (манхэттенское расстояние, *City-block (Manhattan) distance),*расстояние Чебышева (Chebychev *distance metric),* степенное расстояние(Power...), процент несогласия (Percent*disagreement).*

Методкластеризации (Amalgamation *(linkage) rule).*Здесьвозможнысследующиеварианты: одиночнаясвязь (методближайшегососеда) (Single *Linkage),*полнаясвязь (методнаиболееудаленныхсоседей) (Complete *Linkage),*невзвешенноепопарноесреднее (Unweighted *pair-group average),*взвешенноепопарноесреднее (Weighted *pair-group average),*невзвешенныйцентроидныйметод (Unweighted *pair-group centroid),*взвешенныйцентроидныйметод (медиана) (Weighted *pair-group centroid (median)),*методУорда (Ward's *method).*

В результате кластеризации строится горизонтальная или вертикальная дендрограмма – график, на котором определены расстояниямежду объектами и кластерами при их последовательном объединении.Древовидная структура графика позволяет определить кластеры в зависимости от выбранного порога – заданного расстояния между кластерами. Кроме того, выводится матрица расстояний между исходнымиобъектами (Distance *matrix);* средние и среднеквадратичные отклонениядля каждого исходного объекта (Distiptive *statistics).*

Контрольные вопросы

1. В чём состоит сущность кластерного анализа?

2. Как провести кластерный анализ в программе Statistica методомk-средних?

3. Какие иерархические методы кластерного анализа реализованыв программе Statistica и как провести анализ?

4. Почему необходимо нормировать данные при проведении кластерного анализа?

Список использованной литературы.

1. Боровиков В.П. Программа Statistica для студентов и инженеров /В.П. Боровиков. – М.: Компьютер пресс, 2000. – 301 с.

2. Боровиков В.П. Statistica: искусство анализа данных на компьютере / В.П. Боровиков. – СПб.: Питер, 2001. – 656 с.

3. Боровиков В.П. Популярное введение в программуSTATISTICA / В.П. Боровиков. – М.: Компьютер пресс, 1998. – 267 с.

4. Вуколов Э.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетовSTATISTICA и EXCEL: учебное пособие / Э.А. Вуколов. – М.: ФОРУМ:ИНФРА-М, 2004. – 464 с. – (Профессиональное образование). – ISBN5-8199-0141-X (ФОРУМ). – ISBN 5-16-002003-9 (ИНФРА-М).

5. Электронный учебник по статистике [Электронный ресурс]. –М.: StatSoft, Inc. – 2001. – Режим доступа: http://www.statsoft.ru/home/textbook/default.htm.

6. Казанцева Н.Н. Статистический контроль и статистические методы управления качеством: учебное пособие / Н.Н. Казанцева. – Томск:Изд-во ТПУ, 2004. – 116 с.

8. Использование пакета Statistica 5.0 для статистической обработки опытных данных: методические указания / Сост.: С.В. Кабанов. – Саратов: Изд-во Сарат. гос. агр. ун-та, 2000. – 90 с.

9. Берестнева О.Г. Компьютерный анализ данных: учебное пособие / О.Г. Берестнева, Е.А. Муратова, А.М. Уразаев. – Томск: Изд-воТПУ, 2003. – 204 с. : ил. – Библиогр.: С. 200–201. – ISBN 5-98298-004-8.

10. Горицкий Ю.А. Практикум по статистике с пакетомSTATISTICA. Учебное пособие по курсу «Математическая статистика» /Ю.А. Горицкий. – М.: Изд-во МЭИ, 2000. – 44 с. – ISBN 5-7046-0573-7.

11. Каримов Р.Н. Обработка экспериментальной информации. Ч. 1.Разведочный анализ. Анализ качественных данных / Р.Н. Каримов. –Саратов: Саратовский госуд. техн. ун-т, 2002. – 112 с.

12. Каширина И.Б. Экономико-математическая модель прогнозирования спроса на образовательные услуги / И.Б. Каширина, В.Г. Мысник // Моделирование систем. – 2002. – № 2 (4). – С. 46–53.

13. Eickelmann N. Statistical Process Control: What You Don't Measure Can Hurt You! / N. Eickelmann, A. Anant // IEEE Software. – 2003. – № 3. – P. 49–51.

14. Волкова Н.А. Кластерный анализ результатов социологическогоопроса работников предприятия / Н.А. Волкова, О.В. Стукач // ВестникУльяновского государственного технического университета. – 2005. –№ 2. – С. 68–72.

16. Мартюшева П.В. Кластерный анализ как инструмент менеджмента качества для обработки социологических опросов на промышленном предприятии / П.В. Мартюшева, О.В. Стукач // Доклады томскогогосударственногоуниверситетасистемуправлениярадиоэлектроники. – 2007. – Вып. 1 (15). – С. 71–76. – ISSN 1818-0442.

17. Рыбалко В.В. Параметрическое диагностирование энергетических объектов на основе факторного анализа в среде Statistica / В.В. Рыбалко // Exponenta Pro. – 2004. – № 2 (6). – С. 78–83.

18. Статистические методы повышения качества / Под ред. Х. Кумэ. – Пер. с англ. – М.: Финансы и статистика, 1990. – 304 с.

19. Ефимов В.В. Статистические методы в управлении качествомпродукции: учебное пособие / В.В. Ефимов, Т.В. Барт. – М.: КНОРУС,2006. – 240 с. – ISBN 5-85971-262-6.

Приложение 1.

Вариант	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	10	10	10	10	13	12	10	10	10	10
2	12	12	12	12	12	10	12	12	12	12
3	15	13	13	11	11	11	11	14	11	14
4	14	12	12	10	10	12	10	15	12	15
5	10	10	10	12	10	13	12	12	11	13
6	11	12	10	14	12	13	15	14	10	12
7	10	14	12	13	12	12	14	12	12	10
8	12	15	11	13	11	16	15	10	13	11
9	15	12	10	14	10	14	12	12	12	11
10	14	13	12	12	12	15	13	13	10	10
11	12	12	13	15	15	15	16	12	15	12
12	18	10	15	14	14	14	15	11	12	15
13	17	12	12	12	12	12	12	12	12	14
14	14	15	10	13	11	13	14	12	13	13
15	12	14	14	12	13	12	13	11	12	12
16	10	12	12	12	14	15	12	10	15	10
17	11	13	15	15	15	14	12	12	16	12
18	11	22	12	14	12	12	13	15	13	10
19	12	14	10	12	13	12	10	12	12	9
20	13	10	13	16	12	15	11	13	12	10
21	12	12	12	12	15	13	10	12	15	13
22	13	10	10	15	15	15	9	10	14	12
23	12	12	15	14	14	12	10	12	14	15
24	15	13	18	13	12	12	12	10	11	14

25	14	10	22	12	13	11	10	14	12	10
26	15	12	10	12	12	12	12	12	12	12
27	16	15	12	14	14	15	11	13	15	15
28	18	14	15	15	13	14	12	12	14	14
29	17	12	14	18	12	12	10	10	12	12
30	14	16	12	12	18	10	13	11	12	13
31	12	18	13	14	15	10	12	11	12	10
32	15	14	12	12	15	12	14	10	11	11
33	10	15	10	15	14	11	15	12	12	17
34	14	12	14	12	12	12	13	13	10	18
35	12	13	12	13	12	13	12	12	11	14
36	15	10	15	12	15	12	10	12	10	10
37	11	12	10	16	12	14	12	12	12	10
38	14	14	12	15	13	12	13	15	10	12
39	12	11	11	12	15	10	12	10	9	15
40	10	12	15	14	18	11	11	12	11	14
41	11	15	14	12	12	12	14	13	10	16
42	11	14	12	15	15	10	15	11	11	12
43	10	18	13	12	16	9	17	12	12	14
44	12	12	12	10	14	12	12	13	11	12
45	15	10	10	10	15	14	10	14	10	15
46	14	12	12	12	16	12	10	12	10	10
47	17	11	11	15	17	12	11	12	12	12
48	15	10	10	12	15	12	12	12	14	14
49	12	10	12	13	12	14	11	11	12	12
50	11	12	15	12	15	12	10	12	10	15

Приложение 2.

13,39	13,33	13,56	13,38	13,43	13,37	13,53	13,40	13,25	13,37
13,28	13,34	13,50	13,38	13,38	13,45	13,47	13,62	13,45	13,39
13,53	13,58	13,32	13,27	13,42	13,40	13,57	13,46	13,33	13,40
13,57	13,36	13,43	13,38	13,26	13,52	13,35	13,29	13,48	13,43
13,40	13,39	13,50	13,52	13,39	13,39	13,46	13,29	13,55	13,31
13,29	13,33	13,38	13,61	13,55	13,40	13,20	13,31	13,46	13,13
13,43	13,51	13,50	13,38	13,44	13,62	13,42	13,54	13,31	13,58
13,41	13,49	13,42	13,45	13,34	13,47	13,48	13,59	13,20	14,56
13,55	13,44	13,50	13,40	13,48	13,29	13,31	13,42	13,32	13,48
13,43	13,26	13,58	13,38	13,48	13,45	13,29	13,32	13,24	13,38
13,34	13,14	13,31	13,51	13,59	13,32	13,52	13,57	13,62	13,29
13,23	13,37	13,64	13,30	13,40	13,58	13,24	13,32	13,52	13,50
13,43	13,58	13,63	13,48	13,34	13,37	13,18	13,50	13,45	13,60
13,38	13,33	13,57	13,28	13,32	13,40	13,40	13,33	13,20	13,44
13,34	13,54	13,40	13,47	13,28	13,41	13,39	13,48	13,42	13,46
13,28	13,46	13,37	13,53	13,43	13,30	13,45	13,40	13,45	13,40
13,33	13,39	13,56	13,46	13,26	13,35	13,42	13,36	13,44	13,41
13,43	13,51	13,51	13,24	13,34	13,28	13,37	13,54	13,43	13,35
13,52	13,23	13,48	13,48	13,54	13,41	13,51	13,44	13,36	13,36
13,53	13,44	13,69	13,66	13,32	13,26	13,51	13,38	13,46	13,34

Приложение 3.

Вариант										
1	48	30	43	44	30	34	32	43	40	46
2	25	21	34	49	39	37	45	49	31	49
3	43	46	34	35	42	30	41	34	42	22
4	38	40	26	47	34	42	38	20	38	36
5	30	13	41	40	40	15	35	11	38	45
6	37	12	38	36	14	39	32	54	43	39
7	23	30	32	36	32	34	49	18	49	50
8	37	20	44	28	44	35	45	34	33	41
9	43	45	50	14	33	39	41	39	46	31
10	40	52	44	39	35	54	33	42	42	36
11	44	51	45	19	34	44	40	37	43	32
12	33	42	40	35	37	13	48	48	50	32
13	40	48	45	23	36	36	42	40	37	30
14	44	50	46	39	31	48	44	42	36	51
15	44	50	54	37	33	34	42	43	43	47
16	33	48	18	42	15	32	34	14	39	45
17	48	26	31	34	38	36	46	49	40	48
18	42	47	35	34	41	33	41	35	43	42
19	39	37	47	27	33	22	37	19	19	37
20	43	41	30	39	38	36	36	34	42	46
21	39	44	37	35	43	38	33	47	45	38
22	37	48	38	52	40	45	44	42	38	40
23	44	46	37	34	41	37	41	39	30	38
24	32	41	48	36	51	36	33	39	45	40
25	34	41	38	34	33	27	51	45	27	38
26	42	37	46	41	47	36	30	45	41	40
27	37	37	39	42	48	41	36	39	33	47
28	43	49	27	31	41	46	40	36	36	42
29	41	46	33	37	47	35	31	29	30	36
30	48	38	37	34	40	34	36	50	48	39
31	30	38	43	41	44	45	38	37	46	50

Приложение 4.

1 Argument	2 Var-1	3 Var-2	4 Var-3	5 Var-4	6 Var-3	7 Var-6	8 Var-7	9 Var-8	10 Var-9	11 Var-10	12 Var-11	13 Var-12
0,5	0,0964	-0,2756	-0,4202	1,3092	1,1256	2,2374	2,0417	0,9551	0,0906	0,1987	-0,4134	1,2553
1	-0,6562	-1,1101	-1,2482	1,0551	0,1209	1,3682	1,0008	0,0966	-0,6298	-0,3378	-1,199	0,9388
1,5	-1,185	-1,4185	-1,6357	1,4138	-0,8598	0,4669	0,1831	-0,7179	-1,0409	-0,4488	-1,4956	1,2258
2	-1,4474	-1,2462	-1,635	2,1097	-1,5876	-0,2668	-0,0234	-1,415	-1,1012	-0,1356	-1,3778	1,8393
2,5	-1,4406	-0,7769	-1,3254	2,6144	-1,8792	-0,6553	0,4401	-1,868	-0,8583	0,4432	-0,9773	2,2498
3	-1,2017	-0,2728	-0,8379	2,5207	-1,6865	-0,6441	1,2067	-1,9831	-0,4332	1,0401	-0,4877	2,0485
3,5	-0,8018	0,0241	-0,3327	1,836	-1,1244	-0,3347	1,7111	-1,7427	0,0166	1,4219	-0,1132	1,2413
4	-0,3314	-0,0136	0,0425	0,994	-0,4285	0,0548	1,5666	-1,2164	0,3435	1,466	-0,0032	0,2603
4,5	0,1159	-0,3512	0,1939	0,5763	0,1366	0,2806	0,8244	-0,5426	0,4531	1,21	-0,201	-0,315
5	0,4611	-0,8095	0,1071	0,9298	0,3822	0,176	-0,0616	0,1102	0,3302	0,8351	-0,6299	-0,1398
5,5	0,6542	-1,1416	-0,1506	1,942	0,268	-0,2768	-0,5436	0,5913	0,0415	0,5893	-1,12	0,6711
6	0,6814	-1,1383	-0,4527	3,1232	-0,0858	-0,9359	-0,3519	0,8091	-0,2874	0,6825	-1,4691	1,6252
6,5	0,5667	-0,7186	-0,6541	3,9419	-0,4524	-1,5554	0,3219	0,7555	-0,5118	1,1962	-1,514	2,1881
7	0,3642	0,0321	-0,6374	4,1831	-0,5965	-1,8842	0,956	0,5042	-0,5149	2,0454	-1,1896	2,1416
7,5	0,145	0,8913	-0,351	4,0883	-0,3758	-1,7681	1,0539	0,1858	-0,2456	3,0081	-0,5532	1,7236
8	-0,0185	1,5848	0,1719	4,1814	0,196	-1,2146	0,4825	-0,0543	0,2626	3,8102	0,2327	1,4539
8,5	-0,07	1,8936	0,8203	4,9005	0,9495	-0,3917	-0,4366	-0,1022	0,8984	4,2332	0,9554	1,7662
9	0,0187	1,7396	1,4351	6,2876	1,6252	0,4368	-1,1451	0,0883	1,5023	4,1991	1,4235	2,6976
9,5	0,2405	1,214	1,8541	7,9484	1,976	1,0143	-1,2272	0,4804	1,912	3,7997	1,5312	3,8482
10	0,5535	0,5388	1,9589	9,3106	1,8664	1,1911	-0,6913	0,9633	2,0087	3,257	1,2934	4,6398

Приложение 5.

Time	Day	Var 1	Var 2	Var 3	Var 4	Var 5	Var 6	Var 7	Var 8	Var 9	Var 10
1	1	10	10	10	10	13	12	10	10	10	10
2	1	12	12	12	12	12	10	12	12	12	12
3	1	15	13	13	11	11	11	11	14	11	14
4	1	14	12	12	10	10	12	10	15	12	15
5	1	10	10	10	12	10	13	12	12	11	13
1	1	11	12	10	14	12	13	15	14	10	12
2	1	10	14	12	13	12	12	14	12	12	10
3	1	12	15	11	13	11	16	15	10	13	11
4	1	15	12	10	14	10	14	12	12	12	11
5	1	14	13	12	12	12	15	13	13	10	10
1	2	12	12	13	15	15	15	16	12	15	12
2	2	18	10	15	14	14	14	15	11	12	15
3	2	17	12	12	12	12	12	12	12	12	14
4	2	14	15	10	13	11	13	14	12	13	13
5	2	12	14	14	12	13	12	13	11	12	12
1	2	10	12	12	12	14	15	12	10	15	10
2	2	11	13	15	15	15	14	12	12	16	12
3	2	11	22	12	14	12	12	13	15	13	10
4	2	12	14	10	12	13	12	10	12	12	9
5	2	13	10	13	16	12	15	11	13	12	10
1	3	12	12	12	12	15	13	10	12	15	13
2	3	13	10	10	15	15	15	9	10	14	12
3	3	12	12	15	14	14	12	10	12	14	15
4	3	15	13	18	13	12	12	12	10	11	14
5	3	14	10	22	12	13	11	10	14	12	10
1	3	15	12	10	12	12	12	12	12	12	12
2	3	16	15	12	14	14	15	11	13	15	15
3	3	18	14	15	15	13	14	12	12	14	14
4	3	17	12	14	18	12	12	10	10	12	12
5	3	14	16	12	12	18	10	13	11	12	13
1	4	12	18	13	14	15	10	12	11	12	10
2	4	15	14	12	12	15	12	14	10	11	11
3	4	10	15	10	15	14	11	15	12	12	17
4	4	14	12	14	12	12	12	13	13	10	18
5	4	12	13	12	13	12	13	12	12	11	14
1	4	15	10	15	12	15	12	10	12	10	10
2	4	11	12	10	16	12	14	12	12	12	10
3	4	14	14	12	15	13	12	13	15	10	12
4	4	12	11	11	12	15	10	12	10	9	15
5	4	10	12	15	14	18	11	11	12	11	14
1	5	11	15	14	12	12	12	14	13	10	16
2	5	11	14	12	15	15	10	15	11	11	12
3	5	10	18	13	12	16	9	17	12	12	14
4	5	12	12	12	10	14	12	12	13	11	12
5	5	15	10	10	10	15	14	10	14	10	15
1	5	14	12	12	12	16	12	10	12	10	10
2	5	17	11	11	15	17	12	11	12	12	12
3	5	15	10	10	12	15	12	12	12	14	14
4	5	12	10	12	13	12	14	11	11	12	12
5	5	11	12	15	12	15	12	10	12	10	15

Приложение 6.

1 Режим работы	2 Мотивация к труду	3 Размер зарплаты	4 Отношение к существующей системе обучения работников	5 Поощрения за хорошо выполненную работу	6 Перспективы проф. и служебного роста	7 Условия труда	8 Взаимоотношения с товарищами по работе	9 Перспективы повышения квалификации	10 Удовлетворенность работой	11 Наличие "доски почета"	12 Наличие доски "они позорят наше предприятие"	13 Социальная защищенность	14 Ощущение экономического благосостояния	15 Желание сменить работу
1	4	1	7	1	1	7	10	7	7	4	1	1	1	1
1	4	1	10	1	1	7	10	7	7	1	1	1	1	1
10	4	1	7	1	4	10	10	7	7	10	1	1	4	4
10	4	1	7	1	4	10	10	7	4	10	1	7	1	4
4	7	4	7	7	7	7	7	4	4	1	1	4	1	4
10	10	10	10	10	10	10	10	10	10	10	4	10	10	10
10	10	10	10	4	10	10	10	10	10	10	7	10	7	4
10	1	1	4	4	4	7	10	1	4	10	10	7	1	7
7	7	1	4	4	7	10	10	7	7	10	1	4	4	7
10	7	1	10	1	10	4	10	7	10	10	4	1	4	7
10	1	1	10	1	1	10	7	1	10	10	10	1	1	7
10	1	1	1	1	10	10	10	10	4	10	10	1	1	4
10	1	1	10	1	10	10	10	1	10	10	7	4	10	10
10	4	1	7	7	4	10	10	7	7	10	7	10	1	4
7	7	7	4	10	10	10	10	4	7	10	4	7	7	4
10	1	1	10	1	1	1	10	1	7	10	10	1	1	4
10	10	4	10	10	1	10	10	1	10	10	1	10	4	4
10	1	1	1	1	1	10	10	1	7	10	1	1	1	10
10	7	7	10	10	10	10	10	10	10	10	10	10	7	10
7	4	4	10	4	4	7	10	4	1	10	1	7	4	10
10	4	4	7	1	10	7	10	10	7	4	4	4	1	7
10	7	1	10	1	10	10	10	10	7	7	1	1	7	7
10	10	4	4	4	10	10	10	10	10	10	10	10	10	10
4	10	4	10	4	7	4	7	10	10	4	4	4	10	1
10	7	1	10	4	4	7	10	10	4	10	10	4	4	7

10	10	4	10	10	7	7	10	10	7	10	10	10	10	10
10	4	1	7	7	10	1	10	10	7	10	10	10	4	10
10	7	1	10	1	1	4	10	1	10	10	1	1	4	7
10	10	1	1	10	1	1	10	1	10	1	1	1	1	4
7	7	4	10	1	1	10	10	4	7	10	1	4	4	7
7	4	1	4	1	4	7	10	4	10	7	4	7	4	7
10	4	4	1	4	1	7	10	4	10	7	7	1	4	10
10	10	7	7	1	7	10	10	4	10	10	10	4	4	10
10	4	1	4	4	1	1	10	1	10	7	4	7	4	4
1	7	7	4	10	7	7	4	7	7	7	4	7	7	7
4	1	1	1	4	4	1	10	4	4	1	1	1	1	10
7	7	7	10	7	10	7	10	4	7	10	7	10	10	10
10	7	1	7	7	10	7	7	10	7	7	7	7	4	7
10	7	7	4	7	4	7	7	1	7	7	1	4	7	1
10	7	7	4	7	7	1	10	7	7	4	1	4	7	1
10	10	7	7	10	10	7	10	1	7	10	10	7	7	10

Приложение 7.

месяц	окна, $м^2$	окна, тыс.руб.	двери, $м^2$	двери, тыс.руб.	подоконники, $м^2$	подоконники, тыс.руб.	погонаж, $м^2$	погонаж, тыс.руб.	прочие, $м^2$	монтаж, тыс.руб.	итог, тыс.руб.
янв-2013	136,0	585,2	60,8	133,5	2,2	2,7	882,6	38,4	0,0	6,6	766,4
фев-2013	128,8	509,0	49,6	104,4	8,8	10,6	1 651,2	39,1	24,4	10,5	698,0
мар-2013	104,5	362,6	90,4	308,9	71,7	101,5	3 706,4	48,8	9,5	41,9	873,2
апр-2013	89,9	274,1	113,9	383,7	62,5	65,6	3 250,2	56,5	38,2	51,5	869,5
май-2013	297,8	608,1	34,7	70,0	18,2	22,7	2 745,8	65,6	0,0	75,2	841,5
июн-2013	195,1	454,4	21,4	40,8	37,9	46,1	1 237,4	33,8	131,8	76,1	783,0
июл-2013	171,1	628,8	30,2	59,0	18,3	21,9	3 961,6	67,3	85,1	66,1	928,1
авг-2013	132,3	480,0	57,0	116,5	21,4	25,7	3 228,8	71,7	1,3	20,9	716,1
сен-2013	175,5	642,7	40,5	92,5	42,3	50,7	1 107,0	61,8	6,8	40,9	895,4
окт-2013	147,2	560,0	51,3	112,1	83,6	100,7	3 654,1	59,9	27,3	25,9	885,9
ноя-2013	181,1	629,2	44,5	107,2	31,8	34,6	5 057,2	173,7	0,0	52,0	996,7
дек-2013	184,5	559,0	83,5	188,2	64,2	86,1	4 309,0	139,2	8,4	18,4	999,3
янв-2014	123,8	477,4	102,5	216,3	3,2	4,2	654,3	52,5	0,0	16,8	767,2
фев-2014	172,6	708,6	52,4	161,6	1,1	1,5	1 443,2	54,0	4,0	260,1	1 189,7
мар-2014	153,9	818,6	64,5	191,1	10,7	15,9	813,6	43,0	30,1	50,2	1 148,9
апр-2014	163,2	834,5	90,6	276,3	8,6	16,6	1 414,0	61,5	2,9	99,3	1 291,2
май-2014	217,4	1 138,2	73,9	136,5	56,5	109,9	509,5	23,8	13,0	121,8	1 543,2
июн-2014	199,1	669,5	111,9	368,2	43,5	89,2	813,6	105,0	72,9	109,7	1 414,5
июл-2014	226,2	781,6	87,9	264,0	25,0	49,0	1 569,6	151,4	2,5	83,5	1 332,0
авг-2014	91,2	626,7	83,7	280,1	13,7	32,9	941,9	69,5	9,3	117,9	1 136,4
сен-2014	95,7	578,9	89,3	266,3	18,0	35,2	468,8	80,8	0,0	109,4	1 070,6
окт-2014	115,9	610,8	79,9	279,7	26,4	51,6	682,2	78,7	9,9	99,7	1 130,4
ноя-2014	34,5	169,3	77,6	393,9	70,9	56,2	1 776,4	67,0	32,3	71,0	789,7
дек-2014	56,4	300,3	38,3	139,7	29,4	67,4	2 876,0	117,2	65,5	97,2	787,2
янв-2015	25,8	192,0	48,7	258,9	32,0	64,7	2 767,2	90,0	86,6	30,3	722,5
фев-2015	53,4	444,1	82,9	286,9	10,8	21,1	1 020,4	77,8	50,3	62,6	942,8
мар-2015	32,0	293,9	80,3	291,3	0,0	0,0	1 608,7	225,4	21,9	64,9	897,4
апр-2015	12,1	111,5	94,5	395,5	14,7	28,9	1 890,1	96,7	182,7	118,2	933,5
май-2015	84,2	408,6	60,7	326,0	11,4	25,1	1 564,2	161,4	18,8	113,8	1 053,7
июн-2015	196,7	1 020,6	57,3	332,9	7,1	13,8	1 901,1	89,0	44,0	67,8	1 568,1

Проверочные тесты

1. *Выбрать правильные ответы (ответ):*
Статистические таблицы обязательно содержат итоговые показатели и состоят из

a. графического образа
b. сказуемого
c. кумуляты
d. полигона
e. подлежащего

2. *Выбрать правильные ответы (ответ):*
Вторичная группировка проводится двумя способами, такими как:

a. компактирование первоначальных интервалов
b. укрупненная перегруппировка
c. укрупнение первоначальных интервалов
d. долевая перегруппировка

3. *Выбрать правильные ответы (ответ):*
Научно организованный сбор сведений об изучаемых явлениях и процессах - это

a. статистическое наблюдение
b. статистическое решение
c. содержание программы наблюдения
d. статистический эксперимент

4. *Впишите пропущенные слова:*
Завершите предложение:

Сведения, собранные в процессе _____ статистического наблюдения, подвергаются научной обработке и _____.

5. *Выбрать правильные ответы (ответ):*
В зависимости от характера исходных данных и методологии исчисления статистические показатели могут быть выражены в форме

a. суммарных величин
b. абсолютных величин
c. относительных величин
d. индивидуальных величин

e. средних величин

6. *Выбрать правильные ответы (ответ):*
Отдельные значения варьирующего признака называются:

a. вариантами
b. частотами
c. модами

7. *Выбрать правильные ответы (ответ):*
Медианным является первый интервал, в котором сумма накопленных частот не превысит половину общего числа наблюдений.

a. Верно
b. Неверно

8. *Впишите пропущенные слова:*
Основная задача динамических методов заключается в _____ разновременных затрат и результатов путем приведения (_____) их ценности к единому моменту времени (обычно — к началу расчетного периода).

9. *Выбрать правильные ответы (ответ):*
К статистическим методам эффективности проекта относятся:

a. показатель источников инвестирования
b. расчет показателей рентабельности
c. срок окупаемости инвестиции
d. показатель социальной эффективности
e. анализ точки безубыточности проекта

10. *Выбрать правильные ответы (ответ):*
Общая вариация складывается из

a. систематической и случайной
b. симметричной и ассиметричной
c. ранговой и дифференциальной

11. *Выбрать правильные ответы (ответ):*
В зависимости от характера статистических данных применяют различные виды средних величин. Укажите каких.

a. Алгебраические
b. Геометрические

c. Арифметические
d. Гармонические

12. *Выбрать правильные ответы (ответ):*
Финансовое состояние предприятия —

a. способность платить по своим краткосрочным обязательствам
b. способность своевременно производить платежи по своим обязательствам
c. это совокупность показателей, отражающих наличие, размещение и использование финансовых ресурсов

13. *Выбрать правильные ответы (ответ):*
В статистике используется комплекс электронной обработки информации на базе пакета прикладных программ (ППП), который представляет собой:

a. обработку, контроль, корректировку и печать сводных таблиц
b. комплекс программных средств решения регламентных задач сводно-группировочного характера
c. совокупность программных, технических и организационных средств, предназначенная для решения задач формирования конкретных форм статистической отчетности

14. *Впишите число:*

На основе пакета прикладных программ (ППП) разработаны и внедрены рабочие проекты системных и локальных комплексов электронной обработки информации для автоматизированного решения более _____регламентных статистических задач.

Ответы размещены на сайте дисциплины: