МЕЖГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«БЕЛОРУССКО-РОССИЙСКИЙ УНИВЕРСИТЕТ»

Кафедра «Маркетинг и менеджмент»

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические рекомендации к лабораторным работам для студентов направления подготовки 38.04.02 «Менеджмент» дневной формы обучения

Могилев 2019

УДК 519.2
ББК 22.17
М 27

Рекомендовано к изданию
учебно-методическим отделом
Белорусско-Российского университета

Одобрено кафедрой «Маркетинг и менеджмент» «27» июня 2019 г., протокол № 12

Составитель канд. экон. наук, доц. А. В. Александров

Рецензент канд. экон. наук, доц. Т. В. Романькова

Методические рекомендации содержат задания для проведения лабораторных работ по дисциплине «Математическая статистика» для студентов направления подготовки 38.04.02 «Менеджмент» дневной формы обучения.

Учебно-методическое издание

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Ответственный за выпуск А. В. Александров
Технический редактор А. А. Подошевко
Компьютерная верстка Н. П. Полевничая

Подписано в печать . Формат 60×84/16. Бумага офсетная. Гарнитура Таймс. Печать трафаретная. Усл. печ. л. . Уч.-изд. л. . Тираж 21 экз. Заказ №

Издатель и полиграфическое исполнение:
Межгосударственное образовательное учреждение высшего образования
«Белорусско-Российский университет».
Свидетельство о государственной регистрации издателя,
изготовителя, распространителя печатных изданий
№ 1/156 от 07.03.2019.
Пр-т Мира, 43, 212022, Могилев.

Содержание

Введение

Выполнение лабораторных работ по дисциплине «Математическая статистика» формирует у обучающихся навыки математического анализа статистических данных исследований, а также интерпретации его результатов.

Перечень используемого оборудования для проведения лабораторных работ включает персональный компьютер с установленными программами MS Excel и MS Word.

В качестве методических рекомендаций по использованию встроенных функций и инструментов MS Excel следует пользоваться соответствующими разделами справки MS Excel.

Результаты выполнения лабораторной работы отражаются в рабочих окнах соответствующей программы (MS Excel).

Отчет по лабораторной работе представляется в электронной форме в виде документа MS Word и содержит:

1) титульный лист с указанием наименований университета и кафедры, названий учебной дисциплины и лабораторной работы, группы, фамилии, имени и отчества обучающегося, выполнившего лабораторную работу;
2) цель и задачи лабораторной работы;
3) перечень использованного оборудования и программного обеспечения, исходные данные;
4) порядок выполнения работы, который включает изложение последовательности выполненных действий (проведенных расчетов);
5) результаты в виде таблиц, полученных в ходе выполнения работы;
6) выводы по работе (краткое описание полученных таблиц).

Отчет по лабораторной работе составляется каждым обучающимся. Обучающийся, выполнивший работу и оформивший по ней отчет, допускается к защите лабораторной работы.

1 Построение рядов и функций распределения

Цель работы: приобрести навыки построения вариационных рядов и функций распределения, расчета статистических характеристик вариационного ряда.

Задачи работы:

– изучить методику построения вариационного ряда и расчета его характеристик;

– изучить встроенные возможности MS Excel по построению и вычислению характеристик вариационного ряда;

– изучить возможности MS Excel по построению полигонов частот и гистограмм.

Краткие теоретические сведения

Генеральной совокупностью называют множество всех значений измеряемой случайной величины.

Выборочной совокупностью или просто выборкой называют некоторое подмножество генеральной совокупности, полученное в результате произведенного эксперимента (наблюдения).

Объемом выборки называют число элементов в выборке.

Вариантой называют наблюдаемое значение x_i признака ξ.

Вариационным рядом называют последовательность вариант, записанных в возрастающем порядке.

Частотой n_i называют число наблюдений варианты. Сумма частот равна объему выборки: $\sum_{i=1}^{M} n_i = n,$ где M – число вариант.

Относительной частотой называют отношение частоты к объему выборки: $p_i{}^* = n_i / n$. Очевидно, что $\sum_{i=1}^{M} p_i^* = 1.$

Статистическим распределением выборки называют перечень вариант и соответствующих им частот или относительных частот.

Эмпирической функцией распределения называют функцию $F^*(x)$, определяющую для каждого значения x относительную частоту события $\xi < x$:

$$F^*(x) = \frac{n_x}{n},$$

где n_x – число вариант, меньших x.

Полигоном частот (или относительных частот) называют ломаную, отрезки которой соединяют точки (x_1, n_1), (x_2, n_2), ... (или точки $(x_i, p_i{}^*)$).

Гистограммой частот называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длины h_k, а высоты равны отношению n_k/h_k (плотность частоты). Площадь частичного k-го прямоугольника равна h_k (n_k / h_k) = n_k – сумме частот вариант, попавших в k-й интервал. Площадь гистограммы частот равна сумме всех частот, то есть объему выборки. Гистограмма относительных частот состоит из прямоугольников с высотами p_k* / h_k.

Выборочным средним называют величину

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{k=1}^{M} x_k n_k = \sum_{k=1}^{M} x_k p_k^*.$$

Выборочная дисперсия вычисляется по формуле

$$D_в = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{k=1}^{M}(x_k - \bar{x})^2 n_k = \sum_{k=1}^{M}(x_k - \bar{x})^2\, p_k^*.$$

Иногда более удобна другая формула

$$D_в = \overline{x^2} - \overline{x}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\frac{1}{n}\sum_{i=1}^{n} x_i)^2.$$

Исправленной выборочной дисперсией называется величина

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

Квадратные корни из величин выборочной и исправленной выборочной дисперсий называются выборочным средним квадратическим отклонением и исправленным выборочным средним квадратическим отклонением соответственно.

Задания

1 Прибор измеряет жирность молока с точностью до 0,2 %. Измерялась жирность молока 20 коров. Данные измерения приведены в процентах: 3,8; 3,4; 4,2; 3,4; 3,6; 3,6; 4,0; 4,2; 4,0; 3,6; 3,6; 3,8; 4,0; 4,2; 3,8; 3,8; 3,6; 3,8; 3,8; 4,0. Построить вариационный ряд, статистический ряд, полигон относительных частот, эмпирическую функцию распределения. Вычислить выборочное среднее, выборочную и исправленную дисперсии, соответствующие средние квадратические отклонения.

2 При измерении длины стержня были получены следующие данные (в миллиметрах):

9,68; 9,81; 9,77; 9,60; 9,61; 9,55; 9,74; 9,48; 9,72; 9,70;
9,52; 9,63; 9,68; 9,88; 9,47; 9,44; 9,82; 9,71; 9,84; 9,57;
9,79; 9,43; 9,59; 9,50; 9,78; 9,64; 9,72; 9,71; 9,58; 9,61.

Разбив все значения на пять интервалов, построить гистограмму относительных частот. Найти выборочное среднее, выборочную и исправленную дисперсии, соответствующие средние квадратические отклонения. За варианты принять середины интервалов.

Порядок выполнения работы

1 Внести исходные данные на лист MS Excel.

2 Вычислить требуемые характеристики, воспользовавшись расчетными формулами, а также встроенными функциями ЧАСТОТА, СРЗНАЧ, ДИСП.Г, ДИСП.В, СТАНДОТКЛОН.Г, СТАНДОТКЛОН.В.

3 Построить требуемые графические объекты, воспользовавшись пунктом меню ВСТАВКА/ДИАГРАММЫ, а также инструментом ГИСТОГРАММА надстройки АНАЛИЗ ДАННЫХ.

Контрольные вопросы

1 Дайте определение генеральной совокупности, выборки, размаха выборки и объема выборки.

2 Что называется вариационным и статистическим рядами, функцией распределения и статистической функцией распределения?

3 Какими свойствами обладает статистическая функция распределения?

4 Дайте определение группированного статистического ряда. Как строится гистограмма?

5 Как рассчитываются статистические показатели вариационного ряда?

2 Получение точечных оценок

Цель работы: овладеть навыками получения точечных оценок.

Задачи работы:

– изучить методику получения точечных оценок методом моментов;

– изучить методику получения точечных оценок методом максимального правдоподобия.

Краткие теоретические сведения

Наиболее часто применяемыми числовыми характеристиками случайной величины ξ являются начальные и центральные моменты различного порядка. Для дискретной случайной величины моменты порядка k определяются следующими формулами:

$$\alpha_k = \sum_{i=1}^{n} x_i^k p_i;$$

$$\mu_k = \sum_{i=1}^{n} \left(x_i - m_\xi\right)^k p_i;$$

для непрерывной случайной величины ξ

$$\alpha_k = \int_{-\infty}^{\infty} x^k f(x)\,dx;$$

$$\mu_k = \int_{-\infty}^{\infty} \left(x - m_\xi\right)^k f(x)\,dx.$$

Чаще всего используется первый начальный момент $a_1 = m_\xi$, называемый *математическим ожиданием случайной величины* ξ, и второй центральный момент $\mu_2 = D_\xi$, называемый *дисперсией*. Матожидание – это среднее значение случайной величины, его называют еще центром распределения, дисперсия характеризует разброс случайной величины относительно центра распределения. Часто вместо дисперсии используют среднее квадратичное отклонение $\sigma_\xi = \sqrt{D_\xi}$.

Если закон распределения случайной величины неизвестен, то не сможем вычислить числовые характеристики. В этом случае их заменяют оценками, полученными как функции выборки $x = \left(x_1, x_2, ..., x_n\right)$. Всякую функцию $t_n(x)$ от выборки называют статистикой. Подходящую статистику используют в качестве оценки числовой характеристики. Чаще всего оценками начальных и центральных моментов служат соответствующие выборочные начальные и центральные моменты:

$$a_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k;$$

$$m_k = \frac{1}{n}\sum_{i=1}^{n} \left(x_i - Mx\right)^k.$$

Таким образом, оценкой математического ожидания служит выборочное среднее $Mx = \frac{1}{n}\sum_{i=1}^{n} x_i$, но в качестве оценки можно взять и, например, величину $0{,}5 \cdot \left(x_{\max} + x_{\min}\right)$ и другие величины.

Чтобы иметь практическую ценность, оценка некоторого параметра θ должна удовлетворять следующим требованиям.

1 Оценка $t_n(x)$ должна приближаться к оцениваемому параметру θ по мере увеличения объема выборки. Если оценка стремится по вероятности к оцениваемому параметру, то она называется состоятельной.

2 Оценка не должна содержать систематической ошибки. Это означает,

что ее математическое ожидание должно совпадать с оцениваемым параметром θ, т. е. $M\left[t_n(x)\right]=\theta$. Такая оценка называется несмещенной.

3 Из всех состоятельных и несмещенных оценок предпочтительнее та, которая имеет наименьшую дисперсию. Такая оценка называется эффективной.

Например, среднее выборочное Mx является состоятельной оценкой математического ожидания, а $0{,}5\cdot\left(x_{\max}+x_{\min}\right)$ – несостоятельной. Второй выборочный центральный момент

$$m_2=S^2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-Mx\right)^2$$

является состоятельной оценкой дисперсии, но эта оценка смещенная. Несмещенные следующие оценки:

$$\tilde{S}^2=\frac{1}{(n-1)}\sum_{i=1}^{n}\left(x_i-Mx\right)^2;$$

$$S_*^2=\frac{1}{(n-1)}\sum_{i=1}^{n}\left(x_i-m_\xi\right)^2.$$

Если случайная величина распределена по нормальному закону, то оценка S_*^2 является и эффективной.

Пусть закон распределения известен, но зависит от одного или нескольких неизвестных параметров. Например, $f(x,\theta)$ – известная плотность распределения, а $\theta=(\theta_1, \theta_2, \ldots, \theta_s)$ – неизвестный параметр. Требуется по выборке $x=\left(x_1,x_2,..,x_n\right)$ оценить параметр θ.

Существует несколько методов оценки параметра θ. Рассмотрим два из них – метод моментов и метод функции правдоподобия.

Метод моментов заключается в том, что теоретический момент k-го порядка $\alpha_k=\alpha_k(\theta)$ приравнивают к соответствующему выборочному моменту α_k. Из полученного уравнения $\alpha_k(\theta)=\alpha_k$ находят неизвестный параметр θ. Например, случайная величина ξ (время безотказной работы радиоаппаратуры) распределена по экспоненциальному закону

$$f(t)=\frac{1}{T}e^{-\frac{t}{T}},\ t\geq 0,$$

где T – неизвестный параметр. Оценим T по методу моментов. Для этого найдем первый начальный момент

$$\alpha_1=\int_{-\infty}^{\infty}t\,f(t)\,dt=\frac{1}{\mathrm{T}}\int_{0}^{\infty}te^{-\frac{t}{\mathrm{T}}}dt=T.$$

Так как первый выборочный момент равен Mx, то из равенства $\alpha_1 = a_1$ получим $T = Mx$. Таким образом, оценкой неизвестного параметра T, найденной по методу моментов, является среднее выборочное Mx.

Пусть $L(u, \theta)$ – плотность распределения выборочного вектора $x = (x_1, x_2, .., x_n)$, $\theta = (\theta_1, \theta_2, \ldots, \theta_s)$ – неизвестный параметр. $L(u, \theta)$ – функция двух аргументов, неслучайного θ и случайного $x = (x_1, x_2, .., x_n)$, называется функцией правдоподобия. Так как $L(u, \theta)$ – плотность распределения, то оценка параметра θ, доставляющая максимум функции правдоподобия, является наиболее вероятной. Отсюда

$$\frac{\partial L(x, \theta)}{\partial \theta} = 0$$

или

$$\frac{\partial}{\partial \theta}[\ln L\ (x, \theta)] = 0$$

есть необходимые условия существования максимума. Оценка, полученная из этих условий, называется оценкой наибольшего правдоподобия.

Пусть $x = (x_1, x_2, .., x_n)$ – случайная выборка из генеральной совокупности, распределенной по нормальному закону

$$f\ (x, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)^2}{2\sigma^2}}$$

где θ – неизвестный параметр, $\theta = (\mu, \sigma)$.

Запишем функцию правдоподобия. Так как x_i – независимые случайные величины, распределенные по тому же закону, а плотность распределения вектора равна произведению плотностей составляющих вектора, то функция правдоподобия будет следующей:

$$L(x, \theta) = \prod_{i=1}^{n} f(x_i, \theta) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right].$$

Пусть ξ – дискретная случайная величина, закон распределения которой зависит от неизвестного параметра $p(\xi = x_i) = p_i(\theta)$. Будем рассматривать выборку $x = (x_1, x_2, .., x_n)$ как реализацию того, что случайная величина приняла последовательно значения $x_1, x_2, .., x_n$. Вероятность этого равна произведению вероятностей. Следовательно, функция правдоподобия будет

$$L(x, \theta) = \prod_{i=1}^{n} P(x_i, \theta).$$

Например, для дискретной случайной величины, распределенной по закону Пуассона

$$p_k = p(\xi = k) = \frac{\mu^k}{k!} e^{-\mu}; \; k = 0, 1, 2, 3, \ldots ,$$

функция правдоподобия может быть записана в виде

$$L(x, \mu) = \mu^{\sum_{i=1}^{n} x_i} \left(\prod_{i=1}^{n} x_i! \right)^{-1} e^{-n\mu} .$$

Здесь x_i – целые неотрицательные числа. Однако при больших n вычисления по этой формуле могут приводить к переполнениям разрядной сетки.

Задания

1 В таблице 2.1 представлены данные о распределении случайной величины X, равномерно распределенной в диапазоне значений $[a, b]$. Параметры a и b, от которых существенно зависит функция распределения, неизвестны. Необходимо методом моментов выполнить оценку для границ диапазона a и b, на котором распределена случайная величина.

Таблица 2.1 – Исходные данные

Опыт	Значение
1	0,716
2	0,894
3	0,899
4	0,009
5	0,207
6	0,255
7	0,422
8	0,350
9	0,697
10	0,445

Рекомендация. Поскольку случайная величина распределена равномерно в диапазоне значений от a до b, то функция плотности распределения случайной величины имеет вид $f(x) = 1 / (b - a)$. Тогда начальный момент первого порядка определяется по формуле $(a + b) / 2$, второго порядка – по формуле $(a^2 + ab + b^2) / 3$.

2 По представленной выборке вычислить оценки максимального правдоподобия для математического ожидания a и дисперсии σ^2 из условия максимума функции правдоподобия вида

$$-\frac{n}{2}\ln(2\pi)-n\ln(\sigma)-\sum_{i=1}^{n}\frac{(x_i-a)^2}{2\sigma^2},$$

предполагая при этом, что выборка порождена случайной величиной, подчиняющейся нормальному распределению.

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

Рекомендация. Использовать ограничение $\sigma \geq 0{,}0000001$, чтобы $\ln(\sigma)$ не был равен $-\infty$.

Порядок выполнения работы

1 Внести исходные данные на лист MS Excel.

2 Задать ячейки с искомыми параметрами, равными единице.

3 Сформировать систему уравнений (для задания 1) или целевую функцию (для задания 2), выполнив необходимые предварительные вычисления.

4 Используя надстройку ПОИСК РЕШЕНИЯ, вычислить искомые параметры.

Контрольные вопросы

1 Назовите выборочные числовые характеристики.

2 Что такое статистики и для чего они служат?

3 Какими свойствами должны обладать оценки?

4 Приведите примеры состоятельной, несмещенной и эффективной оценок.

5 Что такое функция правдоподобия? В чем сущность метода наибольшего правдоподобия?

3 Построение доверительных интервалов

Цель работы: овладеть навыками получения интервальных оценок.

Задачи работы:

– изучить методику получения интервальных оценок;

– изучить встроенные возможности MS Excel по получению интервальных оценок.

Краткие теоретические сведения

В предыдущей работе были рассмотрены методы, дающие оценку параметра в виде некоторого числа или точки на числовой оси. Такие оценки называют точечными. Точечная оценка без указания степени точности и надежности не имеет практического значения, т. к представляет собой только возможное значение случайной величины, т. е. сама точечная оценка является величиной случайной. Можно доказать, что в выборке объема n из генеральной совокупности, распределенной по нормальному закону $N(a,\sigma)$, среднее выборочное Mx распределено также по нормальному закону $N\left(a,\sigma/\sqrt{n}\right)$. Величина nS_*^2/σ^2 распределена по закону χ^2 с n степенями свободы, а $t_n=(Mx-a)\sqrt{n-1}/S$ – по закону Стьюдента с $n-1$ степенью свободы.

Чтобы получить представление о точности и надежности оценки $\tilde{\theta}$ для параметра θ, возьмем достаточно большую вероятность β и найдем такое $\delta>0$, для которого $P\left(\left|\tilde{\theta}-\theta\right|<\delta\right)=\beta$ или $P\left(-\delta<\theta-\tilde{\theta}<\delta\right)=P\left(\tilde{\theta}-\delta<\theta<\delta+\tilde{\theta}\right)=\beta$.

Последнее равенство означает, что точное, но неизвестное значение параметра θ с вероятностью β накрывается интервалом $l=\left(\tilde{\theta}-\delta,\tilde{\theta}+\delta\right)$. Этот интервал называют доверительным, а вероятность β – доверительной вероятностью или надежностью оценки. Очевидно, чем меньше δ для заданного β, тем точнее оценка.

В общем случае интервал, образованный статистиками $U(x)$ и $V(x)$, называется доверительным для оцениваемого параметра θ, если выполняется равенство $P\left(U(x)<\theta<V(x)\right)=\beta$. Здесь x – выборочный вектор, надежность β выбирается близкой к единице. Концы интервала называются доверительными границами.

Порядок нахождения доверительного интервала следующий. Подыскивают подходящую статистику $t_n(x,\theta)$, зависящую от параметра θ, но распределение которой от этого параметра не зависит. Задают надежность β и по закону распределения статистики $t_n(x,\theta)$ находят доверительные границы из последнего условия. Затем полученное неравенство решают относительно θ.

Таким образом, интервальные оценки математического ожидания нормального распределения:

– при известной генеральной дисперсии $D=\sigma^2$

$$\overline{X}_в-\frac{x_\gamma\sigma}{\sqrt{n}}<a<\overline{X}_в+\frac{x_\gamma\sigma}{\sqrt{n}},$$

где число x_y определяется из равенства $\Phi(x_y)=\gamma/2$, т. е. по таблице функции Лапласа находят значение аргумента x_y, которому соответствует значение функции Лапласа γ/2;

– при неизвестной генеральной дисперсии

$$\left(\bar{X}_в - \frac{t(\gamma,n)\sqrt{D_в}}{\sqrt{n-1}},\ \bar{X}_в + \frac{t(\gamma,n)\sqrt{D_в}}{\sqrt{n-1}}\right)$$

или

$$\left(\bar{X}_в - \frac{t(\gamma,n)S}{\sqrt{n}},\ \bar{X}_в + \frac{t(\gamma,n)S}{\sqrt{n}}\right),$$

где значение числа $t(\gamma,n)$ определяется по таблице критических точек распределения Стьюдента при уровне вероятности $\alpha = 1 - \gamma$ и числе степеней свободы $n - 1$.

Интервальная оценка дисперсии нормального распределения

$$\left(\frac{nD_в}{\chi^2_{пр,\gamma}},\ \frac{nD_в}{\chi^2_{лев,\gamma}}\right)$$

или

$$\left(\frac{n-1}{\chi^2_{пр,\gamma}}S^2, \frac{n-1}{\chi^2_{лев,\gamma}}S^2\right),$$

где значения $\chi^2_{лев,\gamma}$, $\chi^2_{пр,\gamma}$ находятся по таблице критических точек распределения χ^2 при числе степеней свободы $n - 1$ и уровнях вероятности $(1 + \gamma)/2$ и $(1 - \gamma)/2$ соответственно.

Интервальная оценка вероятности события

$$P\left(p_{лев,\gamma} < p < p_{пр,\gamma}\right) = \gamma.$$

При большом числе испытаний

$$p_{лев,\gamma} = p_1; \qquad p_{пр,\gamma} = p_2;$$

$$p_1 \approx p^* - x_\gamma\sqrt{p^*(1-p^*)/n}; \quad p_2 \approx p^* + x_\gamma\sqrt{p^*(1-p^*)/n}.$$

При малом числе испытаний – корни уравнений

$$\sum_{x=0}^{m-1} c_n^x p_{лев,\gamma}^x (1 - p_{лев,\gamma})^{n-x} = \frac{1+\gamma}{2};$$

$$\sum_{x=0}^{m} c_n^x p_{пр,\gamma}^x (1 - p_{пр,\gamma})^{n-x} = \frac{1-\gamma}{2}.$$

Задания

1 Выборка из большой партии электроламп содержит 100 ламп. Средняя продолжительность горения лампы выборки оказалась равной 1 000 ч. Найти с надежностью 0,95 доверительный интервал для средней продолжительности горения лампы всей партии, если известно, что среднее квадратическое отклонение продолжительности горения такой лампы – 40 ч. Предполагается, что продолжительность горения ламп распределена нормально.

2 Одним и тем же прибором со средним квадратическим отклонением случайных ошибок измерений 40 м произведено пять равноточных измерений расстояния от орудия до цели. Найти доверительный интервал для оценки истинного расстояния до цели с надежностью 0,9, зная среднее результатов измерений – 2 000 м.

3 По выборке объемом n = 20 из нормально распределенной генеральной совокупности вычислено значение дисперсии выборки, равное 1,5. Определить интервальную оценку для генеральной дисперсии с надежностью 0,96.

4 Из генеральной совокупности нормально распределенного признака извлечена выборка объемом n = 50:

2; -8; -4; 1; -9; -5; 2; 7; -15; -4;
-3; -5; -2; -10; -1; -7; 2; 3; -8; -6;
4; -1; 6; -6; -10; -4; -2; -2; 9; -13;
2; -1; 8; 0; -6; -6; -4; -2; 0; -13;
0; 1; -3; 2; 2; -9; -1; -3; 1; 2.

Построить статистический ряд и для полученного распределения:

а) найти выборочное среднее, выборочную и исправленную дисперсии;

б) построить доверительный интервал для среднего значения с надежностью 0,95;

в) построить доверительный интервал для среднего квадратического отклонения с надежностью 0,9.

Порядок выполнения работы

1 Вычислить доверительные интервалы по соответствующим формулам. Для определения значений критических точек воспользоваться встроенными функциями НОРМ.СТ.ОБР, СТЬЮДЕНТ.ОБР.2Х или ХИ2.ОБР / ХИ2.ОБР.ПХ.

2 Вычислить доверительные интервалы с помощью встроенных функций ДОВЕРИТ.НОРМ, ДОВЕРИТ.СТЬЮДЕНТ.

3 При выполнении задания 4 использовать знания, полученные на лабораторной работе № 1.

Контрольные вопросы

1 Что называется доверительным интервалом и доверительной вероятностью?

2 Дайте общую схему построения доверительного интервала.

3 Как изменяется доверительный интервал с увеличением надежности?

С увеличением объема выборки?

4 Как изменяется доверительный интервал в зависимости от того, известны ли другие параметры точно или нет?

4 Проверка гипотез о средних значениях

Цель работы: овладеть навыками проверки гипотез о средних значениях.

Задачи работы:
– изучить методику проверки гипотез о средних значениях;
– изучить встроенные возможности MS Excel по проверке гипотез о средних значениях.

Краткие теоретические сведения

Статистической называют гипотезу о виде неизвестного распределения или о параметрах известного распределения.

Нулевой (основной) называют выдвинутую гипотезу H_0.

Конкурирующей (альтернативной) называют гипотезу H_1, которая противоречит нулевой.

Простой называют гипотезу, содержащую только одно предположение.

Сложной называют гипотезу, которая состоит из конечного или бесконечного числа простых гипотез.

В итоге проверки гипотезы могут быть допущены ошибки двух родов.

Ошибка первого рода состоит в том, что будет отвергнута правильная нулевая гипотеза. Вероятность ошибки первого рода называют уровнем значимости и обозначают через α.

Ошибка второго рода состоит в том, что будет принята неправильная нулевая гипотеза. Вероятность ошибки второго рода обозначают через β.

Статистическим критерием (или просто критерием) называют случайную величину K, которая служит для проверки гипотезы. Для проверки гипотезы по данным выборок вычисляют частное наблюдаемое или эмпирическое значение критерия $K_{набл}$.

Множество всех возможных значений критерия разбивают на два непересекающихся подмножества: одно из них содержит значения критерия, при которых нулевая гипотеза отвергается, – критическая область, а другое – при которых она принимается – область принятия гипотезы. Точки, отделяющие одну область от другой, называют критическими точками $k_{кр}$. Гипотеза принимается, если наблюдаемое значение критерия оказывается в области принятия гипотезы. Поскольку K – одномерная случайная величина, то это условие формулируется в виде неравенства, например $K_{набл} < k_{кр}$.

Проверка гипотезы о числовом значении математического ожидания при известной дисперсии.

Проверяемая гипотеза задается в виде

$$H_0 : a = a_0 ,$$

а конкурирующая

$$H_1 : a \neq a_0 .$$

В качестве критерия используется величина

$$K = \frac{\overline{X}_в - a_0}{\sigma / \sqrt{n}},$$

значение которой является случайной величиной и подчиняется нормальному распределению $N(0,1)$.

Проверка гипотезы о числовом значении математического ожидания при неизвестной дисперсии.

В этом случае за основу проверки гипотезы положен критерий

$$K = \frac{\overline{X}_в - a_0}{S / \sqrt{n}}.$$

Этот критерий имеет t-распределение с числом степеней свободы $k = n - 1$.

Проверка гипотезы о равенстве математических ожиданий (*об однородности математических ожиданий*) *случайных величин, соответствующих двум выборкам.*

Проверяемая гипотеза задается в виде

$$H_0 : \mu_1 = \mu_2 ,$$

а конкурирующая

$$H_1 : \mu_1 \neq \mu_2 .$$

В общем случае гипотеза о равенстве математических ожиданий имеет вид:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_m ;$$

при конкурирующей

$$H_1 : \mu_{i_1} \neq \mu_{i_2} .$$

Для проверки гипотезы H_0 может использоваться целый ряд критериев.

Условием применения параметрических критериев является принадлежность наблюдений нормальному закону. К ним, например, относятся: критерий сравнения двух выборочных средних при известных дисперсиях; сравнения двух выборочных средних при неизвестных, но равных дисперсиях (критерий Стьюдента); сравнения двух выборочных средних при неизвестных и неравных дисперсиях (проблема Беренса-Фишера).

Применение критерия сравнения двух выборочных средних при неизвестных, но равных дисперсиях предусматривает вычисление статистики t:

$$t = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sqrt{\left[\frac{n_1 + n_2}{n_1 n_2}\right]\left[\frac{Q_1 + Q_2}{n_1 + n_2 - 2}\right]}},$$

где n_i – объем i-й выборки.

$$Q_i = \sum_{j=1}^{n_i}\left(x_{ij} - \bar{x}_i\right)^2; \quad \bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}.$$

В случае принадлежности выборок нормальному закону эта статистика подчиняется распределению Стьюдента с числом степеней свободы $\nu = n_1 + n_2 - 2$.

Задания

1 Из нормальной генеральной совокупности с известным средним квадратическим отклонением σ = 40 извлечена выборка объемом n = 64 и по ней найдена выборочная средняя $x_{ср}$ = 136,5. Требуется при уровне значимости 2α = 0,05 проверить нулевую гипотезу H_0: $a = a_0$ = 130 при конкурирующей гипотезе H_1: $a \neq 130$.

2 Двумя методами проведены измерения одной и той же физической величины, Первый метод дал следующие результаты: 9,6, 10,0, 9,8, 10,2, 10,6, 10,0; второй метод: 10,4, 9,7, 10,0, 10,3, 9,9, 10,0, 10,1, 9,9. Можно ли считать, что оба метода обеспечивают одинаковую точность измерений, если принять уровень значимости 2α = 0,01? Предполагается, что результаты измерений распределены нормально и выборки независимы.

Порядок выполнения работы

1 Провести проверку гипотез, рассчитав значения статистических критериев по формулам.

2 Провести проверку гипотез с помощью встроенной функции Z.TEST (задание 1) и инструментом ДВУХВЫБОРОЧНЫЙ Z-ТЕСТ ДЛЯ СРЕДНИХ надстройки АНАЛИЗ ДАННЫХ (задание 2).

Контрольные вопросы

1 Основные понятия статистической проверки гипотез.
2 Этапы статистической проверки гипотез.
3 Проверка гипотезы о числовом значении математического ожидания при известной дисперсии.
4 Проверка гипотезы о числовом значении математического ожидания при неизвестной дисперсии.
5 Проверка гипотезы о равенстве математических ожиданий случайных величин, соответствующих двум выборкам.

5 Проверка гипотез о дисперсии/вариации

Цель работы: овладеть навыками проверки гипотез о дисперсии/вариации.

Задачи работы:
– изучить методику проверки гипотез о дисперсии/вариации;
– изучить встроенные возможности MS Excel по проверке гипотез о дисперсии/вариации.

Краткие теоретические сведения

Полагаем, что X является случайной величиной, имеющей нормальное распределение, причем числовое значение дисперсии σ^2 неизвестно. Выборочная оценка $S^2 = \sum_{i=1}^{n}(X_i \quad X_в)^2 / (n \quad 1)$ дает приближенное представление о σ^2. Используя эту оценку, проверяется гипотеза

$$H_0 : \sigma^2 = \sigma_0^2 ,$$

где σ_0^2 – заранее заданное число.

В качестве критерия используется случайная величина:

$$K = \frac{(n-1)S^2}{\sigma_0^2}.$$

При выполнении гипотезы эта величина подчиняется χ^2 -распределению с числом степени свободы $k = n - 1$.

Проверяемая гипотеза о постоянстве дисперсии m выборок имеет вид:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_m^2,$$

а конкурирующая с ней гипотеза

$$H_1 : \sigma^2_{i_1} \neq \sigma^2_{i_2},$$

где неравенство выполняется, по крайней мере, для одной пары индексов i_1, i_2.

Для проверки такого вида гипотез применяется критерий Бартлетта.

Статистика критерия Бартлетта вычисляется в соответствии с соотношением

$$\chi^2 = M\left[1 + \frac{1}{3(m-1)}\left(\sum_{i=1}^{m}\frac{1}{\nu_i} - \frac{1}{N}\right)\right]^{-1},$$

где n_i – объемы выборок, $\nu_i = n_i$, если математическое ожидание известно, и $\nu_i = n_i - 1$, если неизвестно.

$$N = \sum_{i=1}^{m}\nu_i,$$

$$M = N\ln\left(\frac{1}{N}\sum_{i=1}^{m}\nu_i S_i^2\right) - \sum_{i=1}^{m}\nu_i \ln S_i^2,$$

где S_i^2 – оценки выборочных дисперсий.

При неизвестном математическом ожидании оценки

$$S_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}\left(X_{ji} - \overline{X}_i\right)^2,$$

где $\overline{X}_i = \dfrac{1}{n_i}\sum_{j=1}^{n_i} X_{ji}$;

$n_i - 1 = \nu_i$.

Если гипотеза H_0 верна, все $\nu_i > 3$ и выборки извлекаются из нормальной генеральной совокупности, то статистика приближенно подчиняется χ^2_{m-1}-распределению.

Задания

1 Из нормальной генеральной совокупности извлечена выборка объемом $n = 21$ и по ней найдена исправленная выборочная дисперсия $s^2 = 16{,}2$. Требуется при уровне значимости $2\alpha = 0{,}01$ проверить нулевую гипотезу H_0: $\sigma^2 = \sigma^2_0 = 15$, приняв в качестве конкурирующей гипотезы H_1: $\sigma^2 \neq 15$.

2 По двум независимым выборкам, объемы которых $n = 14$ и $m = 10$, извлеченным из нормальных генеральных совокупностей ξ и η, найдены исправленные выборочные дисперсии $s^2_\xi = 0{,}84$ и $s^2_\eta = 2{,}52$. При уровне значимости $2\alpha = 0{,}1$ проверить нулевую гипотезу H_0: $D\xi = D\eta$ о равенстве генеральных дисперсий при конкурирующей гипотезе H_1: $D\xi \neq D\eta$.

Порядок выполнения работы

1 Провести проверку гипотез, рассчитав значения статистических критериев по формулам.

2 В задании 2 провести проверку гипотез с помощью встроенной функции F.TEST и инструментом ДВУХВЫБОРОЧНЫЙ F-ТЕСТ ДЛЯ ДИСПЕРСИИ надстройки АНАЛИЗ ДАННЫХ.

Контрольные вопросы

1 Проверка гипотез о числовом значении дисперсии.

2 Проверка гипотез о постоянстве дисперсии нескольких выборок.

6 Проверка гипотез о вероятности

Цель работы: овладеть навыками проверки гипотез о вероятности.

Задачи работы: изучить методику проверки гипотез о вероятности.

Краткие теоретические сведения

Предположим, что A – случайное событие, вероятность p появления которого в единичном испытании неизвестна. Выдвинем гипотезу

$$H_0 : p = p_0$$

о том, что вероятность p равна числу p_0.

Альтернативная гипотеза H_1 имеет вид:

$$H_1 : p \neq p_0 .$$

В основе проверки гипотезы должно лежать сравнение числа p_0 с приближенными значениями вероятности p, найденным по опытным данным. Хорошим приближением к p является относительная частота $\omega = m/n$, где n – число независимых испытаний, проводимых в одинаковых условиях; m – число испытаний (из n приведенных), в которых произошло событие A. Поскольку A – случайное событие, то число m – случайная величина. Поэтому рассмотрим два случая.

Случай большого числа наблюдений.

При большом n распределение величины

$$\frac{\omega - p}{\sqrt{p(1-p)/n}}$$

можно аппроксимировать нормальным распределением. Если нулевая гипотеза справедлива, то распределение критерия

$$\frac{\omega - p_0}{\sqrt{p_0(1-p_0)/n}}$$

можно аппроксимировать нормальным распределением.

Случай малого числа наблюдений.

При малом числе наблюдений допущение о нормальном распределении критерия несправедливо. В этом случае проверка гипотезы проводится следующим образом.

Полагая $\gamma = 1 - \alpha$ и зная *n*, *m*, определяются табличные значения p_1, p_2. Если $p_0 < p_1$ или $p_0 > p_2$, то принимаем гипотезу $H_1 : p \neq p_0$; если $p_1 < p_0 < p_2$, то гипотезу $H_0 : p = p_0$.

Задания

1 Партия принимается, если вероятность того, что изделие окажется бракованным, не превышает $p_0 = 0{,}02$. Среди случайно отобранных n = 1 000 деталей оказалась m = 40 бракованных. Можно ли при уровне значимости α = 0,01 принять партию?

2 В $n = 5$ опытах событие A произошло $m = 4$ раза. Можно ли принять вероятность p равной 0,2 при уровне значимости $\alpha = 0{,}025$?

Порядок выполнения работы

1 Провести проверку гипотез, рассчитав значения статистических критериев по формулам.

2 Для получения критических значений критериев использовать соответствующие встроенные функции MS Excel.

Контрольные вопросы

1 Проверка гипотез о вероятности при большом числе наблюдений.

2 Проверка гипотез о вероятности при малом числе наблюдений.

7 Непараметрические статистические тесты

Цель работы: овладеть навыками проведения непараметрических статистических тестов.

Задачи работы: изучить методику проведения непараметрических статистических тестов.

Краткие теоретические сведения

Непараметрические критерии (тесты) позволяют исследовать данные без каких-либо допущений о характере распределения переменных. При этом обрабатываются не значения переменных, а их ранги или частоты. Непараметрические тесты можно применять при наличии в данных «выбросов» и неоднородных данных.

Непараметрические критерии используются для следующих переменных:

– для количественных переменных, распределение которых не подчиняется нормальному закону распределения;

– для переменных, измеренных в порядковой шкале;

– для переменных, измеренных в номинальной шкале.

Рассмотрим один из непараметрических аналогов t-теста, которым является *знаковый ранговый тест Уилкоксона*, в котором вычисляется ранг всех абсолютных значений исходных данных по возрастанию, затем ранг умножается на знак исходного значения. При равенстве двух абсолютных значений им присваивается средний ранг.

В знаковом ранговом тесте Уилкоксона делается только одно предположение о том, что распределение данных симметрично относительно медианы. Если в рамках нулевой гипотезы считать медиану равной 0, то при этом предполагается, что количество положительных рангов равно количеству отрицательных рангов, а их сумма равна 0.

Гипотезы:

– нулевая H_0: медиана разниц парных данных = 0;

– альтернативная H_1: медиана разниц парных данных ≠ 0.

Для вычисления статистики упорядочивают $m + n$ значений объединенной выборки, определяют сумму рангов R_1, соответствующую элементам первой выборки, и сумму рангов второй R_2. Вычисляются

$$U_1 = mn + \frac{m(m-1)}{2} - R_1;$$

$$U_2 = mn + \frac{n(n-1)}{2} - R_2.$$

Статистика критерия имеет вид: $U = \min\{U_1, U_2\}$.

Для достаточно больших выборок ($m+n>60$), когда объемы выборок не слишком малы ($m \geq 8, n \geq 8$), используется статистика

$$\tilde{z} = \frac{\left| U - \frac{mn}{2} \right|}{\sqrt{\frac{mn(m+n+1)}{12}}},$$

которая приближенно распределена в соответствии со стандартным нормальным законом.

Еще один непараметрический тест – *критерий знаков*. В нем вообще игнорируются значения данных, а учитывается только количество положительных и отрицательных значений. При этом проверяется точность равенства этих двух количеств.

Гипотезы:

– нулевая H_0: вероятность отрицательного значения равна вероятности положительного значения;

– альтернативная гипотеза H_1: вероятность отрицательного значения не равна вероятности положительного значения.

В качестве статистики критерия используется стандартизованная частота

$$Z = \frac{H - 1/2}{\sqrt{1/4n}} = 2\sqrt{n}\,(H - 1/2),$$

которая при условии истинности нулевой гипотезы имеет нормальное распределение.

Задания

1 Имеется 15 наблюдений, на основании которых необходимо определить, демонстрирует ли данная выборка уменьшение процентной доли: 0,05; 0,01; -0,025; -0,002; 0,058; 0,0258; -0,23; 0,065; 0,01; 0,09; 0,01; 0,06; 0; -0,06; 0,1.

2 В группе спасателей (n = 15) был проведен тренинг по формированию стрессоустойчивости (таблица 7.1).

Таблица 7.1 – Исходные данные

N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
До воздействия	12	13	6	13	14	12	10	17	14	15	13	12	16	14	12
После воздействия	16	18	17	20	15	15	17	17	16	16	14	10	23	20	11

Необходимо оценить достоверность сдвига исследуемого параметра.

Порядок выполнения работы

1 Внести исходные данные на лист MS Excel.

2 Провести проверку гипотез, рассчитав значения статистических критериев по формулам.

3 Для получения критических значений критериев использовать соответствующие встроенные функции MS Excel.

Контрольные вопросы

1 Область применения непараметрических тестов.
2 Ранговый критерий Уилкоксона.
3 Критерий знаков.

8 Корреляционно-регрессионный анализ

Цель работы: овладеть навыками корреляционно-регрессионного анализа.

Задачи работы:
– изучить методику корреляционно-регрессионного анализа;
– изучить встроенные возможности MS Excel по проведению корреляционно-регрессионного анализа.

Краткие теоретические сведения

Основная задача *корреляционного анализа* состоит в выявлении связи между случайными переменными путем точечной и интервальной оценки различных (парных, множественных, частных) коэффициентов корреляции.

Для проверки гипотезы о наличии линейной корреляционной связи наибольшее распространение имеет коэффициент линейной корреляции (Пирсона), предполагающий нормальный закон распределения наблюдений.

$$r_{xy} = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\sigma_x \sigma_y} = \frac{\sum (x-\overline{x})(y-\overline{y})}{n\sigma_x \sigma_y} = \frac{\sum (x-\overline{x})(y-\overline{y})}{\sqrt{\sum (x-\overline{x})^2 (y-\overline{y})^2}}.$$

Для двумерной нормально распределенной случайной величины XY при отсутствии линейной корреляции между X и Y коэффициент корреляции равен нулю. Поэтому процедура проверки заключается в выборочной оценке коэффициента корреляции и оценке значимости его отличия от нуля.

Коэффициент корреляции – параметр, характеризующий степень линейной взаимосвязи между двумя выборками. Коэффициент корреляции изменяется от –1 (строгая обратная линейная зависимость) до 1 (строгая прямая пропорциональная зависимость). При значении 0 линейной зависимости между двумя выборками нет.

На практике коэффициент корреляции принимает некоторые промежуточные значения между 1 и –1. Для оценки степени взаимосвязи можно руководствоваться следующей классификацией корреляционных связей по абсолютной величине коэффициента корреляции:

– очень сильная, практически линейная зависимость между параметрами при $r > 0{,}95$;

– сильная (тесная) при коэффициенте корреляции $r > 0{,}7$;

– средняя при $0{,}50 < r < 0{,}69$;

– умеренная при $0{,}30 < r < 0{,}49$;

– слабая при $0{,}20 < r < 0{,}29$;

– очень слабая при $r < 0{,}19$. В этих случаях обычно считают, что линейную взаимосвязь между параметрами выявить не удалось.

Вычислив выборочный коэффициент корреляции, необходимо оценить его статистическую значимость. Для того чтобы понять, насколько значимо отличие выборочного коэффициента корреляции от 0, строят доверительный интервал $(r - t\sigma_r;\, r + t\sigma_r)$. Средняя ошибка коэффициента корреляции вычисляется по формуле

$$\sigma_r = \sqrt{\frac{1-r^2}{n-2}}\,.$$

Если 0 не попадает в доверительный интервал, то коэффициент корреляции статистически значим.

Регрессионный анализ устанавливает формы зависимости между случайной величиной Y (зависимой) и значениями одной или нескольких переменных величин (независимых), причем значения последних считаются точно заданными. Такая зависимость обычно определяется некоторой математической моделью (уравнением регрессии), содержащей несколько неизвестных параметров. В ходе регрессионного анализа на основании выборочных данных находят оценки этих параметров, определяются статистические ошибки оценок или границы доверительных интервалов и проверяется соответствие (адекватность) принятой математической модели экспериментальным данным.

В *линейном регрессионном анализе* связь между случайными величинами предполагается линейной. В самом простом случае в парной линейной регрессионной модели имеются две переменные: X и Y. И требуется по n парам наблюдений (X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n) построить (подобрать) прямую линию, называемую линией регрессии, которая «наилучшим образом» приближает наблюдаемые значения. Уравнение этой линии $y = ax + b$ является регрессионным уравнением. С помощью регрессионного уравнения можно предсказать ожидаемое значение зависимой величины y, соответствующее заданному значению независимой переменной x.

В случае, когда рассматривается зависимость между одной зависимой переменной Y и несколькими независимыми X_1, X_2, ..., X_m, говорят о *множественной линейной регрессии*.

Тогда регрессионное уравнение имеет вид:

$$y = a_0 + a_1x_1 + a_2x_2 + \ldots + a_mx_m,$$

где a_0, a_1, a_2, …, a_m – требующие определения коэффициенты регрессии.

Коэффициенты уравнения регрессии определяются при помощи метода наименьших квадратов, добиваясь минимально возможной суммы квадратов расхождений реальных значений переменной Y и вычисленных по регрессионному уравнению.

Сущность данного метода заключается в нахождении параметров модели, при которых сумма квадратов отклонений эмпирических (фактических) значений результирующего признака от теоретических, полученных по выбранному уравнению регрессии, является минимальной, т. е.

$$S = \sum_{i=1}^{n}\left(y_i^{ð} - y_i\right)^2 = \sum_{i}^{n}(y_i^p - a_0 - a_1x)^2 \to \min,$$

где y_i^p – значение, вычисленное по уравнению регрессии;

$\left(y_i^p \quad y_i\right)$ – отклонение ε (ошибка, остаток);

n – количество пар исходных данных.

Мерой эффективности регрессионной модели является коэффициент детерминации R^2:

$$R^2 = \frac{\sum_{i=1}^{n}\left(y_i^p - \overline{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}.$$

Коэффициент детерминации может принимать значения между 0 и 1 и определяет, с какой степенью точности полученное регрессионное уравнение описывает (аппроксимирует) исходные данные. Исследуется также значимость регрессионной модели с помощью F-критерия (Фишера) и достоверность отличия коэффициентов a_0, a_1, a_2, …, a_m от нуля проверяется с помощью критерия Стьюдента.

Задание

По данным о совокупной выручке и затратах на рекламу (таблица 8.1):

– графически изобразить связь в виде диаграммы рассеяния;

– построить корреляционную таблицу для выявления наличия связи;

– измерить степень тесноты связи с помощью линейного коэффициента корреляции;

– построить уравнение;

– графически отобразить эмпирическую и теоретическую линии регрессии;

– оценить достоверность полученного уравнения корреляционно-регрессионной зависимости.

Таблица 8.1 – Исходные данные

Совокупная выручка, млн ден. ед.	Затраты на рекламу, тыс. ден. ед.
2,88	253
3,34	352
3,47	550
4,21	583
3,94	605
4,49	638
4,50	649
4,62	682
4,60	759
4,66	825

Порядок выполнения работы

1 Внести исходные данные на лист MS Excel.

2 Для построения диаграммы рассеяния воспользоваться пунктом меню ВСТАВКА/ДИАГРАММЫ/ТОЧЕЧНАЯ.

3 Вычислить коэффициент корреляции и параметры уравнения регрессии по формулам.

4 Вычислить коэффициент корреляции и параметры уравнения регрессии с использованием встроенных функций КОРРЕЛ, ОТРЕЗОК, НАКЛОН.

5 Провести корреляционно-регрессионный анализ, воспользовавшись инструментом ЛИНИЯ ТРЕНДА меню работы с диаграммами.

6 Провести корреляционно-регрессионный анализ, воспользовавшись инструментами КОРРЕЛЯЦИЯ и РЕГРЕССИЯ надстройки АНАЛИЗ ДАННЫХ.

Контрольные вопросы

1 Методика корреляционного анализа.

2 Методика регрессионного анализа.

9 Анализ динамических рядов

Цель работы: овладеть навыками анализа динамических рядов.

Задачи работы:

– изучить методику анализа динамических рядов;

– изучить встроенные возможности MS Excel по анализу динамических рядов.

Краткие теоретические сведения

Ряд динамики (временной ряд, хронологический ряд) – это ряд значений изучаемого показателя, изменяющихся во времени и расположенных в хронологическом порядке.

Каждый ряд динамики состоит из двух элементов:

1) уровень ряда динамики – это числовое значение статистического показателя, являющегося количественной оценкой изучаемого явления во времени;

2) показатель времени ряда динамики – это определенные моменты (даты) или периоды (годы, месяцы и т. п.) времени, для которых зафиксированы значения уровней ряда динамики.

Интервальный ряд динамики – это ряд, уровни которого характеризуют величину изучаемого явления за определенные интервалы (периоды) времени.

Моментный ряд динамики – это ряд, уровни которого отражают состояние изучаемого явления на определенные моменты времени.

Возможны два варианта сравнения уровней рядов динамики:

1) каждый i-й уровень ряда сравнивают с одним и тем же уровнем, выбранным в качестве базы сравнения. Полученные показатели называются базисными;

2) каждый i-й уровень ряда сравнивают с предшествующим уровнем. Рассчитанные показатели называются цепными.

Аналитические показатели ряда динамики:

– абсолютный прирост:

а) базисные показатели

$$\Delta y^{б} = y_i - y_1;$$

б) цепные показатели

$$\Delta y^{ц} = y_i - y_{i-1};$$

– коэффициент роста:

а) базисные показатели

$$K_p^{б} = \frac{y_i}{y_1};$$

б) цепные показатели

$$K_p^{ц} = \frac{y_i}{y_{i\text{-}1}};$$

– темп роста:
а) базисные показатели

$$T_p^{б} = \frac{y_i}{y_1} \cdot 100 = K_p^{б} \cdot 100;$$

б) цепные показатели

$$T_p^{ц} = \frac{y_i}{y_{i\text{-}1}} \cdot 100 = K_p^{ц} \cdot 100;$$

– темп прироста:
а) базисные показатели

$$T_{пр}^{б} = \frac{y_i\text{-}y_1}{y_1} \cdot 100 = \frac{\Delta y^{б}}{y_1} \cdot 100 = T_p^{б} - 100;$$

б) цепные показатели

$$T_{пр}^{б} = \frac{y_i\text{-}y_{i\text{-}1}}{y_{i\text{-}1}} \cdot 100 = \frac{\Delta y^{ц}}{y_{i\text{-}1}} \cdot 100 = T_p^{ц} - 100;$$

– абсолютное значение 1 % прироста

$$A_{1\,\%} = \frac{\Delta y^{ц}}{T_{пр}^{ц}}$$

или

$$A_{1\,\%} = y_{i\text{-}1} / 100.$$

Для обобщения характеристики динамики исследуемого явления за ряд периодов определяют средние показатели ряда динамики.

1 Средний уровень ряда динамики исчисляется в зависимости от вида ряда.

В интервальном ряду динамики:

– в ряду с равными интервалами

$$\bar{y} = \frac{\sum y_i}{n};$$

– в ряду с неравными интервалами

$$\bar{y} = \frac{\sum y_i \cdot t_i}{\sum t_i},$$

где y – уровни ряда динамики, сохраняющиеся без изменения в течение периода времени t;

t – веса, длительность интервала времени между смежными датами.

В моментном ряду динамики:

– если приводятся данные только на начало и на конец периода,

$$\bar{y} = (y_1 + y_2)/2;$$

– если моменты времени расположены через равные промежутки,

$$\bar{y} = \frac{\frac{1}{2}y_1 + y_2 + \ldots + y_{n-1} + \frac{1}{2}y_n}{n-1};$$

– если моменты времени расположены через неравные промежутки,

$$\bar{y} = \frac{\sum y_i \cdot t_i}{\sum t_i}.$$

2 Средний абсолютный прирост (абсолютное изменение)

$$\overline{\Delta y} = \frac{\sum \Delta y^{ц}}{n} = \frac{\Delta y^{б}}{n}.$$

3 Средний коэффициент роста

$$\overline{K_p} = \sqrt[n-1]{\frac{y_n}{y_1}} = \sqrt[n]{K_1^{ц} \cdot K_2^{ц} \cdot \ldots \cdot K_n^{ц}} = \sqrt[n]{\Pi K_p^{ц}}.$$

4 Средний темп роста

$$\overline{T_p} = \overline{K_p} - 100\ \%.$$

5 Средний темп прироста

$$\overline{T_{пр}} = \overline{T_p} - 100\ \%.$$

Одна из важнейших задач статистики заключается в определении в рядах динамики общей тенденции развития явления. Основной тенденцией развития (трендом) называется плавное и устойчивое изменение уровня явления во времени, свободное от случайных колебаний.

Чтобы выявить общую тенденцию в изменении уровней ряда, ряды динамики подвергаются обработке следующими методами:

1) метод укрупнения интервалов, основан на укрупнении периодов времени, к которым относятся уровни ряда динамики;

2) метод скользящей (подвижной) средней, заключается в том, что исчисляется средний уровень из определенного числа первых по счету уровней ряда, затем – из такого же числа уровней, но начиная со второго по счету, далее – начиная с третьего и т. д.;

3) аналитическое выравнивание ряда динамики – общая тенденция развития рассчитывается как функция времени:

$$y_t = f(t),$$

где y_t – уровни динамического ряда, вычисленные по соответствующему аналитическому уравнению на момент времени t.

Основа большинства методов прогнозирования – экстраполяция тенденции, связанная с распространением закономерностей, действующих в изучаемом периоде, за его пределы или, другими словами, это получение представлений о будущем на основе информации, относящейся к прошлому и настоящему.

Различают следующие методы экстраполяции:

– наивные, или простейшие (по среднему абсолютному приросту или среднему коэффициенту роста);

– по уравнению тренда.

Задания

1 Имеются данные о производстве изделия «А» за 5 лет (таблица 9.1).

Таблица 9.1 – Данные о производстве изделия «А»

Год	1	2	3	4	5
Объем выпуска, тыс. шт.	810	822	800	870	915

Рассчитать:

– показатели динамики цепным и базисным способами: абсолютный прирост; темп роста; темп прироста;

– среднегодовые показатели: среднегодовой выпуск продукции; средний абсолютный прирост; среднегодовой темп роста; среднегодовой темп прироста.

Графически изобразить динамику выпуска продукции. Сделать выводы.

2 Используя взаимосвязь показателей динамики, определить уровни ряда динамики и недостающие цепные показатели динамики в таблице 9.2.

Таблица 9.2 – Анализ динамики

Год	Произведено продукции, млрд р.	По сравнению с предыдущим годом			
		Абсолютный прирост, млрд р.	Темп роста, %	Темп прироста, %	Абсолютное значение 1 % прироста
1	92,5				
2		4,8			
3			104,0		
4				5,8	
5					
6		7,0			1,15

3 Имеются следующие данные об отправке грузов железнодорожным транспортом в регионе за год (таблица 9.3).

Таблица 9.3 – Данные об отправке грузов

Месяц	Отправлено груза, млн т	Месяц	Отправлено груза, млн т	Месяц	Отправлено груза, млн т
Январь	92	Май	90	Сентябрь	85
Февраль	85	Июнь	87	Октябрь	90
Март	93	Июль	86	Ноябрь	86
Апрель	92	Август	88	Декабрь	88

Рассчитать показатели изменения уровней ряда динамики (по цепной и базисной схемам) и средние показатели ряда динамики.

Произвести сглаживание ряда динамики методом укрупнения динамического ряда и по скользящей средней (по трем точкам).

Изобразить графически фактические и сглаженные уровни ряда. Сделать выводы о характере общей тенденции показателей.

4 Имеются данные, характеризующие динамику производства продукции предприятия по месяцам (таблица 9.4).

Таблица 9.4 – Динамика производства продукции по месяцам

Месяц	1	2	3	4	5	6	7	8	9	10	11	12
Выпуск продукции, шт.	75	105	114	129	138	129	152	138	142	155	157	157

Провести сглаживание ряда динамики. Сделать выводы о тенденции.

Предполагая, что выявленная закономерность сохранится и в дальнейшем, спрогнозировать объем производства на следующий квартал, используя в качестве закономерности:

а) средний абсолютный прирост;

б) средний темп роста;

в) трендовую модель по уравнению прямой.

Порядок выполнения работы

1 Внести исходные данные на лист MS Excel.

2 Вычислить все требуемые показатели по соответствующим формулам.

3 Для сглаживания динамического ряда воспользоваться также инструментом СГЛАЖИВАНИЕ надстройки АНАЛИЗ ДАННЫХ.

4 Для построения прогноза воспользоваться также инструментом ЛИНИЯ ТРЕНДА меню работы с диаграммами.

Контрольные вопросы

1 С какой целью анализируются данные рядов динамики?

2 Что такое правила построения рядов динамики и чем они характеризуются?

3 Каковы основные показатели динамики? Как они рассчитываются?

4 Назовите виды колебаний уровней временного ряда.

5 Что такое средний уровень ряда? Как он исчисляется?

6 Как может быть выявлена основная тенденция в изменениях уровней ряда динамики?

7 Назовите преимущества и роль аналитического выравнивания уровней временного ряда.

8 Как можно рассчитать скользящую среднюю и для каких целей она может быть использована?

Список литературы

1 **Белько, И. В.** Теория вероятностей, математическая статистика, математическое программирование: учебное пособие / И. В. Белько, И. М. Морозова, Е. А. Криштапович. – Москва: ИНФРА-М; Новое знание, 2016. – 299 с.

2 **Постовалов, С. Н.** Математическая статистика: конспект лекций / С. Н. Постовалов, Е. В. Чимитова, В. С. Карманов. – Новосибирск: НГПУ, 2014. – 140 с.

3 **Сапожников, П. Н.** Теория вероятностей, математическая статистика в примерах, задачах и тестах: учебное пособие / П. Н. Сапожников, А. А. Макаров, М. В. Радионова. – Москва: КУРС; ИНФРА-М, 2016. – 496 с.

4 **Соколов, Г. А.** Основы математической статистики: учебник / Г. А. Соколов. – 2-е изд. – Москва: ИНФРА-М, 2014. – 368 с.

5 **Хуснутдинов, Р. Ш.** Математическая статистика: учебное пособие / Р. Ш. Хуснутдинов. – Москва: ИНФРА-М, 2015. – 205 с.