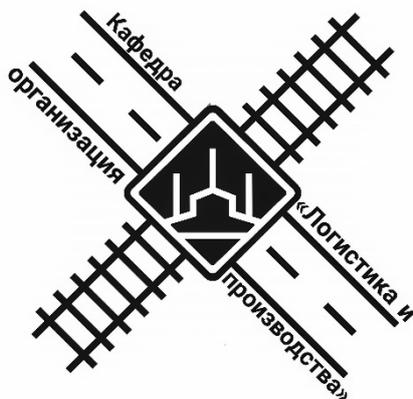


МЕЖГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«БЕЛОРУССКО-РОССИЙСКИЙ УНИВЕРСИТЕТ»

Кафедра «Логистика и организация производства»

КОЛИЧЕСТВЕННЫЕ МЕТОДЫ АНАЛИЗА В БИЗНЕСЕ

*Методические рекомендации к лабораторным работам
для студентов специальности 1-25 80 01 «Экономика»
очной и заочной форм обучения*



Могилев 2020

УДК 005.511
ББК 65.23
К60

Рекомендовано к изданию
учебно-методическим отделом
Белорусско-Российского университета

Одобрено кафедрой «Логистика и организация производства»
«5» октября 2020 г., протокол № 4

Составитель ст. преподаватель Т. А. Бородич

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

В методических указаниях представлены материалы к лабораторным работам для студентов специальности 1-25 80 01 «Экономика» очной и заочной форм обучения.

Учебно-методическое издание

КОЛИЧЕСТВЕННЫЕ МЕТОДЫ АНАЛИЗА В БИЗНЕСЕ

Ответственный за выпуск	М. Н. Гриневич
Корректор	А. А. Подошевка
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 16 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.

© Белорусско-Российский
университет, 2020

Содержание

1 Формирование таблицы распределения частот и построение гистограммы по величине запрашиваемых параметров.....	4
2 Проведение количественного анализа с помощью сводных таблиц в EXCEL.....	7
3 Первичная обработка опытных данных при помощи модуля BasicStatistics/Tables в MS STATISTICA	12
4 Решение задач на определение вероятности событий	18
5 Решение задач на применение нормального, биномиального, пуассоновского распределения средствами EXCEL.....	20
6 Решение задач на применение нормального, биномиального, пуассоновского распределения.....	23
7 Решение задач на построение выборок из генеральной совокупности средствами EXCEL.....	26
8 Построение доверительного интервала для разности средних и разности долей средствами EXCEL и в MS STATISTICA	27
9 Проверка гипотез для различных параметров средствами EXCEL и в MS STATISTICA	30
Список литературы	33

1 Формирование таблицы распределения частот и построение гистограммы по величине запрашиваемых параметров

Цель работы – изучить особенности построения гистограмм средствами MS Excel.

Задачи работы:

- научиться извлекать из данных первоначальную, полезную информацию путем построения графиков и диаграмм;
- изучить различные встроенные средства MS Excel для построения диаграмм.

Задание 1

По исходным данным о банках, представленным в таблице 1.1, произвести группировку 40 банков по кредитам частных лиц и по собственному капиталу.

Построить гистограмму частот и кумуляту.

Таблица 1.1 – Характеристика банков

Номер банка	Собственный капитал, %	Кредиты частным лицам, тыс. р.	Номер банка	Собственный капитал, %	Кредиты частным лицам, тыс. р.
1	10	308437	21	23	2811
2	16	5205	22	26	46
3	8	5084	23	15	171
4	13	1361	24	23	245
5	11	5768	25	19	1773
6	8	4466	26	13	3993
7	3	1392	27	13	3254
8	12	7266	28	12	764
9	8	4119	29	9	1218
10	8	10828	30	10	906
11	10	2719	31	11	1226
12	11	3576	32	14	4
13	16	8170	33	11	22267
14	36	511	34	11	31
15	9	822	35	15	23
16	15	1693	36	12	311
17	11	476	37	30	610
18	14	421	38	21	1178
19	19	38799	39	9	2600
20	9	993	40	16	1

Задание 2

По исходным данным о предприятиях, представленным в таблице 1.2, произвести группировку 20 предприятий по объему выполненных работ.

Таблица 1.2 – Характеристика предприятий

Номер предприятия	Объем выполненных работ, д. е.	Затраты, д. е.	Номер предприятия	Объем выполненных работ, д. е.	Затраты, д. е.
1	18,7	3,3	11	17,8	2,1
2	17,5	2,4	12	11,9	4,0
3	10,6	3,2	13	12,3	4,1
4	17,8	2,2	14	15,7	1,5
5	10,5	3,3	15	18,1	2,5
6	18,5	3,0	16	9,8	3,2
7	16,4	2,1	17	13,9	4,7
8	17,5	2,4	18	12,6	4,5
9	15,5	1,5	19	9,6	3,0
10	17,0	2,7	20	8,6	2,8

Охарактеризовать каждую группу и совокупность в целом: числом предприятий; объемом затрат – всего и в среднем на одно предприятие.

Построить гистограмму частот и кумуляту.

Методические указания к выполнению заданий

При проведении группировок важное значение имеет правильное решение о том, на какое число групп следует подразделять совокупность. Если признак атрибутивный, то число групп, на которое следует подразделять совокупность, определяется числом качественных градаций этого признака. В случае, если группировочный признак количественного порядка непрерывный или дискретный, с большим размахом вариации, то число групп может быть определено по формуле Стерджесса

$$n = 1 + 3,322 \lg N, \quad (1.1)$$

где n – число групп;

N – численность совокупности.

Для характеристики групп должны быть образованы интервалы, которые могут быть равные и неравные, открытые и закрытые. При равных интервалах их величина определяется по формуле

$$i = (X_{\max} - X_{\min}) / n = R / n, \quad (1.2)$$

где n – число групп;

X_{\max} – максимальное значение;

X_{\min} – минимальное значение;

R – размах вариации.

При построении равных интервалов определяются нижняя и верхняя границы каждого из них, причем считается, что пределы наблюдения могут входить «включительно» или «исключительно». На практике применяются оба метода, но все же предпочтительнее принцип «исключительно».

Для графического изображения интервального вариационного ряда применяется гистограмма. При ее построении на оси абсцисс откладываются интервалы ряда, высота которых равна частотам, отложенным на оси ординат. Над осью абсцисс строятся прямоугольники, площадь которых соответствует величинам произведений интервалов на их частоты.

В практике экономической работы возникает потребность в преобразовании рядов распределения в кумулятивные ряды, строящиеся по накопленным частотам (с их помощью можно определить структурные средние, проследить за процессом концентрации изучаемого явления). Используя накопленные частоты, строят график в виде кумуляты (кривой сумм). При графическом изображении кумуляты накопленные частоты наносят на поле графика в виде перпендикуляров к оси абсцисс в верхних границах интервалов. Длина этих линий равна величине накопленных частот в конкретном интервале. Соединяя затем эти перпендикуляры, получаем ломаную линию.

Строится вариационный ряд по группировочному признаку. Упорядочиваются данные по величине возрастания результативного признака с помощью функции «Сортировка».

По формуле (1.1), определяется количество интервалов, а по формуле (1.2) – высота интервала i .

Верхняя граница каждого интервала определяется по следующей формуле (полученные значения используются в меню «Гистограмма» как карман интервалов):

$$x = X_{\min} + i \cdot k, \quad (1.3)$$

где k – порядковый номер интервала.

Значения частот определяются при помощи процедуры «Гистограмма» (закладка «Данные», раздел «Анализ данных»):

1) входной интервал – исходные значения результативного признака;

2) интервал карманов – значения границ интервалов группировочного признака.

Форма представления отчета: предоставить преподавателю задания по теме, выполненные на компьютере.

Вопросы для защиты работы

- 1 Дайте понятие гистограммы, кармана частот, интервал.
- 2 Виды графических методов исследования данных и особенности их применения.

2 Проведение количественного анализа с помощью сводных таблиц в EXCEL

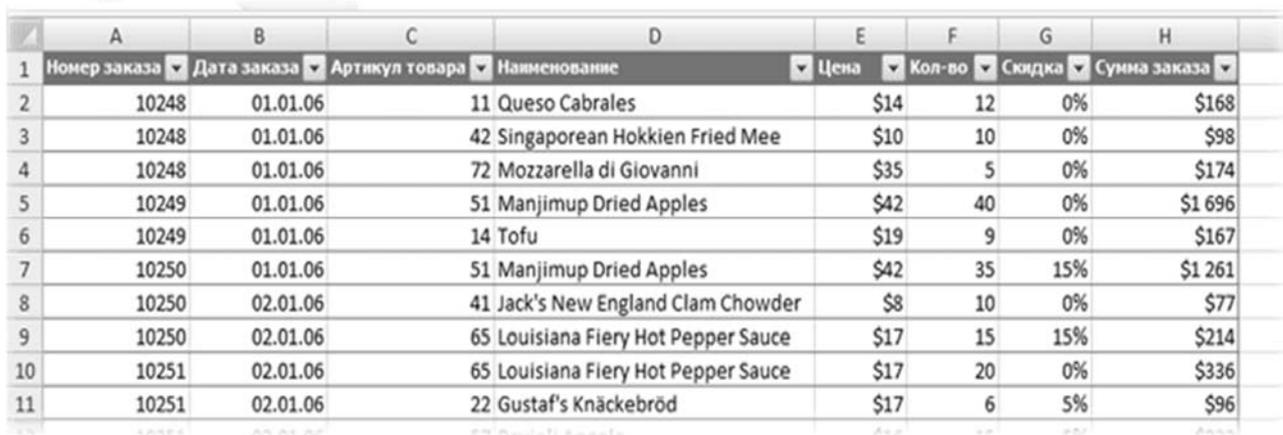
Цель работы – изучить особенности анализа с помощью сводных таблиц MS Excel.

Задачи работы:

- научиться извлекать из данных полезную информацию путем вычисления определенных характеристик, построения сводных таблиц;
- изучить различные способы построения сводных таблиц в MS Excel.

Задание

Провести группировку данных с помощью инструмента «Сводные таблицы» (рисунок 2.1)



	A	B	C	D	E	F	G	H
1	Номер заказа	Дата заказа	Артикул товара	Наименование	Цена	Кол-во	Скидка	Сумма заказа
2	10248	01.01.06	11	Queso Cabrales	\$14	12	0%	\$168
3	10248	01.01.06	42	Singaporean Hokkien Fried Mee	\$10	10	0%	\$98
4	10248	01.01.06	72	Mozzarella di Giovanni	\$35	5	0%	\$174
5	10249	01.01.06	51	Manjimup Dried Apples	\$42	40	0%	\$1 696
6	10249	01.01.06	14	Tofu	\$19	9	0%	\$167
7	10250	01.01.06	51	Manjimup Dried Apples	\$42	35	15%	\$1 261
8	10250	02.01.06	41	Jack's New England Clam Chowder	\$8	10	0%	\$77
9	10250	02.01.06	65	Louisiana Fiery Hot Pepper Sauce	\$17	15	15%	\$214
10	10251	02.01.06	65	Louisiana Fiery Hot Pepper Sauce	\$17	20	0%	\$336
11	10251	02.01.06	22	Gustaf's Knäckebröd	\$17	6	5%	\$96

Рисунок 2.1 – Исходные данные

Сгруппировать данные по сумме заказа, дате заказа.

Рассчитать:

- выручку по каждому дню: общую, среднюю, максимальную и минимальную;
- среднюю цену.

Методические указания к выполнению задания

Обрабатывать большие объемы информации и составлять сложные многоуровневые отчеты достаточно непросто без использования средств автоматиза-

ции. Excel как раз и является инструментом, позволяющим упростить эти задачи, путем создания сводных (перекрестных) таблиц данных (Pivot table).

Сводная таблица в Excel используется для:

- выявления взаимосвязей в большом наборе данных;
- группировки данных по различным признакам и отслеживания тенденции изменений в группах;
- нахождения повторяющихся элементов, детализации и т. п.;
- создания удобных для чтения отчетов, что является самым главным.

Создавать сводные таблицы можно двумя способами. Рассмотрим каждый из них.

Способ 1. Создание сводных таблиц, используя стандартный инструмент Excel «Сводная таблица».

Перед тем как создавать отчет сводной таблицы, определимся, что будет использоваться в качестве источника данных. Рассмотрим вариант с источником, находящимся в этом же документе.

1 Для начала создайте простую таблицу с перечислением элементов, которые вам нужно использовать в отчете. Верхняя строка обязательно должна содержать заголовки столбцов.

2 Откройте вкладку «Вставка» и выберите из раздела «Таблицы» инструмент «Сводная таблица».

Если вместе со сводной таблицей нужно создать и сводную диаграмму – нажмите на стрелку в нижнем правом углу значка «Сводная таблица» и выберите пункт «Сводная диаграмма».

3 В открывшемся диалоговом окне «Создание сводной таблицы» выберите только что созданную таблицу с данными или ее диапазон. Для этого выделите нужную область.

В качестве данных для анализа можно указать внешний источник: установите переключатель в соответствующее поле и выберите нужное подключение из списка доступных.

4 Далее нужно будет указать, где размещать отчет сводной таблицы. Удобнее всего это делать на новом листе.

5 После подтверждения действия нажатием кнопки «ОК», будет создан и открыт макет отчета (рисунок 2.2).

В правой половине окна создается панель основных инструментов управления – «Список полей сводной таблицы». Все поля (заголовки столбцов в таблице исходных данных) будут перечислены в области «Выберите поля для добавления в отчет». Отметьте необходимые пункты и отчет сводной таблицы с выбранными полями будет создан.

Расположением полей можно управлять – делать их названиями строк или столбцов, перетаскивая в соответствующие окна, а также и сортировать в удобном порядке. Можно фильтровать отдельные пункты, перетаскивая соответствующее поле в окно «Фильтр». В окно «Значение» помещается то поле, по которому производятся расчеты и подводятся итоги. Другие опции для редактирования отчетов доступны из меню «Работа со сводными таблицами» на вкладках «Параметры» и «Конструктор». Почти каждый из инструментов этих вкладок

имеет массу настроек и дополнительных функций.

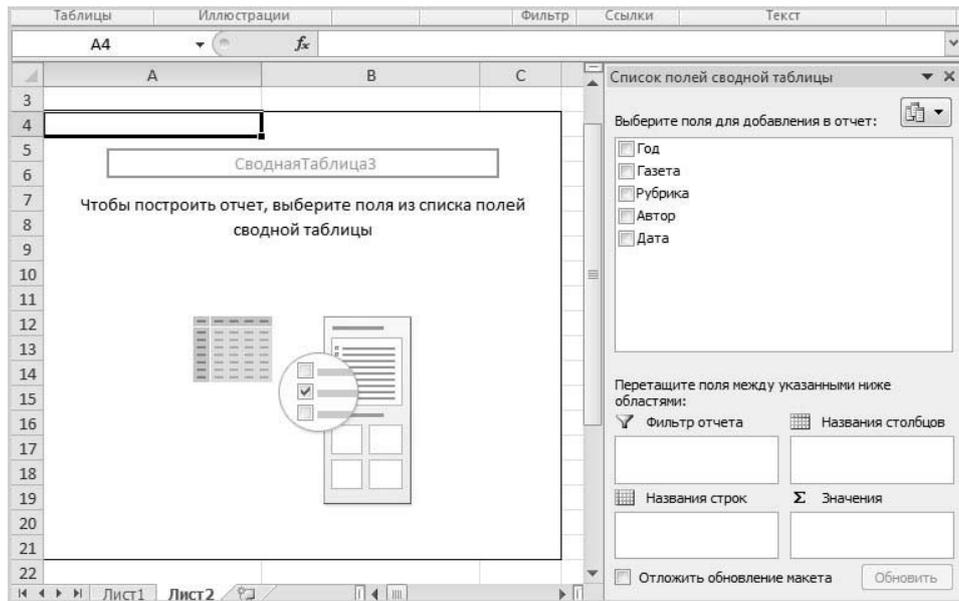


Рисунок 2.2 – Макет отчета

Способ 2. Создание сводной таблицы с использованием инструмента «Мастер сводных таблиц и диаграмм».

Чтобы применить этот способ, придется сделать доступным инструмент, который по умолчанию на ленте не отображается. Откройте вкладку «Файл» – «Параметры» – «Панель быстрого доступа». В списке «Выбрать команды из» отметьте пункт «Команды на ленте». А ниже, из перечня команд, выберите «Мастер сводных таблиц и диаграмм». Нажмите кнопку «Добавить». Иконка мастера появится сверху, на панели быстрого доступа.

Для создания сводных таблиц с помощью Мастера сводных таблиц выполните следующее.

1 Кликните по иконке мастера в панели быстрого доступа. В диалоговом окне поставьте переключатель на нужный вам пункт списка источников данных:

- «в списке или базе данных Microsoft Excel» – источником будет база данных рабочего листа, если таковая имеется;
- «во внешнем источнике данных» – если существует подключение к внешней базе, которое нужно будет выбрать из доступных;
- «в нескольких диапазонах консолидации» – если требуется объединение данных из разных источников;
- «данные в другой сводной таблице или сводной диаграмме» – в качестве источника берется уже существующая сводная таблица или диаграмма.

2 После этого выбирается вид создаваемого отчета – «сводная таблица» или «сводная диаграмма (с таблицей)».

Если в качестве источника выбран текущий документ, где уже есть простая таблица с элементами будущего отчета, задайте диапазон охвата – выделите курсором нужную область. Далее выберите место размещения таблицы – на но-

вом или на текущем листе, и нажмите «Готово». Сводная таблица будет создана. Если же необходимо консолидировать данные из нескольких источников, поставьте переключатель в соответствующую область и выберите тип отчета. А после нужно будет указать, каким образом создавать поля страницы будущей сводной таблицы: одно поле или несколько полей.

При выборе «Создать поля страницы» прежде всего придется указать диапазоны источников данных: выделите первый диапазон, нажмите «Добавить», потом следующий и т. д.

Для удобства диапазонам можно присваивать имена. Для этого выделите один из них в списке и укажите число создаваемых для него полей страницы, потом задайте каждому полю имя (метку). После этого выделите диапазон и т. д.

После завершения нажмите кнопку «Далее», выберите месторасположение будущей сводной таблицы – на текущем листе или на другом, нажмите «Готово» и ваш отчет, собранный из нескольких источников, будет создан.

При выборе внешнего источника данных используется приложение Microsoft Query, входящее в комплект поставки Excel или, если требуется подключиться к данным Office, используются опции вкладки «Данные».

Если в документе уже присутствует отчет сводной таблицы или сводная диаграмма – в качестве источника можно использовать их. Для этого достаточно указать их расположение и выбрать нужный диапазон данных, после чего будет создана новая сводная таблица.

Вычисление значений в сводной таблице.

В сводных таблицах можно использовать функции сведения в полях значений для объединения значений из исходных данных. Если функции сведения и настраиваемые вычисления не дают желаемых результатов, вы можете создать собственные формулы в вычисляемых полях и вычисляемых объектах. Например, можно добавить вычисляемый объект с формулой расчета комиссионных за продажу, которые могут быть разными в различных регионах. Эти комиссионные будут автоматически включены в промежуточные и общие итоги в сводной таблице.

Если есть несколько вычисляемых объектов или формул, можно настроить порядок вычислений.

1 Добавление вычисляемого поля:

- щелкните сводную таблицу;
- отобразится вкладка «Работа со сводными таблицами» с дополнительными вкладками **Анализ** и **Конструктор**;
- на вкладке **Анализ** в группе **Вычисления** выберите команду **Поля, элементы и наборы**, а затем – пункт **Вычисляемое поле**;
- в поле **Имя** введите имя для поля;
- в поле **Формула** введите формулу для поля. Чтобы использовать в формуле данные из другого поля, щелкните его в списке **Поля** и нажмите кнопку **Добавить поле**;
- нажмите кнопку **Добавить**.

2 Добавление вычисляемого объекта в поле:

- щелкните сводную таблицу;

- отобразится вкладка «Работа со сводными таблицами» с дополнительными вкладками **Анализ** и **Конструктор**;
- если элементы в поле сгруппированы, на вкладке **Анализ** в группе **Группировать** выберите команду **Разгруппировать**;
- щелкните поле, в которое нужно добавить вычисляемый объект;
- на вкладке **Анализ** в группе **Вычисления** выберите команду **Поля, элементы и наборы**, а затем – пункт **Вычисляемый объект**;
- в поле **Имя** введите имя вычисляемого объекта;
- в поле **Формула** введите формулу для этого объекта. Чтобы использовать в формуле данные из имеющегося элемента, щелкните его в списке **Элементы** и нажмите кнопку **Добавить элемент** (он должен находиться в том же поле, что и вычисляемый объект);
- нажмите кнопку **Добавить**.

3 Ввод разных формул для вычисляемых объектов в отдельных ячейках:

- щелкните ячейку, для которой нужно изменить формулу;
- чтобы изменить формулу для нескольких ячеек, нажмите клавишу **CTRL** и, удерживая ее, выделите остальные ячейки;
- в строке формул внесите изменения в формулу.

4 Изменение порядка вычислений для нескольких вычисляемых объектов или формул:

- щелкните сводную таблицу. Отобразится вкладка «Работа со сводными таблицами» с дополнительными вкладками **Анализ** и **Конструктор**;
- на вкладке **Анализ** в группе **Вычисления** выберите команду **Поля, элементы и наборы**, а затем – пункт **Порядок вычислений**;
- щелкните одну из формул и нажмите кнопку **Вверх** или **Вниз**;
- повторите эти действия для других формул, чтобы расположить их в нужном порядке.

5 Просмотр всех формул, используемых в сводной таблице:

- щелкните сводную таблицу. Отобразится вкладка «Работа со сводными таблицами» с дополнительными вкладками **Анализ** и **Конструктор**;
- на вкладке **Анализ** в группе **Вычисления** выберите команду **Поля, элементы и наборы**, а затем – пункт **Вывести формулы**.

6 Редактирование формулы в сводной таблице:

- перед тем как редактировать формулу, определите, к чему она относится: к вычисляемому полю или вычисляемому объекту. Если формула находится в вычисляемом объекте, также определите, является ли она единственной для него;
- в случае вычисляемых объектов можно изменять отдельные формулы для конкретных ячеек одного объекта.

7 Редактирование формулы вычисляемого поля:

- щелкните сводную таблицу;
- отобразится вкладка «Работа со сводными таблицами» с дополнительными вкладками **Анализ** и **Конструктор**;
- на вкладке **Анализ** в группе **Вычисления** выберите команду **Поля, элементы и наборы**, а затем – пункт **Вычисляемое поле**;

- в списке **Имя** выберите вычисляемое поле, для которого нужно изменить формулу;
- в поле **Формула** измените формулу;
- нажмите кнопку **Изменить**.

Форма представления отчета: предоставить преподавателю задания по теме, выполненные на компьютере.

Вопросы для защиты работы

- 1 Дайте понятие сводной таблицы.
- 2 Перечислите способы построения сводных таблиц.

3 Первичная обработка опытных данных при помощи модуля BasicStatistics / Tables в MS STATISTICA

Цель работы – изучить особенности первичной обработки данных при помощи модуля BasicStatistics / Tables в MS STATISTICA.

Задачи работы:

- изучить методы первичной обработки данных и вычисляемые при этом статистические характеристики;
- получить практический навык работы в MS STATISTICA.

Задание

По данным, представленным в таблице 3.1, провести первичную обработку переменных:

Y – индекс человеческого развития;

X1 – ВВП 2020 г., % к 2015 г.;

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

X3 – ожидаемая продолжительность жизни при рождении 2020 г., число лет;

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

X5 – расходы домашних хозяйств, % к ВВП;

X6 – валовое накопление, % к ВВП.

Таблица 3.1 – Исходные данные

Страна	Y	X1	X2	X3	X4	X5	X6
Австрия	0,904	115	3343	77,0	75,5	56,1	25,2
Австралия	0,922	123	3001	78,2	78,5	61,8	21,8
Республика Беларусь	0,763	74	3101	68,0	78,4	59,1	25,7
Бельгия	0,923	111	3543	77,2	77,7	63,3	17,8
Великобритания	0,918	113	3237	77,2	84,4	64,1	15,9

Окончание таблицы 3.1

Страна	Y	X1	X2	X3	X4	X5	X6
Германия	0,906	110	3330	77,2	75,9	57,0	22,4
Дания	0,905	119	3808	75,7	76,0	50,7	20,6
Индия	0,545	146	2415	62,6	67,5	57,1	25,2
Испания	0,894	113	3295	78,0	78,2	62,0	20,7
Италия	0,900	108	3504	78,2	78,1	61,8	17,5
Канада	0,932	113	3056	79,0	78,6	58,6	19,7
Казахстан	0,740	71	3007	67,6	84,0	71,7	18,5
Китай	0,701	210	2844	69,8	59,2	48,0	42,4
Латвия	0,744	94	2861	68,4	90,2	63,9	23,0
Нидерланды	0,921	118	3259	77,9	72,8	59,1	20,2
Норвегия	0,927	130	3350	78,1	67,7	47,5	25,2
Польша	0,802	127	3344	72,5	82,6	65,3	22,4
Россия	0,747	61	2704	66,6	74,4	53,2	22,7
США	0,927	117	3642	76,7	83,3	67,9	18,1
Украина	0,721	46	2753	68,8	83,7	61,7	20,1
Финляндия	0,913	107	2916	76,8	73,8	52,9	17,3
Франция	0,918	110	3551	78,1	79,2	59,9	16,8
Чехия	0,833	99,2	3177	73,9	71,5	51,5	29,9
Швейцария	0,914	101	3280	78,6	75,3	61,2	20,3
Швеция	0,923	105	3160	78,5	79,0	53,1	14,1

Методические указания к выполнению задания

Расчет описательных статистик производится при помощи модуля Basic Statistic/Tables. В этом модуле объединены наиболее часто использующиеся на начальном этапе обработки данных процедуры.

В стартовой панели модуля приводится перечень статистических процедур этого модуля (рисунок 3.1):

Descriptive statistics – описательные статистики;

Correlation matrices – корреляционные матрицы;

t-test for independent samples – t-тест для независимых выборок;

t-test for dependent samples – t-тест для зависимых выборок;

Breakdown & one-way ANOVA – классификация и однофакторный дисперсионный анализ и др.

1 Процедура Descriptive statistics (описательные статистики).

После выбора процедуры Descriptive statistics на экране появится одноименное диалоговое окно (рисунок 3.2).

В этом окне при помощи кнопки **Variables** следует выбрать переменные для анализа.

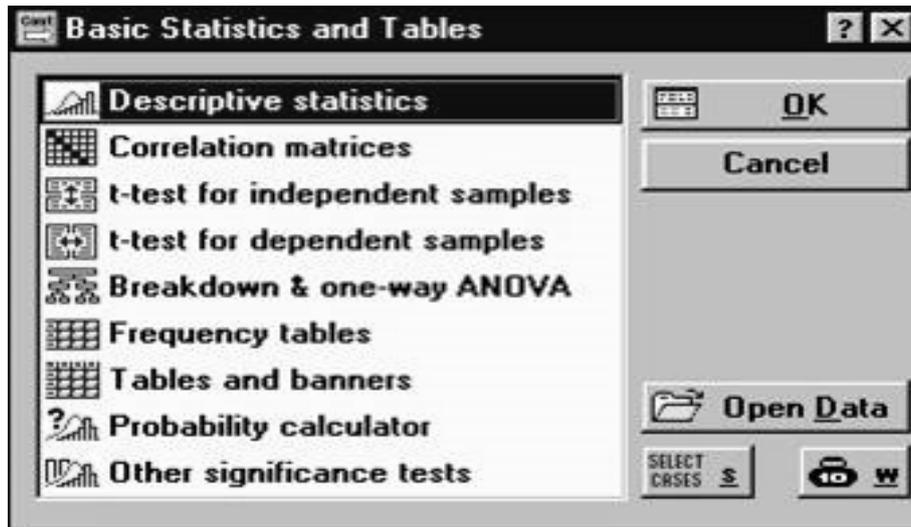


Рисунок 3.1 – Стартовое окно модуля с перечнем статистических процедур

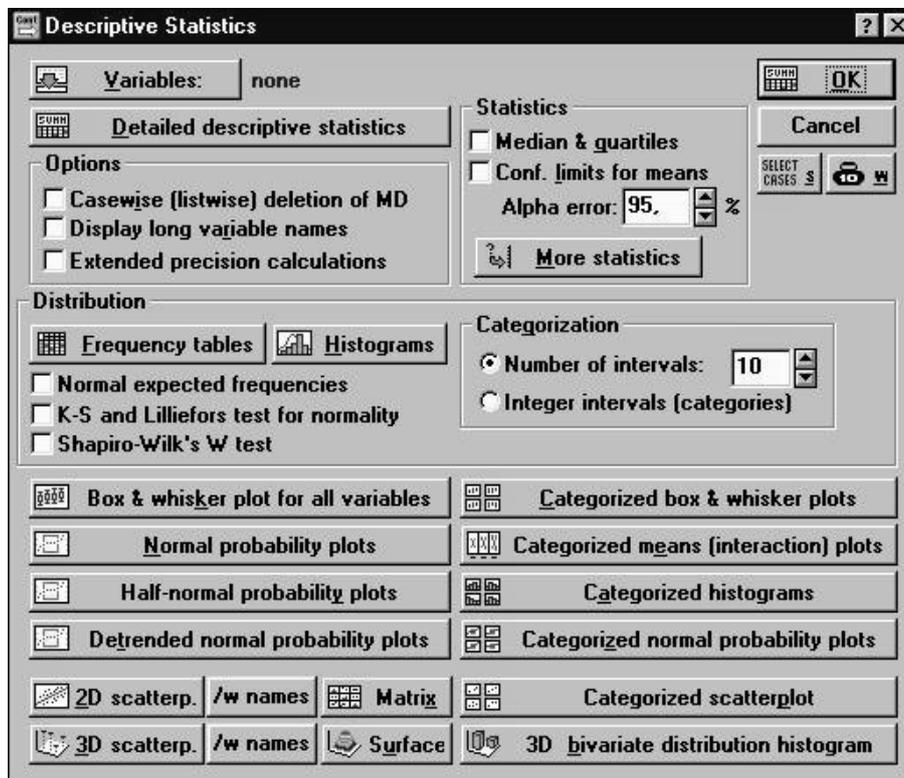


Рисунок 3.2 – Диалоговое окно Descriptive statistics

На первом этапе обработки данных часто возникает необходимость в их группировке. Группировка позволяет представить первичные данные в компактном виде, выявить закономерности варьирования изучаемого признака.

Для построения гистограмм и таблиц частот используется группа кнопок **Distribution** окна Descriptive statistics. Число классов (интервалов) группировки данных устанавливается при помощи счетчика переключателя **Number of intervals** окна Descriptive statistics. Справа от кнопок Distribution находятся две опции **Categorization** (Группировка), позволяющие задать число интервалов группировки или установить величину интервала равную целому

числу. Если заактивировать переключатель **Integer intervals (categories)**, то классы (интервалы) группировки будут представлять из себя целые числа.

Представить распределение переменных на гистограммах можно с помощью кнопки **Histograms** окна **Descriptive statistics**.

На гистограмму при необходимости можно наложить плотность нормального распределения, проверить близость распределения к нормальному виду при помощи критериев Колмогорова-Смирнова, Лилиефорса; вычислить статистику Шапиро-Уилкса. Для этого в группе опций **Distribution** необходимо установить флажок напротив соответствующих статистик. Значения статистик показываются прямо на гистограммах.

О нормальности распределения можно судить по графику нормальной вероятности. Его легко построить при помощи опции **Normal probability plots** окна **Descriptive statistics**. Чем ближе распределение к нормальному виду, тем лучше значения ложатся на прямую линию. Этот метод оценки является фактически глазомерным. В сомнительных случаях проверку на нормальность можно продолжить с использованием специальных статистических критериев (Колмогорова-Смирнова, омега-квадрат (w^2)). Однако детальная проверка гипотезы о нормальности выборки требует довольно значительных объемов выборки (по мнению некоторых авторов не менее 100 наблюдений).

Чтобы выбрать статистики, подлежащие вычислению, удобнее всего воспользоваться кнопкой **More statistics**:

Valid N – объем выборки;

Mean – средняя арифметическая.

Среднее значение случайной величины представляет собой наиболее типичное, наиболее вероятное ее значение, своеобразный центр, вокруг которого разбросаны все значения признака;

Sum – сумма;

Median – медиана.

Медианой является такое значение случайной величины, которое разделяет все случаи выборки на две равные по численности части;

Standard Deviation – стандартное отклонение.

Стандартное отклонение (или среднее квадратическое отклонение) является мерой изменчивости (вариации) признака. Оно показывает на какую величину в среднем отклоняются случаи от среднего значения признака. Особенно большое значение имеет при исследовании нормальных распределений. В нормальном распределении 68 % всех случаев лежит в интервале \pm одного отклонения от среднего, 95 % – \pm двух стандартных отклонений от среднего и 99,7 % всех случаев – в интервале \pm трех стандартных отклонений от среднего;

Variance – дисперсия.

Дисперсия является мерой изменчивости, вариации признака и представляет собой средний квадрат отклонений случаев от среднего значения признака. В отличие от других показателей вариации дисперсия может быть разложена на составные части, что позволяет тем самым оценить влияние различных факторов на вариацию признака. Дисперсия – один из важнейших показателей, характеризующих явление или процесс, один из основных критериев возмож-

ности создания достаточно точных моделей.

Standard error of mean – стандартная ошибка среднего.

Стандартная ошибка среднего – это величина, на которую отличается среднее значение выборки от среднего значения генеральной совокупности при условии, что распределение близко к нормальному. С вероятностью 0,68 можно утверждать, что среднее значение генеральной совокупности лежит в интервале \pm одной стандартной ошибки от среднего, с вероятностью 0,95 – \pm двух стандартных ошибок от среднего и с вероятностью 0,99 – \pm трех стандартных ошибок от среднего;

95 % confidence limits of mean – 95-процентный доверительный интервал для среднего.

Интервал, в который с вероятностью 0,95 попадает среднее значение признака генеральной совокупности;

Minimum, maximum – минимальное и максимальное значения;

Lower, upper quartiles – нижний и верхний квартили.

Верхний квартиль – это такое значение случайной величины, больше которого по величине 25 % случаев выборки. Нижний квартиль – это такое значение случайной величины, меньше которого по величине 25 % случаев выборки;

Range – размах.

Расстояние между наибольшим (maximum) и наименьшим (minimum) значениями признака;

Quartile range – интерквартильная широта.

Расстояние между нижним и верхним квартилями;

Skewness – асимметрия.

Асимметрия характеризует степень смещения вариационного ряда относительно среднего значения по величине и направлению. В симметричной кривой коэффициент асимметрии равен нулю. Если правая ветвь кривой, начиная от вершины больше левой (правосторонняя асимметрия), то коэффициент асимметрии больше нуля. Если левая ветвь кривой больше правой (левосторонняя асимметрия), то коэффициент асимметрии меньше нуля. Асимметрия менее 0,5 считается малой;

Standard error of Skewness – стандартная ошибка асимметрии;

Kurtosis – эксцесс.

Эксцесс характеризует степень концентрации случаев вокруг среднего значения и является своеобразной мерой крутости кривой. В кривой нормального распределения эксцесс равен нулю. Если эксцесс больше нуля, то кривая распределения характеризуется островершинностью, т. е. является более крутой по сравнению с нормальной, а случаи более густо группируются вокруг среднего. При отрицательном эксцессе кривая является более плосковершинной, т. е. более пологой по сравнению с нормальным распределением. Отрицательным пределом величины эксцесса является число -2, положительного предела – нет;

Standard error of Kurtosis – стандартная ошибка эксцесса.

Напротив статистик, подлежащих вычислению, следует поставить флажок.

После нажатия на кнопку ОК окна Descriptive statistics на экране появится

таблица с результатами расчетов описательных статистик.

2 Процедура *Correlation matrices* (Корреляционные матрицы).

Эта процедура предназначена для проведения корреляционного анализа, установления тесноты линейной связи между переменными.

В стартовом окне процедуры **Pearson Product-Moment Correlation** (Корреляция Пирсона) (рисунок 3.3) для расчета квадратной матрицы используется кнопка **One variable list (square matrix)**.

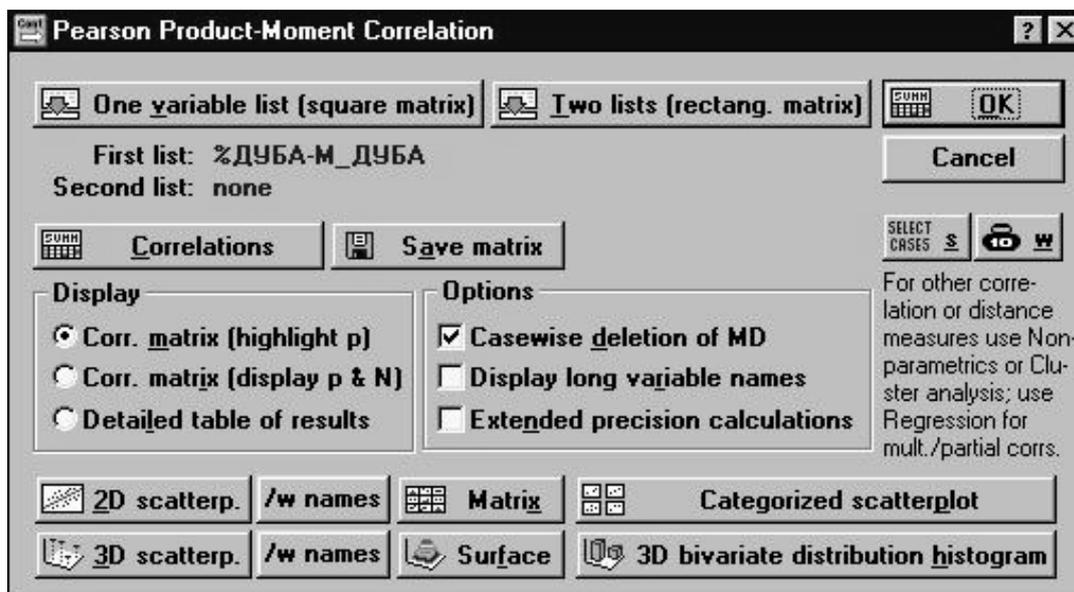


Рисунок 3.3 – Окно Pearson Product-Moment Correlation

В списке переменных выбирают переменные, между которыми будут рассчитаны парные коэффициенты корреляции Пирсона. После нажатия на кнопку ОК или Correlations на экране появится корреляционная матрица.

Коэффициент корреляции – это показатель, оценивающий тесноту линейной связи между признаками. Он может принимать значения от -1 до +1. Знак «-» означает, что связь обратная, «+» – прямая. Чем ближе коэффициент к 1, тем теснее линейная связь. При величине коэффициента корреляции (по Дворецкому) менее 0,3 связь оценивается как слабая, от 0,31 до 0,5 – умеренная, от 0,51 до 0,7 – значительная, от 0,71 до 0,9 – тесная, 0,91 и выше – очень тесная. Для практических целей Дворецкий рекомендует использовать значительные, тесные и очень тесные связи.

Процедура *Correlation matrices* сразу же дает возможность проверить достоверность рассчитанных коэффициентов корреляции. Значение коэффициента корреляции может быть высоким, но не достоверным, случайным. Чтобы увидеть вероятность нулевой гипотезы (p), гласящей о том что коэффициент корреляции равен 0, нужно в опции **Display** окна **Pearson Product-Moment Correlation** установить переключатель на вторую строку **Corr. matrix (display p & N)**. Но даже если этого не делать и оставить переключатель в первом положении **Corr. matrix (highlight p)**, статистически значимые на 5-процентном уровне коэффициенты корреляции будут выделены в корреляционной матрице

на экране монитора цветом, а при распечатке помечены звездочкой. Третье положение переключателя опции Display – **Detail table of results** позволяет посмотреть результаты корреляционного анализа в деталях. Флажок опции **Casewise deletion of MD** устанавливается для исключения из обработки всей строки файла данных, в которой есть хотя бы одно пропущенное значение.

Форма представления отчета: предоставить преподавателю задания по теме, выполненные на компьютере.

Вопросы для защиты работы

1 Дайте определения следующих статистических характеристик: дисперсия, мода, медиана, среднеквадратическое отклонение, коэффициент корреляции.

2 Перечислите инструменты первичной обработки данных в MS STATISTICA.

4 Решение задач на определение вероятности событий

Цель работы – научиться находить наиболее вероятные события и вероятность появления события в повторных независимых испытаниях.

Задачи работы:

- изучить методы нахождения вероятностей событий;
- получить практический навык нахождения вероятностей событий с помощью MS Excel.

Задача 1. Трёх стюардесс для рейса выбирают по жребию из 25 девушек, среди которых пять блондинок, 15 шатенок и пять брюнеток. Какова вероятность того, что среди выбранных девушек все будут иметь разный цвет волос?

Задача 2. В ящике лежат 15 игрушек, среди которых четыре с дефектами. Найти вероятность того, что среди семи наудачу вынутых игрушек одна окажется с дефектом.

Задача 3. Среди 17 желающих поехать на модный курорт 10 женщин. Определить вероятность того, что среди 12 случайным образом купивших путёвки оказались семь женщин?

Задача 4. Шкаф состоит из пяти крупных деталей. Вероятности брака при изготовлении каждой детали равны 0,1; 0,05; 0,03; 0,02; 0,04 соответственно.

Какова вероятность того, что изделие будет бракованным, если для этого достаточно наличие в сборке одной бракованной детали.

Задача 5. Заготовки деталей поступают из двух цехов предприятия: 60 % из первого и 40 % из второго. Заготовки первого цеха содержат 5 % брака, а второго – 3 %. Найти вероятность того, что наугад взятая заготовка будет без дефекта.

Задача 6. Сколько n -значных чисел можно составить из цифр 1, 2, 3, 4, ..., $n = \text{mod}(m + N, 5) + 5$, если каждая цифра входит в запись числа только один раз?

Задача 7. Сколько шифровок без повторений можно составить из $k = \text{mod}(m, 3) + 2$ неповторяющихся символов, используя алфавит из $n = \text{mod}(m + N, 7) + 5$ символов?

Задача 8. Сколькими способами можно выбрать $k = \text{mod}(m, 4) + 4$ мячей из корзины, содержащей $n = \text{mod}(m + N, 6) + 8$ мячей?

Задача 9. В партии из $m + N$ деталей $m - 3$ стандартные. Найти вероятность того, что $m - 5$ взятые наугад детали окажутся стандартными.

При решении задач использовать:

- m – порядковый номер студента в списке группы;
- N – номер группы (уточнить у преподавателя).

Методические указания по решению задач

1 Сколько n -значных чисел можно составить из цифр 1, 2, 3, 4, ..., $n = \text{mod}(m + N, 5) + 5$, если каждая цифра входит в запись числа только один раз?

Чтобы вычислить n , используется встроенная функция EXCEL: ОСТАТ(число; делитель).

Поскольку порядок цифр важен и в числе используются все цифры, искомое число – число перестановок. Для вычисления числа перестановок $P_n = n!$ применяется встроенная математическая функция EXCEL: ФАКТР(число).

2 Сколько шифровок без повторений можно составить из $k = \text{mod}(m, 3) + 2$ неповторяющихся символов, используя алфавит из $n = \text{mod}(m + N, 7) + 5$ символов?

Параметры k и n рассчитываются с помощью функции ОСТАТ(число; делитель). Так как порядок символов в шифровке важен, искомое количество – это число размещений k символов из n . Для расчета A_n^k используется встроенная статистическая функция EXCEL ПЕРЕСТ(число; число выбранных).

3 Сколькими способами можно выбрать $k = \text{mod}(m, 4) + 4$ мячей из корзины, содержащей $n = \text{mod}(m + N, 6) + 8$ мячей?

Параметры k и n помощью функции ОСТАТ(число; делитель). Поскольку порядок выбора мячей не важен, искомое число способов – это число сочетаний из n по k . Для вычисления числа C_n^k используется встроенная функция EXCEL: ЧИСЛКОМБ(число; число выбранных).

4 В партии из $m + N$ деталей $m - 3$ стандартные. Найти вероятность того, что $m - 5$ взятые наугад детали окажутся стандартными.

Для расчета используется формула классической вероятности

$$p = m / n, \quad (4.1)$$

где m и n – число благоприятных и всевозможных событий соответственно.

Поскольку в данной задаче порядок взятых деталей не важен, искомое число событий – число сочетаний. Расчет числа событий проводится с использованием функции ЧИСЛКОМБ (число; число выбранных).

Форма представления отчета: предоставить преподавателю задания по теме, выполненные на компьютере.

Вопросы для защиты работы

- 1 Какие события называются случайными?
- 2 Как определяется классическая вероятность?
- 3 Какие события несовместны?
- 4 Вероятность какого события равна нулю?

5 Решение задач на применение нормального, биномиального, пуассоновского распределения средствами EXCEL

Цель работы – научиться использовать функции Excel для решения задач на применение нормального, биномиального, пуассоновского распределения.

Задачи работы:

- изучить особенности и характеристики нормального, биномиального, пуассоновского распределения;
- получить практический навык использования статистических функций в MS Excel.

Задача 1. Вероятность поражения мишени стрелком при каждом выстреле одинакова и равна $p = 0,1 \cdot (\text{mod}(m + N, 4) + 4)$. Стрелок производит $n = \text{mod}(m + N, 10) + 20$ выстрелов.

Найти:

- 1) вероятность того, что стрелок поразит мишень ровно $k = \text{mod}(m + N, 5) + 10$ раз, используя:
 - а) формулу Бернулли;
 - б) локальную теорему Лапласа;
- 2) вероятность того, что стрелок поразит мишень не менее $k_1 = \text{mod}(m + N, 5) + 10$ раз и не более $k_2 = \text{mod}(m + N, 5) + 12$ раз, используя:

- а) формулу Бернулли;
- б) интегральную теорему Муавра-Лапласа.

Сравнить полученные результаты.

При решении задач использовать:

- m – порядковый номер студента в списке группы;
- N – номер группы (уточнить у преподавателя).

Задача 2. Завод отправил на базу $n = 100 \cdot (\text{mod}(m + N, 3) + 1)$ стандартных изделий. Вероятность того, что в пути изделие повредится, равна $p = 0,01 \cdot (\text{mod}(m, 2) + 1)$. Используя формулу Пуассона, найти вероятность того, что на базу прибудет ровно $k = \text{mod}(m, 5) + 1$ нестандартных изделия.

Задача 3. Используя функции Excel, вычислить следующие вероятности и границы сначала для стандартизованного, а потом для общего нормального распределения.

1 Для $N(0,1)$ (стандартизованное нормальное распределение) вычислить:

- $P(Z < -2)$;
- $P(Z > 1)$;
- $P(-0,4 < Z < 1,6)$;
- 5-процентную границу;
- 75-процентную границу;
- 99-процентную границу.

2 Для $N(75,8)$ вычислить:

- $P(X < 70)$;
- $P(X > 73)$;
- $P(75 < X < 85)$;
- 5-процентную границу;
- 60-процентную границу;
- 97-процентную границу.

Методические указания по решению задач

1 Вероятность поражения мишени стрелком при каждом выстреле одинакова и равна $p = 0,1 \cdot (\text{mod}(m + N, 4) + 4)$. Стрелок производит $n = \text{mod}(m + N, 10) + 20$ выстрелов. Найти вероятность того, что стрелок поразит мишень ровно $k = \text{mod}(m + N, 5) + 10$ раз, используя: формулу Бернулли, локальную теорему Лапласа.

Точное значение вероятности находится по формуле Бернулли:

$$P_n(k) = C_n^k p^k q^{n-k}, \quad (5.1)$$

где $q = 1 - p$, $k = 0, 1, \dots, n$.

Параметры p , q , n и k рассчитываются с помощью встроенной функции

EXCEL: ОСТАТ(число; делитель). Для расчета вероятности $P_n(k)$ используется встроенная функция EXCEL: БИНОМРАСП($k; n; p; 0$). Приближенное значение вероятности находится с помощью локальной теоремы Лапласа по формуле

$$P_n(k) = \frac{1}{\sqrt{npq}} \cdot \varphi(x), \quad (5.2)$$

где $x = \frac{k - np}{\sqrt{npq}}$.

При вычислении квадратного корня используется встроенная математическая функция EXCEL: =КОРЕНЬ(число).

Для расчета значения функции $\varphi(x)$ используется встроенная статистическая функция EXCEL: =НОРМРАСП($x; 0; 1; 0$).

2 Вероятность поражения мишени стрелком при каждом выстреле одинакова и равна $p = 0,1 \cdot (\text{mod}(m + N, 4) + 4)$. Стрелок производит $n = \text{mod}(m + N, 10) + 20$ выстрелов. Найти вероятность того, что стрелок поразит мишень не менее $k_1 = \text{mod}(m + N, 5) + 10$ раз и не более $k_2 = \text{mod}(m + N, 5) + 2$ раз, используя: формулу Бернулли, интегральную теорему Муавра-Лапласа. Сравнить полученные результаты.

Для нахождения точного значения вероятности трижды БИНОМРАСП($k; n; p; 0$) и результаты складываются.

Приближенное значение вероятности находится с помощью интегральной теоремы Лапласа по формуле

$$P_n(k_1 \leq k \leq k_2) = \Phi(x_2) - \Phi(x_1), \quad (5.3)$$

где $x_1 = \frac{k_1 - np}{\sqrt{npq}}$; $x_2 = \frac{k_2 - np}{\sqrt{npq}}$.

Для расчета значения функции $\Phi(x)$ используется встроенная статистическая функция EXCEL: =НОРМРАСП($x; 0; 1; 1$) - 0,5.

3 Завод отправил на базу $n = 100 \cdot (\text{mod}(m + N, 3) + 1)$ стандартных изделий. Вероятность того, что в пути изделие повредится, равна $p = 0,01 \cdot (\text{mod}(m, 2) + 1)$. Используя формулу Пуассона, найти вероятность того, что на базу прибудет ровно $k = \text{mod}(m, 5) + 1$ нестандартных изделий.

Приближенное значение вероятности найдем с помощью формулы Пуассона:

$$P_n(k) = \frac{\lambda^k \cdot e^{-\lambda}}{k!}, \quad (5.4)$$

где $\lambda = np$.

Для расчета значения $P_n(k)$ используется встроенная статистическая функция EXCEL: ПУАССОН($k; \lambda; 0$).

4 Нормальное распределение.

При изучении нормального распределения, как правило, решаются две задачи – поиск вероятностей по заданным границам и поиск границ вероятностных интервалов по заданным вероятностям. Первая задача решается с помощью функций НОРМРАСП (NORMDIST) (для произвольных нормальных распределений) и НОРМСТРАСП (NORMSDIST) (для стандартизованного нормального распределения). Вторая (обратная) задача решается с помощью функций НОРМОБР (NORMINV) и НОРМСТОБР (NORMSINV).

Функции НОРМРАСП(x ; μ ; σ ; 1) и НОРМСТРАСП(x) возвращают вероятности того, что значение меньше x .

Функции =НОРМОБР(p ; μ ; σ) и =НОРМСТОБР(p), где p – заданная вероятность, возвращают правые границы бесконечных (от $-\infty$) интервалов, вероятность попадания в которые равна p .

Форма представления отчета: предоставить преподавателю задания по теме, выполненные на компьютере.

Вопросы для защиты работы

- 1 Биноминальное распределение: понятие, вид, свойства.
- 2 Нормальное распределение: понятие, вид, свойства.
- 3 Распределение Пуассона: понятие, вид, свойства.

6 Решение задач на применение нормального, биномиального, пуассоновского распределения

Цель работы – научиться использовать функции Excel для решения задач на применение нормального, биномиального, пуассоновского распределения.

Задачи работы:

- изучить примеры применения нормального, биномиального, пуассоновского распределения для решения прикладных экономических задач;
- получить практический навык использования статистических функций в MS Excel.

Задача 1. Отдел по работе с персоналом крупной коммуникационной компании «N» пересматривает свою стратегию найма. Каждый кандидат проходит экзамен, результаты которого учитываются при решении вопроса о найме. Обработка результатов показала, что количество баллов имеет приблизительно нормальное распределение с математическим ожиданием 525 и стандартным отклонением 55. На первом шаге всех делят на три категории: автоматически зачисляемые (более 600 баллов), автоматически не зачисляемые (менее 425 баллов) и те, с кем проводится второй тур. Менеджер по персоналу

хочет оценить процент по категориям. Он также хочет вычислить новые границы, при которых автоматически отсекалось бы 10 % и принималось бы 15 %.

Задача 2. Клиент инвестирует 10 000 р. в определенные акции. Исследуя предысторию этих акций и посоветовавшись со своим брокером, он приходит к выводу, что доходность удовлетворяет нормальному закону с математическим ожиданием 10 % и стандартным отклонением 4 %. Доходы от акций будут обложены налогом по ставке 33 %. Вычислить вероятность того, что в качестве налогов будет уплачено, по крайней мере, 400 р. Вычислить 90-процентную верхнюю границу для чистого дохода.

Задача 3. Покупатели супермаркета «Перекресток» тратят на покупки различные суммы. Анализ данных показал, что суммы покупок распределены по нормальному закону с математическим ожиданием 85 р. и стандартным отклонением 30 р. Если предположить, что в некоторый день магазин посещает 500 покупателей, то вычислите математическое ожидание и стандартное отклонение для количества покупателей, потративших по крайней мере 100 р. Вычислите также вероятность, что по крайней мере 30 % покупателей потратили не менее 100 р.

Задача 4. Компания «ТВ-С» является региональной базой по продаже телевизоров различных марок. Одна из самых сложных менеджерских задач – это определить величину запаса по каждому типу. С одной стороны, хочется иметь много, чтобы удовлетворить запрос любого покупателя. С другой, излишние запасы – это замороженные деньги и занимаемые складские площади. Основная трудность возникает из-за неопределенности спроса, который меняется случайным образом из месяца в месяц. Известен только среднемесячный спрос – 17 телевизоров. Приглашенный консультант предполагает использовать вероятностную модель. Как он мог бы действовать?

Методические указания по решению задач

Каждое распределение вероятности определяется некоторыми параметрами, которые служат обобщающими величинами (например, среднее, дисперсия), характеризующими данное распределение. С помощью соответствующей статистики можно произвести оценку этих параметров в выборке. В зависимости от того, является ли случайная переменная дискретной или непрерывной, распределение вероятности может быть либо дискретным, либо непрерывным.

Основные дискретные распределения и их параметры представлены в таблице 6.1.

Таблица 6.1 – Параметры дискретных распределений

Распределение	Вероятность	Параметры
Равномерное дискретное	$\frac{1}{N}, k = 1, 2, \dots, N$	$N = 1, 2, \dots$
Бернулли	$p_1 = p, p_0 = q$	$0 \leq p \leq 1, q = 1 - p$
Биномиальное	$C_n^k p^k q^{n-k}, k = 0, 1, \dots, n$	$0 \leq p \leq 1, q = 1 - p, n = 1, 2, \dots$
Пуассона	$\frac{e^{-\lambda} \lambda^k}{k!}, k = 0, 1, \dots$	$\lambda > 0$
Геометрическое распределение	$q^{k-1} p, k = 1, 2, \dots$	$0 \leq p \leq 1, q = 1 - p$
Отрицательно-биномиальное	$C_{k-1}^{r-1} p^r q^{k-r}, k = r, r+1, \dots$	$0 \leq p \leq 1, q = 1 - p, r = 1, 2, \dots$

Нормально распределенные случайные величины широко распространены на практике, например, результаты измерения любой физической величины подчиняются нормальному закону распределения. Нормальным называется распределение вероятностей непрерывной случайной величины, которое описывается плотностью

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x - \bar{x})^2}{2\sigma^2}}, \quad (6.1)$$

где σ – дисперсия;

\bar{x} – среднее значение случайной величины x .

Форма представления отчета: предоставить преподавателю задания по теме, выполненные на компьютере.

Вопросы для защиты работы

- 1 Биномиальное распределение и распределение Пуассона.
- 2 Равномерное дискретное распределение.

7 Решение задач на построение выборок из генеральной совокупности средствами EXCEL

Цель работы – научиться использовать средства Excel для построения выборок из генеральной совокупности.

Задачи работы:

- изучить способы построения выборок из генеральной совокупности;
- получить практический навык использования статистических функций в MS Excel для построения выборок.

Задача. Исходные данные представлены в таблице 7.1. Построить случайную выборку размера 10, вычислить статистические характеристики: среднее, медиану, среднеквадратическое отклонение.

Таблица 7.1 – Исходные данные

Семья	Доход, у. е.	Семья	Доход, у. е.	Семья	Доход, у.е.	Семья	Доход, у. е.
1	43 300	11	44 100	21	56 400	31	33 000
2	44 300	12	51 500	22	33 600	32	31 700
3	34 600	13	35 900	23	38 100	33	48 600
4	38 000	14	35 600	24	42 500	34	39 300
5	44 700	15	43 000	25	44 900	35	33 000
6	45 600	16	38 600	26	35 200	36	36 300
7	42 700	17	32 400	27	60 800	37	28 400
8	36 900	18	22 900	28	42 500	38	46 900
9	38 400	19	48 100	29	47 600	39	37 300
10	33 700	20	31 900	30	36 100	40	41 000

Методические указания по решению задачи

Простейший подход. Главный принцип простейшего подхода – это равновероятность всех возможных выборок. Пусть N – количество элементов во всей совокупности, а n – размер выборки. Если эти значения малы и количество этих выборок невелико, то теоретически можно разбить интервал от 0 до 1 на N равных подынтервалов, каждый из которых соответствовал бы одному из элементов исходной совокупности, и воспользоваться случайной функцией СЛЧИС (RAND), которая генерирует значения случайной величины, равномерно распределенной на интервале (0;1).

Метод стратификации. Рассмотрим данный метод, который также иногда называют методом пропорциональных частичных выборок. Предположим, что все множество исходных данных, состоящее из N элементов, разбито на I непересекающихся подмножеств, состоящих из N_i элементов, так что:

$$N = N_1 + N_2 + \dots + N_i. \quad (7.1)$$

Для того чтобы получить выборку размера n , нам необходимо выбрать n_i представителей из каждой i -й подгруппы так, чтобы:

$$n = n_1 + n_2 + \dots + n_i. \quad (7.2)$$

Количества n_i вычисляют, округляя величины $n \cdot N_i / N$. После того как величины n_i определены, можно применить метод простых случайных выборок.

Используя логические функции Excel ЕСЛИ (условие; значение если истина; значение если ложь), производят разбивку на группы.

Используя функцию СЧЕТ (COUNT), подсчитывают количество значений в каждой группе.

Вычисляют количество выбираемых элементов в каждой подгруппе, используя функцию округления ОКРУГЛ (ROUND).

Форма представления отчета: предоставить преподавателю задания по теме, выполненные на компьютере.

Вопросы для защиты работы

- 1 Определение размера выборки.
- 2 Показатели выборочного наблюдения и генеральной совокупности.
- 3 Методы, виды и способы отбора единиц из генеральной совокупности.

8 Построение доверительного интервала для разности средних и разности долей средствами EXCEL и в MS STATISTICA

Цель работы – научиться использовать средства Excel и MS STATISTICA для построения доверительного интервала для разности средних и разности долей.

Задачи работы:

- изучить особенности построения доверительных интервалов для разности средних и разности долей;
- получить практический навык использования статистических функций в MS Excel при построении доверительного интервала для разности средних и разности долей.

Задача 1. В ресторане быстрого обслуживания планируется расширить ассортимент новым видом сэндвича. Для того чтобы оценить спрос на него, менеджер случайным образом выбрал 40 посетителей из тех, кто уже попробовал его и предложил им оценить их отношение к новому продукту в баллах

от 1 до 10. Менеджер хочет оценить ожидаемую долю клиентов, которые оценивают новый продукт не менее чем в 6 баллов (он ожидает, что именно эти клиенты и будут потребителями нового продукта).

Задача 2. Станок должен производить детали диаметром 10 см. Однако в силу различных обстоятельств происходят ошибки. Контролера по качеству волнуют два обстоятельства: во-первых, среднее значение должно равняться 10 см; во-вторых, даже в этом случае, если отклонения будут велики, то многие детали будут забракованы. Ежедневно он делает выборку из 50 деталей. Какие выводы может дать такая выборка?

Задача 3. Магазин, готовясь к большой распродаже, предпринял следующие маркетинговые исследования. Были выбраны 300 лучших покупателей, которые в свою очередь были случайным образом поделены на две группы по 150 членов в каждой. Всем из отобранных покупателей были разосланы приглашения для участия в распродаже, но только для членов первой группы был приложен купон, дающий право на скидку 5 %. В ходе распродажи покупки всех 300 отобранных покупателей фиксировались. Каким образом менеджер может интерпретировать полученные результаты и сделать заключение об эффективности предоставления купонов?

Задача 4. При изучении покупательского спроса произведено 5-процентное выборочное бесповторное обследование розничной продажи пылесосов при собственно случайном способе отбора. Полученное распределение реализованных пылесосов по уровню цен представлено в таблице 8.1.

Таблица 8.1 – Результаты выборки

Распределение по уровню цен, р.	До 1000	1000...1500	1500...2000	2000...2500	Св. 2500
Количество пылесосов, шт.	448	134	107	97	23

На основе приведенных данных вычислить: среднюю цену пылесоса (способом момента); среднее квадратическое отклонение; коэффициент вариации.

По результатам выборочного наблюдения определить:

– с вероятностью 0,954 возможные значения доли продажи пылесосов по цене свыше 1500 р. в общем объеме продажи пылесосов;

– с вероятностью 0,997 ($t = 3$) возможные границы средней цены пылесоса.

Методические указания по решению задач

Доверительные интервалы для генеральной средней можно установить на основе соотношений

$$\tilde{x} - \Delta_x \leq \bar{x} \leq \tilde{x} + \Delta_x, \quad (8.1)$$

где \tilde{x} и \bar{x} – генеральная и выборочная средние соответственно;

Δ_x – предельная ошибка выборочной средней.

Доверительные интервалы для генеральной доли устанавливаются на основе соотношений

$$p = w \pm \Delta_p; \quad w - \Delta_p \leq p \leq w + \Delta_p. \quad (8.2)$$

В качестве оценки стандартного отклонения используется s . Функцией плотности распределения оценки s является функция χ -квадрат, которая, как и t -распределение, имеет $n - 1$ степень свободы. Имеются специальные функции для работы с этим распределением ХИ2РАСП (CHIDIST) и ХИ2ОБР (CHIINV).

Разность средних значений будет иметь t -распределение с $n_1 + n_2 - 2$ степенями свободы. Доверительный интервал для $\mu_1 - \mu_2$ выражается соотношением:

$$\bar{X}_1 - \bar{X}_2 \pm t_\alpha \cdot SE(\bar{X}_1 - \bar{X}_2), \quad (8.3)$$

где

$$SE(\bar{X}_1 - \bar{X}_2) = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}; \quad (8.4)$$

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}. \quad (8.5)$$

Форма представления отчета: предоставить преподавателю задания по теме, выполненные на компьютере.

Вопросы для защиты работы

- 1 Показатели выборочного наблюдения и генеральной совокупности.
- 2 Методы, виды и способы отбора единиц из генеральной совокупности.

9 Проверка гипотез для различных параметров средствами EXCEL и в MS STATISTICA

Цель работы – научиться использовать средства Excel и MS STATISTICA для проверки гипотез для различных параметров.

Задачи работы:

- изучить способы проверки статистических гипотез;
- получить практический навык использования средств Excel и MS STATISTICA для проверки гипотез.

Задача 1. В таблице 9.1 представлены изменения выработки на одного основного рабочего. Проверить гипотезу, что выработка рабочих цеха распределена по нормальному закону.

Таблица 9.1 – Исходные данные

Выработка, %	94...100	100...106	106...112	112...118	118...124	124...130	130...136	136...142
Число рабочих	5	17	11	23	29	20	18	12

Задача 2. Менеджер ресторана недавно начал экспериментировать с новым способом приготовления пиццы. С его точки зрения, пицца, приготовленная новым способом, вкуснее, но для принятия окончательного решения о переходе на новый способ он хочет провести анализ мнения посетителей и планирует следующий эксперимент. Для 100 клиентов, заказавших пиццу на дом, он отправляет ее в двух вариантах приготовления и просит оценить в баллах свое мнение. Баллы от -10 до 10. Если клиент имеет сильное предпочтение к старому способу, то -10, к новому 10, если все равно, то ноль и т. д. (таблица 9.2).

Таблица 9.2 – Исходные данные

Клиент	Рейтинг	Клиент	Рейтинг	Клиент	Рейтинг	Клиент	Рейтинг
1	-7	11	3	21	2	31	-3
2	7	12	-4	22	0	32	3
3	-2	13	8	23	2	33	10
4	4	14	-5	24	9	34	-5
5	7	15	7	25	-5	35	1
6	6	16	-5	26	2	36	1
7	0	17	-1	27	4	37	7
8	2	18	7	28	2	38	3
9	8	19	3	29	10	39	5
10	2	20	4	30	-2	40	-6

После сбора результатов какова должна быть процедура обработки? Есть ли основание отвергать гипотезу (склонность к старому способу приготовления)?

Методические указания по решению задач

Проверка гипотез для математического ожидания.

Вычисляем t -статистику:

$$t = \frac{\bar{X} - \mu_0}{s / \sqrt{n}}. \quad (9.1)$$

$H_0: \mu \leq 0$ (нулевая гипотеза), $H_1: \mu > 0$ (альтернативная).

Использование функции СТЬЮДРАСП (x , степени свободы, хвосты) показывает вероятность результата для нулевой гипотезы.

Проверка гипотез для доли совокупности.

Выборочная доля имеет распределение, близкое к нормальному, поэтому определяем z -статистику по формуле

$$z = \frac{p^* - p}{\sqrt{\frac{pq}{n}}}, \quad (9.2)$$

где p^* – относительная частота оценка параметра p вычисляемая как: $p^* = x / n$;
 $q = 1 - p$.

Для проверки гипотезы $H_0: p = p_0$ можно использовать доверительные интервалы для параметра p . При этом гипотеза H_0 не отвергается (принимается) на уровне значимости α , если соответствующий односторонний или двусторонний доверительный интервал накрывает значение p_0 ; в противном случае гипотеза H_0 отклоняется.

Проверка гипотез для разности между долями совокупности.

Пусть p_1 и p_2 – значения долей, а \bar{p}_1 и \bar{p}_2 – их разностные оценки. Стандартное отклонение находим по формуле

$$SE(\bar{p}_1 - \bar{p}_2) = \sqrt{\bar{p}_c(1 - \bar{p}_c)(1/n_1 + 1/n_2)}, \quad (9.3)$$

где \bar{p}_c – выборочная пропорция, полученная по объединению выборок.

$$z = \frac{\bar{p}_1 - \bar{p}_2}{SE(\bar{p}_1 - \bar{p}_2)}. \quad (9.4)$$

Далее используется функция НОРМСТРАСП (z).

Проверка гипотезы о нормальном законе распределения проводится с помощью критерия согласия Пирсона. Критерий основан на сравнении эмпирических и теоретических частот. Эмпирические частоты получают в результате эксперимента, а теоретические частоты рассчитывают по формулам (таблица 9.3). Предварительно необходимо найти точечные оценки параметров предполагаемого распределения по выборке, а затем использовать их для вычисления теоретических частот. Критерий имеет χ^2 -распределение с $ck = m - r - 1$ степенями свободы, где m – число интервалов вариационного ряда; r – число параметров теоретического распределения, вычисленных по экспериментальным данным. Малочисленные эмпирические частоты ($ni_{эм} < 5$) следует объединить с соседним интервалом, в этом случае и соответствующие им теоретические частоты также складываются. Если производилось объединение частот, то при определении числа степеней свободы следует в качестве m принять число интервалов, оставшихся после объединения частот.

Таблица 9.3 – Определение теоретических частот основных распределений

Распределение	Параметры распределения	Теоретическая частота	Критерий
Нормальное	$\mu = \bar{x}_B$ $\sigma = \sigma_B$	$n_{i\text{теор}} = \frac{nh}{\sigma} \cdot \varphi(u_i),$ $u_i = \frac{x_i - \mu}{\sigma}, \varphi(u_i)$	$\chi^2 = \sum_{i=1}^m \frac{(n_{i\text{эксп}} - n_{i\text{теор}})^2}{n_{i\text{теор}}}$. Критическая область – правосторонняя: $\chi_{кр}^2(\alpha; k), k = m - r - 1$
Показательное	$\lambda = \frac{1}{\bar{x}_B}$	$n_{i\text{теор}} = n(e^{-\lambda x_i} - e^{-\lambda x_{i+1}})$	
Равномерное	$a = \bar{x}_B - \sigma_B \sqrt{3},$ $b = \bar{x}_B + \sigma_B \sqrt{3}$	$n_{i\text{теор}} = \frac{nh}{b - a}$	

Форма представления отчета: предоставить преподавателю задания по теме, выполненные на компьютере.

Вопросы для защиты работы

- 1 Критерий согласия Пирсона.
- 2 Методы проверки гипотез.

Список литературы

1 **Бродецкий, Г. Л.** Экономико-математические методы и модели в логистике: процедуры оптимизации: учебник / Г. Л. Бродецкий, Д. А. Гусев. – 2-е изд., стер. – Москва: Академия, 2014. – 288 с.

2 **Ефимова, М. Р.** Общая теория статистики: учебник / М. Р. Ефимова, Е. В. Петрова, В. Н. Румянцев. – Москва : ИНФРА-М, 2013. – 416 с.

3 **Захаренков, С. Н.** Статистика: учебное пособие / С. Н. Захаренков, В. А. Тарловская. – Минск: Современная школа, 2010. – 272 с.

4 **Иода, Е. В.** Статистика: учебное пособие / Е. В. Иода. – Москва: Вузовский учебник; ИНФРА-М, 2012. – 303 с.

5 **Колесникова, И. В.** Статистика. Практикум: учебное пособие для вузов / И. В. Колесникова, Г. В. Круглякова. – Минск : Вышэйшая школа, 2011. – 285 с.

6 **Косоруков, О. А.** Методы количественного анализа в бизнесе: учебник / О. А. Косоруков. – Москва: ИНФРА-М, 2005. – 367 с.

7 **Мхитарян, В. С.** Анализ данных в MS Excel: учебное пособие / В. С. Мхитарян, В. Ф. Шишов, А. Ю. Козлов. – Москва: КУРС, 2019. – 368 с.