**Практическое задание 1**

**Статистический анализ данных в Еxcel**

**Цель работы:** научиться обрабатывать статистические данные с помощью встроенных функций Excel; изучить возможности Пакета анализа и его некоторые инструменты: Генерация случайных чисел, Гистограмма, Описательная статистика.

**1. Основные статистические характеристики.**

Электронные таблицы *Excel* имеют огромный набор средств для анализа статистических данных. Наиболее часто используемые статистические функции встроены в основное ядро программы, то есть эти функции доступны с момента запуска программы. Другие более специализированные функции входят в дополнительную подпрограмму, называемую пакетом анализа. Команды и функции пакета анализа называют Инструментами анализа. Мы ограничимся изучением нескольких основных встроенных статистических функций и наиболее полезных инструментов анализа из пакета.

**Среднее значение.**

Функция СРЗНАЧ вычисляет выборочное (или генеральное) среднее, то есть среднее арифметическое значение признака выборочной (или генеральной) совокупности. Аргументом функции СРЗНАЧ является набор чисел, как правило, задаваемый в виде интервала ячеек, например, =СРЗНАЧ (А3:А201).

**Дисперсия и среднее квадратическое отклонение.**

Для оценки разброса данных используются такие статистические характеристики, как дисперсия *D* и среднее квадратическое (или стандартное) отклонение http://www.pandia.ru/text/77/196/images/image001_246.gif. Стандартное отклонение есть квадратный корень из дисперсии: http://www.pandia.ru/text/77/196/images/image002_176.gif. Большое стандартное отклонение указывает на то, что значения измерения сильно разбросаны относительно среднего, а малое – на то, что значения сосредоточены около среднего.

В Excel имеются функции, отдельно вычисляющие выборочную дисперсию *Dв* и стандартное отклонение http://www.pandia.ru/text/77/196/images/image001_246.gif*в* и генеральные дисперсию *D*г и стандартное отклонение http://www.pandia.ru/text/77/196/images/image001_246.gifг. Поэтому, прежде чем вычислять дисперсию и стандартное отклонение, следует четко определиться, являются ли ваши данные генеральной совокупностью или выборочной. В зависимости от этого нужно использовать для расчета *D*г и http://www.pandia.ru/text/77/196/images/image001_246.gifг , *D*в и http://www.pandia.ru/text/77/196/images/image001_246.gifв.

Для вычисления выборочной дисперсии *Dв* и выборочного стандартного отклонения http://www.pandia.ru/text/77/196/images/image001_246.gif*в* имеются функции ДИСП и СТАНДОТКЛОН. Для вычисления генеральной дисперсии *D*г и генерального стандартного отклонения http://www.pandia.ru/text/77/196/images/image001_246.gifг имеются функции ДИСПР и СТАНДОТКЛОНП, соответственно.

**Объем совокупности.**

Объем совокупности выборочной или генеральной – это число элементов совокупности. Функция СЧЕТ определяет количество ячеек в заданном диапазоне, которые содержат числовые данные. Пустые ячейки или ячейки, содержащие текст, функция СЧЕТ пропускает. Для определения количества непустых ячеек, независимо от их содержимого, используется функция СЧЕТ3. Ее аргументом является интервал ячеек.

**Мода и медиана.**

*Мода* – это значение признака, которое чаще других встречается в совокупности данных. Она вычисляется функцией МОДА.

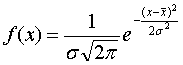
*Медиана* – это значение признака, которое разделяет совокупность на две равные по числу элементов части. Она вычисляется функцией МЕДИАНА.

**Размах варьирования. Наибольшее и наименьшее значения.**

Размах варьирования *R* – это разность между наибольшим *x*max и наименьшим xmin значениями признака совокупности (генеральной или выборочной): *R*= *x*max – xmin. Для нахождения наибольшего значения *x*max имеется функция МАКС, а для наименьшего xmin – функция МИН

**Отклонение случайного распределения от нормального.**

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 http://www.pandia.ru/text/77/196/images/image003_142.gif

,

где http://www.pandia.ru/text/77/196/images/image005_88.gif дисперсия, http://www.pandia.ru/text/77/196/images/image006_77.gif - среднее значение случайной величины http://www.pandia.ru/text/77/196/images/image007_68.gif.

Для оценки отклонения распределения данных эксперимента от нормального распределения используются такие характеристики как асимметрия *А* и эксцесс *Е*. Для нормального распределения *А*=0 и *Е*=0.

Асимметрия показывает, на сколько распределение данных несимметрично относительно нормального распределения: если *А*>0, то большая часть данных имеет значения, превышающие среднее http://www.pandia.ru/text/77/196/images/image006_77.gif; если *А*<0, то большая часть данных имеет значения, меньшие среднего http://www.pandia.ru/text/77/196/images/image006_77.gif. Асимметрия вычисляется функцией СКОС.

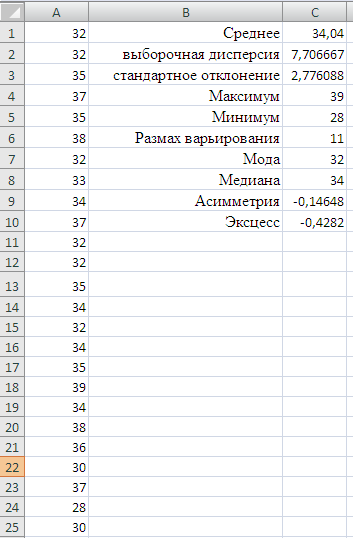
Эксцесс оценивает «крутость», т. е. величину большего или меньшего подъема максимума распределения экспериментальных данных по сравнению с максимумом нормального распределения. Если *Е*>0, то максимум экспериментального распределения выше нормального; если *Е*<0, то максимум экспериментального распределения ниже нормального. Эксцесс вычисляется функцией ЭКСЦЕСС.http://www.pandia.ru/text/77/196/images/image003_142.gif

**Задание 1**

Одним и тем же вольтметром было измерено 25 раз напряжение на участке цепи. В результате опытов получены следующие значения напряжения в вольтах:

32, 32, 35, 37, 35, 38, 32, 33, 34, 37, 32, 32, 35, 34, 32, 34, 35, 39, 34, 38, 36, 30, 37, 28, 30.

С помощью статистических функций найдите выборочные среднюю, дисперсию, стандартное отклонение, размах варьирования, моду, медиану. Проверить отклонение от нормального распределения, вычислив асимметрию и эксцесс.



**Задание 2**

Найдите выборочные среднюю, дисперсию, стандартное отклонение с помощью формул.

**3. Инструменты статистического анализа: *Генерация случайных чисел, Гистограмма, Описательная статистика*.**

**Загрузка *Пакета анализа*.**

*Пакет анализа* без дополнительных установок автоматически не загружается при запуске *Excel*. Он входит в так называемую *Надстройку* – набор дополнительных подпрограмм, к которым относятся, например, *Мастер диаграмм* и *Мастер функций*. Для загрузки *Пакет анализа* необходимо:

1)  в *Основном меню* выбрать пункт *Параметры Excel*;

2)  выбрать вкладку *Надстройки*;

3)  в появившемся списке *Надстроек* выбрать Пакет анализа и активизировать переключатель *Пакет анализа*-*VBA* и нажать ОК.

После этого на ленте Данные добавится пункт *Анализ данных*. К этому пункту следует обращаться для вызова *Пакета анализа*.

**Инструмент: *Генерация случайных чисел*.**

В*Excel* имеется встроенная функция СЛЧИСЛ (или RAND) для генерации равномерно распределенных случайных чисел в интервале [0,1].

Пакет анализа позволяет генерировать случайные числа с различными типами распределений: равномерное, нормальное, Бернулли, биномиальное, Пуассона и дискретное (определенное пользователем). Для генерации случайных чисел следует:

1)  выбрать команду*Анализ данных*;

2)  в появившемся диалоговом окне *Анализ данных* в группе *Инструменты анализа* выбрать пункт*Генерация случайных величин*и нажать *ОК*;

3)  в появившемся диалоговом окне *Генерация случайных чисел*следует заполнить поля ввода:

-  в полях *Число переменных*и *Число случайных чисел* указать нужное количество столбцов и сколько чисел вы хотите получить в каждом столбце;

-  в поле *Распределение*следует выбрать один из имеющихся типов распределения случайных чисел;

-  в группе *Параметры*следует указать диапазон чисел, т. е. min и max числа распределения для *Равномерного распределения*; или среднее значение и стандартное отклонение для *Нормального распределения* и т. д.

-  поле *Случайное*рассеивание заполняется только в том случае, если вам необходимо несколько раз воспроизводить одну и туже последовательность случайных чисел;

-  в поле *Выходной интервал*указывается место, куда следует поместить последовательность чисел, как правило, это интервал ячеек (или столбец целиком).

**Инструмент: *Гистограмма*.**

Графическое представление результатов обработки статистических данных обычно оформляется в виде гистограммы. Совокупность данных разбивается на частичные интервалы, называемые нормальными. Интервалы разбиения могут быть любой ширины, но обязательно они должны следовать в порядке возрастания. Интервалы разбиения откладываются по оси абсцисс гистограммы. На оси ординат гистограммы откладывается число значений, попавших в интервал разбиения. Это число значений признака совокупности называется частотой.

Для построения гистограммы:

1)  в начале следует задать частичные интервалы разбиения;

2)  на вкладке Данныевыбрать команду *Анализ данных*и указать инструмент анализа –*Гистограмма* и нажать *ОК*;

3)  в диалоговом окне *Гистограмма*следует указать:

-  в группе *Входные данные*в поле *Входной интервал* – интервал ячеек с данными, а в поле *Интервал карманов* – интервал ячеек с частичными интервалами разбиения;

-  в группе *Параметры вывода*указывается интервал ячеек для вывода частот и отмечается галочкой переключатель *Вывод графика*.

После нажатия *ОК* инструмент *Гистограмма* выводит два столбца: карман и частота. Сама гистограмма выводится правее столбца частот. Форматирование гистограммы производится так же, как и любой диаграммы в *Excel.*

**Инструмент: *Описательная статистика*.**

В пакете анализа *Excel*содержится инструмент *Описательная статистика*, который создает таблицу основных статистических характеристик для совокупности данных.

В этой таблице будут содержаться следующие характеристики: среднее, стандартная ошибка, дисперсия, стандартное отклонение, мода, медиана, размах варьирования интервала, максимальное и минимальное значения, асимметрия, эксцесс, объем совокупности, сумму всех элементов совокупности, доверительный интервал (уровень надежности).

Инструмент *Описательная статистика* существенно упрощает статистический анализ тем, что нет необходимости вызывать каждую функцию для расчета статистических характеристик отдельно.

Для того, чтобы вызвать *Описательную статистику*, следует:

1)  выбрать команду *Анализ данных*;

2)  в списке *Инструменты анализа* диалогового окна *Анализ данных*выбрать инструмент

*Описательная статистика* и нажать *ОК*;

3)  в появившемся диалоговом окне *Описательная статистика* необходимо:

-  в группе *Входные данные* в поле *Входной интервал* указать интервал ячеек, содержащих данные;

-  если первая строка во входном диапазоне содержит заголовок столбца, то в *поле Метки в первой строке* следует поставить галочку;

-  активизировать переключатель *Итоговая статистика*, если нужен полный список характеристик;

-  активизировать переключатель *Уровень надежности* и указать надежность в %, если необходимо вычислить доверительный интервал.

**Задание 3.**

Вычислите характеристики случайной величины из задания 1 с помощью инструмента *Описательная статистика*

**Задание 4.**

Сгенерировать 500 случайных чисел, распределенных нормально. Построить гистограмму и полный список статистических характеристик с помощью *инструмента Описательная статистика*.

1. Выполните команду *Анализ данных/Генерация случайных чисел*;

2..В диалоговом окне *Генерация случайных* чисел введите в поле число переменных: 1; в поле Число случайных чисел 500; выберите *Распределение Нормальное*; задайте любое среднее значение (желательно около 100) и небольшое стандартное отклонение (не больше 10); в поле Выходной интервал укажите абсолютный адрес столбца $A$2. Нажмите ОК.

3.  Теперь постройте гистограмму по совокупности случайных чисел. Сначала нужно задать интервалы решения. Пусть длины интервалов будут одинаковыми и равны 3. Для автоматического составления интервалов разбиения наберите в ячейку В2 начальное число, например, 75 для наших случайных чисел. Затем выполните команду *Правка/Заполнить/Прогрессия*. В появившемся диалоговом окне заполните данные:

-  в группе переключателей поле *Расположение* установите *по столбцам*;

-  в поле *Шаг* наберите 3;

-  в поле *Предельное значение*наберите 125;

-  в группе переключателей *Тип* установите *арифметическая* и нажмите ОК.

В результате столбец В будет содержать интервалы разбиения (карманы).

4.  Выполните команду *Анализ данных/Гистограмма*. В появившемся диалоговом окне *Гистограмма* заполните:

-  входной интервал появится, если щелкнуть мышью по столбцу А;

-  интервал карманов появится, если щелкнуть мышью по столбцу В;

-  поставьте галочку в поле метки;

-  укажите столбец С в поле *Выходной интервал*;

-  активизируйте переключатель *Вывод графика*; если это поле не содержит галочки, нажмите ОК.

5.  В результате вычисления получатся столбец под названием *Карман*, который дублирует ваш столбец интервалов разбиения, и столбец под название *Частота* с рассчитанными частотами. После того, как появилась гистограмма, измените ее размеры с помощью мыши так, чтобы хорошо были видны все столбцы и подписи.

6.  Теперь осталось получить таблицу статистических характеристик с *помощью Описательной статистики*. Выполните команду  *Анализ данных/Описательная статистика*. В появившемся диалоговом окне *Описательная статистика* укажите:

-  в поле *Входной интервал* появится адрес, если выделить мышью интервал сданными или с клавиатуры набрать адрес $A$2: $A$501;

-  в поле *Группирование* активизировать переключатель *по столбцам*;

-  активизировать переключатель *Метки в первой строке*;

-  в группе *Параметры вывода* укажите *Выходной интервал*, щелкнув мышью по какой-либо пустой ячейке ниже столбца частот, например, по С 25;

- активизируйте переключатель *Итоговая статистика* (если в этом поле нет галочки);

-  активизируйте переключатель *Уровня надежности* и установите 95%;

-  снимите галочки с полей*наименьший* и *наибольший*и нажмите ОК.