УДК 519.2:004

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА В ЕХСЕЬ

к. ф.-м. н., доц. Цыбрий Л. В.

ГВУЗ «Приднепровская государственная академия строительства и архитектуры», Украина, г. Днепропетровск

Современные информационные технологии И просто возможности Excel позволяют отказаться от введения условных вариант и так называемых упрощенных методов расчета сводных характеристик выборки. Возможностей Excel вполне достаточно для создания простого и группированного статистического ряда и решения основных задач математической статистики: оценки параметров распределения как точечных, так и интервальных, и проверки их значимости; выравнивания статистических распределений; проверки выравнивающей кривой гистограмме; соответствия проверки статистических гипотез с помощью надстроек Excel; корреляционного анализа и построения функции регрессии.

Для простого статистического ряда возможности Мастера диаграмм позволяют построить полигон р* и эмпирическую функцию распределения F* как точечные диаграммы, для чего сформировать на рабочем листе Excel таблицу



Рис.1. График F*

Для группированного статистического ряда график F* строится как точечная диаграмма, состоящая из прямых, соединяющих значения

F* в правых концах разрядов, вычисленных предварительно на рабочее листе Excel. Для построения гистограммы следует использовать инструмент Гистограмма надстройки Excel Пакет анализа:



Рис. 2. Гистограмма

При обработке статистических данных в среде Excel появляется возможность находить оценки параметров распределения ПО первичной статистической совокупности без построения статистического ряда. Для этого достаточно, разместив данные на рабочем листе Excel, задать выполнение специальных функций от статистических данных как аргументов. Так для вычисления выборочного среднего **X**, выборочной дисперсии $\mathbf{D}_{\mathbf{x}}^*$ и выборочного среднего квадратического отклонения S следует задать функции:

=CP3HAЧ(B36:B65); = ДИСП(B36:B65); =CTAHДOTКЛOH(B36:B65),

аргументы которых указывают диапазон исходных данных.

Доверительный интервал для математического ожидания генеральной совокупности определяется по такой формуле:

$\mathbf{X} - \mathbf{t}_{\gamma} \cdot \mathbf{S} / \sqrt{\mathbf{n}} \le \mathbf{X} + \mathbf{t}_{\gamma} \cdot \mathbf{S} / \sqrt{\mathbf{n}}$

 $t_{\gamma} = t_{\gamma}$ (α ; n-1) – критическое значение t-распределения, которое находится по таблице, но может быть найдено средствами Excel с помощью статистической функции

СТЬЮДРАСПОБР(а ; n-1),

где $\alpha = 1 - \gamma, \gamma - доверительная вероятность,$ **n**– объем выборки.

Основные выборочные числовые характеристики можно получить с помощью инструмента Описательная статистика надстройки Excel Пакет анализа.

Получив статистические оценки параметров распределения и подставив их в аналитическую формулу соответствующего закона распределения, мы сможем заменить статистическое распределение решена его аналитическим описанием, т.е. будет задача выравнивания статистического распределения. Предположение о виде закона распределения или о значениях его параметров является статистической гипотезой, правдоподобие которой еще следует Проверка проверить. гипотезы 0 предполагаемом законе распределения производится при помощи критерия согласия. При использовании критерия Пирсона χ^2 критическое значение χ^2 может быть вычислено с помощью функции XИ2ОБР(a;n-1), где a – уровень значимости. Все вычисления выполняются в среде Excel. . Гистограмма относительных частот (h - плотность относительной частоты) и выравнивающая ее кривая распределения f(x) должны хорошо согласовываться, что можно проверить, построив гистограмму и выравнивающую кривую:



Рис. З.Выравнивание статистического распределения

Если две или более переменных рассматриваются как случайные величины, образующие систему, возникает задача оценки тесноты связи между ними и разработка статистической модели этой связи. Для решения этой задачи используется корреляционный анализ, в основе которого лежат числовые характеристики системы и ее компонент. Вычисления числовых характеристик и выборочного коэффициента корреляции по корреляционной таблице выполняются средствами Excel. Проверка значимости коэффициента корреляции генеральной совокупности выполняется с помощью критерия Стьюдента, критическое значение которого может быть найдено как значение статистической функции **СТЬЮДРАСПОБР(***a*; n-1),

Оценку функции регрессии Y на X при известных выборочных параметрах принято представлять как линейную функцию g(x). В случае системы двух дискретных случайных величин оценку функции регрессии Y на X можно представить как оценку условного математического ожидания $M^*(Y | x)$, т.е. как выборочное условное среднее \overline{y}_x .



Рис. 4. Функция регрессии

Как видно по рисунку, прямая среднеквадратической регрессии Y на X g(x) достаточно близка к условным средним. Отличие вызвано тем, что функция регрессии g(x) заранее предполагается линейной. Это упрощает задачу, но не всегда соответствует статистическим данным.

В случае первичного статистического распределения, когда систему образуют пары значений, соответствующие одному одновременному наблюдению, корреляционно-регрессионный анализ может быть выполнен полностью средствами Excel и надстройками. В двумерного случая линейной регрессии среде Excel для предусмотрено несколько инструментов: статистические функции (КОРРЕЛ, ЛИНЕЙН, ТЕНДЕНЦИЯ и др.); инструмент Регрессия

надстройки **Пакет анализа**; графические средства при работе с диаграммой – построение линии тренда.

Используя инструмент Регрессия, можно:

найти выборочный коэффициент корреляции и оценить его значимость; найти выборочный коэффициент линейной регрессии Y на X и оценить его значимость; найти выборочное уравнение линейной регрессии Y на X и оценить его значимость.

Кроме того, в среде Excel можно построить диаграмму разброса статистических данных; построить выборочную прямую регрессии с помощью линии тренда и оценить ее значимость.

С помощью инструментов Пакета анализа можно проверять статистические гипотезы о математическом ожидании и других характеристиках генеральной совокупности по одной и нескольким выборкам, выполнять однофакторный и двухфакторный дисперсионные анализы. Для этого необходимо разместить на рабочем листе Excel статистические данные, вызвать команды Сервис – Анализ данных и в появившемся диалоговом окне Инструменты анализа отметить необходимый инструмент.

Изложенные методики использования современных информационно-компьютерных технологий должны использоваться при изучении теории вероятностей и математической статистики. Задачи и методы теории вероятностей и математической статистики, как никакие другие близки к проблемам научных и инженерных исследований. Отсюда следует необходимость обучения студентов не только основам статистических исследований, но и информационнокомпьютерным технологиям их выполнения.

ИСПОЛЬЗОВАННАЯ ЛИТЕРАТУРА

- 1. Васильев А.Н. Научные вычисления в Microsoft Excel. М.: Издательский дом «Вильямс», 2004. 512 с.
- 2. Ларсен Рональд У. Инженерные расчеты в Excel. : Пер. с англ. М.: Издательский дом «Вильямс», 2004. 544 с.
- Левин Дэвид М., Стефан Дэвид, Кребилт Тимоти С., Беренсон Марк Л. Статистика для менеджеров с использованием Microsoft Excel.: Пер с англ. – М.: Издательский дом «Вильямс», 2004. – 1312 с.
- Уокенбах Джон. Ехсеl 2003. Библия пользователя: Пер. с англ. М.: Издательский дом «Вильямс», 2005. – 768 с.
- 5. Кошелев В.Е. Excel 2007. М.: ООО «Бином-Пресс», 2008.–544 с.