СОДЕРЖАНИЕ

ВВЕДЕНИЕ

. ПЕРВИЧНЫЙ АНАЛИЗ ДАННЫХ

.1 Построение статистического ряда

.2 Графическое представление данных

.3 Эмпирическая функция распределения

. ОСНОВНЫЕ ХАРАКТЕРИСТИКИ СТАТИСТИЧЕСКИХ ДАННЫХ. ТОЧЕЧНЫЕ ОЦЕНКИ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ

.1 Основные характеристики выборочных данных. Свойства полученных оценок

.2 Закон распределения

.3 Метод максимального правдоподобия

.4 Плотность вероятности и функция распределения

.5 Доверительные интервалы для неизвестного математического ожидания и для среднего квадратического отклонения

. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

.1 Критерий согласия 

.2 Сравнительный анализ теоритического и эмпирического распределения

ЗАКЛЮЧЕНИЕ

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

ВВЕДЕНИЕ

статистический распределение математический отклонение

Математическая статистика - раздел математики, в котором изучаются методы сбора, систематизации и обработки результатов наблюдений массовых случайных явлений для выявления существующих закономерностей.

Целью данной курсовой работы является детальный анализ исследование доли занятого населения в общей численности экономически активного населения по Северо-Кавказскому федеральному округу РФ. Для достижения выше поставленной цели предполагается решить следующие задачи:

· Построить интервальный или дискретный статистический ряд.

· Построить полигон или гистограмму, в зависимости от того, дискретна или непрерывна изучаемая случайная величина.

· Найти эмпирическую функцию распределения и построить ее график.

· Вычислить основные характеристики выборочных данных. Указать какими свойствами обладают полученные оценки.

· На основе полигона или гистограммы, полученных выборочных характеристик сделать предварительный выбор закона распределения.

· Найти точечные оценки параметров закона распределения случайной величины методом максимального правдоподобия.

· Используя точечные оценки параметров, записать плотность вероятности и функцию распределения.

· В случае нормальности распределения построить доверительные интервалы с надежностью 0,95:

а) для математического ожидания, считая  известным, равным ;

б) для математического ожидания, считая дисперсию неизвестной;

в) для среднего квадратического отклонения.

· Проверить с помощью критерия согласия , согласуется ли гипотеза о виде распределения с опытными данными, уровень значимости .

· Для непрерывной случайной величины построить график функции плотности вероятности и сравнить его с гистограммой, для дискретной случайной величины построить многоугольник распределения и сравнить его с полигоном.

Объектом данной курсовой работы является статистика доли занятого населения и численности экономически активного населения.

Предметом выступают статистические показатели численности занятого населения и численности экономически активного населения за 2000-2011 гг. по Северо-Кавказскому федеральному округу РФ.

Структура курсовой работы. Курсовая работа состоит из введения, трех глав, заключения и списка использованной литературы.

1. ПЕРВИЧНЫЙ АНАЛИЗ ДАННЫХ

.1 Построение статистического ряда

Исходные данные к курсовой работе взяты на сайте: http://www.gks.ru/Официальная статистика\Рынок труда, занятость и заработная плата\Трудовые ресурсы.

Для построения статистического ряда необходимо найти долю населения. Воспользуемся формулой:



где  - численность занятого населения;  - численность экономически активного населения.

Таблица 1 - Исходные данные.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 |
| Численность занятого населения | 2604 | 2653,4 | 2762,8 | 2776,8 | 2745,6 | 2960,5 | 3124,2 | 3387,9 | 3588 | 3664,8 | 3616,1 | 3768,5 |
| Численность экономически активного населения | 3272,3 | 3264,1 | 3344,3 | 3351,6 | 3396,8 | 3587,7 | 4051,0 | 4207,9 | 4279,3 | 4387,3 | 4351,1 | 4431,2 |
| Доля населения | 0,7958 | 0,8129 | 0,8261 | 0,8284 | 0,8083 | 0,8252 | 0,7712 | 0,8051 | 0,8385 | 0,8353 | 0,8311 | 0,8505 |

Мы получили выборку значений непрерывной случайной величины, где отдельные значения случайной величины как угодно мало отличаться друг от друга и поэтому в совокупности наблюдаемых данных одинаковые значения величины встречаться редко, а частоты вариантов мало отличаются друг от друга, следовательно, построим интервальный статистический ряд.

Интервальный статистический ряд - это упорядоченная совокупность интервалов варьирования значений случайной величины с соответствующими частотами или частостями попаданий в каждый из них значений величины.

Для исследования полученных данных используем инструменты анализа MS Excel.

Полученные значения выборки вставляем в диапазон A1:А12. Выполняем ранжирование: скопируем эту выборку в диапазон С2:С13 и выполним сортировку по возрастанию.

Теперь построим интервальный статистический ряд. Введем формулы в ячейки рабочего листа согласно рисунку 1. Количество интервалов при вычислении карманов (границ интервалов) получилось на единицу больше, чем вычислено в ячейке F4. Это результат того, что при вычислении границ интервалов начальная точка получена по формуле: 

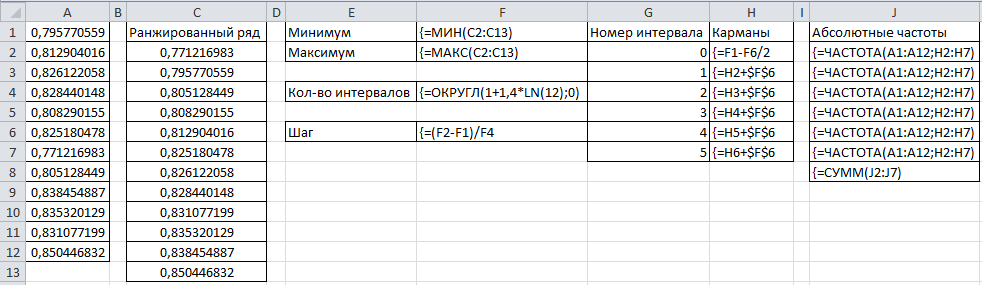


Рис. 1

Расчет данных представлен на рисунке 2.

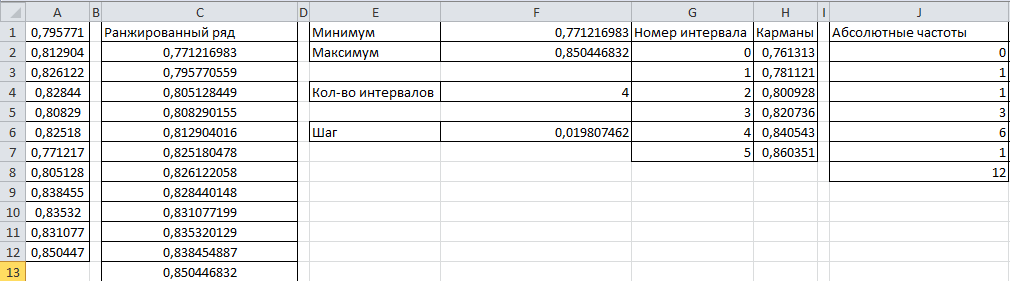


Рис. 2

Таким образом, построили следующий статистический ряд (таблица 2).

Таблица 2 - Интервальный статистический ряд

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Интервал | 0,76-0,78 | 0,78-0,80 | 0,80-0,82 | 0,82-0,84 | 0,84-0,86 |
| Частота | 1 | 1 | 3 | 6 | 1 |

.2 Графическое представление данных

Гистограмма - служит только для представления интервальных статистических рядов и представляет собой столбчатую диаграмму, состоящую из прямоугольников, основания которых - частичные интервалы длины h, а высоты - абсолютные  или относительные частоты .

Выполним команду Данные/Анализ данных/Гистограмма

В появившемся диалоговом окне заполним поля согласно рисунку 3.

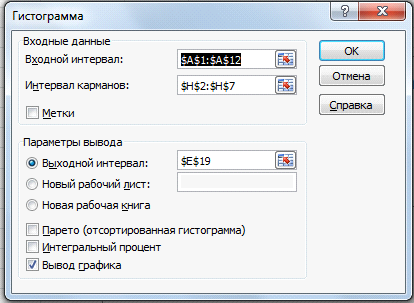


Рис. 3

На рабочем листе получили таблицу частот (рисунок 4) и гистограмму (рисунок 5).

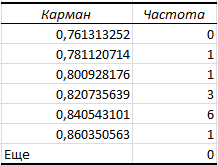


Рис. 4



Рис. 5

.3 Эмпирическая функция распределения

Эмпирической функцией распределения (функцией распределения выборки) называют функцию, определяющую для каждого значения х относительную частоту события , т.е.



Для построения графика эмпирической функции найдем накопительную частость.

Введем формулы в ячейки рабочего листа согласно рисунку 6.

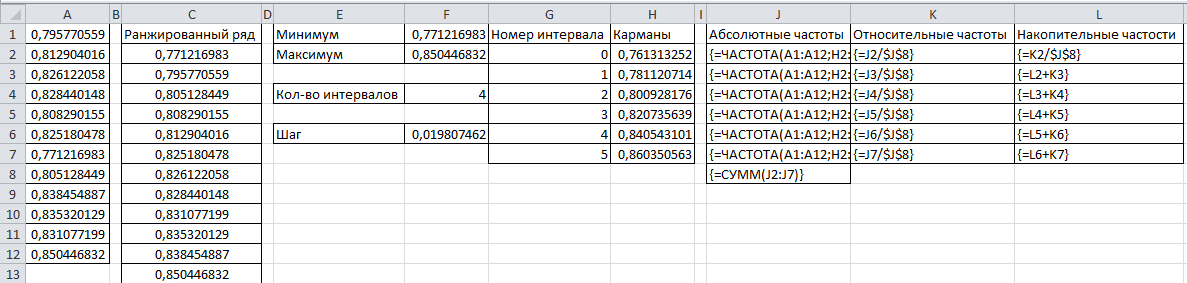


Рис. 6

Получим массив накопленных частостей (рисунок 7).

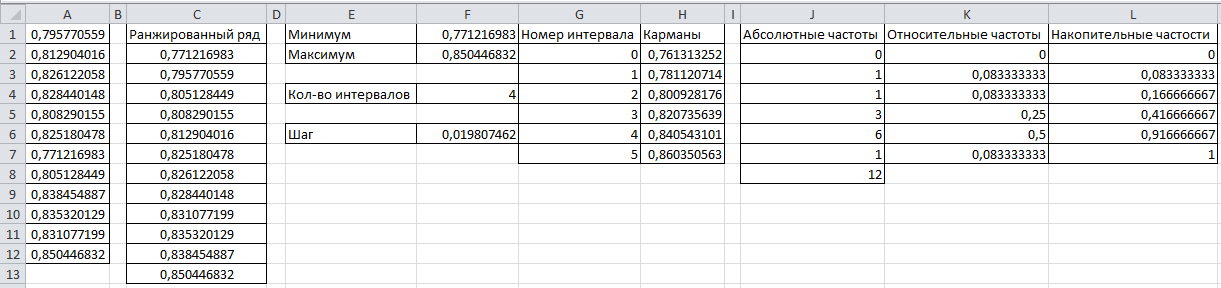


Рис. 7

Теперь построим эмпирическую функцию. Выделим диапазон L2:L7, затем выполним команду Вставка/График.

Эмпирическая функция представлена на рисунке 8.

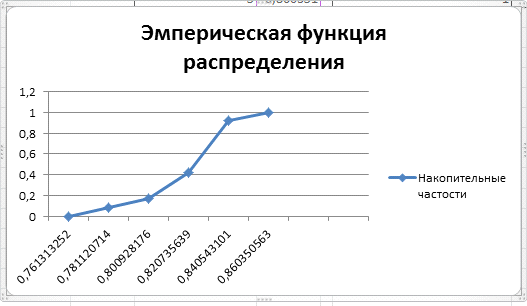


Рис. 8

График эмпирической функции для интервального вариационного ряда есть непрерывная линия.

2. ОСНОВНЫЕ ХАРАКТЕРИСТИКИ СТАТИСТИЧЕСКИХ ДАННЫХ. ТОЧЕЧНЫЕ ОЦЕНКИ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ

2.1 Основные характеристики выборочных данных. Свойства полученных оценок

После того, как мы построили интервальный статистический ряд, сделаем описательную статистику. Для этого выполним команду Данные/Анализ данных/Описательная статистика. Заполним поля как указанно на рисунке 9.

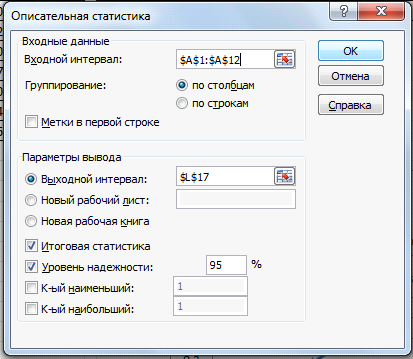


Рис. 9

Итог описательной статистики представлен на рисунке 10.

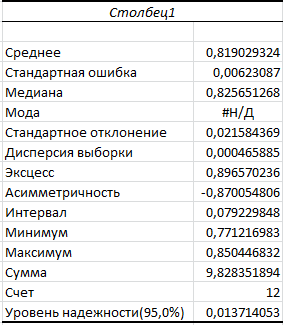


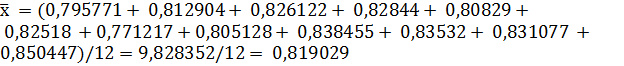
Рис. 10

Найдем числовые характеристики:

Средним арифметическим  наблюдаемых значений случайной величины Х (выборочным средним) называется частное от деления суммы всех этих значений на их число, т.е.



где  - значение признака у i-го объекта, n - число объектов в совокупности.



Основные свойства среднего арифметического:

. Если индивидуальные значения признака (варианты), уменьшить (увеличить) в n раз, то среднее значение нового признака соответственно уменьшится или увеличится во столько же.

. Если все варианты усредняемого признака уменьшить (увеличить) на число А, то средняя арифметическая соответственно изменится на это же число.

. Если вес всех усредняемых вариантов уменьшить (увеличить) в k раз, то средняя арифметическая не изменится.

. Сумма отклонений отдельных значений признака от средней арифметической равна нулю.

Медианой вариационного ряда называется значение признака, приходящееся на середину ряда.

Для вариационного ряда с нечетным числом членов медиана равна серединному варианту, а для ряда с четным числом членов полусумме двух серединных вариантов.

 = (0, 825180478+ 0, 826122058)/2= 1, 651302536/2= 0,825651268

Свойства медианы:

. Медиана не зависит от тех значений признака, которые расположены по обе стороны от нее.

. Аналитические операции с медианой весьма ограничены, поэтому при объединении двух распределений с известными медианами невозможно заранее предсказать величину медианы нового распределения.

. Медиана обладает свойством минимальности. Его суть заключается в том, что сумма абсолютных отклонений значений х, от медианы представляет собой минимальную величину по сравнению с отклонением X от любой другой величины

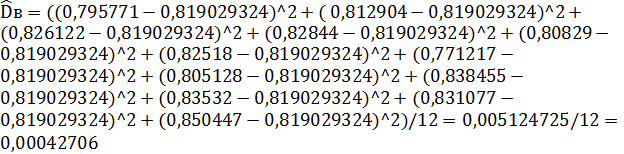
Модой вариационного ряда называется вариант, имеющий наибольшую частоту. Моду определить невозможно.

Простейшим показателем вариации является вариационный размах R, равный разности между наибольшим и наименьшим вариантами ряда:

 =0,850447- 0,771217= 0,079229848

Выборочной дисперсией значений случайной величины X называется среднее арифметическое квадратов отклонений наблюдаемых значений этой величины от их среднего арифметического (обозначение ):





Основные свойства выборочной дисперсии:

. Дисперсия постоянной величины равна нулю

. Если все результаты наблюдений увеличить (уменьшить) на одно и то же число С, то дисперсия и среднее квадратическое отклонение не изменятся

. Если все результаты наблюдений умножить на одно и то же число, то имеет место равенство 

. Если все частоты вариантов умножить на одно и то же число, то выборочные дисперсия и среднее квадратическое отклонение не изменятся.

. Выборочная дисперсия равна разности между средним арифметическим квадратов наблюдений над случайной величиной и квадратом ее среднего арифметического, т. е.



Выборочным средним квадратическим отклонением называется арифметический квадратный корень из выборочной дисперсии (обозначение).





Исправленная выборочная дисперсия.





Исправленным выборочным средним квадратическим отклонением.





.2 Закон распределения

На основе гистограммы сделаем предварительный выбор закона распределения.

Набор данных можно считать нормально распределенным, если форма гистограммы напоминает колокол, в котором большинство значений сконцентрировано в средней части, а остальные распределены равномерно с затуханием по обе стороны от центра. На графике отчетливо видно, что набор данных имеет симметричное распределение. Поэтому можно предположить, что имеющийся набор данных можно назвать нормально распределенной случайной величиной.

Нормальное распределение - это непрерывное распределение, имеющее графическое представление в виде симметричной колоколообразной кривой.

2.3 Метод максимального правдоподобия

Метод максимального правдоподобия основывается на представлении выборки объема n, взятой из генеральной совокупности, т.е. подчинённой закону распределения с некоторыми параметрами где  может быть одним параметром или вектором.

Для непрерывной случайной величины строится функция правдоподобия, множителями которой является плотности вероятности совместного появления результата, а для дискретной случайной величины эти множители являются вероятностями появления значений 



Естественно потребовать от оценки параметра  свойства максимальности функции правдоподобия, т.е. найти такие оценки, которые бы максимизировали функцию правдоподобия 



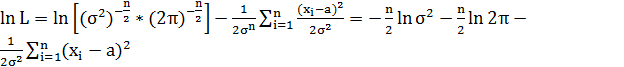
Поскольку максимизация функции L и ln L достигается в одной и той же точке, т.е.



По выборке нормально распределенной величины, оценим параметры (а, ):

;

;

 ;











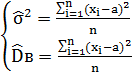
;





;

Т.к. , а , следовательно, .

 , следовательно, 

Т.к. , следовательно, оценка равна .

.4 Плотность вероятности и функция распределения

Используя точечные оценки параметров нормального закона распределения  и  запишем плотность вероятности и функцию распределения.









.5 Доверительные интервалы для неизвестного математического ожидания и среднего квадратического отклонения

А) для математического ожидания, считая  известным, равным 

Если известно среднее квадратическое отклонение , то доверительный интервал для математического ожидания имеет вид:



где а - оцениваемое математическое ожидание, х - выборочное среднее, п - объем выборки, t - такое значение аргумента функции Лапласа , при котором .

Т.к. предельная ошибка выборки вычисляется по формуле:



Следовательно, доверительный интервал для математического ожидания имеет вид:



Найдем доверительный интервал для математического ожидания нормально распределенной случайной величины, если объем выборки n=12,

, .





Б) для математического ожидания, считая дисперсию неизвестной

Найдем доверительный интервал для математического ожидания нормально распределенной случайной величины при неизвестной дисперсии, если объем выборки n=12, , s=, t=2,2.

Если дисперсия неизвестна, то доверительный интервал для математического ожидания имеет вид:







В) для среднего квадратического отклонения

Найдем доверительный интервал для среднего квадратического отклонения нормально распределенной случайной величины, если объем выборки n=12, , .

Доверительный интервал дисперсии имеет вид:









3. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

.1 Критерия согласия 

. Основная гипотеза:  - исследуемая случайная величина Х подчиняется нормальному закону распределения.

. Восстанавливаем теоретическое распределение. Выполним следующие шаги:

· Записываем границы интервалов;

· Находим интегральную функцию распределения на концах интервалов;

· Находим теоретическую вероятность;

· Рассчитываем критерий.

Итог указанных шагов представлены на рисунке 11

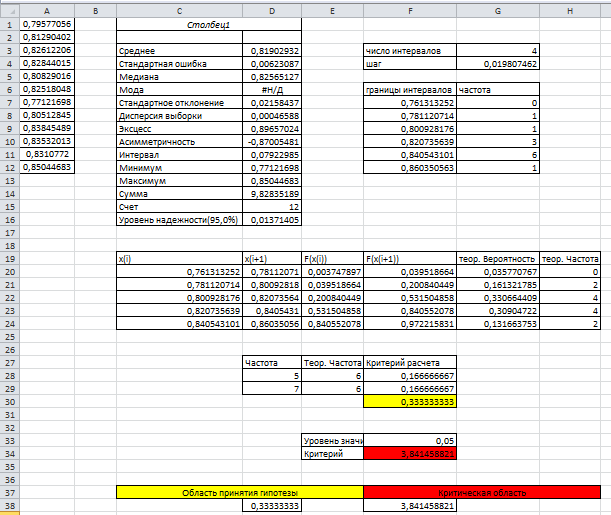


Рис. 11

Формулы всех параметров представлены на рисунке 12.

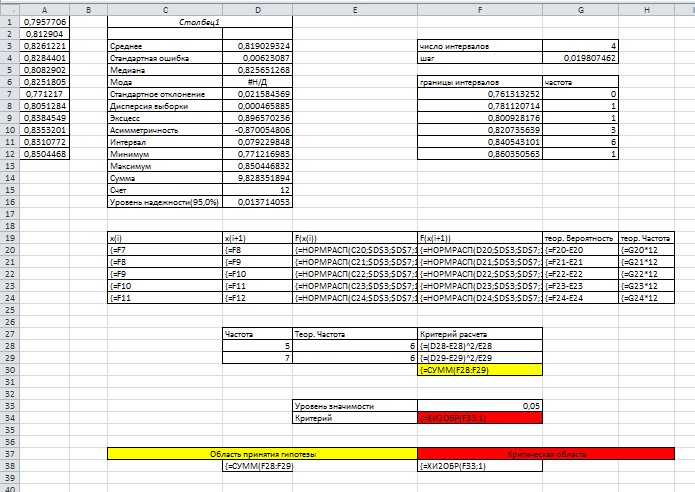


Рис. 12

Критической областью называется множество возможных значений статистического критерия, при которых основная гипотеза отвергается.

Областью принятия гипотезы или областью допустимых значений называется множество возможных значений статистического критерия, при которых основная гипотеза принимается.

Согласование выборочных исходных данных.

 >, это значит, что гипотеза  о нормальном законе распределения принимается.

3.2 Сравнительный анализ теоритического и эмпирического распределения

Для непрерывной случайной величины построим график функции плотности вероятности.

Плотность нормального распределения - колоколообразная кривая, симметричная относительно некоторой вертикальной оси, но она может быть смещена по горизонтали относительно оси Оу. Значения х могут быть разного знака. Выражение для плотности нормального распределения имеет вид:

На рисунке 13 изображен график функции плотности вероятности и полигон частот.

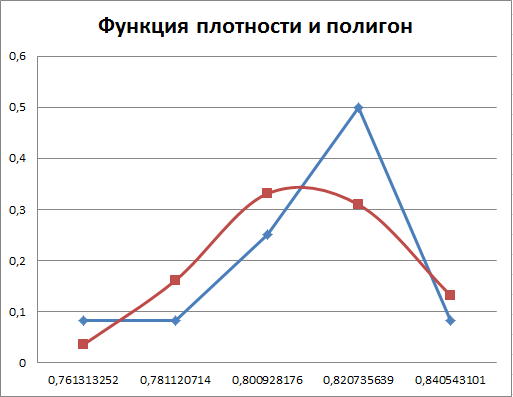


Рис. 13

На графике видно, что и эмпирическое и теоретическое распределение подчиняется нормальному закону распределения.

ЗАКЛЮЧЕНИЕ

Все задачи математической статистики касаются вопросов обработки наблюдений над массовыми случайными явлениями, но в зависимости от характера решаемого практического вопроса и от объема имеющегося экспериментального материала эти задачи могут принимать ту или иную форму.

Математическая (или теоретическая) статистика опирается на методы и понятия теории вероятностей, но решает в каком-то смысле обратные задачи.

Если мы наблюдаем одновременно проявление двух (или более) признаков, т.е. имеем набор значений нескольких случайных величин - что можно сказать об их зависимости? Есть она или нет? А если есть, то какова эта зависимость? Часто бывает возможно высказать некие предположения о распределении, спрятанном в "черном ящике", или о его свойствах. В этом случае по опытным данным требуется подтвердить или опровергнуть эти предположения ("гипотезы"). При этом надо помнить, что ответ "да" или "нет" может быть дан лишь с определенной степенью достоверности, и чем дольше мы можем продолжать эксперимент, тем точнее могут быть выводы. Наиболее благоприятной для исследования оказывается ситуация, когда можно уверенно утверждать о некоторых свойствах наблюдаемого эксперимента - например, о наличии функциональной зависимости между наблюдаемыми величинами, о нормальности распределения, о его симметричности, о наличии у распределения плотности или о его дискретном характере, и т.д.

· Итак, о (математической) статистике имеет смысл вспоминать, если имеется случайный эксперимент, свойства которого частично или полностью неизвестны,

· мы умеем воспроизводить этот эксперимент в одних и тех же условиях некоторое (а лучше - какое угодно) число раз.

СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

1. Граничина О.А. Математико-статистические методы психолого-педагогических исследований. - 2012.- С. 115

. Вентцель Е.С. "Теория вероятности и математическая статистика" - 2003

. Методические указание по математической статистики. - 2010. - С. 75

. http://michael983.narod.ru/t/8.htm

. http://www.gks.ru