**Календарный план занятий**

**Занятие 1**. Первоначальная обработка статистических данных. Задача №1.

Центральная предельная теорема (повторение).

**Лекция 1**. Основные распределения математической статистики.

**Занятие 2.** Свойства основных распределений математической статистики.

**Лекция 2**. Виды сходимости случайных величин и их использование в предельных теоремах теории вероятностей.

**Занятие 3**. Свойства выборочной функции распределения. Моделирование выборок из дискретных и абсолютно непрерывных законов распределения. Задачи №2-3.

**Лекция 3.** Основные задачи математической статистики. Выборочные характеристики и их свойства. Порядковые статистики. Метод моментов.

**Занятие 4**. Порядковые статистики. Метод моментов.

**Лекция 4**. Метод максимального правдоподобия. Свойства точечных оценок. Неравенство Рао-Крамера. Информация Фишера.

**Занятие 5**. Оценки максимального правдоподобия. Самостоятельная работа №1.

**Лекция 5.** Достаточные статистики. Критерий факторизации. Интервальные оценки. Принцип построения доверительных интервалов.

**Занятие 6**. Достаточные статистики. Подготовка к контрольной работе №1.

**Лекция 6** Доверительные интервалы для параметров нормального распределения. Доверительный интервал для параметра показательного закона.

**Занятие 6**. Контрольная работа №1.

**Лекция 7**. Связь между эффективными оценками и достаточными статистиками. Асимптотическая эффективность и асимптотическая нормальность. Приближенные доверительные интервалы. Теорема Колмогорова-Блэкуэлла.

**7 неделя. Коллоквиум по модулю 1.**

**КСР№4**. Коллоквиум (продолжение). Переписывание контрольной работы.

**Лекция 8**. Задача проверки гипотез. Критерий отношения правдоподобия. Ошибки 1-го и 2-го рода.

**Занятие 8**. Критерий отношения правдоподобия. Задача №5.

**Лекция 9-10**. Непараметрические гипотезы. Критерий Пирсона. Проверка гипотезы о виде распределения. Проверка гипотезы о независимости признаков.

Критерии Колмогорова и Смирнова.

**Занятие 9**. Непараметрические критерии. Проверка гипотез о виде распределения.

**КСР№5**. Самостоятельная работа №2. Задача №6.

**Занятие 10**. Критерий Пирсона. Проверка о независимости признаков.

**Лекция 11.** Обработка двумерной выборки. Законы распределения выборочных характеристик в случае гауссовского распределения. Проверка гипотез о коэффициенте корреляции.

**Занятие 11**. Обработка двумерной выборки. Коэффициент корреляции.

**Лекция 12**. Двумерная выборка. Функция регрессии. Корреляционное отношение. Линейность функции регрессии в гауссовском случае. Метод наименьших квадратов.

**Занятие 12.** Линейная регрессионная модель.

**КСР№6**. Самостоятельная работа №2. Задача №7.

**Лекции13-14**. Линейная модель регрессии. Оценки коэффициентов регрессии и их свойства. Статистический анализ регрессионной модели. Проверка гипотез в регрессионной модели.

**Занятие 13.** Свойства оценок коэффициентов регрессии

**Занятие 14.** Рубежный контроль№2.

**Лекция 15**. Коллоквиум по модулю.

**Занятие 15. Коллоквиум.**

**Занятие 1.**

Первоначальная обработка статистических данных. Задача №1.

**Определение 1**. Все значения, которые может принимать случайная величина , называют выборочным пространством, или генеральной совокупностью.

Выборкой объема называют возможных значений случайной величины:

- реализация выборки.

**Определение 2.** Статистикой, называется любая функция от выборки, не зависящая от неизвестных параметров распределения.

**Примеры**. 1) - выборочное среднее

(или - для статистического ряда)

2)– выборочная дисперсия

- для статистического ряда

**Определение 3**. Если выборку упорядочить в порядке возрастания, то получим вариационный ряд: ; при этом - -ая порядковая статистика

3) - статистика (функция от выборки)

Рассмотрим вариационный ряд , построенный по выборке

из распределения

**Определение 4.** Выборочной (эмпирической ) функцией распределения называется

, , ,

То есть , если ровно наблюденных значений меньше .

- статистика (при каждом ), то есть функция от выборки

**Пример**. Рассмотрим выборку: 1,3,4,6,3,1,7,2,3,7

Строим статистический ряд и находим эмпирическую функцию распределения:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 6 | 7 |
| Относительные  частоты |  |  |  |  |  |  |
| Относительные  накопленные частоты |  |  |  |  |  |  |

**Группировка выборки**

Если выборка очень большая и не содержит повторяющихся значений, то ее группируют следующим образом.

**Определение 5.** Размахом выборки называется разность между крайними членами вариационного ряда: .

Группируем выборку, выбирая число интервалов группировки по правилу Стерджеса

Составляем таблицу

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Интервалы группировки |  |  |  |  |  |
| Середины интервалов |  |  |  |  |  |
| Частоты |  |  |  |  |  |
| Относительные частоты |  |  |  |  |  |
| Относительные накопленные частоты |  |  |  |  |  |

Частоты равны количеству элементов выборки, попавших в интервал.

Строим гистограмму относительных частот

**Определение 6.** Гистограммой относительных частот называется объединение прямоугольников, построенных на интервалах группировки так, что площадь каждого равна относительной частоте попадания элементов выборки в соответствующий интервал.

(Горизонтальные стороны прямоугольников равны , а вертикальные - .)

**Определение 7.** Выборочной функцией распределения группированной выборки называется

, , .

- статистика (при каждом ), то есть функция от группированной выборки

- ступенчатая функция, имеющая скачки в серединах интервалов группировки

Величина скачков равна относительной частоте .

**Задача 1**. Первоначальная обработка статистических данных

По данной выборке

1. Найдите крайние члены вариационного ряда и размах выборки

2. Осуществите группировку данных (количество интервалов находим по правилу Стерджеса)

3. По сгруппированным данным постройте гистограмму относительных частот

4. Найдите эмпирическую функцию распределения и постройте ее график

5. Вычислите выборочное среднее и выборочную дисперсию.

**Примечание.**

Варианты, помеченные буквой «с», подчиняются «неудобному» закону и

выбираются только по желанию студента.

**Демонстрационный вариант**:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1.913 | 2.875 | 3.302 | 2.338 | 3.967 | 0.702 | 0.655 | 2.038 | 0.342 | 0.331 |
| 8.936 | 7.5 | 0.86 | 1.763 | 3.297 | 2.003 | 3.036 | 6.432 | 3.267 | 2.78 |
| 0.679 | 2.7 | 0.927 | 3.094 | 2.928 | 2.274 | 4.5 | 3.229 | 5.441 | 2.86 |
| 8.084 | 4.336 | 3.673 | 1.261 | 1.217 | 0.383 | 6.351 | 0.561 | 0.276 | 3.415 |
| 2.626 | 1.753 | 2.088 | 0.552 | 4.465 | 5.842 | 6.888 | 1.189 | 0.454 | 5.123 |
| 3.136 | 0.25 | 3.536 | 0.369 | 0.859 | 8.418 | 3.623 | 2.261 | 2.289 | 2.373 |
| 3.884 | 3.107 | 0.02 | 0.354 | 6.632 | 4.586 | 1.594 | 2.683 | 10.39 | 0.648 |
| 0.471 | 10.102 | 0.094 | 0.192 | 0.471 | 6.658 | 4.263 | 0.049 | 4.102 | 0.818 |
| 0.617 | 1.39 | 1.527 | 5.405 | 2.492 | 5.335 | 0.521 | 1.716 | 0.489 | 0.228 |
| 2.839 | 2.647 | 1.243 | 5.501 | 1.115 | 7.52 | 4.539 | 1.494 | 0.865 | 0.869 |

Пример обработки данных в среде MATHCAD

Вводим данные вашего варианта



Находим крайние члены вариационного ряда и размах выборки

















Находим число интервалов

















Осуществляем группировку и строим гистограмму относительных частот, используя встроенные функции













Построение эмпирической функции распределения



















Вычисление выборочных характеристик

**Центральная предельная теорема (повторение)**

**Теорема 1.** Если случайные величины - независимы, одинаково распределены и имеют конечные то

**Теорема 2.** Если случайные величины - независимы, имеют конечные моменты , пусть , , , причем ,то

**Задачи для решения в аудитории.**

1. Случайные величины  независимы и распределены по закону Пуассона с параметром . Положим , . Найдите .

**Решение.** Как известно,. Используем теорему 3

1. Случайные величины  независимы и распределены по нормальному закону с параметрами . Положим , . Найдите .

**Ответ**. .

1. Случайные величины  независимы и распределены по биномиальному закону распределения с параметрами , . Положим , . Найдите вероятность .

**Ответ*.***.

1. Случайные величины  независимы и распределены равномерно на интервале . Положим , . Найдите вероятность **Ответ*.*** .
2. Случайные величины  независимы и распределены по закону Пуассона с параметром . Положим , . Найдите вероятность .

**Ответ.** .

1. На отрезке  случайным образом выбраны  числа, т.е. рассматриваются  независимых и равномерно распределенных случайных величин . Найдите вероятность того, что их сумма заключена между числами  и .

**Ответ.** .

1. Пусть  последовательность независимых стандартных случайных величин (т.е. ). Используя центральную предельную теорему, найдите вероятность того, что случайная величина  примет значение больше, чем .

**Ответ**. .

1. Вероятность искажения одного сигнала равна . Пользуясь центральной предельной теоремой, найдите вероятность того, что из  переданных сигналов будет искажено: а) более ; б) менее .

**Ответ***.* а) ; б).

1. Случайные величины  независимы и распределены равномерно на интервале . Положим , . Найдите .

**Ответ**. .

1. Урожайность куста картофеля равна 0 кг с вероятностью 0,1, 1 кг с вероятностью 0,2, 1,5 кг с вероятностью 0,2, 2 кг с вероятностью 0,3 и 2,5 кг с вероятностью 0,2. Какое наименьшее число клубней надо посадить, чтобы с вероятностью не менее 0,975 урожай был не менее 1 тонны?

**Решение**. Рассмотрим независимыеслучайные величины , распределенные по закону: , , , ,

Тогда нужно найти такое , что .

Воспользуемся ЦПТ:

**Задачи для самостоятельного решения.**

1. Складываются 300 независимых СВ , равномерно распределенных на отрезке . Найдите а) приближенное выражение плотности СВ ;

б) вероятность события .

**Ответ**.

1. Игральная кость подбрасывается  раз. Найдите вероятность того, что суммарное число очков будет находиться в пределах от  до .

**Ответ**..

1. Случайные величины  независимы и распределены по показательному закону распределения параметром . Положим , . Найдите вероятность .

**Ответ*.*** .

1. Случайные величины  независимы и распределены по биномиальному закону распределения с параметрами , . Положим , . Найдите вероятность 

**Ответ*.*** .

1. В поселке 2500 жителей. Каждый из них примерно 6 раз в месяц ездит в город на поезде, который ходит раз в сутки. Какой наименьшей вместимостью должен обладать поезд, чтобы он переполнялся в среднем не чаще, чем 1 раз в 100 дней?

**Ответ**.

**Лекция 1. Основные распределения математической статистики**

**1. Нормальный закон распределения .**

**2. Гамма-распределение.**

,

Свойства гамма-функции ,

1.

2.

3.

В частности,

**Замечание**.

1.

2. Если , то гамма-распределение совпадает с показательным с параметром :

, .

**3. Бэта-распределение.**

, , где

**Частные случаи.**

**1**. - равномерное распределение на ,

поскольку.

**2.** – закон арксинуса,

поскольку.

**Свойства бэта-функции**

1.  *.*

**Доказательство.**

2. *.*

**Доказательство**

В самом деле, , откуда , или , , домножив на , получаем

.

Далее интегрируем по интервалу обе части по переменной :

**4. Хи-квадрат распределение с степенями свободы**

**Свойства.**

1. - совпадает с распределением квадрата стандартной нормально распределенной случайной величины

**Доказательство**.

Пусть и .

Тогда .

Для продолжаем цепочку: .

Осталось продифференцировать полученное выражение при :

.

2. Пусть независимы и распределены по стандартному нормальному закону. Тогда случайная величина распределена по закону Хи-квадрат с степенями свободы.

**Доказательство** проведем по индукции.

Индукционный шаг:

Использовано значение интеграла , соотношения и .

3.

**5**. **Распределение Фишера-Снедекора .**

**Теорема**. Если случайные величины , , , , независимы и распределены по стандартному нормальному закону , то плотность распределения случайной величины

**Доказательство**.

Найдем сначала плотность распределения частного двух независимых случайных величин и , распределенных по закону и соответственно.

Согласно стандартной процедуре для этого находим и дифференцируем функцию распределения.

Здесь все преобразования проведены для , то есть вычисление двойного интеграла проводится в первой координатной четверти: При очевидно .

Дифференцируем по переменной под знаком интеграла:

Откуда

Далее, поскольку , то ее плотность можно пересчитать по стандартной формуле: . Итак,

**6. Распределение Стьюдента с степенями свободы.**

Пусть , , независимы и распределены по стандартному нормальному закону. Тогда случайная величина имеет плотность распределения   
 (распределение Стьюдента с степенями свободы) **Доказательство**.

Поскольку случайная величина представима в виде , где и независимы, случайная величина распределена по стандартному нормальному закону, а подчинена закону Хи-квадрат с степенями свободы, то найдем ее распределение по стандартной процедуре:

Далее,

**Частный случай**.

– распределение Коши с параметром .

**Справка.** Плотность Коши .

**Занятие 2.**

**Свойства основных распределений математической статистики.**

**Задачи для решения в аудитории.**

* 1. Найдите математическое ожидание и дисперсию случайной величины , имеющей гамма – распределение с плотностью: , .

**Ответ.** , **.**

* 1. Найдите математическое ожидание и дисперсию случайной величины , имеющей бэта – распределение с плотностью: , .

**Ответ**. , **.**

* 1. Докажите, что при плотность распределения Стьюдента поточечно сходится к стандартной нормальной плотности.

**Решение**. Рассмотрим

согласно второму замечательному пределу.

То есть достаточно проверить, что .

? а как это сделать просто?

Если на пальцах, то

* 1. Если случайные величины , , , , независимы и распределены по стандартному нормальному закону , то случайная величина распределена по закону .

**Решение.** Введем обозначения , .

Эти случайные величины независимы и распределены по законам и соответственно. Тогда для

Дифференцируем по под знаком интеграла:

* 1. Найдите математическое ожидание и дисперсию случайной величины , имеющей распределение Фишера: , .

**Решение.**

***-*** существует при

**Ответ**. ,

* 1. Докажите, что сумма независимых случайных величин, распределенных по закону и соответственно, распределена по закону .

**Решение**. Вычислим свертку гамма-плотностей ():

**Задачи для самостоятельного решения.**

* 1. Докажите, что параметр – несущественный (масштабный), поскольку если случайная величина имеет гамма-распределение , то имеет плотность .
  2. Найдите математическое ожидание и дисперсию случайной величины , имеющей Хи-квадрат распределение с степенями свободы с плотностью:

**Ответ.** , .

* 1. Найдите математическое ожидание и дисперсию случайной величины, имеющей распределение Стьюдента с плотностью.

При каких значениях параметра эти характеристики существуют?

**Ответ**. при , при **.**

* 1. Найдите закон распределения случайной величины , если случайные величины , , независимы и распределены по стандартному нормальному закону .

**Ответ**. .

* 1. Пусть случайные величины и независимы и распределены по Коши с параметрами и . Докажите, что случайная величина также распределена по Коши с параметром .

**Лекция 2.** Виды сходимости случайных величин**.** Предельные теоремы теории вероятностей. Закон больших чисел.

Пусть на заданы случайная величина и последовательность случайных величин .

**1. Сходимость почти наверное (с вероятностью 1).**

**Определение 1.** , если .

**Пример**. Пусть - алгебра измеримых подмножеств,

. Определим .

Тогда , , то есть

**Определение 2.** , если при любом

**Утверждение 1.** Определение 1 эквивалентно определению 2.

**Доказательство**.

Далее, по определению предела,

Что означает

Если по условию , то для всех , откуда с учетом , получаем, что для всех

Использовано свойство непрерывности вероятности:

Если так, что , то

где

Обратная импликация тоже имеет место, поскольку

**2. Сходимость по вероятности**.

**Определение 3**. , если при любом

**Утверждение 2.**  .

**Доказательство**. Очевидно.

**Замечание**. Из сходимости по вероятности, вообще говоря, не следует сходимость почти наверное.

**Пример**. Пусть - алгебра измеримых подмножеств, .

Определим для последовательность . Тогда однако .

**Утверждение 3.** Если последовательность монотонно возрастает (убывает) и , то .

**Доказательство**. Пусть монотонно убывает и (иначе рассмотрим последовательность ).

Тогда .

**Утверждение 4**. **(б/д)**. Если , то из последовательности можно выбрать подпоследовательность такую, что .

**3. Сходимость по распределению (слабая)**.

**Определение 4**. , или , если для любой непрерывной ограниченной функции

**Определение 5.** , если в каждой точке непрерывности ).

Следовательно, .

**Факт без доказательства**. Определение 4 и определение 5 эквивалентны.

**Утверждение 5.**  .

**Доказательство**. Обозначим

Если , то , следовательно

, откуда . Далее,

, что влечет . Следовательно,

, откуда выводим, что

.

Остается заметить, что в каждой точке непрерывности , и, следовательно, существует .

**Утверждение 6.**  .

**Доказательство**. По условию при любом

, . Следовательно,

.

Таким образом, , чтд.

**4. Сходимость в среднем порядка** .

**Определение 6**. , если .

**Утверждение 7**. .

**Доказательство**. Применим неравенство Чебышева:.

|  |  |  |
| --- | --- | --- |
| \ | 0 | 1 |
|  |  |  |

**Пример (факультативно).** Пусть - последовательность независимых случайных величин, распределенных по закону

Тогда =.

С другой стороны,

.

**Предельные теоремы**

**Теорема 1 (неравенства Чебышева).** Для любого имеют место неравенства:

**Доказательство**.

Второе неравенство получается из первого подстановкой вместо :

**Следствие**.

**Теорема 2 (закон больших чисел Чебышева).** Пусть - независимые случайные величины и существует такая константа , что все ,

Тогда при любом

**Доказательство**. Если , то

. Следовательно,

- используется сходимость по вероятности:

**Теорема 3(ЦПТ).** Если случайные величины - независимы, одинаково распределены и имеют конечные то

**-**используется сходимость по распределению

**Теорема 4 (Ляпунов).** Если случайные величины - независимы, имеют конечные моменты , пусть , , , причем ,то

**-**используется сходимость по распределению:

**Теорема 5(теорема Хинчина ).** Пусть - одинаково распределенные случайные величины с конечным математическим ожиданием . Тогда при любом

-используется сходимость по вероятности:

**Теорема 6(теорема Маркова).** Пусть - последовательность случайных величин, удовлетворяющая условию . Тогда при любом

*-* используется сходимость по вероятности:

**Теорема 7(необходимое и достаточное условие).** Пусть - последовательность случайных величин с . Тогда закон больших чисел имеет место тогда и только тогда, когда . Тогда при любом

-используется сходимость по вероятности:

**Теорема 8**. (усиленный закон больших чисел Колмогорова) Пусть - независимые и одинаково распределенные случайные величины. Для того, чтобы

необходимо и достаточно, чтобы существовало конечное .

- используется сходимость почти наверное:.

**Резюме**.

**Занятие 3.**

**Разбор текущего домашнего задания №2+долги.**

**(задача 2.12)** Найдите математическое ожидание и дисперсию случайной величины, имеющей распределение Стьюдента с плотностью: .

При каких значениях параметра эти характеристики существуют?

**Ответ**. при , при **.**

**Решение.** , так как при интеграл сходится и подынтегральная функция нечетная.

Далее, заметим, что

Следовательно,

**Определение 5.** Выборочной(эмпирической) функцией распределения называется

, , ,

То есть , если ровно наблюденных значений меньше .

- статистика (при каждом ), то есть функция от выборки

**Задачи для решения в аудитории**

1. Докажите, используя центральную предельную теорему и задачи 2.1 и 2.6, что

.

**Решение***.* Рассмотрим независимыеслучайные величины , распределенные по закону . Согласно задаче 1.1

По центральной предельной теореме случайная величина

в пределе имеет стандартное нормальное распределение.

С другой стороны, согласно задаче 1.6. , распределение суммы будет Пересчитаем плотность по формуле .

Получаем: .

Осталось заменить в этой формуле переменную на .

1. Докажите следующие свойства эмпирической функции распределения.

1.

2. ,

3. не убывает

4. непрерывна слева

1. Докажите **Теорему:** Если взята выборка объема из генеральной совокупности, имеющей функцию распределения то дискретная случайная величина, закон распределения которой имеет вид:

*,*

**Доказательство**. При каждом индикаторы – независимые случайные величины с законом распределения:

Следовательно, –распределена по закону Бернулли, то есть

.

1. Вычислите

1.

2.

1. Докажите, что если – выборка неограниченного объема, то .

**Доказательство**. Рассмотрим

Следовательно, , а значит и .

**Теорема 4 (Гливенко-Кантелли (б/д))** При  .

**Задача 2**. Моделирование и обработка выборки из дискретного закона распределения.

**Задание.**

1. Для данного смоделируйте выборку из заданного дискретного закона распределения: .

2. Для полученной выборки постройте статистический ряд. Найдите эмпирическую функцию распределения Постройте на одном рисунке графики и .

3. Вычислите выборочное среднее и выборочную дисперсию и сравните с истинными значениями этих характеристик.

**Объяснения**

**Принцип моделирования выборки из дискретного распределения.**

Пусть моделируемый закон имеет ряд распределения

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Значения СВ |  |  | … |  |  |
| Вероятности |  |  | … |  |  |

Интервал разбиваем на интервалов следующим образом

**Алгоритм**

1. Берем - случайную точку на

2. Если , то

Если попадает в интервал , , то

Если , то .

**Задача 3**. Моделирование выборки из абсолютно непрерывного закона распределения методом обратных функций.

**Задание.**

1. Для данного методом обратных функций смоделируйте выборку из закона распределения с заданной плотностью .

2. Для полученной выборки найдите гистограмму относительных частот. Постройте на одном рисунке графики теоретической плотности и гистограмму относительных частот.

3. Вычислите выборочное среднее и выборочную дисперсию и сравните с истинными значениями этих характеристик.

**Объяснения к задаче 3.**

**Метод обратной функции моделирования абсолютно непрерывной СВ**

Пусть функция распределения строго монотонно возрастает

**Теорема**. Пусть случайная величина равномерно распределена на , пусть обратная функция к функции распределения . Тогда случайная величина распределена по закону .

**Доказательство**. Для равномерно распределенной на случайной величины верно, что . Следовательно, поскольку , то для всех значений

.

**Алгоритм**

1. - вспомогательное случайное значение на

2. - значение моделируемой случайной величины

**Лекция 3**. Основные определения и задачи математической статистики. Выборочные характеристики и их свойства. Теорема Гливенко-Кантелли. Порядковые статистики.

**Задача оценки параметров**

Пусть - выборка из известного распределения , зависящего от неизвестного параметра .

**Задача:** оценить значение параметра .

**Определение 1.**  - оценка параметра (функция от выборки)

**Определение 2** . Оценка параметра называется несмещенной, если .

Выборочные характеристики и их свойства

**Примеры**.

1) Выборочное среднее является несмещенной оценкой математического ожидания.

2) Выборочная дисперсия является несмещенной оценкой дисперсии .

**Теорема 1.** (свойства выборочного среднего и выборочной дисперсии)

Если - выборка из распределения с , то

1. , .

, .

**Доказательство**.

1.

2.

**Определение 3**. Оценка параметра называется состоятельной, если

**Примеры**.

1) Выборочное среднее является состоятельной оценкой математического ожидания в силу ЗБЧ (теорема Хинчина):

2) Выборочная дисперсия является состоятельной оценкой дисперсии , поскольку

Первое слагаемое в квадратных скобках сходится по вероятности к в силу ЗБЧ, а второе по пункту 1) к . Следовательно, .

Поскольку то.

**Определение 4** . Оценка параметра называется асимптотически несмещенной, если .

**Теорема 2.** Если оценка асимптотически несмещенная и ее дисперсия стремится к нулю, то она состоятельна.

**Доказательство**. По условию и .

Рассмотрим разность

Второе слагаемое в правой части по условию стремится к нулю,

Первое в силу неравенства Чебышева оценивается

И тоже стремится к нулю по вероятности. Следовательно, .

**Порядковые статистики.**

Рассмотрим вариационный ряд , построенный по выборке

из распределения

**Определение**Члены вариационного ряда называются порядковыми статистиками: - -ая порядковая статистика

**Теорема 3.** Если независимая выборка взята из генеральной совокупности с функцией распределения , то функции распределения крайних членов вариационного ряда и их совместная функция распределения имеют вид:

**Следствие**. Если выборка взята из абсолютно непрерывного закона с плотностью , то плотности распределения крайних членов вариационного ряда и их совместная плотность имеют вид:

**Теорема 4**. Если независимая выборка взята из генеральной совокупности с плотностью распределения , то плотность распределения - ой порядковой статистики имеет вид:

**Пример**. Найдем математическое ожидание и дисперсию , если выборка получена из равномерного распределения на .

Следовательно, – асимптотически несмещенная оценка параметра .  
**Метод моментов.**

Пусть –неизвестный параметр для известного закона распределения.

Вычисляются эмпирических и теоретических моментов и решается система из уравнений с неизвестными.

**Примеры.**  1. Пуассоновский закон с неизвестным параметром .

Тогда – математическое ожидание (первый теоретический момент)

- выборочное среднее (первый эмпирический момент)

Следовательно, – оценка параметра , построенная по выборке

**2.** Пусть  –выборка из нормального распределения с неизвестными параметрами и .

Теоретические моменты: ,

Приравниваем к эмпирическим моментам и находим оценки:

**Занятие 4**. Свойства выборочных характеристик. Порядковые статистики. Метод моментов.

**Задачи для решения в аудитории.**

1. Докажите, что оценка неизвестного параметра , построенная по выборке из распределения с плотностью , является асимптотически несмещенной и состоятельной.

**Решение**. Функция распределения выборки ,

Математическое ожидание

По теореме 2 из лекции3 статистика является состоятельной оценкой параметра .

1. Постройте по методу моментов оценку неизвестного параметра по выборке из распределения с плотностью .

**Ответ**. .

1. Найдите оценку параметров равномерного распределения на отрезке по методу моментов.

**Решение**.

**-** оценки по методу моментов

1. Выборка – имеет плотность распределения

При заданных значениях параметров и найдите оценку параметра .

Таблица частот

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| интервалы | 0-0,4 | 0,4-0,8 | 0,8-1,2 | 1,2-1,6 | 1,6-2 | 2-2,4 | 2,4-2,8 | 2,8-3,2 | 3,2-3,6 | 3,6-4 |
| Часто  ты | 217 | 175 | 151 | 122 | 96 | 86 | 78 | 57 | 11 | 7 |

**Ответ.**

1. Найдите математическое ожидание и дисперсию , если выборка получена из равномерного распределения на .

**Задачи для самостоятельного решения.**

1. Постройте по методу моментов оценку неизвестного параметра по выборке из распределения с плотностью .

**Ответ**. .

1. Постройте оценки по методу моментов для параметров и по выборке из гамма-распределения с плотностью .

**Ответ**.

1. Дана выборка объема из распределения Парето с плотностью

. В качестве оценки неизвестного параметра используется . Докажите, что эта оценка является асимптотически несмещенной.

1. По выборке из генеральной совокупности, распределенной по закону Бернулли с вероятностью успеха , в качестве точечной оценки математического ожидания используется . Докажите, что эта оценка несмещенная и состоятельная.
2. По выборке из генеральной совокупности, распределенной по показательному закону с параметром , в качестве точечной оценки математического ожидания используется . Докажите, что эта оценка смещенная.

**Решение 4.9**. Плотность показательного закона .

Функция распределения показательного закона .

Математическое ожидание . Найдем .

Следовательно, .

**Лекция 4**. Метод максимального правдоподобия. Свойства точечных оценок. Неравенство Рао-Крамера. Информация Фишера.

## Метод максимального правдоподобия.

Дискретный закон, зависящий от неизвестного параметра

**Определение**. 1.Функцией правдоподобия случайной выборки из дискретного закона, зависящего от неизвестного параметра, называется

2.Функцией правдоподобия случайной выборки из абсолютно непрерывного закона с плотностью , зависящей от неизвестного параметра, называется

**Определение**. Значение **,** при котором достигается максимальное значение функции правдоподобия, называется оценкой максимального правдоподобия.

**Примеры**. 1. Пуассоновский закон с неизвестным параметром .

Поскольку , то функция правдоподобия имеет вид:

то естьоценка максимального правдоподобия совпадает с оценкой по методу моментов.

2. Нормальный закон с неизвестными параметрами и .

Составим функцию правдоподобия

Критические точки: ,

Находим производные второго порядка

Следовательно, – выполнены достаточные условия максимума

Следовательно,

, - ОМП.

3. Равномерное распределение на неизвестном интервале

Составим функцию правдоподобия

Функция правдоподобия примет свое максимальное значение, если разность

, при условии, что все , то есть , - оценки максимального правдоподобия

**Определение**. Оценка называется асимптотически нормальной, если

## Сравнение оценок.

**Определение**. Если и  **–** две несмещенные оценки параметра , и . Тогда оценка , чем оценка .

**Пример**. Рассмотрим выборку из равномерного закона , где – неизвестный параметр, – известный параметр.

Рассмотрим две оценки и

Найдем и , используя их законы распределения.

В самом деле, , где

.  
Найдем плотности:

Следовательно,

Таким образом, , и обе оценки несмещенные.

Сравним дисперсии.

Аналогично,

Далее, поскольку

то

Итого, что означает большую эффективность оценки .

**Теорема** (неравенство Рао-Крамера)

Пусть – выборка из параметрической модели с плотностью , пусть допускает дифференцирование по параметру. Тогда каждая несмещенная оценка параметра удовлетворяет неравенству:

где – информация Фишера относительно семейства , содержащаяся в одном наблюдении.

**Замечание.** Для дискретной модели вместо используем .

**Доказательство**. Пусть . Тогда, дифференцируя по параметру очевидные интегральные соотношения, получаем:

Следовательно, , откуда с учетом неравенства Коши-Буняковского , выводим, что

Поскольку

то неравенство доказано.

**Определение.** Оценка называется эффективной, если для нее неравенство Рао-Крамера превращается в равенство, то есть ее эффективность

**Пример**. Пусть - выборка из нормального распределения , причем - известно, – неизвестно.

Рассмотрим оценку неизвестного параметра . Известно, что , . Вычислим информацию Фишера :

Следовательно, – оценка эффективна

**Занятие 5**. Точечные оценки параметров. Самостоятельная работа №1.

**Задачи для решения в аудитории**

1. Найдите оценку параметров равномерного распределения на отрезке по методу максимального правдоподобия

**Решение**. Метод максимального правдоподобия

Функция правдоподобия примет свое максимальное значение, если разность

, при условии, что все , то есть , - оценки максимального правдоподобия

1. Постройте ОМП для параметра по выборке из распределения Парето с плотностью .

**Решение**. Функция правдоподобия

возрастает по при всех наборах и любом - ОМП

1. Покажите, что оценка из предыдущей задачи смещенная.

**Решение**. Поскольку , то

1. Методом максимального правдоподобия найдите оценку неизвестного параметра θ, если плотность распределения с.в. имеет вид и по наблюдениям получены следующие данные: 2,4; 3,5; 3,2; 3,4; 2,5; 2,4; 3,1; 3,4; 3,8; 2,6.

**Ответ**.

1. Покажите, что оценка максимального правдоподобия параметра , полученная по выборке из пуассоновского закона, является эффективной.

**Решение**. Поскольку , то функция правдоподобия имеет вид:

то естьоценка максимального правдоподобия совпадает с оценкой по методу моментов.

Вычислим информацию Фишера:

, ,

Следовательно,

**Задачи для самостоятельного решения**

1. Постройте ОМП для параметра по выборке из логнормального распределения с плотностью . Параметр считать известным. Вычислите информацию Фишера и покажите, что полученная оценка эффективна.

**Ответ**.

1. Постройте ОМП для параметра по выборке из биномиального закона распределения: .

**Ответ**.

1. Постройте ОМП для параметра по выборке из распределения с плотностью . Параметр считать известным.

А) Покажите, что полученная оценка смещенная.

Б) Постройте несмещенную оценку параметра .

В) Вычислите информацию Фишера и покажите, что асимптотически эффективна.

**? Ответ**.

1. Найдите ОМП двумерного параметра , если плотность распределения имеет вид и по наблюдениям получены следующие данные: -1,04; -0,52; 0,0; 0,78; 1,0.

**Самостоятельная работа №1 «Методы получения оценок»**

**Вариант 1**

1. Найдите методом моментов оценку неизвестного параметра θ, если выборка взята из равномерного на распределения. Докажите, что полученная оценка состоятельна.

2. Методом максимального правдоподобия найдите оценку неизвестного параметра θ, если плотность распределения с.в. имеет вид и по наблюдениям получены следующие данные: 6; 12; 15; 24; 30.

**Вариант 2**

1. Методом максимального правдоподобия найдите оценку неизвестного параметра θ, если плотность распределения с.в. имеет вид и по наблюдениям получены следующие данные: 1,4; 3,5; 3,2; 4,4; 2,5; 3,4; 2,1; 2,4; 3,8; 2,6.

2. Найдите методом моментов оценку неизвестного параметра θ, если выборка взята из распределения с плотностью . Докажите, что полученная оценка несмещенная.

**Вариант 3**

1. Методом максимального правдоподобия найдите оценку неизвестного параметра θ, если плотность распределения с.в. имеет вид и по наблюдениям получены следующие данные: 2,4; 3,5; 3,2; 3,4; 2,5; 2,4; 3,1; 3,4; 3,8; 2,6.

2. Найдите методом моментов оценку неизвестного параметра θ, если выборка взята из распределения с плотностью . Докажите, что полученная оценка несмещенная.

**Вариант 4**

**1.** Методом максимального правдоподобия найдите оценку неизвестного параметра θ, если плотность распределения с.в. имеет вид и по наблюдениям получены следующие данные: 10; 14; 16; 18; 20.

2. Постройте оценку по методу моментов для параметра по выборке из биномиального закона распределения: . Докажите, что полученная оценка несмещенная и состоятельная.

**Вариант 5**

**1.** Найдите ОМП неизвестного параметра , если выборка взята из распределения с плотностью параметр известен. Докажите, что полученная оценка несмещенная.

2. Методом моментов найдите оценку неизвестного параметра θ, если плотность распределения с.в. имеет вид и по наблюдениям получены следующие данные: 1; 2; 3; 4; 5.

**Вариант 6**

1. Найдите методом моментов оценку двумерного параметра по выборке , полученной из закона распределения с плотностью .

2. Методом максимального правдоподобия найдите оценку неизвестного параметра θ, если плотность распределения с.в. имеет вид и по наблюдениям получены следующие данные: 4; 7; 9; 12; 14.

**Ответы.**

**Вариант 1. 1.**

**Вариант 2. 1.**  .

**Вариант 3. 1.**  . **2.**

**Вариант 4. 1.**

**Вариант 5. 1.**

**Вариант 6. 1.**  . **2.**

|  |  |
| --- | --- |
| **Вариант 1 СР№1-2016-ФН**  **1.** Найдите методом моментов оценку неизвестного параметра θ, если выборка взята из равномерного на распределения. Докажите, что полученная оценка состоятельна.  **2.** Методом максимального правдоподобия найдите оценку неизвестного параметра θ, если плотность распределения имеет вид и по наблюдениям получены следующие данные:  6; 12; 15; 24; 30. | **Вариант 2 СР№1-2016-ФН**  **1**. Методом максимального правдоподобия найдите оценку неизвестного параметра θ, если плотность распределения имеет вид и по наблюдениям получены следующие данные: 1,4; 3,5; 3,2; 4,4; 2,5; 3,4; 2,1; 2,4; 3,8; 2,6.  **2.** Найдите методом моментов оценку неизвестного параметра θ, если выборка взята из распределения с плотностью . Докажите, что полученная оценка несмещенная. |
| **Вариант 3 СР№1-2016-ФН**  **1.** Методом максимального правдоподобия найдите оценку неизвестного параметра θ, если плотность распределения имеет вид , и по наблюдениям получены данные: 2,4; 3,5; 3,2; 3,4; 2,5; 2,4; 3,1; 3,4; 3,8; 2,6.  **2**. Найдите методом моментов оценку неизвестного параметра θ, если выборка взята из распределения с плотностью . Докажите, что полученная оценка несмещенная. | **Вариант 4 СР№1-2016-ФН**  **1.** Методом максимального правдоподобия найдите оценку неизвестного параметра θ, если плотность распределения с.в. имеет вид и по наблюдениям получены следующие данные:10;14;16;18; 20.    **2**. Постройте оценку по методу моментов для параметра по выборке из биномиального закона распределения: . Докажите, что полученная оценка несмещенная и состоятельная. |
| **Вариант 5 СР№1-2016-ФН**  **1.** Найдите ОМП неизвестного параметра , если выборка взята из распределения с плотностью параметр известен. Докажите, что полученная оценка несмещенная.  **2.** Методом моментов найдите оценку неизвестного параметра θ, если плотность распределения с.в. имеет вид и по наблюдениям получены следующие данные: 1; 2; 3; 4; 5. | **Вариант 6 СР№1-2016-ФН**  **1.** Найдите методом моментов оценку двумерного параметра по выборке , полученной из закона распределения с плотностью  .  **2.** Методом максимального правдоподобия найдите оценку неизвестного параметра θ, если плотность распределения с.в. имеет вид и по наблюдениям получены следующие данные: 4; 7; 9; 12; 14. |

**Лекция 5.** Достаточные статистики. Критерий факторизации. Совместный закон распределения выборочного среднего и выборочной дисперсии

## Достаточные статистики

Пусть - выборка из распределения, зависящего от параметра (реализация случайного вектора с независимыми компонентами)

**Определение**. Вектор-функция называется достаточной статистикой для оценки параметра , если условная функция распределения не зависит от при любых значениях .

**Замечание.**

Статистика является достаточной для оценки параметра , если

**Дискретный случай**

Условные вероятности не зависят от при любых значениях .

**Непрерывный случай**

Условные плотности не зависят от при любых значениях .

**Пример**. Пусть – реализации бернуллиевской случайной величины (принимающей значения 1 и 0 с вероятностями и ).

Докажем по определению, что является достаточной статистикой для оценки параметра . Найдем условные вероятности

Вероятность, стоящая в знаменателе, по теореме Бернулли равна

Числитель равен нулю, если

А если , то

Таким образом, условная вероятность в любом случае не зависит от :

**Теорема (критерий факторизации)**.

Статистика является достаточной для оценки параметра тогда и только тогда, когда функция правдоподобия имеет вид:

**Доказательство** для дискретного случая.

Пусть Тогда

Зафиксируем и рассмотрим случай Тогда в числителе стоит , а в знаменателе

Следовательно,

Обратно, если - достаточная статистика и , то

=

В этом произведении первый множитель не зависит от , а второй зависит от , но от выборки зависит только через статистику

**Примеры.**

**1.** Пусть - выборка из закона Пуассона с параметром .

Выпишем функцию правдоподобия:

Применяем критерий факторизации: , а

- зависит от выборки только через , то есть - достаточная статистика для оценки параметра .

**2.** Пусть - выборка из нормального закона распределения .

Выпишем функцию правдоподобия:

Она зависит от выборки только через и , то есть - достаточная статистика.

**Совместный закон распределения выборочного среднего и выборочной дисперсии**

**для нормально распределенной генеральной совокупности.**

**Теорема.** Если –выборка из нормального распределения с параметрами и , то и независимы, и имеет распределение .

**Лемма.** Если случайный вектор имеет плотность , – невырожденная матрица, а , то плотность .

**Доказательство леммы**.

С другой стороны,

Следовательно, .

**Доказательство теоремы**. Рассмотрим центрированную и нормированную выборку .

Компоненты вектора независимы и распределены по закону , следовательно, .

Произведем ортогональное преобразование , причем положим

, то есть матрица перехода имеет вид .

При этом , ,

.

Кроме того,

что означает, что вектор также имеет независимые компоненты, распределенные по закону .

Тогда статистика представима в виде:

, откуда следует независимость и , а также вид распределения .

**Занятие 6**. Достаточные статистики.

**Задачи для решения в аудитории**

Пусть - выборка из показательного закона распределения, (плотности )

Выпишем функцию правдоподобия:

Она зависит от выборки только через , то есть - достаточная статистика.

**Задачи для самостоятельного решения**

**Лекции №6-7**

**Доверительные интервалы**

**Определение.** Доверительным интервалом для одномерного параметра с доверительной вероятностью называется любой интервал , содержащий истинное значение параметра с вероятностью .

## Принцип построения доверительных интервалов:

Пусть – выборка из некоторого закона распределения, зависящего от параметра .

1. Находим статистику , закон распределения которой не зависит от неизвестного параметра .

2. Находим квантили и функции распределения уровней и , то есть такие значения, что монотонно возрастающая функция достигает указанных уровней в этих точках: , . При этом

3. Неравенство разрешается относительно параметра

Обозначив , , получаем доверительный интервал уровня

## Построение доверительных интервалов для параметров нормального распределения

1. Доверительный интервал для (неизвестного) математического ожидания при известном с.к.о. .

Статистика распределена по нормальному закону с параметрами

Следовательно,

Далее, для квантилей в силу симметрии имеет место соотношение . Разрешаем относительно параметра неравенство :

Таким образом, –доверительный интервал для параметра с доверительной вероятностью .

1. Доверительный интервал для (неизвестного) с.к.о. при известном математическом ожидании .

Рассмотрим статистику . Поскольку , то , , следовательно, распределена по закону .

Находим квантили и , разрешаем неравенство

относительно :

, – доверительный интервал для при известном

1. Доверительный интервал для (неизвестного) с.к.о. при неизвестном математическом ожидании .

Рассмотрим статистику .

По теореме статистика распределена по закону .

Находим квантили и , разрешаем неравенство

относительно :

, - доверительный инт ервал для при неизвестном .

**Примечание**. Доверительный интервал для при неизвестном математическом ожидании шире, чем при известном, что объяснятся меньшим количеством информации.

1. Доверительный интервал для (неизвестного) математического ожидания при неизвестном с.к.о. .

По теореме, статистика и независима от статистики , распределенной по закону . Таким образом, отношение распределено по закону Стьюдента с .

Строим доверительный интервал:

**Примеры**. Пусть

Построим доверительные интервалы

## **5. Построение доверительного интервала для разности математических ожиданий двух независимых выборок из нормального распределения при известных значениях СКО**

Пусть –выборка из нормального распределения с параметрами и ,

–выборка из нормального распределения с параметрами и .

Рассмотрим статистику , распределение которой нормально с параметрами

, .

Следовательно, статистика распределена по стандартному нормальному закону .

Строим доверительный интервал:

## **6. Построение доверительного интервала для разности математических ожиданий двух независимых выборок из нормального распределения при неизвестном СКО**

Пусть –выборка из нормального распределения с параметрами и ,

–выборка из нормального распределения с параметрами и .

Параметр неизвестен, но известно, что распределено по стандартному закону и не зависит от статистики , подчиняющейся распределению .

После очевидных преобразований

, получаем распределение

Стьюдента .

Найдем квантили и и построим доверительный интервал:

Следовательно, с вероятностью

, где .

## **7. Построение доверительного интервала для отношения дисперсий двух независимых выборок из нормального распределения**

Пусть –выборка из нормального распределения с параметрами и ,

–выборка из нормального распределения с параметрами и .

Рассмотрим статистику: .

Отношение, стоящее справа имеет распределение Фишера .

Находим квантили и распределения Фишера и

строим доверительный интервал для отношения :

## **8. Построение доверительного интервала для отношения дисперсий двух независимых выборок из нормального распределения при известных средних**

В случае известных средних значений, используя статистики и , получаем аналогично

**Теорема Колмогорова-Блэкуэллла .**

Пусть –выборка из распределения - достаточная статистика для неизвестного параметра , пусть - несмещенная оценка параметра с конечной дисперсией. Тогда также является несмещенной оценкой и .

**Доказательство**.

**Пример.** Пусть –выборка из нормального распределения

Рассмотрим в качестве оценки неизвестного параметра статистику .

Достаточной статистикой в этом примере является .

Найдем , вычислив условную плотность .

По определению . Поскольку , то знаменатель . Чтобы найти числитель, выпишем характеристики нормального вектора : ,

, . Следовательно, .

В итоге

Из этой формулы видно, что условное математическое ожидание , что означает .

## Лекция 9. Задача проверки гипотез. Виды гипотез: параметрические, непараметрические, простые, сложные. Основная гипотеза и альтернатива. Ошибки 1-го и 2-го рода. Оптимальный критерий Неймана-Пирсона.

## Проверка гипотез.

Пусть –выборка из некоторого закона распределения, возможно зависящего от некоторых параметров.

**Гипотезы –** предположения относительно параметров, вида распределения и т.д.

**Определения.**

Гипотеза называется параметрической, если делается предположения о значениях параметров.

Гипотеза называется непараметрической, если делается предположения о виде закона распределения, независимости признаков.

Гипотеза называется простой, если она однозначно определяет закон распределения.

Гипотеза называется сложной, если она неоднозначно определяет закон распределения.

**Примеры**.

выборка из распределения Пуассона с параметром – простая параметрическая

выборка из нормального распределения – сложная непараметрическая

выборка из распределения с параметром - простая параметрическая, если .

выборка из нормального распределения с параметром – сложная параметрическая

монета правильная – простая гипотеза

монета неправильная – сложная гипотеза

**Постановка задачи:**

Одна из гипотез – основная, и вместе с ней рассматривается конкурирующая гипотеза, или альтернатива

Если выборка из известного закона распределения с параметром

Альтернатива : - простая

– двусторонняя

- односторнние

**Определение.** Критерием называется правило, по которому принимается решение, о принятии гипотезы , или отклонении в пользу альтернативы , на основе анализа выборки.

Критерий задается с помощью критического множества и статистики

Если **z ,** то гипотеза  **отклоняется,**

Если **z ,** то гипотеза  **принимается**

**Определения.**

**z –** ошибка первого рода – вероятность отклонения верной гипотезы (уровень значимости)

**z –** ошибка второго рода **–**вероятность принятия неверной гипотезы в случае простой альтернативы

- функция мощности критерия

=z – вероятность отвергнуть неверную гипотезу

Пусть - простая гипотеза

Вид критического множества для

: – двусторонняя

альтернатива

: – правосторонняя

альтернатива

**Примеры**.

1) Пусть сопротивление резисторов – случайная величина, подчиняющаяся нормальному закону. Произведено измерений, в результате чего оказалось, что кОм.

Проверьте на уровне значимости гипотезу против двусторонней альтернативы : в двух ситуациях: а) ; б)

А) Как известно, статистика имеет стандартное нормальное распределение, то есть , или .

То есть критическое множество в этом случае имеет вид

Поскольку , то гипотеза отвергается.

Б) В этом случае используем статистику , распределенную по закону Стьюдента с 35 степенями свободы. Тогда

Следовательно, получаем критическое множество

В этом случае гипотеза принимается, поскольку .

2) В условиях предыдущего примера проверьте на том же уровне значимости гипотезу против левосторонней альтернативы : в двух ситуациях: а) ; б)

А) Решение этого примера построено на той же статистике , однако критическое множество будет состоять из значений этой статистики, говорящих в пользу альтернативы, то есть

Поскольку , то гипотеза отвергается.

Б) Аналогично предыдущему случаю, основную гипотезу нужно отвергать в случае малых значений . Произведем вычисления:

В этом случае гипотеза принимается, поскольку .

3) В условиях предыдущего примера проверьте на том же уровне значимости ту же гипотезу против простой альтернативы : , если .

Поскольку , критическое множество совпадет с найденным в примере 2a):

, что дает основания отвергнуть и принять . Вычислим ошибку 2-го рода:

.

4) В условиях предыдущего примера для того же уровня значимости найдите ошибку второго рода при проверке гипотезы против простой альтернативы : , если .

Поскольку , критическое множество будет тем же: , что дает основания отвергнуть и принять . Вычислим ошибку 2-го рода:

.

**Выводы**.

Для построенного критерия в случае простой альтернативы ошибка второго рода вычисляется однозначно.

Большая ошибка второго рода говорит о низкой мощности критерия, он плохо различает близкие гипотезы.

**Возможные задачи:**

1. Для данной альтернативы найти максимально мощный (оптимальный) критерий , то есть

2. Для построенного критерия выбрать такую альтернативу, чтобы достигнуть максимальной мощности.

3. Найти необходимый объем выборки, обеспечивающий заданную мощность критерия при заданных основной и альтернативной гипотезах.

**Улучшение критерия за счет увеличения объема наблюдений**.

**Дано** ,

,

, .

**Найти**: минимальный объем выборки.

В рассматриваемой ситуации критическое множество имеет вид:

По условию

Решаем систему: , .

**Оптимальный критерий Неймана-Пирсона.**

С каждым критерием свяжем функцию . Тогда

.

**Теорема (критерий отношения правдоподобия).** Рассмотрим множество критериев уровня для проверки гипотезы против простой альтернативы , . Тогда для любого существуют такие числа и , что критерий с функцией является оптимальным во множестве критериев уровня .

**Доказательство**.

Рассмотрим функцию распределения случайной величины

Проведем доказательство для случая, когда непрерывная функция.

Тогда выберем - квантиль уровня функции , то есть , в этом случае положим .

Уровень значимости полученного критерия будет равен , поскольку

.

Рассмотрим любой другой критерий с тем же уровнем значимости. Тогда интеграл

будет неотрицательным. В самом деле, на множестве значение , а значит и , откуда следует, что первое слагаемое неотрицательно. Аналогично второй интеграл также неотрицателен, поскольку на множестве получаем , что влечет , и снова под интегралом неотрицательная функция.

Следовательно, .

В этом неравенстве интеграл слева равен , а интеграл справа –

, откуда и следует доказательство теоремы.

**Занятие 8.** Проверка параметрических гипотез. Критерий отношения правдоподобия.

**Задачи для решения в аудитории.**

1. Пример 4.19\*. Постройте на уровне значимости критерий для проверки гипотезы против простой альтернативы : , если выборка подчинена нормальному закону, Найдите ошибку 2-го рода.

**Ответ**. , .

1. Постройте на уровне значимости критерий для проверки гипотезы против простой альтернативы : , если выборка подчинена показательному закону Проверьте гипотезу против , если . Найдите ошибку 2-го рода.

**Решение**. Отношение правдоподобия ,

**Ответ**. , принимается, .

1. Постройте на уровне значимости критерий для проверки гипотезы против простой альтернативы : , если выборка подчинена нормальному закону, Найдите ошибку 2-го рода.

**Ответ**. , .

1. Найдите минимальное число опытов, позволяющее при проверке гипотезы против простой альтернативы : обеспечить ошибку второго рода , если выборка подчинена нормальному закону, , .

**Ответ**. .

1. Проверьте на уровне значимости гипотезу против двусторонней альтернативы : , если выборка подчинена нормальному закону,

.

**Ответ**. , отклоняется.

1. Пример 4.28. До наладки станка была проверена точность изготовления 10 деталей и найдена оценка дисперсии . После наладки проверено еще 15 изделий и получена оценка . Можно ли считать, что после наладки точность повысилась? Принять . Контролируемый признак имеет нормальное распределение.

**Решение**.

:

Используем статистику Фишера , распределенную по закону Фишера с параметрами 9 и 14. Критическое множество при гипотезе имеет вид:

. Поскольку , то

**Ответ**. , принимается.

1. Пример 4.25. Ведутся наблюдения за состоянием технологического процесса. Разладка оборудования приводит к изменению контролируемого признака, имеющего нормальное распределение с дисперсией . По результатам двух выборок объема Найдены м и м. Проверьте стабильность технологического процесса на уровне .

**Ответ**. Процесс стабилен, так как гипотеза о равенстве средних принимается.

1. Пример 4.26 Давление в камере дважды измерялось двумя манометрами. По результатам 10 замеров получены следующие данные: Есть ли основания считать, что давление не изменилось? .

**Ответ**. Гипотезу о неизменности давления отвергаем.

**Самостоятельная работа.**

**1**. задача 4.16. Найдите минимальное число опытов, позволяющее при проверке гипотезы против простой альтернативы : обеспечить ошибку второго рода , если выборка подчинена нормальному закону, , .

**Ответ**. .

**2**. Пусть размер шариков, изготовляемых станком-автоматом, – случайная величина, подчиняющаяся нормальному закону. Произведено измерений, в результате чего оказалось, что м. Проверьте на уровне значимости гипотезу против односторонней альтернативы в двух ситуациях: а) ; б)

**Ответ**. а) Гипотеза принимается; б) Гипотеза принимается.

**3**. По двум независимым выборкам, объёмы которых , извлеченным из нормальных генеральных совокупностей, найдены исправленные выборочные дисперсии . При уровне значимости проверьте гипотезу о равенстве генеральных дисперсий при альтернативной гипотезе .

**4.** По двум независимым выборкам, извлечённым из нормальных генеральных совокупностей, найдены выборочные средние . Генеральные дисперсии известны: При уровне значимости 0.05 проверьте гипотезу при альтернативной гипотезе . Объёмы выборок .

**5.** По двум независимым выборкам объёмов , извлеченным из нормальных генеральных совокупностей, найдены исправленные выборочные дисперсии . При уровне значимости 0.05 проверьте гипотезу о равенстве генеральных дисперсий при альтернативной гипотезе .

6. Из нормальной генеральной совокупности с известным средним квадратичным отклонением извлечена выборка объёмом , и по ней найдено выборочное среднее . При уровне значимости 0.05 проверьте гипотезу при альтернативной гипотезе .

**7**. По двум независимым выборкам объёмов , извлеченным из нормальных генеральных совокупностей, найдены выборочные средние и исправленные выборочные дисперсии . При уровне значимости 0.05 проверьте гипотезу при альтернативной гипотезе .

8. Найдите минимальное число опытов, позволяющее при проверке гипотезы против простой альтернативы : обеспечить ошибку второго рода , если выборка подчинена нормальному закону, , .

**9**. Постройте на уровне значимости критерий для проверки гипотезы против простой альтернативы : , если выборка подчинена показательному закону Проверьте гипотезу против , если . Найдите ошибку 2-го рода.

**10**. пример 4.27. Из партии болтов взята выборка объема и рассчитана выборочная дисперсия длины болта. Можно ли считать, что станок обеспечивает допустимый разброс длины, если нормативное значение квадрата отклонения равно 400. Принять 0.05.

**Лекция 10**

**Определение объема выборки. Критерий Вальда.**

**Занятие 9**

**Запасные задачи**

По двум независимым выборкам объёмов , извлеченным из нормальных генеральных совокупностей, найдены исправленные выборочные дисперсии . При уровне значимости 0.05 проверьте гипотезу о равенстве генеральных дисперсий при альтернативной гипотезе .

Из нормальной генеральной совокупности с известным средним квадратичным отклонением извлечена выборка объёмом , и по ней найдено выборочное среднее . При уровне значимости 0.05 проверьте гипотезу при альтернативной гипотезе .

**Еще задачи к занятию 9:**

Пример 4.21

Пример 4.9

Пример 4.7+4.22

Пример 4.29

Пример 4.15+задача 4.36

Пример 4.16+задача 4.37

|  |  |
| --- | --- |
| Самостоятельная работа  «Параметрические гипотезы»-2016  **Вариант 1**  1. По двум независимым выборкам, объёмы которых , извлеченным из нормальных генеральных совокупностей, найдены исправленные выборочные дисперсии . При уровне значимости проверьте гипотезу о равенстве генеральных дисперсий при альтернативной гипотезе .  2. Найдите минимальное число опытов, позволяющее при проверке гипотезы  против простой альтернативы : обеспечить ошибку второго рода , если выборка подчинена нормальному закону, , . | Самостоятельная работа  «Параметрические гипотезы»-2016  **Вариант 2**  1. Пусть размер шариков, изготовляемых станком-автоматом, – случайная величина, подчиняющаяся нормальному закону. Произведено измерений, в результате чего оказалось, что м. Проверьте на уровне значимости гипотезу против односторонней альтернативы в двух ситуациях: а) ; б)  **2.** Постройте на уровне значимости критерий для проверки гипотезы  против простой альтернативы : , если выборка подчинена показательному закону Проверьте гипотезу против , если . Найдите ошибку 2-го рода. |
| Самостоятельная работа  «Параметрические гипотезы»-2016  **Вариант 3**  1. Найдите минимальное число опытов, позволяющее при проверке гипотезы  против простой альтернативы : обеспечить ошибку второго рода  , если выборка подчинена нормальному закону, , .  2. По двум независимым выборкам объёмов , извлеченным из нормальных генеральных совокупностей, найдены выборочные средние и исправленные выборочные дисперсии  . При уровне значимости 0.05 проверьте гипотезу при альтернативной гипотезе . | Самостоятельная работа  «Параметрические гипотезы»-2016  **Вариант 4**  1. По двум независимым выборкам, извлечённым из нормальных генеральных совокупностей, найдены выборочные средние . Генеральные дисперсии известны: При уровне значимости 0.05 проверьте гипотезу при альтернативной гипотезе . Объёмы выборок .  . Из партии болтов взята выборка объема и рассчитана выборочная дисперсия длины болта. Можно ли считать, что станок обеспечивает допустимый разброс длины, если нормативное значение квадрата отклонения равно 400. Принять 0.05. |

**Лекция 9**. Непараметрические гипотезы. Критерий согласия Пирсона. Проверка гипотезы о виде распределения.

Рассмотрим выборку из некоторого закона распределения и гипотезы

выборка из распределения с законом распределения (c плотностью )

выборка не подчиняется этому закону

Переходим к группированной выборке.

В дискретном случае:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Значения |  |  |  |  |
| Эмпирические  частоты |  |  |  |  |
| Теоретические  частоты |  |  |  |  |

Во второй строке представлены эмпирические частоты выпадения значений из верхней строки; в третьей теоретические частоты при гипотезе

В непрерывном случае:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Интервалы |  |  |  |  |
| Эмпирические  частоты |  |  |  |  |
| Теоретические  частоты |  |  |  |  |

Во второй строке представлены частоты попадания значений в интервалы из верхней строки; в третьей теоретические частоты при гипотезе

Вычисляется статистика . Ее распределение при больших приближенно равно , где – число параметров распределения, оцениваемых по выборке.

**Теорема Пирсона**. Статистика , где - число параметров распределения , оцениваемых по выборке.

**Доказательство**. Для частного случая .

При справедливости основной гипотезы вектор имеет полиномиальное распределение: где

Заметим, что

Найдем характеристическую функцию вектора :

Далее определим случайный вектор , положив .

Его характеристическая функция может быть выражена через следующим образом:

Исследуем асимптотическое поведение при больших значениях

Докажем, что представляет собой вырожденную квадратичную форму.

Для этого рассмотрим произвольное ортогональное преобразование с фиксированной первой строкой вида: .

Тогда для случайного вектора получим:

Следовательно, , , поскольку

Таким образом, характеристическая функция вектора имеет вид: , что означает, что в пределе вектор нормален с нулевым математическим ожиданием и матрицей ковариаций , то есть .

Следовательно, в пределе подчиняется распределению . Теорема доказана.

**Выводы**:

Большие значения статистики говорят о расхождении практики с предположением . Критическое множество имеет вид: .

**Критерий**

гипотеза отклоняется

гипотеза принимается

**Пример 1**. Имеется выборка объема

: выборка взята из равномерного распределения

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | 2 |  | 4 | 5 | 6 | 7 | 8 | 9 |  |
|  | 11 | 12 |  | 13 | 7 | 10 | 11 | 9 | 14 | 8 |
|  | 10,4 | 10,4 | 10,4 | 10,4 | 10,4 | 10,4 | 10,4 | 10,4 | 10,4 | 10,4 |

Вычисляем

По таблице находим

Поскольку гипотеза принимается

**Пример 2**. Опыт Менделя 556 горошин, полученных при скрещивании во втором поколении

: Гипотеза Менделя о доминантных признаках согласуется с практикой

|  |  |  |  |
| --- | --- | --- | --- |
| Признаки | Частоты | Вероятность |  |
| Круглые и желтые | 315 | 9/16 | 312,75 |
| Круглые и зеленые | 108 | 3/16 | 104,25 |
| Морщинистые и желтые | 101 | 3/16 | 104,25 |
| Морщинистые и зеленые | 32 | 1/16 | 34,75 |

Гипотеза Менделя подтверждается

**Замечание**. Если , то этот столбец объединяется с соседним.

**Пример**. Бомбардировки Лондона, на 576 участков по 0,25 упало 537 снарядов. Проверить гипотезу о пуассоновском распределении числа снарядов, упавших на участок.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | 1 |  | 3 | 4 |  |
|  | 229 | 211 |  | 35 | 7 | 1 |
|  | 226,74 | 211,39 | 98,54 | 30,62 | 7,14 | 1,57 |

Оцениваем , находим

*, ,*,54…

(два последних интервала объединяются)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | 1 |  | 3 |  |
|  | 229 | 211 |  | 35 | 8 |
|  | 226,74 | 211,39 | 98,54 | 30,62 | 8,71 |

Находим – гипотеза принимается.

**Занятие 9.** **Критерий Пирсона. Проверка гипотезы о виде распределения.**

**Задачи для решения в аудитории.**

1. (ЗАДАЧА 1, Щетинин) Цифры 0, 1, 2, … 9 среди первых 800 первых десятичных знаков числа π появились 74, 92, 83, 79, 80, 73, 75, 76, 91 раз соответственно. Проверить с помощью критерия χ2 гипотезу согласия этих данных с законом равномерного распределения на уровне значимости α = 0, 1.

**Ответ**. Гипотеза принимается,

1. (ЗАДАЧА 10, Щетинин). Семь монет подбрасывались одновременно 1536 раз, причем каждый раз отмечалось число выпавших гербов. В таблице приведены числа случаев, когда число выпавших гербов было равно

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|  | 12 | 78 | 270 | 456 | 386 | 252 | 69 | 13 |

Пользуясь критерием , проверить согласие гипотезы о биномиальном законе распределения с опытными данными. Учесть, что вероятность выпадения герба при бросании каждой из монет равна 0, 5. Уровень значимости принять равным α = 0,05.

**Ответ**. Гипотеза принимается,

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|  | 12 | 78 | 270 | 456 | 386 | 252 | 69 | 13 |
|  | 12 | 84 | 252 | 420 | 420 | 252 | 84 | 12 |

1. (ЗАДАЧА 7, Щетинин). Ниже приводятся результаты опыта (данные Уэлдона) с подбрасыванием костей. Выпадение 6 очков на одной грани при 4096 подбрасываниях 12 костей.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Число выпадений | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 и более | Итого |
| Число случаев | 447 | 1145 | 1181 | 796 | 380 | 115 | 24 | 8 | 4096 |

Проверить с помощью критерия χ2 гипотезу о правильности костей. Уровень значимости принять равным α = 0,05.

**Ответ**.

1. В итоге регистрации прихода посетителей выставки получена таблица:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Интервал времени | 12–13 | 13–14 | 14–15 | 15–16 | 16–17 |
| Число посетителей | 250 | 157 | 99 | 54 | 30 |

С помощью критерия , проверить гипотезу о том, что время прихода посетителей на выставку распределено по показательному закону. Принять α=0,1.

**Ответ.** Гипотеза отвергается, так как

(По выборке оценивался один параметр )

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Интервал времени | 12–13 | 13–14 | 14–15 | 15–16 | 16–17 | Итого |
| Число посетителей | 250 | 157 | 99 | 54 | 30 | 590 |
| Вероятности | 0,469 | 0,249 | 0,132 | 0,08 | 0,08 | 1 |
| Теоретические частоты | 277 | 147 | 78 | 41 | 47 | 590 |

1. При испытании радиоэлектронной аппаратуры фиксировалось число отказов. Результаты 60 испытаний приводятся ниже:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Число отказов | 0 | 1 | 2 | 3 |
| Число испытаний | 42 | 10 | 5 | 3 |

С помощью критерия , проверить, гипотезу о том, что число отказов имеет распределение Пуассона. Принять α=0,05.

**Ответ**.

1. С помощью критерия Пирсона проверьте гипотезу о нормальном распределении, если имеются следующие данные (:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Интервал |  |  |  |  |  |  |  |
| Эмпир.  частота | 6 | 8 | 15 | 40 | 16 | 8 | 7 |
| Теор.  Частота |  |  |  |  |  |  |  |

**Ответ.** Гипотеза отвергается,

**Самостоятельно**

1. При 50 подбрасываниях монеты герб появился 20 раз. Можно ли считать монету симметричной? .

**Ответ**. ДА, .

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Район | 1 | 2 | 3 | 4 | 5 |
| Объем сбыта | 110 | 130 | 70 | 90 | 100 |

1. Ниже приводятся данные о фактических объемах сбыта (в условных единицах) в пяти районах:

С помощью критерия , проверить, согласуются ли эти результаты с предположением о том, что сбыт продукции в этих районах должен быть одинаковым? Принять α=0,01.

**Ответ**. Гипотеза отклоняется,

**Задачи для самостоятельного решения.**

1. Число выпадений герба при 20 подбрасываниях двух монет распределилось следующим образом: . Согласуются ли эти результаты с предположением о симметричности монет и независимости опытов? .

**Ответ**. ДА, .

1. (ЗАДАЧА 6, Щетинин). По официальным данным шведской статистики в Швеции в 1935 г. родились 88 273 ребенка, причем в январе родилось 7280 человек, в феврале – 6957, в марте – 7883, в апреле – 7884, в мае – 7892, в июне – 7609, в июле – 7585, в августе – 7393, в сентябре – 7203, в октябре – 6903, в ноябре – 6552, в декабре – 7132 человека. Используя критерий χ2, проверить гипотезу о том, что день рождения наудачу выбранного человека с равной вероятностью приходится на любой из 365 дней года. Уровень значимости принять равным α = 0,05.

**Ответ**.

1. Число деталей, поступающих на конвейер в течение 600 двухминутных интервалов: .

Проверить гипотезу о пуассоновском распределении числа деталей при .

**Ответ**. не принимается, (последние 5 интервалов объединяются).

1. Имеются данные о 120 отклонений размера вала от номинального значения (мкм). С помощью критерия *Х*2 проверить гипотезу о том, что результаты получены из нормального распределения генеральной совокупности. Принять α=0,1.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Частота | 36 | 29 | 19 | 18 | 18 |
| Середина интервала | -0,04 | -0,02 | 0,00 | 0,02 | 0,04 |

**Ответ**.

**Лекция 10.**

**Проверка гипотезы о независимости признаков**

Статистика , где

В пределе распределена по ,

: признаки 1 и 2 независимы

**Критерий**

гипотеза отклоняется

гипотеза принимается

**Пример**. 100 студентов, опрос на тему, мешает ли курение учебе.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Курс/ответ | I курс | II курс | III курс |  |
| Нет | 15 | 10 | 0 | 25 |
| Не знаю | 8 | 5 | 7 | 20 |
| Да | 0 | 30 | 25 | 55 |
|  | 23 | 45 | 32 | 100 |

Вычисляем

Находим

Гипотеза о независимости мнения студента от курса отклоняется

**Проверка гипотезы о независимости признаков (таблица сопряженности признаков)**

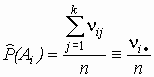
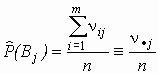
Предположим, имеется большая совокупность объектов, каждый из которых обладает двумя признаками *А* и *В*; признак *А* имеет *m* уровней: *A*1, …, *A*m, а признак *В* – *k* уровней: *B*1, …, *Bk* . Пусть уровень *Аi* встречается с вероятностью *P(Ai),* а уровень *Bj* – c вероятностью *P*(*Bj*). Признаки *А* и *В* независимы, если

*P(Ai Bj) = P(Ai) P(Bj), i =* 1*, …, m, j =* 1*, …, k* ,                                                                     (10)

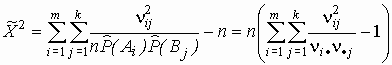
т.е. вероятность встретить комбинацию *Ai Bj* равна произведению вероятностей. Пусть признаки определены на *n* объектах, случайно извлеченных из совокупности; n *ij* – число объектов, имеющих комбинацию *Ai Bj*, http://www.exponenta.ru/educat/systemat/goritskii/part2/lr5/Image138.gif=*n*. По совокупности наблюдений {n *ij* } (таблица *m x k*) требуется проверить гипотезу *Н* о независимости признаков *А* и *В*. Задача сводится к случаю с неизвестными параметрами; ими являются вероятности

*P(Ai), i =* 1*, …, m;  P(Bj), j =* 1*, …, k,*

всего *(m-*1*) + (k-*1*)*; их оценки:

, 

(в обозначениях точка означает суммирование по соответствующему индексу), и статистика (6) принимает вид:

.                             (11)

Если гипотеза *Н* верна, то по теореме Фишера http://www.exponenta.ru/educat/systemat/goritskii/part2/lr5/Image142.gifасимптотически распределена по закону хи-квадрат с числом степеней свободы

*f = mk –* 1 *– (m –* 1*) – (k –* 1*) = (m –* 1*)(k –* 1*),*

и потому, если

image143.gif (1150 bytes),                                           (12)

то гипотезу о независимости признаков следует отклонить.

Ясно, что по (11) – (12) можно проверять независимость двух случайных величин, разбив диапазоны их значений на *m* и *k* частей.

**Критерий согласия Колмогорова.**

**Пример.**

Пассажир, приходящий в случайные моменты времени на автобусную остановку, в течение пяти поездок фиксировал своё время ожидания автобуса: 5,1; 3,7; 1,2; 9,2; 4,8 мин. Проверить гипотезу о том, что время ожидания автобуса равномерно распределено на отрезке [0; 10] на уровне значимости 0,05.

**Занятие 10.**

**Тема**. **Критерий Пирсона. Проверка о независимости признаков.**

**Задачи для решения в аудитории.**

1. =9.6
2. Величина контрольного размера 70 деталей, изготовленных на одном станке (мм):

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Интервал | 2,9-3,9 | 3,9-4,9 | 4,9-5,9 | 5,9-6,9 | 6,9-7,9 |
| Частота | 4 | 16 | 25 | 19 | 6 |

С помощью критерия проверить гипотезу о том, что результаты получены из нормального распределения генеральной совокупности. Принять α=0,1.

**Ответ**.

1. С помощью критерия Пирсона проверьте гипотезу о нормальном распределении, если имеются следующие данные (:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Интервал |  |  |  |  |  |  |  |
| Эмпир.  частота | 6 | 8 | 15 | 40 | 16 | 8 | 7 |
| Теор.  Частота |  |  |  |  |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Результат\способ | A | B | C |  |
| Благоприятный | 11 | 17 | 16 |  |
| Неблагоприятный | 20 | 23 | 19 |  |
|  |  |  |  |  |

**Ответ**

1. Утверждается, что результат действия лекарства не зависит от способа применения. Проверить это утверждение по следующим данным ():

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Результат\поставщик | A | B | C | Всего |
| Годные | 29 | 38 | 53 | 120 |
| Негодные | 1 | 2 | 7 | 10 |
|  | 30 | 40 | 60 | 130 |

**Ответ**.

1. Комплектующие одного наименования поступают с трех предприятий. Можно ли считать, что качество не зависит от поставщика? .

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Недостаток\Дефект |  |  |  | Сумма |
|  | 45 | 26 | 12 | 83 |
|  | 32 | 50 | 21 | 103 |
|  | 4 | 10 | 17 | 31 |
| Сумма | 81 | 86 | 50 | 217 |

**Ответ.**

1. Данные , собранные по ряду школ, относительно физических недостатков школьников (*P*1, *P*2, *P*3 - признак *А*) и дефектов речи (*S*1, *S*2, *S*3 - признак *В*) приведены в таблице.

**Решение.** Для проверки гипотезы о независимости этих двух признаков вычислим статистику: 8; число степеней свободы , минимальный уровень значимости Это значит, что при независимых признаках вероятность получить значение такое же, как в опыте или большее, меньше 0.01, и потому гипотезу о независимости следует отклонить.

1. Пользуясь Критерием согласия Колмогорова, установить, согласуются ли данные наблюдений с предположением о том, что рост мужчин является случайной величиной, распределенной по нормальному закону.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Рост в см. http://abc.vvsu.ru/Books/u_statist_g/obj.files/image162.gif | Число мужчин | Fn(x) | http://abc.vvsu.ru/Books/u_statist_g/obj.files/image164.gif | http://abc.vvsu.ru/Books/u_statist_g/obj.files/image166.gif | http://abc.vvsu.ru/Books/u_statist_g/obj.files/image168.gif | |Fn(x)-F(x)| |
| Менее 143 | 0 | 0 | -3,725 | -0,4999 | 0,0001 | 0,0001 |
| 143-146 | 1 | 0,001 | -2,229 | -0,4994 | 0,0006 | 0,0004 |
| 146-149 | 2 | 0,003 | -2,733 | -0,4969 | 0,0031 | 0,0001 |
| 149-151 | 8 | 0,011 | -2,237 | -0,4874 | 0,0126 | 0,0016 |
| 152-155 | 26 | 0,037 | -1,741 | -0,4592 | 0,0408 | 0,0038 |
| 155-158 | 65 | 0,102 | -1,245 | -0,3934 | 0,1066 | 0,0046 |
| 158-161 | 120 | 0,222 | -0,749 | -0,2730 | 0,2270 | 0,0050 |
| 161-164 | 181 | 0,403 | -0,253 | -0,0998 | 0,4002 | 0,0028 |
| 164-167 | 201 | 0,604 | 0,243 | 0,0960 | 0,5960 | 0,0080 |
| 167-170 | 170 | 0,774 | 0,739 | 0,2700 | 0,7700 | 0,0040 |
| 170-173 | 120 | 0,894 | 1,235 | 0,3916 | 0,8916 | 0,0024 |
| 173-176 | 64 | 0,958 | 1,731 | 0,4582 | 0,9582 | 0,0002 |
| 176-179 | 28 | 0,986 | 2,227 | 0,4870 | 0,9870 | 0,0010 |
| 179-182 | 10 | 0,996 | 2,723 | 0,4968 | 0,9968 | 0,0008 |
| 182-185 | 3 | 0,999 | 3,219 | 0,4994 | 0,9994 | 0,0004 |
| 185-188 | 1 | 1,000 | 3,715 | 0,4999 | 0,9999 | 0,0001 |
| S | 1000 | - | - | - | - | - |

**Решение.**

k0=min|Fn(x)-F(x)|=0,008 (см), n=1000, http://abc.vvsu.ru/Books/u_statist_g/obj.files/image170.gif

 По таблице значений

http://abc.vvsu.ru/Books/u_statist_g/obj.files/image172.gif. http://abc.vvsu.ru/Books/u_statist_g/obj.files/image174.gif – не мала.

 Расхождение между F(x) и Fn(x) незначительное.

 Рост мужчин является случайной величиной, распределенной по нормальному закону.

**Самостоятельная работа (занятие 10) Вариант 1**

1. При исследовании 200 партий (выявлялось число нестандартных деталей в партии) были получены следующие данные: . Проверьте гипотезу о том, что число нестандартных деталей в партии подчинено распределению Пуассона, .
2. В таблице приведены данные о коэффициентах соотношения собственных и заемных средств на 100 малых предприятиях региона. На уровне значимости проверить гипотезу о том, что коэффициенты можно описать нормальным распределением.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Интервал |  |  |  |  |  |  |  |  |
| Середина | 5,1 | 5,2 | 5,3 | 5,4 | 5,5 | 5,6 | 5,7 | 5,8 |
| Частоты | 5 | 8 | 12 | 20 | 26 | 15 | 10 | 4 |

**Самостоятельная работа (занятие 10) Вариант 2**

1. В таблице представлены данные о количестве сделок, заключенных инвесторами на фондовой бирже за квартал. Проверить, используя критерий Пирсона, что на уровне значимости число сделок, заключённых одним инвестором за квартал, распределено по закону Пуассона с параметром .

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|  | 112 | 168 | 130 | 68 | 32 | 5 | 1 | 1 |

|  |  |  |  |
| --- | --- | --- | --- |
| Брови\Волосы | Светлые | Темные | Сумма |
| Светлые | 30472 | 3238 | 33710 |
| Темные | 3364 | 9468 | 12832 |
| Сумма | 33836 | 12706 | 46542 |

1. В таблице приведены данные о распределении цвета волос на голове и бровей у 46542 человек. На уровне значимости проверить гипотезу независимости признаков.

**Самостоятельная работа (занятие 10) Вариант 3**

1. При 120 бросаниях игральной кости шестерка выпала 24 раз, пятерка 19 раз, четверка 22 раз, тройка 22 раза, двойка 17 раз, единица 16 раз. С помощью критерия проверить, согласуется ли этот результат с утверждением, что кость правильная? Принять α=0,05.
2. В таблице приведены данные о распределении 1000 экземпляров северной сосны по диаметру ствола. На уровне проверьте гипотезу о нормальном законе распределения диаметра.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Диаметр ствола, см | 14-18 | 18-22 | 22-26 | 26-30 | 30-34 | 34-38 | 38-42 | 42-46 | 46-50 | 50-54 | 54-58 |
| Количество сосен | 16 | 35 | 109 | 183 | 214 | 197 | 115 | 71 | 36 | 19 | 5 |

**Самостоятельная работа (занятие 10) Вариант 4**

1. Имеются следующие статистические данные о числе вызовов бригад скорой помощи в час в некотором населенном пункте в течение 100 часов:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Число вызовов | 0 | 1 | 2 | 3 | 4 | 5 |
| Частота | 6 | 16 | 36 | 27 | 11 | 4 |

На уровне значимости α=0,1 проверить гипотезу о том, что число вызовов бригад скорой помощи за 1 час имеет распределение Пуассона, используя критерий Пирсона.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  | 1768 | 807 | 180 | 47 |
|  | 946 | 1387 | 746 | 53 |
|  | 115 | 438 | 288 | 16 |

1. В эксперименте каждый индивидуум классифицировался по двум признакам: цвету глаз и цвету волос; при этом по первому признаку ξ индивидуум относился к одной из трех категорий, , а по второму η – к одной из четырех категорий . соответствующие данные для *n* = 6800 индивидуумов приведены в таблице. Проверьте гипотезу о независимости признаков науровне.

**Самостоятельная работа (занятие 10) Вариант 4**

1. Имеются следующие статистические данные о числе вызовов бригад скорой помощи в час в некотором населенном пункте в течение 100 часов:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Число вызовов | 0 | 1 | 2 | 3 | 4 | 5 |
| Частота | 6 | 16 | 36 | 27 | 11 | 4 |

На уровне значимости α=0,1 проверить гипотезу о том, что число вызовов бригад скорой помощи за 1 час имеет распределение Пуассона, используя критерий Пирсона.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  | 1768 | 807 | 180 | 47 |
|  | 946 | 1387 | 746 | 53 |
|  | 115 | 438 | 288 | 16 |

1. В эксперименте каждый индивидуум классифицировался по двум признакам: цвету глаз и цвету волос; при этом по первому признаку ξ индивидуум относился к одной из трех категорий, , а по второму η – к одной из четырех категорий . соответствующие данные для *n* = 6800 индивидуумов приведены в таблице. Проверьте гипотезу о независимости признаков науровне.

**Самостоятельная работа (занятие 10) Вариант 4**

1. Имеются следующие статистические данные о числе вызовов бригад скорой помощи в час в некотором населенном пункте в течение 100 часов:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Число вызовов | 0 | 1 | 2 | 3 | 4 | 5 |
| Частота | 6 | 16 | 36 | 27 | 11 | 4 |

На уровне значимости α=0,1 проверить гипотезу о том, что число вызовов бригад скорой помощи за 1 час имеет распределение Пуассона, используя критерий Пирсона.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  | 1768 | 807 | 180 | 47 |
|  | 946 | 1387 | 746 | 53 |
|  | 115 | 438 | 288 | 16 |

1. В эксперименте каждый индивидуум классифицировался по двум признакам: цвету глаз и цвету волос; при этом по первому признаку ξ индивидуум относился к одной из трех категорий, , а по второму η – к одной из четырех категорий . соответствующие данные для *n* = 6800 индивидуумов приведены в таблице. Проверьте гипотезу о независимости признаков науровне.**Лекции 13-14**. Линейная регрессионная модель (общий случай). Оценки коэффициентов регрессии и их свойства. Статистический анализ регрессионной модели. Проверка гипотез в регрессионной модели.

Рассматривается двумерная выборка:

Условное математическое ожидание (функция регрессии)-

**Свойства**. 1. – несмещенность

2. - наилучшее в смысле средне

квадратического приближение через

Следовательно, **,** где ошибка имеет нулевое математическое ожидание и минимальную возможную дисперсию:

.

В частности, если - гауссовский вектор, то функция регрессии – линейная: , где .

Далее будет рассматриваться общий случай линейной регрессионной модели:

*-* выборка из – мерного распределения,

Векторы – векторы факторов (элементы факторного пространства, или входные переменные)

- отклик, или выходная переменная, измеряемая с ошибкой , – базисные функции, характеризующие модель

**Примеры**.

1. Простейшая линейная регрессия: ,

- единственный фактор ,

– двумерный вектор параметров регрессии ,

- базисные функции.

2. Простейшая квадратическая регрессия: ,

- единственный фактор ,

– трехмерный вектор параметров регрессии ,

- базисные функции.

3. Простейшая линейная двухфакторная регрессия: ,

- двумерный фактор ,

– трехмерный вектор параметров регрессии ,

- базисные функции.

Матричная запись: , где

- вектор откликов,

- вектор параметров регрессии,

- вектор ошибок, удовлетворяющий условиям ()

- матрица базисных функций.

Условия (): 1. для всех .

2. для всех .

3. для всех .

4. Значения входных переменных , считаются неслучайными

(измеренными без ошибок), откуда в частности следует, что

.

**Примеры**. Найдем матрицу для рассмотренных примеров.

1. Простейшая линейная регрессия: .

- матрица размера .

2. Простейшая квадратическая регрессия: ,

- матрица размера .

3. Простейшая линейная двухфакторная регрессия: ,

- матрица размера .

**Теорема 1**. Если выполнены условия ()и матрица является невырожденной, то вектор является оценкой МНК параметра , обладающей свойствами несмещености и оптимальности в классе несмещенных, линейных по оценок.

**Доказательство***.*

1) Найдем МНК параметра , для чего минимизируем сумму квадратов ошибок:

Дифференцируем по каждому из параметров критические точки ( точки минимума):

Эту систему условий можно записать в матричном виде разрешить относительно :

.

2) Проверим несмещенность:

.

3) Докажем эффективность, то есть минимальность дисперсии.

Пусть – произвольная линейная по несмещенная оценка параметра , – произвольная матрица размера .

В силу несмещенности получаем, , откуда - единичная матрица размера .

Далее,

.

Матрица представляется в виде:

Представление верно, поскольку второе и третье слагаемые в предпоследней формуле равны нулю:

,

(матрицы и симметричны).

Диагональные элементы матрицы вида неотрицательны, следовательно,

будет иметь минимальные элементы на диагонали тогда и только тогда, когда , то есть для ОНК.

**Примеры.** Выпишем ОНК для примеров.

1) Простейшая линейная регрессия: .

.

Получаем

2) Простейшая квадратическая регрессия: .

3. Простейшая линейная двухфакторная регрессия: .

**Теорема 2**. Если выполнены условия ()и матрица является невырожденной, то матрица ковариаций оценки имеет вид:

**Доказательство.**

**Следствия.**

1. В модели простейшей линейной регрессии матрица ковариаций имеет вид:

Или

2. Если в модели простейшей линейной регрессии ошибки распределены нормально, а известный параметр, то доверительные интервалы с доверительной вероятностью для параметров линейной регрессии имеют вид:

**Доказательство**. В условиях следствия и имеют нормальное распределение как линейные комбинации нормально распределенных случайных величин . Следовательно, с учетом следствия 1,

, , откуда и получаются приведенные доверительные интервалы.

**Определение.** Случайная величина называется оценкой среднего значения отклика.

Случайная величина называется остаточной суммой квадратов.

В случае простейшей линейной регрессии

- оценка среднего значения отклика,

- остаточная сумма квадратов.

**Теорема 3**. Если выполнены условия ()и ранг матрицы максимален, то несмещенная оценка остаточной дисперсии имеет вид:

**Доказательство.** Поскольку , то

Следовательно, имеет место представление

Далее, вычислим математическое ожидание левой части:

и математическое ожидание последнего слагаемого справа

Следовательно, Чтд.

**Теорема 4**. Если выполнены условия ()и ранг матрицы максимален, то

распределена по закону .

**Доказательство.** Аналогично предыдущей теореме имеет место матричное равенство:

Дальше черновик

, следовательно случайная величина распределена по закону .

По теореме 2 , что означает, что