**Datacamp**

**Part I Основные понятие математической статистики**

В этой главе мы познакомимся с базовыми понятиями, формулами из математической статистики, которые необходимы для построение доверительных интервалов и их реализациями на языке python.

Понятия из мат. статистики необходимы в большинстве экономических (и не только) исследований, при этом исследуя большие базы данных считать, например, мат ожидания, дисперсии и т.д просто нереально, и в таком случае python приходит на помощь не только как быстрый калькулятор, но и как помощник в визуализации данных.

Начнем эту главу с определения таких понятий как выборка, эмпирическая функция распределения и выборочные характеристики.

Проведя n раз эксперимент в одинаковых условиях получим числа – значения наблюдаемой величины в первом, втором и т.д. экспериментах.

***Выборкой*** объёма n из распределения называется набор из n независимых и одинаково распределенных случайных величин, имеющих распределение .

Если элементы выборки упорядочить по возрастанию на каждом элементарном исходе, то полученный набор случайных величин будет называться ***вариационным рядом*.**

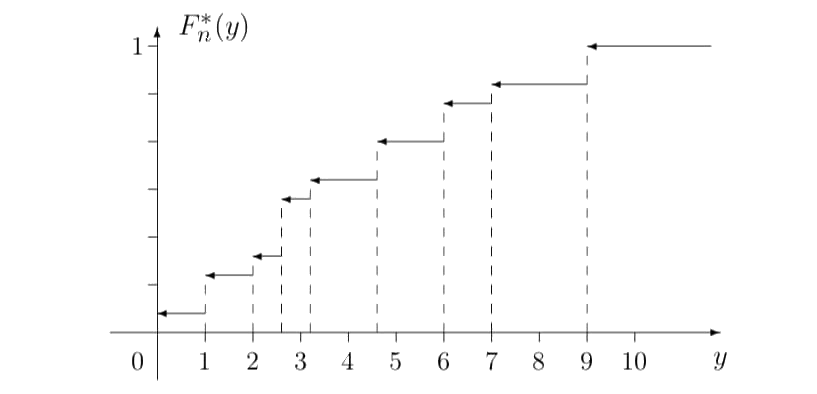
А называется k-ой **порядковой статистикой**

**Эмпирическая функция распределения** (выборочная функция распределения) — естественное приближение теоретической функции распределения данной случайной величины, построенное по выборке.

Эмпирической функцией распределения, построенной по выборке объёма n, называется случайная функция X равная

= ,

где = - индикатор события {}



**Выборочные моменты**

|  |  |  |
| --- | --- | --- |
| Истинные моменты | Оценки для истинных моментов | Функция на python |
|  | – выборочное среднее |  |
|  | – выборочная дисперсия (смещенная)  Либо  – несмещенная выборочная дисперсия |  |
|  | – выборочный начальный момент k-ого порядка |  |

**Тестовый вопрос**

Пусть дана выборка из нормального распределения: (1, 5, 2, 12, 3). Чему равна смещенная и несмещенная дисперсии?

а) Смещенная – 19,3, несмещенная – 15, 44   
  
б) Смещенная – 6,528, несмещенная – 8,16   
  
**в) Смещенная – 15,44, несмещенная – 19, 3**

**Задача 1**

Дана реализация случайной выборки x = (2, 8, 20, 65, 80,23, 60,90, 37, 2, 23, 90, 2). Для данной реализации:

1. Найдите вариационный ряд
2. Найдите мат ожидание, дисперсию, стандартное отклонение и медиану (с использованием numpy и используя формулы и функцию sum)
3. Найдите выборочный момент 4 порядка
4. Постройте гистограмму для данной выборки

**Задача 2**

Для проведения исследования, ученые собрали данных о выборах в Аргентине в период 1983-2007 гг, когда предыдущий президент переизбирался на второй или третий и т.д срок и посмотрели на разницу процента голосов отданных за него, и за его главного конкурента назвав эту переменную winning margin, а также собрали данные об темпах роста ВВП за год до выборов.

|  |  |  |  |
| --- | --- | --- | --- |
| Country | Year | Winning Margin | GdpGrowth-1 |
| Argentina | 1983 | 6.20 | -0.043 |
| Argentina | 1989 | 8.60 | -0.029 |
| Argentina | 1995 | 1.10 | 0.050 |
| Argentina | 1999 | 7.30 | -0.008 |
| Argentina | 2003 | 16.60 | -0.147 |
| Argentina | 2007 | 99.42 | 0.070 |

В рамках этого небольшого датасета, мы научимся вычислять некоторые параметры по реальной выборке.

1. Найдите число наблюдений в данном датасете

2. Постройте вариационный ряд значений winning margin

3. Постройте эмпирическую функцию распределения.

4.Посчитайте среднее, дисперсию и стандартное отклонение как для winning margin, так и для темпов роста ВВП  
  
5. Найдите 25% и 75% квантили для доли GdpGrowth-1

6.Посчитайте корреляцию между Winning margin и GdpGrowtht-1

**Part II Нормальное (гауссовское) распределение**

Определение: Случайная величина имеет нормальное распределение с параметрами , где , >0, если имеет следующую плотность распределения:

f(=,

Графики плотностей нормальных распределений с одним и тем же математическим ожиданием

Изображение выглядит как снимок экрана

Автоматически созданное описание

 Если же мы построим графики плотностей нормального распределения с одной и той же дисперсией , то они будут выглядеть следующим образом

*Изображение выглядит как снимок экрана

Автоматически созданное описание*

При этом, нормальное распределение c параметрами называется стандартным нормальным распределением и его плотность равна (=

– функция распределения стандартного нормального распределения может быть записана в следующем виде:

Значения этой функции при различных действительных вычислены и могут быть найдены в таблице нормального распределения

Важной операцией, связанной с нормальным распределением является **операция стандартизации**, которая означает преобразование произвольного нормального распределения с параметрами µ, σ в стандартное с параметрами 0 и 1. Стандартизация означает, что мы смещаем распределение и изменяем его форму так, чтобы оно стало стандартным

Если мы имеем , то z =

*Центральная предельная теорема (ЦПТ)*

Пусть   - последовательность независимых одинаково распределенных с.в. с конечной дисперсией. Обозначим   и  ,

 Тогда:

,

где функция распределения стандартного нормального закона.

Или очень простыми словами: Сумма большого числа как угодно распределенных независимых случайных величин распределена асимптотически нормально, если только слагаемые вносят равномерно малый вклад в сумму.

Простейшие, но неточные методы моделирования основываются на центральной предельной теореме: если сложить много независимых одинаково распределённых величин с конечной дисперсией, то сумма будет распределена *примерно* нормально.

**Тестовый вопрос**

Задана следующая функция плотности: f(=

Определите, чему равно математическое ожидание и дисперсия

а) mu = 4, = 32

**b) mu = 5, = 16**

c) mu = 5, = 4

**Задача 1**

Итак, в этой задаче мы познакомимся с базовыми командами в python, связанными с нормальным распределением

Для генерации мы будем пользоваться пакетом numpy:

import numpy as np

Чтобы сгенерировать значение из нормального распределения:

a = np.random.normal(mu, sigma, n)

Предположим, что на основании проведенных за несколько лет контрольных работ преподаватель знает, что результаты студентов заданы нормальным распределением. Среднее оценка студентов за контрольную работу равна 5, стандартное отклонение равно 3.

1. Загрузите пакет numpy и сгенерируйте выборку из нормального распределения с параметрами N(5,9) и n=100
2. Предположим, что сгенерированная вами выборка, это результаты студентов за недавно проведенную контрольную работу. Найдите среднее по этой выборке
3. Найдите дисперсию оценок в этой выборке
4. Постройте график выборочной функции плотности
5. Найдите квантили порядка 0.5, 0.95, 0.99

**Задача 2**

В этой задаче мы научимся считать вероятности, если мы работаем с нормальным распределением.

Дневная добыча нефти в некотором месторождении распределена по нормальному закону с математическим ожиданием 100 тысяч баррелей и стандартным отклонением 22 тысячи баррелей.

а) Найдите вероятность того, что в определенный день будут добыты по крайней мере 110 тысяч баррелей нефти.

б) Определите долю рабочих дней, в которые будет добыто от 70 до 105 тысяч баррелей нефти.

в) Найдите вероятность того, что в данный день добыча нефти окажется ниже 65 тысяч баррелей.