Министерство науки и высшего образования Российской Федерации

Федеральное государственное автономное образовательное учреждение высшего образования

«Национальный исследовательский Томский политехнический Университет»



01.03.02 «Прикладная математика и информатика»

**Лабораторная работа № 2**

Оценка закона распределения на основе выборочных данных.

Вариант – 18

по дисциплине:

**Математическая статистика**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Исполнитель:** |  | | | | |
| студент группы | 0В01 |  | Саматов Денис Сергеевич |  | 6.02.2022 |
|  |  |  |  |  |  |
| **Руководитель:** |  | | | | |
| преподаватель |  |  | Шинкеев Михаил Леонидович |  |  |
|  |  |  |  |  |  |

Томск – 2020

***Цель работы:***

Проверка гипотез о законе распределения выборочных данных с помощью критерия хи-квадрат.

***Задание.***

Дана выборка (приложение 3) объемом n =100 из неизвестного распределения F. Требуется, используя критерий хи-квадрат, на уровне значимости 0,05 проверить гипотезы о распределении выборочных данных по законам, указанным в варианте задания. При построении статистики критерия для неизвестных параметров распределения использовать оценки метода максимального правдоподобия (или метода моментов, если есть сложности с нахождением оценок ММП). Для каждой из гипотез указать достигнутый

уровень значимости.

Вариант 18.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 5,70 | 11,67 | 0,99 | 3,27 | 2,27 | 9,66 | 2,21 | 15,04 | 2,68 | 6,21 |
| 10,48 | 9,16 | 3,29 | 12,61 | 7,17 | 6,88 | 4,92 | 3,06 | 4,97 | 5,54 |
| 1,62 | 9,16 | 1,76 | 5,21 | 7,43 | 12,35 | 1,61 | 7,41 | 7,91 | 10,56 |
| 11,98 | 8,90 | 13,34 | 5,51 | 4,12 | 4,94 | 13,62 | 14,67 | 8,36 | 4,74 |
| 11,52 | 8,15 | 3,02 | 5,94 | 9,12 | 2,17 | 4,17 | 4,36 | 3,75 | 1,81 |
| 2,03 | 3,95 | 11,81 | 2,58 | 3,11 | 4,54 | 11,91 | 9,25 | 5,92 | 9,90 |
| 6,11 | 2,00 | 6,83 | 7,71 | 6,30 | 10,88 | 1,48 | 3,73 | 4,51 | 6,58 |
| 4,53 | 6,10 | 15,03 | 15,16 | 8,02 | 6,75 | 6,32 | 5,40 | 2,01 | 3,04 |
| 5,31 | 8,42 | 3,22 | 8,25 | 8,03 | 9,35 | 3,98 | 2,73 | 5,65 | 5,41 |
| 2,61 | 3,70 | 6,22 | 2,30 | 5,27 | 3,80 | 9,70 | 6,36 | 11,70 | 2,54 |
|  |  |  |  |  |  |  |  |  |  |

1.  - закон распределения с плотностью распределения , , где параметр  неизвестен;
2.  - распределение Релея с плотностью , , где параметр  - неизвестен.

***Выполнение задания:***

***Для второго закона распределения:***

Построим статистический ряд, осуществив группировку данных. Находим

xmin = 0,99, xmax = 15,16. Число интервалов группирования m определяем по формуле Стерджесса: m =1+ [log2n] = 8. Для удобства возьмем в качестве нижней границы первого интервала значение x0 = 0, а в качестве верхней границы последнего интервала значение x8 = 16, тогда длина каждого интервала группирования будет равна Δx =16 / 8 = 2. Подсчитывая частоты, получаем следующий ряд:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Интервал | 0 - 2 | 2 - 4 | 4 - 6 | 6 - 8 | 8 - 10 | 10 - 12 | 12 - 14 | 14 - 16 |
| Частота | 7 | 24 | 21 | 16 | 15 | 9 | 4 | 4 |

Видим, что частоты распределены по интервалам крайне неравномерно (для корректного применения критерия хи-квадрат необходимо, чтобы количество значений, попавших в каждую ячейку, было не слишком мало), поэтому делаем перегруппировку данных, добиваясь более равномерного распределения частот по интервалам. В результате получаем следующий ряд:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Интервал | 0 - 3 | 3 - 4 | 4 - 5 | 5 - 6 | 6 - 8 | 8 - 10 | 10 - 13 | 13 - 16 |
| Середина | 1,5 | 3,5 | 4,5 | 5,5 | 7 | 9 | 11,5 | 14,5 |
| Частота | 18 | 13 | 10 | 11 | 16 | 15 | 11 | 6 |
| Относительная частота | 0,18 | 0,13 | 0,10 | 0,11 | 0,16 | 0,15 | 0,11 | 0,06 |
| Плотность  частоты | |  | | --- | | 0,06 | | 0,13 | 0,1 | 0,11 | 0,08 | 0,075 | 0,03666 | 0.02 |

Соответствующая гистограмма приведена на рисунке.

Анализируя гистограмму, видим, что распределение экспериментальных данных похоже на распределение Релея.

Проверим, используя критерий хи-квадрат гипотезу : «Выборочные данные имеют распределение Релея».

Поскольку параметр σ показательного распределения нам неизвестен, необходимо найти оценку данного параметра. В качестве оценки неизвестного параметра σ возьмем оценку, полученную по методу моментов (через второй момент). Так =, то = = 27,29; .

Вычислим значения плотности распределение Релея в точках, соответствующих серединам интервалов группирования, и сравним гистограмму с графиком плотности:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Интервал | 0 - 3 | 3 - 4 | 4 - 5 | 5 - 6 | 6 - 8 | 8 - 10 | 10 - 13 | 13 - 16 |
| Середина | 1,5 | 3,5 | 4,5 | 5,5 | 7 | 9 | 11,5 | 14,5 |
| Плотность частоты | |  | | --- | | 0,06 | | 0,13 | 0,1 | 0,11 | 0,08 | 0,075 | 0,03666 | 0.02 |
| Теоретическая плотность | |  | | --- | | 0,054409 | | 0,105065 | 0,116104 | 0,117437 | 0,104817 | 0,073547 | 0,035628 | 0,010265 |

Насколько статистически незначимо (или значимо) это различие в полученном результате можно определить, используя критерий Пирсона.

Примерим критерий Пирсона для проверки нашей гипотезы о законе распределения выборочных данных. Подсчитаем вероятности  попадания в каждый интервал при условии, что генеральная совокупность имеет распределение Релея с параметром  = 5,22: , F(x)= - функция распределения Релея. Причем для последнего интервала, полагаем  и, соответственно,  и, поскольку теоретически для распределение Релея плотность отлична от нуля на интервале . Далее находим ожидаемые значения -  и нормированные квадраты отклонений  по всем интервалам. Результаты оформляем в виде таблицы:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Интервал | 0 - 3 | 3 - 4 | 4 - 5 | 5 - 6 | 6 - 8 | 8 - 10 | 10 - 13 | 13 - 16 |
| Частота | 18 | 13 | 10 | 11 | 16 | 15 | 11 | 6 |
| Функция распределения F(x) | 0 | 0,2612 | 0,3769 | 0,4940 | 0,7021 | 0,8493 | 0,9591 | 1 |
| Вероятность | 0,1566 | 0,1046 | 0,1156 | 0,1171 | 0,2081 | 0,1471 | 0,1098 | 0,0408 |
| Ожидаемые значения | 15,66 | 10,46 | 11,56 | 11,70 | 20,81 | 14,71 | 10,98 | 4,08 |
|  | 0,34 | 0,61 | 0,21 | 0,04 | 1,11 | 0,01 | 1,75 E-05 | 0,90 |

Находим наблюдаемое значение статистики критерия Пирсона: 

Зададим уровень значимости . Для заданного уровня значимости и числа степеней свободы  ( - так как один параметр распределения - σ мы оценивали по выборке) найдем критическое значение статистики, как критическую точку распределения  уровня  (или что тоже самое – квантиль уровня 0,95): .

Так как , то гипотеза о распределении данных по распределению Релея.

Уточним значение , минимизируя наблюдаемое значение статистики 

Найдем также достигнутый уровень значимости, то есть такое значение , для которого при истинности нашей гипотезы Для 6 степеней свободы и  получим . Наблюдаемый уровень значимости можно трактовать, как вероятность того, насколько возможно данное распределение выборочных данных при истинности нулевой гипотезы. Поскольку в данном случае эта вероятность не слишком велика, то гипотеза о распределении данных по распределение Релея не принимается.

**Для первого закона распределения:**

Проверим, используя критерий хи-квадрат гипотезу : «Выборочные данные имеют заданное распределения».

Поскольку параметр а показательного распределения нам неизвестен, необходимо найти оценку данного параметра. В качестве оценки неизвестного параметра а возьмем оценку, полученную по методу моментов (через первый момент). Так a =, то = = 0.6201.

Вычислим значения плотности, заданного закона распределение в точках, соответствующих серединам интервалов группирования, и сравним гистограмму с графиком плотности:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Интервал | 0 - 3 | 3 - 4 | 4 - 5 | 5 - 6 | 6 - 8 | 8 - 10 | 10 - 13 | 13 - 16 |
| Середина | 1,5 | 3,5 | 4,5 | 5,5 | 7 | 9 | 11,5 | 14,5 |
| Плотность частоты | |  | | --- | | 0,06 | | 0,13 | 0,1 | 0,11 | 0,08 | 0,075 | 0,03666 | 0.02 |
| Теоретическая плотность | |  | | --- | | 0,0328 | | 0,1206 | 0,1378 | 0,1353 | 0,1101 | 0,0677 | 0,0299 | 0,0093 |

Как видно, различие между кривой плотности и гистограммой невелико (при истинности  площадь столбца гистограммы должна приближенно равняться площади под кривой плотности на данном интервале). Насколько статистически незначимо (или значимо) это различие в полученном результате можно определить, используя критерий Пирсона.

Примерим критерий Пирсона для проверки нашей гипотезы о законе распределения выборочных данных. Подсчитаем вероятности  попадания в каждый интервал при условии, что генеральная совокупность имеет распределение Релея с параметром  = : , F(x)= - функция распределения заданного закона. Причем для последнего интервала, полагаем  и, соответственно,  и, поскольку теоретически для заданного закона распределение плотность отлична от нуля на интервале . Далее находим ожидаемые значения -  и нормированные квадраты отклонений  по всем интервалам. Результаты оформляем в виде таблицы:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Интервал | 0 - 3 | 3 - 4 | 4 - 5 | 5 - 6 | 6 - 8 | 8 - 10 | 10 - 13 | 13 - 16 |
| Частота | 18 | 13 | 10 | 11 | 16 | 15 | 11 | 6 |
| Функция распределения F(x) | 0 | 0,2382 | 0,3752 | 0,5101 | 0,7294 | 0,8658 | 0,959 | 1 |
| Вероятность | 0,1185 | 0,1196 | 0,1310 | 0,1348 | 0,2192 | 0,1364 | 0,0934 | 0,0406 |
| Ожидаемые значения | 11,85 | 11,96 | 13,70 | 13,48 | 21,92 | 13,64 | 9,34 | 4,06 |
|  | 3,18 | 0,08 | 1,00 | 0,45 | 1,60 | 0,13 | 0,29 | 0,92 |

Находим наблюдаемое значение статистики критерия Пирсона: 

Зададим уровень значимости . Для заданного уровня значимости и числа степеней свободы  ( - так как один параметр распределения - а мы оценивали по выборке) найдем критическое значение статистики, как критическую точку распределения  уровня  (или что тоже самое – квантиль уровня 0,95): .

Так как , то гипотеза о распределении данных по распределению Релея.

Уточним значение , минимизируя наблюдаемое значение статистики 

Найдем также достигнутый уровень значимости, то есть такое значение , для которого при истинности нашей гипотезы Для 6 степеней свободы и  получим . Наблюдаемый уровень значимости можно трактовать, как вероятность того, насколько возможно данное распределение выборочных данных при истинности нулевой гипотезы. Поскольку в данном случае эта вероятность велика, то гипотеза о распределении данных по заданному закону принимается.