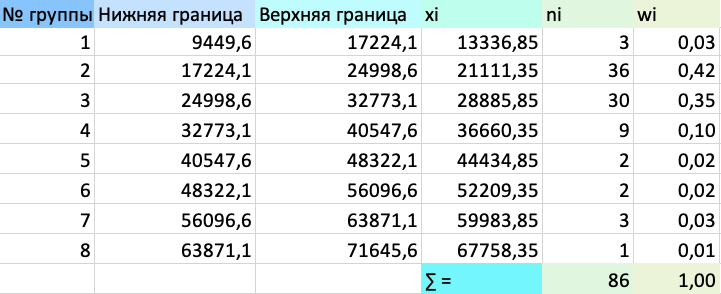
1. Описание данных

**Описание данных:** В качестве примера был выбран средне-душевой доход руб/мес по областям России. Для примера исследуем распределение чисел средне-душевого дохода

1. Интервальный вариационный ряд



1. Выборочные характеристики

Мода: -

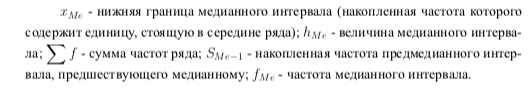


x0 – начало модального интервала; h – величина интервала; n2 –частота, соответствующая модальному интервалу; n1 – предмодальная частота; n3 – послемодальная частота.

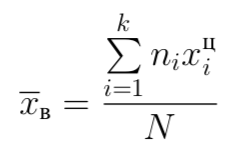
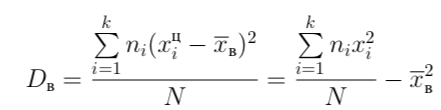
Медиана

|  |
| --- |
|  |

= 27318



Выборочное среднеe, выборочная дисперсия и выборочное среднеквадратическое отклонение для интервального ряда вычисляются по формулам:

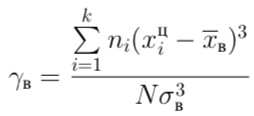
 = 28343,44302  
= 109346313,7

= 10456,8788

где ni — число элементов выборки, попавших в i-ый интервал, xцi — центр i-ого интервала, k — число интервалов.

**Коэффициент асимметрии** — это величина, определяющая степень несимметричности распределения случайной величины.

Выборочный коэффициент асимметрии вычисляется как отношение центрального выборочного момента 3-го порядка к кубу выборочного среднеквадратического отклонения, т.е.

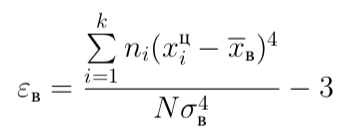
= 2,000283753

**Отрицательный коэффиициент** асимметрии показывает, что распределение более

плотное слева от центра. **Положительный** — что распределение более плотное справа.

**Коэффициент эксцесса** — это величина, которая характеризует степень остроты пика расределения случайной величины. Коэффициенты асимметрии и эксцесса для нормального распределения равны 0.

Выборочный коэффициент эксцесса определяется следующей формулой :

 = 3,943859001

Вычитание 3 требуется для того, чтобы коэффициент эксцесса нормального распре- деления был равен 0.

**Положительный коэффициент** эксцесса показывает, что пик распределения более острый по сравнению с пиком нормального распределения.  
**Отрицательный̆** — пик более сглаженный̆.

1. Проверка на нормальность распределения по критерию Пирсона

Для проверки гипотез о принадлежности выборки некоторому закону распределения воспользуемся критерием согласия Пирсона.

Данный критерий позволяет оценить значимость различий между наблюдаемыми данными и теоретическими ожиданиями, полученными в предположении о верности нулевой гипотезы, при заданном уровне значимости. Статистика критерия вычисляется по следующей формуле:



где mi = piN — теоретическая частота попаданий в i-ый интервал, ni — наблюдаемая частота, k — число интервалов.

Данная величина имеет распределение, близкое к распределению χ2 c (k – 1 степенями свободы.

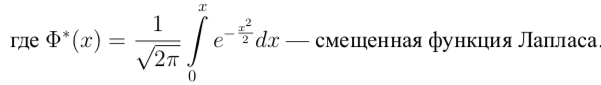
Вычислим теоретические вероятности попадания величин в заданные интервалы для трех видов распределений:

1. H0: распределение случайной величины ξ - нормальное

Вероятность попадания случайной величины, распределенной по нормальному

закону N (m, σ), в интервал (a, b) вычисляется по формуле :





Рассчитаем вероятности для каждого интервала по ф-ле

1)

0,1068

9,187461935

20,22976983

25,5406382

0,2149

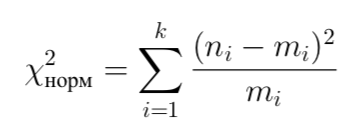
18,48919333

7,674491278

1,826529307

0,249258001

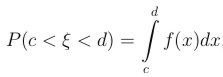
0,019503699

= 105,9699848

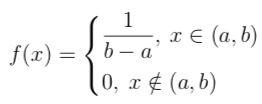
1. H0: распределение случайной величины ξ - равномерное

Вероятность попадания случайной величины, распределенной равномерно на

интервале (c, d), в интервал (a, b) вычисляется по формуле :



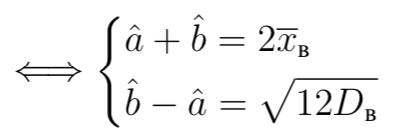
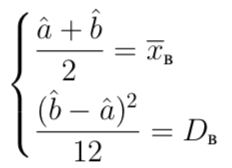
где f(x) — функция плотности вероятности равнормерного распределения

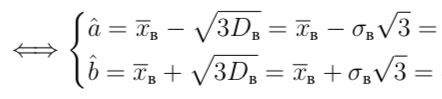


Оценки параметров равномерного распределения найдём методом моментов.

Метод моментов заключается в нахождении состоятельных оценок параметров распределения путем приравнивания теоретических начальных и центральных моментов случайной̆ величины к выборочным моментам. Теоретические значения математического ожидания и дисперсии равномерного распределения равны и

Их выборочными оценками являются выборочное среднее и несмещенная выборочная дисперсия. Получим систему:





a = 10231,5977

B = 46455,28835

Теперь мы можем вычислить точный вид функции плотности вероятности равномерного распределения для нашей выборки:

Вычислим вероятности для каждого из интервалов:

16,8

16,8

18,662

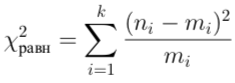
18,662

14,19

0

0

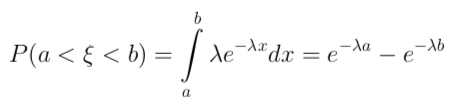
0

 = 56,4057961

1. H0: распределение случайной величины ξ является показательным

Вероятность попадания случайной величины, распределенной показательно, в

интервал (a, b) вычисляется по формуле :



По методу моментов найдем оценку параметра λ. Математическое ожидание

показательного распределения равно Тогда

 = 0.0000352815

Вычислим вероятности для каждого из интервалов:

39,1638792

0,13064725

11,2356639

0,09930588

8,54030525

0,07548308

6,49154464

0,05737521

4,93426764

0,04361129

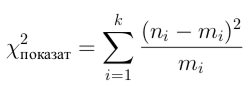
3,7505707

0,03314924

2,85083453

0,10503412

9,03293412

 = 152,58205

Мы получили оценку χ2 для рассматриваемых в данной работе распределений, теперь необходимо принять решение о распределении нашей выборки. Начнем с критерия согласия Пирсона. Для его проверки нам необходимо задать уровень значимости α - вероятность отвергнуть верную нулевую гипотезу, а также вычислить количество степеней свободы df = k − l − 1, где k - количество непересекающихся интервалов, l - количество оцененных по выборке параметров.

Определим уровень α = 0.05:

**Применим критерий к нормальному распределению:**

Рассчитаем df = 8−2−1 = 6

В табличных критических значений Пирсона при данном уровне значимости находим χ20.05 = 12.6 < 105,96 - гипотеза H0 отклоняется

**Применим критерий к равномерному распределению:**

Рассчитаем df = 8−2−1 = 6

В таблице распределений при данном уровне значимости находим χ20.05 = 12.6 < 56,4057961 - гипотеза H0 отвергается

**Применим критерий к показательному распределению:**

Рассчитаем df = 8−1−1 = 7

В таблице распределений при данном уровне значимости находим χ20.05 = 14.1 < 152,58205 - гипотеза H0 отвергается

Вывод : Поскольку значения статистики не попадает в область принятия, то согласно критерию 3 гипотезы отвергается с уровнем доверия 0,95. Значит, по критерию согласия Пирсона для нашей выборки точно не подходит ни одно распределение

**Проверим гипотезу о том, что Х распределено по *нормальному закону с помощью показателей As и Ex*:**  
В случае нормального распределения справедливо следующее условие:

|As| < 3SAs; |A| < 3SAs; |E| < 3SEx

Проверим выполнение этого условия для нашего примера:

SAs=0.58, SEx= 1,10 |2| > 3\*0.58 = 1,74 => Условия не выполняются

As= 2 , Ex= 3,9438 |3,94| > 3\*1,1= 3,3

Проверку выборочной совокупности на близость ее к нормальному распределению можно производить, используя статистики χ2, As и Ex.  
Сначала вычисляют статистику χ2 по формуле:

 = 24,97

Затем при заданном уровне значимости α и числе степеней свободы k = 2 (используют в расчетах две статистики As и Ex) для распределения χ2 Пирсона находят χкр2.

Если выполняется неравенство **χ2 < χкр2**, то гипотезу о нормальном распределении выборочной совокупности принимают.

В противном случае, т.е. когда **χ2>χкр2,** гипотезу о нормальном распределении выборки отвергают.

**Проверка гипотезы о том, что Х распределено по нормальному закону с помощью правила 3-х сигм:**

Если случайная величина распределена нормально, то абсолютная величина её отклонения от математического ожидания не превосходит утроенного среднеквадратического отклонения, т.е. все значения случайной величины должны попасть в интервал:

 => (28343,45 - 3\*10456,9 ; 57.9 + 3\*7.115) = (- 3027,25 ; 59514,15 )

Не Все значения величин попадают в интервал, так как xmin=13 336,80 ; xmax = 71 072,00

**Выводы**:  
Проверка гипотезы по критерию согласия Пирсона показала, что есть основания отвергать гипотезу о нормальном законе распределения.  
Значения As и Ex отличаются от нуля. Поэтому можно предположить , что нет близости данной выборки к нормальному распределению.

1. Разбиение распределения на малые выборки

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| с 71к по 34,6к | | с 34,1к по 27,6к | | с 27к по 25к |
| 71 072,00 | 34 080,70 | | 27 006,40 | |
| 63 035,90 | 33 312,60 | | 26 916,80 | |
| 59 567,00 | 32 692,20 | | 26 853,70 | |
| 57 333,40 | 32 407,70 | | 26 808,90 | |
| 48 852,50 | 31 391,30 | | 26 695,10 | |
| 48 734,20 | 31 375,90 | | 26 480,20 | |
| 44 519,70 | 30 922,40 | | 26 425,70 | |
| 41 116,50 | 30 849,60 | | 26 231,30 | |
| 39 844,70 | 30 448,10 | | 25 780,70 | |
| 39 494,10 | 30 141,10 | | 25 770,80 | |
| 38 395,90 | 29 703,90 | | 25 712,60 | |
| 37 857,30 | 29 130,50 | | 25 667,50 | |
| 37 677,30 | 27 907,10 | | 25 641,80 | |
| 35 952,30 | 27 814,50 | | 25 344,70 | |

Выборочное среднее:

Несмещенная оценка дисперсии(исправленная дисперсия):

Оценка среднеквадратического отклонения:

* 1. Проверим равенство дисперсий

Введем статистику F, которое имеет распределение близкое к распределению Фишера со степенями свободы:

14

14



Уровень значимости

.

Пусть:

Для 1 и 2 выборок:

* Гипотеза

Для 1 и 3 выборок:

* Гипотеза

Для 2 и 3 выборок: :

* Гипотеза

1. Сравнение средних двух больших выборок

|  |  |
| --- | --- |
| до 71к | до 27к |
| 71 072,00 | 27 006,40 |
| 63 035,90 | 26 916,80 |
| 59 567,00 | 26 853,70 |
| 57 333,40 | 26 808,90 |
| 48 852,50 | 26 695,10 |
| 48 734,20 | 26 480,20 |
| 44 519,70 | 26 425,70 |
| 41 116,50 | 26 231,30 |
| 39 844,70 | 25 780,70 |
| 39 494,10 | 25 770,80 |
| 38 395,90 | 25 712,60 |
| 37 857,30 | 25 667,50 |
| 37 677,30 | 25 641,80 |
| 35 952,30 | 25 344,70 |
| 34 696,40 | 25 082,50 |
| 34 080,70 | 25 078,30 |
| 33 312,60 | 25 042,80 |
| 32 692,20 | 24 931,10 |
| 32 407,70 | 24 464,00 |
| 31 391,30 | 24 118,10 |
| 31 375,90 | 24 074,20 |
| 30 922,40 | 24 067,70 |
| 30 849,60 | 24 014,20 |
| 30 448,10 | 23 793,50 |
| 30 141,10 | 23 786,90 |
| 29 703,90 | 23 579,00 |
| 29 130,50 | 23 205,10 |
| 27 907,10 | 23 079,80 |
| 27 814,50 | 23 026,00 |
| 27 629,60 | 22 818,90 |

Выборочное среднее:

38 598,55

25 049,94

Несмещенная оценка дисперсии(исправленная дисперсия):

Пусть , тогда введем статистику .

Критерий двусторонний:



Уровень значимости

***Вывод****:* Среднее двух выборок не равны. Следовательно , можно сделать вывод, о том, что и при больших обьемах Выборки распределения для каждого из чисел различнообразы, разнообразность для малых выборок нельзя обусловить размером выборок

1. Критерий медиан для трех выборок

Критерий не требует предварительных сведений о виде распределения.

Пусть

}

Медиана объединенной выборки:

30849,6

Наблюдаемые частоты

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Группа 1 | Группа 2 | Группа 3 |  | % |
|  | 0 | 8 | 15 | 23 | 51,11 |
|  | 15 | 7 | 0 | 22 | 48,89 |
|  | 15 | 15 | 15 | 45 | 100 |

Теоретические частоты

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Группа 1 | Группа 2 | Группа 3 |  | % |
|  | 8 | 8 | 8 | 24 | 53,33 |
|  | 7 | 7 | 7 | 21 | 46,67 |
|  | 15 | 15 | 15 | 45 | 100 |

Введем статистику

Уровень значимости 0,05

Из таблицы критических значений Пирсона:

Гипотеза отвергается

1. Одновариантный анализ дисперсии Краскэлла-Уэллиса

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| с 71к по 34,6к | ранг | с 34,1к по 27,6к | ранг | с 27к по 25к | ранг |
| 71072 | 45 | 34080,7 | 30 | 27006,4 | 15 |
| 63035,9 | 44 | 33312,6 | 29 | 26916,8 | 14 |
| 59567 | 43 | 32692,2 | 28 | 26853,7 | 13 |
| 57333,4 | 42 | 32407,7 | 27 | 26808,9 | 12 |
| 48852,5 | 41 | 31391,3 | 26 | 26695,1 | 11 |
| 48734,2 | 40 | 31375,9 | 25 | 26480,2 | 10 |
| 44519,7 | 39 | 30922,4 | 24 | 26425,7 | 9 |
| 41116,5 | 38 | 30849,6 | 23 | 26231,3 | 8 |
| 39844,7 | 37 | 30448,1 | 22 | 25780,7 | 7 |
| 39494,1 | 36 | 30141,1 | 21 | 25770,8 | 6 |
| 38395,9 | 35 | 29703,9 | 20 | 25712,6 | 5 |
| 37857,3 | 34 | 29130,5 | 19 | 25667,5 | 4 |
| 37677,3 | 33 | 27907,1 | 18 | 25641,8 | 3 |
| 35952,3 | 32 | 27814,5 | 17 | 25344,7 | 2 |
| 34696,4 | 31 | 27629,6 | 16 | 25082,5 | 1 |
| R1 | 570 | R2 | 345 | R3 | 120 |

Введем статистику:

Уровень значимости

Из таблицы критических значений Пирсона:

*>*5,99 => Гипотеза

1. Двухфакторный критерий Фридмана

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| с 71к по 34,6к | ранг | с 34,1к по 27,6к | ранг | с 27к по 25к | ранг |
| 71072 | 3 | 34080,7 | 2 | 27006,4 | 1 |
| 63035,9 | 3 | 33312,6 | 2 | 26916,8 | 1 |
| 59567 | 3 | 32692,2 | 2 | 26853,7 | 1 |
| 57333,4 | 3 | 32407,7 | 2 | 26808,9 | 1 |
| 48852,5 | 3 | 31391,3 | 2 | 26695,1 | 1 |
| 48734,2 | 3 | 31375,9 | 2 | 26480,2 | 1 |
| 44519,7 | 3 | 30922,4 | 2 | 26425,7 | 1 |
| 41116,5 | 3 | 30849,6 | 2 | 26231,3 | 1 |
| 39844,7 | 3 | 30448,1 | 2 | 25780,7 | 1 |
| 39494,1 | 3 | 30141,1 | 2 | 25770,8 | 1 |
| 38395,9 | 3 | 29703,9 | 2 | 25712,6 | 1 |
| 37857,3 | 3 | 29130,5 | 2 | 25667,5 | 1 |
| 37677,3 | 3 | 27907,1 | 2 | 25641,8 | 1 |
| 35952,3 | 3 | 27814,5 | 2 | 25344,7 | 1 |
| 34696,4 | 3 | 27629,6 | 2 | 25082,5 | 1 |
| R1 | 45 | R2 | 30 | R3 | 15 |

Введем статистику:

k = 3 – число групп

N = 60 – число данных в группе

Уровень значимости

Из табл критических значений Пирсона:

* Гипотеза

1. Вывод

Мы проверяем в медиане, красскелле и в фридмане то , что подвыборки взяты из одной генеральной совокупности

и если отвергаем гипотезы, значит они не принадлежат одной генеральной совокупности.