**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ**

**ФГБОУ ВО «Байкальский государственный университет»**

Кафедра математики и информатики

ИНДИВИДУАЛЬНАЯ РАБОТА

**по курсу «Теория вероятностей и математическая статистика»**

Уровень безработицы

Автор: Дмитриева Анастасия Дмитриевна, БИ-18-1

Руководитель:

к.т.н., доцент Т.И.Ведерникова

Иркутск

2020 год

Занятость - одна из важнейших социально-экономических проблем рыночной экономики. Ее статистическое отражение неоднократно обсуждалось на международных конференциях статистиков труда (1949, 1957, 1982, 1993 гг.), проводимых Международным бюро труда (г. Женева) - основным рабочим органом Международной организации труда (МОТ).

Безработица представляет собой сложное, многоаспектное социально-экономическое явление, когда часть экономически активного населения не занята в общественном производстве, не может реализовать свои физические и умственные способности при помощи рынка труда.

Данная работа должна быть полезна и актуальна в современном мире. Сейчас высшее образование не каждый может получить, a после окончания ВУЗа работу найти очень нелегко? Поэтому вопрос безработицы стоит в обществе остро.

Целью индивидуальной работы «Статистический анализ данных» является расширение и закрепление теоретических знаний и практических навыков по дисциплине «Теория вероятностей и математическая статистика».

Задача работы состоит в формировании вероятностно-статистического мышления, приобретении навыков грамотного использования стохастических моделей, технологий эффективной обработки временных рядов и адекватной интерпретации результатов.

Основным условием применимости вероятностно-статистической методологии является статистическая устойчивость.

В узком смысле под статистической устойчивостью понимается следующее: если в серии экспериментов проводятся наблюдения за некоторым событием, то вероятность его появления остается неизменной; если наблюдения ведутся за случайной величиной, то ее закон распределения вероятностей не изменяется в ходе наблюдений.

Вероятностно-статистическая методология применима к данной совокупности, так как вероятность распределения этих показателей не изменяется в ходе изучения.

Выборка – это совокупность независимых наблюдений за одним и тем же объектом или процессом.

**Исходные данные:**

Исходное множество – генеральная совокупность, а группа индивидуумов, наблюдаемых при наблюдении и повторении – выборка из генеральной совокупности.

|  |  |
| --- | --- |
| **Уровень безработицы** | **Дата** |
| 42,8 | янв.97 |
| 46,7 | фев.97 |
| 49,4 | мар.97 |
| 52,4 | апр.97 |
| 51 | май.97 |
| 49,9 | июн.97 |
| 48,6 | июл.97 |
| 47,7 | авг.97 |
| 46,3 | сен.97 |
| 45,7 | окт.97 |
| 46,7 | ноя.97 |
| 48,9 | дек.97 |
| 50,4 | янв.98 |
| 52,4 | фев.98 |
| 53,7 | мар.98 |
| 54,6 | апр.98 |
| 52,5 | май.98 |
| 49,8 | июн.98 |
| 46,7 | июл.98 |
| 43,6 | авг.98 |
| 40,3 | сен.98 |
| 37,7 | окт.98 |
| 35,9 | ноя.98 |
| 34,5 | дек.98 |
| 33,6 | янв.99 |
| 34,6 | фев.99 |
| 32,8 | мар.99 |
| 32,5 | апр.99 |
| 31,4 | май.99 |
| 39,1 | июн.99 |
| 30 | июл.99 |
| 39,3 | авг.99 |
| 39,2 | сен.99 |
| 59,3 | окт.99 |
| 31,2 | ноя.99 |
| 32,5 | дек.99 |
| 10,7 | янв.00 |
| 15,6 | фев.00 |
| 32,1 | мар.00 |
| 31,3 | апр.00 |
| 29,3 | май.00 |
| 27,1 | июн.00 |
| 26,4 | июл.00 |
| 25,2 | авг.00 |
| 20,5 | сен.00 |
| 23,6 | окт.00 |
| 24,1 | ноя.00 |
| 25,8 | дек.00 |
| 25,2 | янв.01 |
| 26 | фев.01 |
| 26,1 | мар.01 |
| 25,6 | апр.01 |
| 23,6 | май.01 |
| 22,4 | июн.01 |
| 21,9 | июл.01 |
| 21,8 | авг.01 |
| 21,1 | сен.01 |
| 21 | окт.01 |
| 22 | ноя.01 |
| 23 | дек.01 |
| 23,5 | янв.02 |
| 24,9 | фев.02 |
| 25,4 | мар.02 |
| 23,9 | апр.02 |
| 21,9 | май.02 |
| 20,2 | июн.02 |
| 20,3 | июл.02 |
| 20,9 | авг.02 |
| 19,9 | сен.02 |
| 20,2 | окт.02 |
| 21,4 | ноя.02 |
| 22,8 | дек.02 |
| 25,9 | янв.03 |
| 28,9 | фев.03 |
| 29,2 | мар.03 |
| 29,1 | апр.03 |
| 26,1 | май.03 |
| 23,5 | июн.03 |
| 23,4 | июл.03 |
| 23,2 | авг.03 |
| 22,1 | сен.03 |
| 22,6 | окт.03 |
| 23,9 | ноя.03 |
| 26,2 | дек.03 |
| 27,5 | янв.04 |
| 27,2 | фев.04 |
| 27,1 | мар.04 |
| 28,5 | апр.04 |
| 27,4 | май.04 |
| 26 | июн.04 |
| 26,1 | июл.04 |
| 26,4 | авг.04 |
| 25,3 | сен.04 |
| 25,8 | окт.04 |
| 26,6 | ноя.04 |
| 28,2 | дек.04 |
| 28 | янв.05 |
| 27,7 | фев.05 |
| 27,1 | мар.05 |
| 27,2 | апр.05 |
| 26,3 | май.05 |
| 25,7 | июн.05 |
| 26,3 | июл.05 |
| 26,3 | авг.05 |
| 25,1 | сен.05 |
| 25 | окт.05 |
| 26,8 | ноя.05 |
| 30,4 | дек.05 |
| 31 | янв.06 |
| 32,1 | фев.06 |
| 35 | мар.06 |
| 35,8 | апр.06 |
| 33,5 | май.06 |
| 32 | июн.06 |
| 31,7 | июл.06 |
| 30,1 | авг.06 |
| 27,2 | сен.06 |
| 26,6 | окт.06 |
| 29,8 | ноя.06 |
| 33,8 | дек.06 |
| 32,6 | янв.07 |
| 35,4 | фев.07 |
| 34,5 | мар.07 |
| 34,3 | апр.07 |
| 34,9 | май.07 |
| 32,7 | июн.07 |
| 32,1 | июл.07 |
| 30,5 | авг.07 |
| 28,2 | сен.07 |
| 27,7 | окт.07 |
| 30,6 | ноя.07 |
| 33,3 | дек.07 |
| 33,5 | янв.08 |
| 32,8 | фев.08 |
| 32,3 | мар.08 |
| 29,6 | апр.08 |
| 26,4 | май.08 |
| 23,9 | июн.08 |
| 22,7 | июл.08 |
| 21,5 | авг.08 |
| 20,8 | сен.08 |

Вариационный ряд – выборка, упорядоченная по возрастанию.

|  |  |  |
| --- | --- | --- |
|  | **Уровень безработицы** | **Дата** |
| 1 | 10,7 | янв.00 |
| 2 | 15,6 | фев.00 |
| 3 | 19,9 | сен.02 |
| 4 | 20,2 | июн.02 |
| 5 | 20,2 | окт.02 |
| 6 | 20,3 | июл.02 |
| 7 | 20,5 | сен.00 |
| 8 | 20,8 | сен.08 |
| 9 | 20,9 | авг.02 |
| 10 | 21 | окт.01 |
| 11 | 21,1 | сен.01 |
| 12 | 21,4 | ноя.02 |
| 13 | 21,5 | авг.08 |
| 14 | 21,8 | авг.01 |
| 15 | 21,9 | июл.01 |
| 16 | 21,9 | май.02 |
| 17 | 22 | ноя.01 |
| 18 | 22,1 | сен.03 |
| 19 | 22,4 | июн.01 |
| 20 | 22,6 | окт.03 |
| 21 | 22,7 | июл.08 |
| 22 | 22,8 | дек.02 |
| 23 | 23 | дек.01 |
| 24 | 23,2 | авг.03 |
| 25 | 23,4 | июл.03 |
| 26 | 23,5 | янв.02 |
| 27 | 23,5 | июн.03 |
| 28 | 23,6 | окт.00 |
| 29 | 23,6 | май.01 |
| 30 | 23,9 | апр.02 |
| 31 | 23,9 | ноя.03 |
| 32 | 23,9 | июн.08 |
| 33 | 24,1 | ноя.00 |
| 34 | 24,9 | фев.02 |
| 35 | 25 | окт.05 |
| 36 | 25,1 | сен.05 |
| 37 | 25,2 | авг.00 |
| 38 | 25,2 | янв.01 |
| 39 | 25,3 | сен.04 |
| 40 | 25,4 | мар.02 |
| 41 | 25,6 | апр.01 |
| 42 | 25,7 | июн.05 |
| 43 | 25,8 | дек.00 |
| 44 | 25,8 | окт.04 |
| 45 | 25,9 | янв.03 |
| 46 | 26 | фев.01 |
| 47 | 26 | июн.04 |
| 48 | 26,1 | мар.01 |
| 49 | 26,1 | май.03 |
| 50 | 26,1 | июл.04 |
| 51 | 26,2 | дек.03 |
| 52 | 26,3 | май.05 |
| 53 | 26,3 | июл.05 |
| 54 | 26,3 | авг.05 |
| 55 | 26,4 | июл.00 |
| 56 | 26,4 | авг.04 |
| 57 | 26,4 | май.08 |
| 58 | 26,6 | ноя.04 |
| 59 | 26,6 | окт.06 |
| 60 | 26,8 | ноя.05 |
| 61 | 27,1 | июн.00 |
| 62 | 27,1 | мар.04 |
| 63 | 27,1 | мар.05 |
| 64 | 27,2 | фев.04 |
| 65 | 27,2 | апр.05 |
| 66 | 27,2 | сен.06 |
| 67 | 27,4 | май.04 |
| 68 | 27,5 | янв.04 |
| 69 | 27,7 | фев.05 |
| 70 | 27,7 | окт.07 |
| 71 | 28 | янв.05 |
| 72 | 28,2 | дек.04 |
| 73 | 28,2 | сен.07 |
| 74 | 28,5 | апр.04 |
| 75 | 28,9 | фев.03 |
| 76 | 29,1 | апр.03 |
| 77 | 29,2 | мар.03 |
| 78 | 29,3 | май.00 |
| 79 | 29,6 | апр.08 |
| 80 | 29,8 | ноя.06 |
| 81 | 30 | июл.99 |
| 82 | 30,1 | авг.06 |
| 83 | 30,4 | дек.05 |
| 84 | 30,5 | авг.07 |
| 85 | 30,6 | ноя.07 |
| 86 | 31 | янв.06 |
| 87 | 31,2 | ноя.99 |
| 88 | 31,3 | апр.00 |
| 89 | 31,4 | май.99 |
| 90 | 31,7 | июл.06 |
| 91 | 32 | июн.06 |
| 92 | 32,1 | мар.00 |
| 93 | 32,1 | фев.06 |
| 94 | 32,1 | июл.07 |
| 95 | 32,3 | мар.08 |
| 96 | 32,5 | апр.99 |
| 97 | 32,5 | дек.99 |
| 98 | 32,6 | янв.07 |
| 99 | 32,7 | июн.07 |
| 100 | 32,8 | мар.99 |
| 101 | 32,8 | фев.08 |
| 102 | 33,3 | дек.07 |
| 103 | 33,5 | май.06 |
| 104 | 33,5 | янв.08 |
| 105 | 33,6 | янв.99 |
| 106 | 33,8 | дек.06 |
| 107 | 34,3 | апр.07 |
| 108 | 34,5 | дек.98 |
| 109 | 34,5 | мар.07 |
| 110 | 34,6 | фев.99 |
| 111 | 34,9 | май.07 |
| 112 | 35 | мар.06 |
| 113 | 35,4 | фев.07 |
| 114 | 35,8 | апр.06 |
| 115 | 35,9 | ноя.98 |
| 116 | 37,7 | окт.98 |
| 117 | 39,1 | июн.99 |
| 118 | 39,2 | сен.99 |
| 119 | 39,3 | авг.99 |
| 120 | 40,3 | сен.98 |
| 121 | 42,8 | янв.97 |
| 122 | 43,6 | авг.98 |
| 123 | 45,7 | окт.97 |
| 124 | 46,3 | сен.97 |
| 125 | 46,7 | фев.97 |
| 126 | 46,7 | ноя.97 |
| 127 | 46,7 | июл.98 |
| 128 | 47,7 | авг.97 |
| 129 | 48,6 | июл.97 |
| 130 | 48,9 | дек.97 |
| 131 | 49,4 | мар.97 |
| 132 | 49,8 | июн.98 |
| 133 | 49,9 | июн.97 |
| 134 | 50,4 | янв.98 |
| 135 | 51 | май.97 |
| 136 | 52,4 | апр.97 |
| 137 | 52,4 | фев.98 |
| 138 | 52,5 | май.98 |
| 139 | 53,7 | мар.98 |
| 140 | 54,6 | апр.98 |
| 141 | 59,3 | окт.99 |

Вариационный ряд представлен в более читабельном виде.

















Случайная величина – это переменная, которая в результате опыта  может принять то или иное числовое значение в зависимости от случайных обстоятельств, которые выпадают из круга рассматриваемых условий , но сопутствующих каждому испытанию.

Различают дискретные и непрерывные случайные величины.

Случайная величина, значениями которой является любое подмножество множества действительных чисел , называется *непрерывной*.

Множество значений непрерывной случайной величины несчетно.

Вариационные ряды бывают двух типов: интервальные и безынтервальными. В интервальном вариационном ряду частоты, характеризующие повторяемость вариант в выборке, распределяются по интервалам группировки.

Интервальный ряд строится для непрерывных рядов, так как данные по уровню безработицы вообще не являются конечным рядом, поэтому данная выборка будет являться непрерывной случайной величиной.

Теперь выборку нужно разбить на частичные интервалы. Сколько интервалов рассмотреть (если в условии не сказано) предполагается находить по формуле Стерджесса:

http://www.mathprofi.ru/s/intervalnyi_variacionnyi_ryad_clip_image014.gif, где http://www.mathprofi.ru/s/intervalnyi_variacionnyi_ryad_clip_image016.gif – десятичный логарифмот объёма выборки и http://www.mathprofi.ru/s/intervalnyi_variacionnyi_ryad_clip_image018.gif – оптимальное количество интервалов, при этом результат округляют до ближайшего левого целого значения.

K = 1+3,332\*LOG10(141) = 8,161198083

Возьмем k=9\*

\*Полученную по формуле Стерджесса величину округляют обычно до целого большего числа, поскольку количество групп не может быть дробным числом.

Далее нужно найти шаг, то есть величина интервала.

h=(Xмакс-Xмин)/K = (59,3-10,7)/9 = 5,5

Далее подсчитываем частоты по каждому интервалу.

**Примечание: если варианта попадает на «стык» интервалов, то её следует относить в правый интервал**.

Интервальный ряд:



**Гистограмма относительных частот** – это фигура, состоящая из прямоугольников, ширина которых равна длинам частичных интервалов, а высота – соответствующим относительным частотам. Гистограмма – столбиковая диаграмма.

Площадь гистограммы равна единице, и это статистический аналог [**функции плотности распределения**](http://www.mathprofi.ru/nepreryvnaya_sluchaynaya_velichina.html) непрерывной случайной величины. Построенный чертёж даёт наглядное и весьма точное представление о распределении безработицы по всей генеральной совокупности.

**Функция распределения выборки:**

Fx(x) = , где n – объем выборки (кол-во элементов выборки), – кол-во элементов выборки меньших x.



Эмпирическая функция распределения выборки имеет вид:

Fx

**Оценки Случайных величин**

Пусть θ – некоторый параметр распределения FX ( x, θ ) . Информация, необходимая для нахождения оценки θ\* неизвестного параметра θ, содержится в выборке Х1, …, Хn из данного распределения. Таким образом, возникает задача построения оценки θ\* параметра распределения как функции случайной выборки θ\* = θ\* (Х1, …, Хn). Заметим, что оценка параметра распределения является случайной величиной (статистикой). В результате проведения эксперимента (серии n независимых наблюдений) получают реализацию выборки – числа x1, x2, …, xn. При этом оценка θ\* принимает соответствующее числовое значение θ\*e= θ\* (x1, x2,…, xn), которое является приближенным значением неизвестного параметра θ . Оценки указанного типа называют точечными, их применение целесообразно при достаточно больших выборках.

В данной таблице представленные различные расчеты, которые потребуются в дальнейшем.



**Оценка математического ожидания:**

Выборочным средним называется среднее арифметическое значений признака выборочной совокупности.

Для различных значений x1, x2, x3, …, xn признак выборочной совокупности объема n имеем:

g3(x) = \*Σui(сред) \* ni = 30,76914894

**Оценка дисперсии:**

Выборочной дисперсией называется среднее арифметическое квадратов отклонений наблюдаемых значений признака от среднего значения.

Для различных значений x1, x2, x3, …, xn признак выборочной совокупности объема n имеем:

g6(x) = \*Σ(ui2(сред) \* ni) - g32(x) = 87,50452694

**Оценка медианы:**

Медиана (Mе) – это величина изучаемого признака, которая находится в середине упорядоченного вариационного ряда.

Всего элементов n=141, следовательно, (141/2 = 71) элемент в упорядоченном ряду под номером 71 будет иметь значение уровня безработицы соответствующего медиане.

Me = 28, т.е. за весь рассмотренный промежуток времени (выборки) средним значением уровня безработицы является 28.

**Оценка моды:**

Мода (Мо) – это величина изучаемого признака, которая в данной совокупности встречается наиболее часто.

Чтобы определить моду по гистограмме, надо выбрать самый высокий прямоугольник и найти среднее значение этого интервала. В нашем случае самая большая частота 50 на интервале 21,7 - 27,2. Среднее число этого интервала 24,45.

Значит, наиболее часто встречаемый уровень безработицы в выборке 24-25.

**Оценка коэффициента асимметрии:**

Коэффициент асимметрии в теории вероятностей — величина, характеризующая асимметрию распределения данной случайной величины.

Ac(x) = , по формуле µ3(x) = \*Σ(ui(сред) \* ni - (g3(x))³)

Ac(x) = 1,074

**Оценка коэффициента эксцесса:**

Коэффициент эксцесса (коэффициент островершинности) в теории вероятностей — мера остроты пика распределения случайной величины.

Кэк(x) = , по формуле µ4(x) = \*Σ(ui(сред) \* ni - (g3(x))4)

Кэк(x) = 0,634

**Оценка коэффициента вариации:**

Коэффициент вариации, также известный как относительное стандартное отклонение - это стандартная мера дисперсии распределения вероятностей или частотного распределения. Она часто выражается в процентах и определяется как отношение стандартного отклонения σ к среднему μ. КВ широко используются в аналитической химии для выражения точности и повторяемости анализа. Они также часто применяются в инженерии и физике, при проведении исследований по обеспечению качества. Кроме того, КВ используется экономистами и инвесторами в экономических моделях.

Он показывает степень изменчивости по отношению к среднему показателю выборки.

Кв = = 0,304018335

**Предположение о законе распределения генеральной совокупности**

Мое первое предположение было таким: для данной выборки закон распределения вероятностей будет **Нормальный** *(*распределение Гаусса или Гаусса**-**Лапласа*)*. Данный вывод я определила, исходя из следующих рассуждений:

* Медиана, мода и математическое ожидание примерно равны друг другу
* Коэффициенты эксцесса и асимметрии близки к нулю
* Гистограмма примерно соответствует графику нормального распределения

**Оценка параметров предполагаемого распределения**

Для того, чтобы оценить параметры предполагаемого распределения используют методы **точечного оценивания**. В данном случае я использовала *Метод моментов*, он заключается в приравнивании теоретических моментов к соответствующим эмпирическим моментам. Составляется система уравнений, где количество моментов равно количеству параметров.

**Проверка выдвинутой гипотезы о законе распределения вероятностей**

Я выдвинула гипотезу о нормльном распределении:

Проверим гипотезу о том, что CВ Х распределена по нормальному закону с помощью **критерия согласия Пирсона.**

,



где , а – оцененное по выборке значение вероятности для гипотетического распределения.



g40(x) =,

где — вероятность попадания в i-й интервал случайной величины, распределенной по гипотетическому закону.

Проверка гипотезы о теоретическом распределении отвечает на вопрос: можно ли считать расхождение между предполагаемыми теоретическим и эмпирическим распределениями случайным или несущественным, объясняемым случайностью попадания в выборку тех или иных объектов, или же это расхождение говорит о существенном расхождении между распределениями и их несоответствии.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| интервал | -∞÷16.2 | 16.2÷21.7 | 21.7÷27.2 | 27.2÷32.7 | 32.7÷38.2 | 38.2÷43.7 | 43.7÷49.2 | 49.2÷54.7 | 54.7÷∞ |
|  | 2 | 11 | 50 | 35 | 18 | 6 | 8 | 10 | 1 |
|  | 0,060 | 0,106 | 0,185 | 0,230 | 0,205 | 0,130 | 0,059 | 0,019 | 0,005 |
|  | 8,401 | 15,006 | 26,126 | 32,496 | 28,880 | 18,338 | 8,318 | 2,695 | 0,739 |
|  | -6,401 | -4,006 | 23,874 | 2,504 | -10,880 | -12,338 | -0,318 | 7,305 | 0,261 |
|  | 40,974 | 16,050 | 569,985 | 6,269 | 118,381 | 152,231 | 0,101 | 53,367 | 0,068 |
|  | 4,877 | 1,070 | 21,817 | 0,193 | 4,099 | 8,301 | 0,012 | 19,804 | 0,092 |

Для данной гипотезы g40(x)=60.265.

Для того, чтобы H0 принималась, необходимо чтобы значение статистики критерия g40(x) было меньше или равно 

Число степеней свободы  в нашем случае: 9 – 1 – 2 = 6, где r – это число интервалов, а k – количество параметров.

Назначим уровень значимости ε = 0,1. (*Сами берёте это значение*.).

Далее смотрим таблицу распределения хи-квадрат и находим критическое значение 10.645

**Вывод**: гипотеза отвергается, т.к. g40(x)=60.265 > 10.645 , и данные выборки не согласуются с выдвигаемой гипотезой.

Так как наблюдаемое значение критерия Пирсона(g40(x)) больше, чем значение критической точки, при любом уровне значимости, я отвергаю гипотезу о нормальном распределении с выбранными параметрами. Данные не согласуются, гипотеза не выполняется.

**Разбиение выборки на несколько частей**

В 1998 году в России был экономический кризис. Это событие значительно влияет на уровень безработицы – он может повышаться.

В связи с этим, я разбила выборку на две части время до 2000 года (кризис и время выхода из него) и после 2000 года.

**1 Часть выборки (после 2000 года)**

По аналогии я составила интервальный ряд:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| интервалы | 19.9 - 22 | 22 - 24.1 | 24.1 - 26.2 | 26.2 - 28.3 | 28.3 - 30.4 | 30.4 - 32.5 | 32.5 - 34.6 | 34.6 - 36.7 |  |
| кол-во | 13 | 15 | 14 | 21 | 7 | 9 | 9 | 4 | 92 |

И функцию распределения выборки:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **x<=26.1** | **19.9<x<= 22** | **22<x<=24.1** | **24.1 <x<=26.2** | **26.2<x<=28.3** | **28.3<x<=30.4** | **30.4<x<=32.5** | **32.5<x<=34.6** | **34.6<x<=36.7** |
| 0 | 13 | 28 | 42 | 63 | 70 | 79 | 88 | 92 |
| 0 | 0,141 | 0,304 | 0,457 | 0,685 | 0,761 | 0,859 | 0,957 | 1,000 |

Оценки СВ:

|  |  |
| --- | --- |
| **Мат. Ожидание g3** | 27,07671 |
| **Дисперсия g6** | 19,97598 |
| **Медиана** | 26,4 |
| **Мода** | 26,1 |
| **Коэф. Ассиметрии** | 0,253466 |
| **Коэф. Эксцесса** | -1,01876 |
| **Коэф. Вариации** | 0,165066 |

***Выдвижение предположения о нормальном распределении***

Предположительно закон распределения вероятностей будет *Нормальным.* Данный вывод я сделала, исходя из следующих рассуждений:

* Медиана, мода и математическое ожидание примерно равны
* Коэф. эксцесса и коэф. асимметрии близки к нулю

Для того, чтобы оценить параметры предполагаемого распределения я использовала *Метод моментов.*

**Проверка выдвинутой гипотезы о законе распределения вероятностей части выборки после 2000 года**

Я выдвинула гипотезу о нормальном распределении:

Находим оценку g40(x). Результаты вычислений приведены в таблице.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |  |
| интервалы | -ꚙ - 22 | 22 - 24.1 | 24.1 - 26.2 | 26.2 - 28.3 | 28.3 - 30.4 | 30.4 - 32.5 | 32.5 - 34.6 | 34.6 - ꚙ |  |
| кол-во | 13 | 15 | 14 | 21 | 7 | 9 | 9 | 4 |  |
| p | 0,128 | 0,125 | 0,170 | 0,186 | 0,164 | 0,116 | 0,066 | 0,031 | 1 |
|  |  |  |  |  |  |  |  |  |  |
| n\*p | 11,777 | 11,472 | 15,598 | 17,075 | 15,050 | 10,680 | 6,102 | 2,807 |  |
|  |  |  |  |  |  |  |  |  |  |
| |  | | --- | |  | | 1,223 | 3,528 | -1,598 | 3,925 | -8,050 | -1,680 | 2,898 | 1,193 |  |
|  |  |  |  |  |  |  |  |  |  |
| |  | | --- | |  | | 1,497 | 12,447 | 2,553 | 15,405 | 64,800 | 2,822 | 8,399 | 1,424 |  |
|  |  |  |  |  |  |  |  |  |  |
| |  | | --- | |  | | 0,127 | 1,085 | 0,164 | 0,902 | 4,306 | 0,264 | 1,377 | 0,507 | 8,732 |

Для данной гипотезы g40(x)= 8,732.

Для того, чтобы H0 принималась, необходимо чтобы значение статистики критерия g40(x) было меньше или равно 

Число степеней свободы  в нашем случае: 8 – 1 – 2 = 5, где r – это число интервалов, а k – количество параметров.

Далее смотрим таблицу распределения хи-квадрат и находим критическое значение 

**Вывод**: гипотеза подтверждается, т.к. g40(x)= 8,732 < .

Данные выборки согласуются с гипотезой о Нормальном распределении с параметрами:, по критерию Пирсона с уровнем значимости 0,1.

**Доверительные интервалы для оцениваемых параметров**

Задача интервального оценивания в нахождении границ доверительного интервала при заданной доверительной вероятности .

Доверительным называют интервал, который покрывает неизвестный параметр с заданной надёжностью.

*Оценка математического ожидания при неизвестной дисперсии*

Доверительный интервал математического ожидания имеет вид:

,

где  находится по таблицам распределения Стьюдента

 с числом степеней свободы  так, что коэффициент доверия .

Примем уровень значимости (вероятность ошибочно отвергнуть проверяемую гипотезу в случае, если она верна)  = 0,2;  = 1 -  = 0.8.

 находим по таблице распределения Стьюдента с числом степеней свободы n-1 = 8-1 = 7, где n – количество интервалов.  = 1.415.

,



27 – • 1,415 ≤ a ≤ 27 + • 1,415

27 – 1,7• 1,415 ≤ a ≤ 27 + 1,7• 1,415

27 – 2,4≤ a ≤ 27 + 2,4

24,6 ≤ a ≤ 29,4

Доверительный интервал для мат. ожидания [24,6; 29,4].

*Оценка дисперсии при неизвестном мат. ожидании*

Доверительный интервал дисперсии имеет вид:

,

где  и  находятся по таблицам χ2 (хи-квадрат) распределения с  степенями свободы с коэффициентом доверия *γ* так, что .

Примем как и ранее  = 0,2;  = 0,8.

 и  находим по таблице распределения χ2 с числом степеней свободы n-1 = 8-1 = 7, где n – количество интервалов.  = 3,822 и = 9,803.

,

14,28  36,63

Доверительный интервал для дисперсии [14,28; 36,63].

***Выдвижение предположения о равномерном распределении***

Предположительно закон распределения вероятностей будет *Равномерным.* Данный вывод я сделала, исходя из следующих рассуждений:

* Примерно одинаковые вероятности на всем рассматриваемом промежутке

Для того, чтобы оценить параметры предполагаемого распределения я использовала *Метод моментов.*

**Проверка выдвинутой гипотезы о законе распределения вероятностей части выборки после 2000 года**

Я выдвинула гипотезу о равномерном распределении:

Я изменила интервальный ряд, так как в 5 – 8 интервалах количество значений

< 10, то я объединила интервалы с рядом стоящими.

И тогда получается вместо 8 интервалов – 6.

Найдем предполагаемую плотность распределения:

0,064516129

Найдем теоретические частоты по формулам:

ni ' = n• •(xi+1 – xi) = 92• 0,064516129 •(xi+1 – xi)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| интервалы | 19.9 - 22 | 22 - 24.1 | 24.1 - 26.2 | 26.2 - 28.3 | 28.3 - 32.5 | 32.5 - 36.7 |  |
| кол-во (ni) | 13 | 15 | 14 | 21 | 16 | 13 | 92 |
| (хi) | 19,9 | 22 | 24,1 | 26,2 | 28,3 | 32,5 |  |
| (хi+1) | 22 | 24,1 | 26,2 | 28,3 | 32,5 | 36,7 |  |
| Теор. Частоты (ni ') | 12,46452 | 12,46452 | 12,464516 | 12,464516 | 24,929032 | 24,929032 |  |
|  | 0,023005 | 0,515758 | 0,1891538 | 5,8449509 | 3,1981834 | 5,7082766 | 15,479328 |

X² наблюдаемая = Σ = 15,479328

Для того, чтобы H0 принималась, необходимо чтобы значение статистики критерия g40(x) было меньше или равно 

Число степеней свободы  в нашем случае: 6 – 1 – 2 = 3, где r – это число интервалов, а k – количество параметров.

Назначим уровень значимости ε = 0,01. (*Сами берёте это значение*.).

Далее смотрим таблицу распределения хи-квадрат и находим критическое значение 10.645

**Вывод**: гипотеза отвергается, т.к. X² наблюдаемая =15,479328> 11,345, и данные выборки не согласуются с выдвигаемой гипотезой.

Так как наблюдаемое значение критерия Пирсона больше, чем значение критической точки, при любом уровне значимости, я отвергаю гипотезу о равномерном распределении с выбранными параметрами. Данные не согласуются, гипотеза не выполняется.

**2 Часть выборки (до 2000 года включительно)**

По аналогии я составила интервальный ряд:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| интервалы | 10,7 - 17,7 | 17,7 - 24,7 | 24,7 - 31,7 | 31,7 - 38,7 | 38,7 - 45,7 | 45,7 - 52,7 | 52,7 - 59,7 |
| кол-во | 2 | 3 | 9 | 9 | 6 | 16 | 3 |

И функцию распределения выборки:

Оценки СВ:

|  |  |
| --- | --- |
| **Мат. Ожидание g3** | 36,03611 |
| **Дисперсия g6** | 76,6398 |
| **Медиана** | 32,75 |
| **Мода** | 28,3 |
| **Коэф. Ассиметрии** | 0,93638 |
| **Коэф. Эксцесса** | -0,32817 |
| **Коэф. Вариации** | 0,242935 |

Предположительно закон распределения вероятностей будет *Нормальным.* Данный вывод я сделала, исходя из следующих рассуждений:

* Медиана, мода и математическое ожидание примерно равны
* Коэф. эксцесса и коэф. асимметрии близки к нулю

Для того, чтобы оценить параметры предполагаемого распределения я использовала *Метод моментов.*

**Проверка выдвинутой гипотезы о законе распределения вероятностей части выборки после 2000 года**

Я выдвинула гипотезу о нормальном распределении:

Находим оценку g40(x). Результаты вычислений приведены в таблице.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 |  |
| p | 0,037 | 0,083 | 0,167 | 0,232 | 0,226 | 0,152 | 0,072 | 1 |
|  |  |  |  |  |  |  |  |  |
| n\*p | 1,800 | 4,008 | 8,011 | 11,156 | 10,825 | 7,318 | 3,447 |  |
|  |  |  |  |  |  |  |  |  |
| |  | | --- | |  | | 0,200 | -1,008 | 0,989 | -2,156 | -4,825 | 8,682 | -0,447 |  |
|  |  |  |  |  |  |  |  |  |
| |  | | --- | |  | | 0,040 | 1,016 | 0,978 | 4,648 | 23,278 | 75,370 | 0,200 |  |
|  |  |  |  |  |  |  |  |  |
| |  | | --- | |  | | 0,022 | 0,253 | 0,122 | 0,417 | 2,150 | 10,299 | 0,058 | 13,321 |

Для данной гипотезы g40(x)= 13,321.

Для того, чтобы H0 принималась, необходимо чтобы значение статистики критерия g40(x) было меньше или равно 

Число степеней свободы  в нашем случае: 7 – 1 – 2 = 4, где r – это число интервалов, а k – количество параметров.

**Вывод**: гипотеза отвергается, т.к. g40(x)= 13,321>  , и данные выборки не согласуются с выдвигаемой гипотезой.

Так как наблюдаемое значение критерия Пирсона(g40(x)) больше, чем значение критической точки, при любом уровне значимости, я отвергаю гипотезу о нормальном распределении с выбранными параметрами. Данные не согласуются.

**Заключение**

В данной работе представлен разбор такой случайной величины, как уровень безработицы.

Сначала предполагалась рассмотреть все исходные данные как одна СВ, выдвинута предположение о нормальном распределении данной случайной величины, но оно было опровергнуто мною таким методом, как критерий Пирсона.

Далее была проведена аналитическая и историческая работа с выборкой, то есть определение каких-либо экономических событий, которые могли бы отразиться на исходных данных. Такое событие было найдено – это экономический кризис в 1998 году.

Было принято решение разбить исходные данные на 2 части: 1 часть – докризисное, кризисное время начиная от 1997 года до 2000 года включительно;

2 часть – послекризисное время, начиная от 2001 года.

И далее наблюдение продолжалось уже за 2 частью (послекризисное время).

Для 2 части (начиная от 2001 года) я выдвинула гипотезу о Нормальном распределении данной СВ, и она согласуется с ними, также вычислены доверительные интервалы для математического ожидания и дисперсии.