

Работу выполнил:

Бадмаев Валерий, э205

Исследование крестьянских хозяйств Новгорода в начале XX-го века

В самом начале работы проверим данные и создадим новый столбец “посевная площадь(общая)”, сложив надельную, купчую и арендованную площади. Импортируем отредактированную таблицу и устанавливаем необходимые для анализа библиотеки, предварительно настроив рабочую директорию.

Задание 1: Вычисление среднего значения дохода и его дисперсии

Начинаем работу с данными, записав за Income доходы районов. Используя функции `mean()` и `var()`, вычислим среднее значение дохода, а также его дисперсию. Получены результаты: Среднее – 635.888043, дисперсия – 104660.144374.

Задание 2: Построение гистограммы дохода

Далее воспользуемся функцией `hist()`, построим и выведем на экран гистограмму доходов в Новгороде:

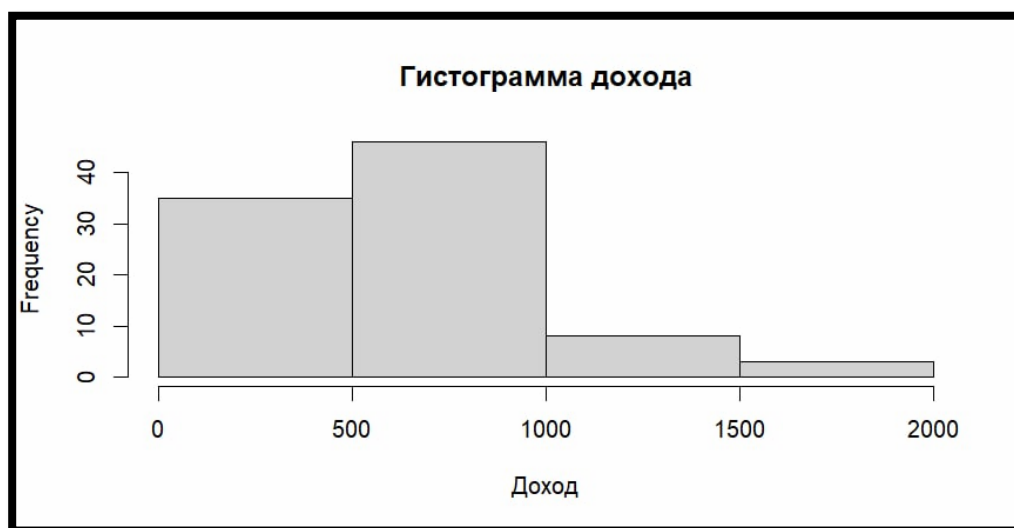


Рисунок 1. Гистограмма доходов в Новгороде

Задание 3: Поиск значений коэффициентов асимметрии и эксцесса дохода

Следующим шагом находим коэффициенты асимметрии и эксцесса, используя функции `skewness()` и `kurtosis()`:

1) $\hat{\alpha} = \frac{\hat{\mu}_3}{\hat{\mu}_2^{3/2}} = 1.569469$. Коэффициент асимметрии – величина, характеризующая асимметрию распределения случайной величины.

Так как коэффициент асимметрии положителен, то правый хвост распределения длиннее левого, и концентрация плотности смещена влево относительно математического ожидания.

2) $\hat{\gamma} = \frac{\hat{\mu}_4}{\hat{\mu}_2^2} - 3 = 6.215781$. Коэффициент эксцесса – мера остроты пика распределения случайной величины.

Так как коэффициент эксцесса больше нуля, то пик распределения около математического ожидания острый.

Задание 4: Проверка гипотезы о том, что доход распределен нормально

Проверяем гипотезу о том, что доход распределен нормально, применив функции `shapiro.test()`, `ks.test()`.

Для `Shapiro.test()` получен $p\text{-value} = 3.778e-07$

Для `ks.test()` получен $p\text{-value} = 0.02239$

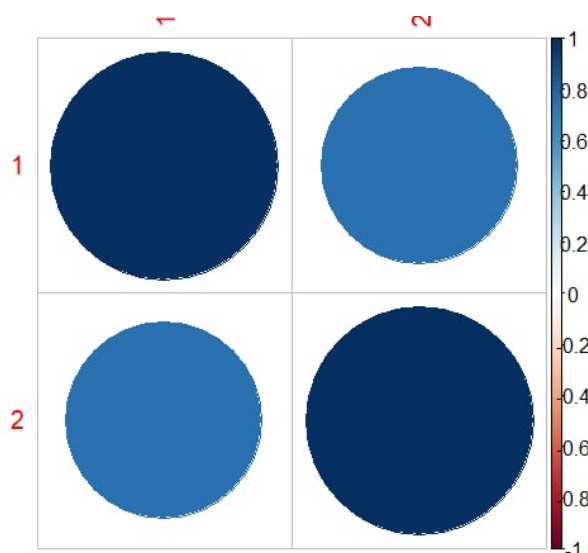
Выводы по полученным значениям $p\text{-value}$:

На уровне значимости 5% согласно результатам теста Шапиро мы отвергаем гипотезу, что доход распределен нормально, так как $0,05 > p\text{-value}$, полученный при тесте Шапиро; а на этом же уровне значимости согласно результатам теста Колмогорова-Смирнова мы отвергаем гипотезу о том, что доход распределен нормально, так как $0,05 > p\text{-value}$, полученный при проведении теста Колмогорова-Смирнова.

Задание 5: Поиск коэффициента корреляции дохода и посевной площади. Проверка его на значимость

Перейдем к вычислению корреляции. Для этого воспользуемся функцией `cor()`. Получено значение 0.74087

Для наглядного представления о корреляции дохода и расхода воспользуемся функцией `corrplot()`, получаем:



Для того, чтобы проверить корреляцию на значимость, используем `cor.test()`, получаем $p\text{-value} < 2.2e-16$. Так как $0,05 > p\text{-value}$, то на 5% уровне значимости мы отвергаем гипотезу о том, что корреляция **не значима**

Задание 6: Проверка гипотезы о равенстве дисперсий дохода и расхода

Далее проверим гипотезу о равенстве дисперсий дохода и расхода. Прибегнем к функции `var.test()`. Результатом выполнения функции стало $p\text{-value} = 0.5713$, из чего можем сделать вывод о том, что на 5% процентном уровне значимости мы принимаем гипотезу о том, что дисперсии доходов и расходов равны.

Задание 7: Построение доверительного интервала для средней стоимости скота в предположении, что стоимость скота распределена показательно

Построим доверительный интервал для средней стоимости скота в предположении, что стоимость скота распределена показательно. Сделаем это двумя способами:

1 способ – точный ДИ:

Чтобы построить точный доверительный интервал воспользуемся связью показательного распределения и распределения хи-квадрат, получаем: интервал $48.96721 < 1/\lambda < 73.76796$.

2 способ – асимптотический ДИ:

Чтобы построить асимптотический доверительный интервал, мы использовали тот факт, что выборочное среднее в асимптотике распределено нормально. Имеем результат: $49.37756 < 1/\lambda < 74.73974$.