Работу выполнил:

Бадмаев Валерий, э205

Исследование крестьянских хозяйств Средней Азии в начале XX-го века

В самом начале работы проверим данные, удалим лишние строки, не относящиеся к таблице. 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 уездов. Используя функции mean() и var(), вычислим среднее значение дохода, а также его дисперсию. Получены результаты: Среднее – 758.926887, дисперсия – 430690.803008.

Задание 2: Построение гистограммы дохода

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

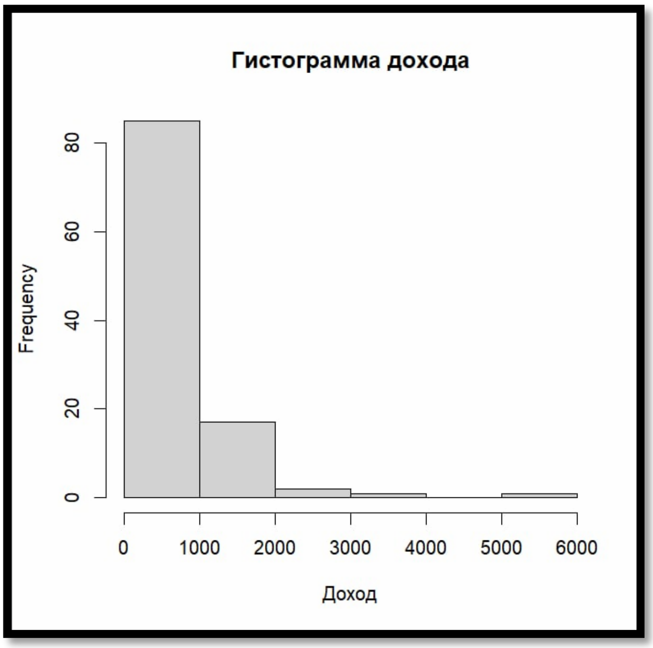


Рисунок 1. Гистограмма доходов в Средней Азии

Задание 3: Поиск значений коэффициентов асимметрии и эксцесса дохода

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

1) $\hat{\alpha} = \frac{\hat{\mu_3}}{\hat{\mu_2}^{\frac{3}{2}}}$ = 3.685155. Коэффициент асимметрии – величина, характеризирующая асимметрию распределения случайной величины.

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

2) $\hat{\varkappa} = \frac{\widehat{\mu_4}}{\widehat{\mu_2}^2} - 3$ = 22.578096. 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-value = 2.323e-14

Для ks.test() получен p-value = 0.0004466

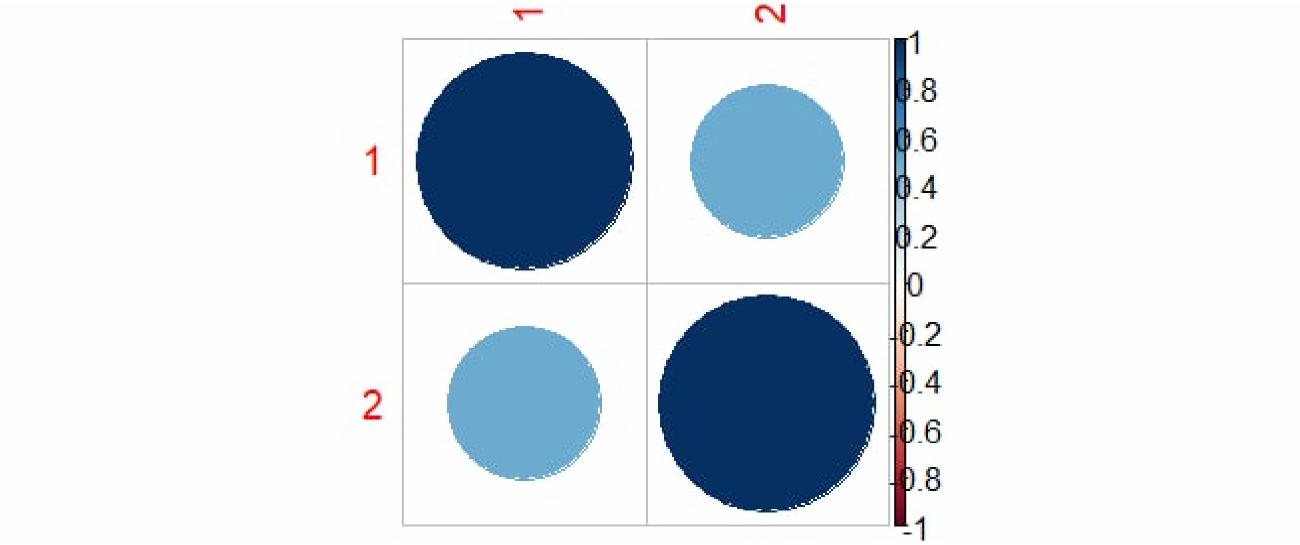
Выводы по полученным значениям p-value:

На уровне значимости 5% согласно результатам теста Шапиро мы отвергаем гипотезу, что доход распределен нормально, так как 0,05 > p-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-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 0.498748

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



Для того, чтобы проверить корреляцию на значимость, используем cor.test(), получаем p-value, равный 5.293e-08. Так как 0,05 > p-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-value < 2.2e-16, из чего можем сделать вывод о том, что на 5% процентном уровне значимости мы отвергаем 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 $307.617322 < 1/\lambda < 450.553654$.

2 способ – асимптотический ДИ:

Чтобы построить асимптотический доверительный интервал, мы использовали тот факт, что выборочное среднее в асимптотике распределено нормально. Имеем результат: $309.884213 < 1/\lambda < 455.610274$.