

Работу выполнил:

Бадмаев Валерий, э205

Исследование крестьянских хозяйств Амурской области 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 экономических агентов. Используя функции `mean()` и `var()`, вычислим среднее значение дохода, а также его дисперсию. Получены результаты: Среднее - 1912.269, дисперсия – 1171135.

Задание 2: Построение гистограммы дохода

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

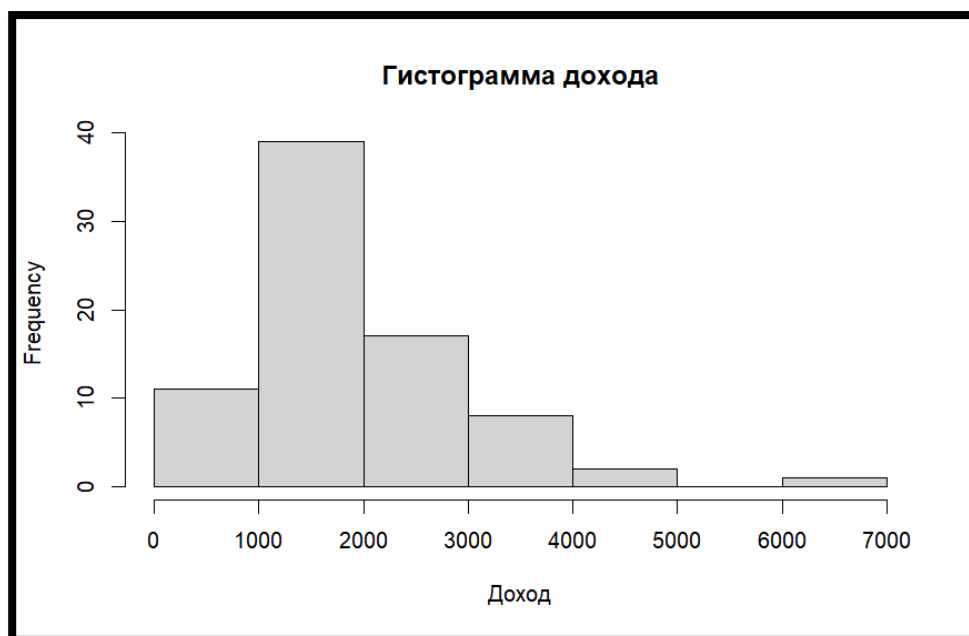


Рисунок 1. Гистограмма доходов в Амурской области

Задание 3: Поиск значений коэффициентов асимметрии и эксцесса дохода

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

1) $\hat{\alpha} = \frac{\hat{\mu}_3}{\hat{\mu}_2^{3/2}} = 1.395323$. 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) $\hat{\gamma} = \frac{\hat{\mu}_4}{\hat{\mu}_2^2} - 3 = 5.970007$. 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 $p\text{-value} = 1.955e-05$

Для `ks.test()` получен $p\text{-value} = 0.08093$

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

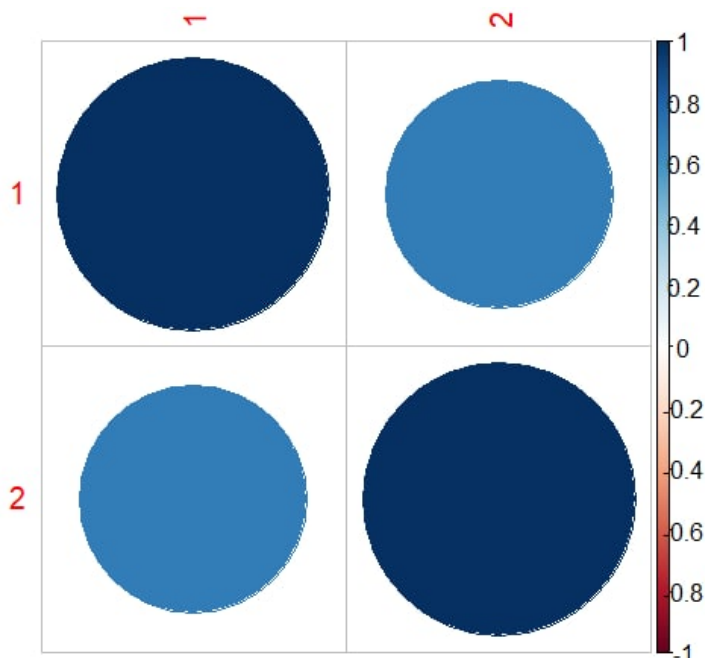
На уровне значимости 5% согласно результатам теста Шапиро мы отвергаем гипотезу, что доход распределен нормально, так как $0,05 > p\text{-value}$, полученный при тесте Шапиро; а на этом же уровне значимости согласно результатам теста Колмогорова-Смирнова мы принимаем гипотезу о том, что доход распределен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

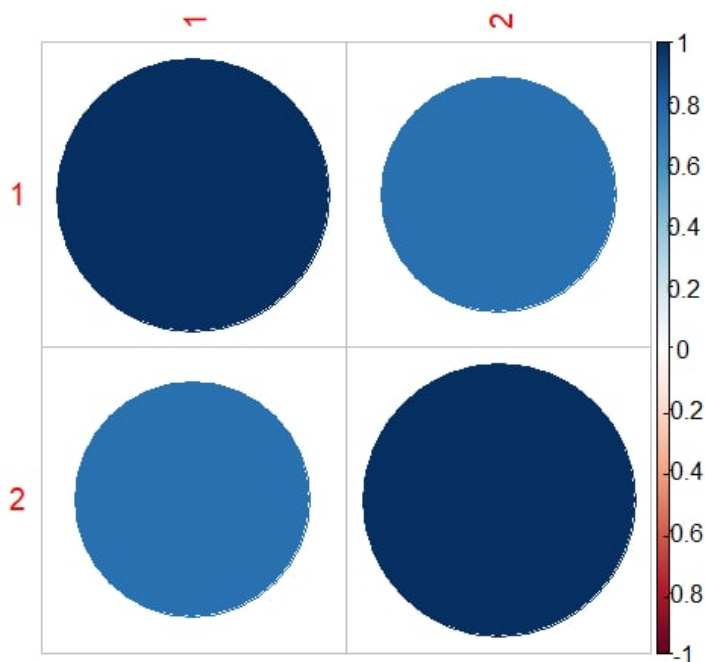
Перейдем к корреляции. В процессе вычисления корреляции дохода и посевной площади столкнулись с проблемой отсутствия данных по посевной площади у одного из субъектов. Было принято решение найти корреляцию двумя способами:

1 способ заключался в том, что мы заменили неизвестное значение посевной площади на ноль и посчитали таким образом корреляцию, равную 0.6982993.

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



2 способ предполагал собой исключение из данных о субъекте с неизвестной посевной площади. Получена корреляция, равная 0.7408966.



Заметим, что при использовании функции `cor.test()` мы получали корреляцию, которая совпадала с результатом вычисления из второго способа.

С помощью функции `cor.test()` мы получили p-value, равное $1.188e-12$, из чего делаем вывод, что на 5% уровне значимости можно утверждать, что корреляция между доходами и расходами 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 = 0.6986, 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 $782.585341 < 1/\lambda < 1221.515979$.

2 способ – асимптотический ДИ:

Чтобы построить асимптотический доверительный интервал, мы использовали тот факт, что выборочное среднее в асимптотике распределено нормально. Имеем результат: $790.195706 < 1/\lambda < 1240.952663$.