

Задание 5

№1

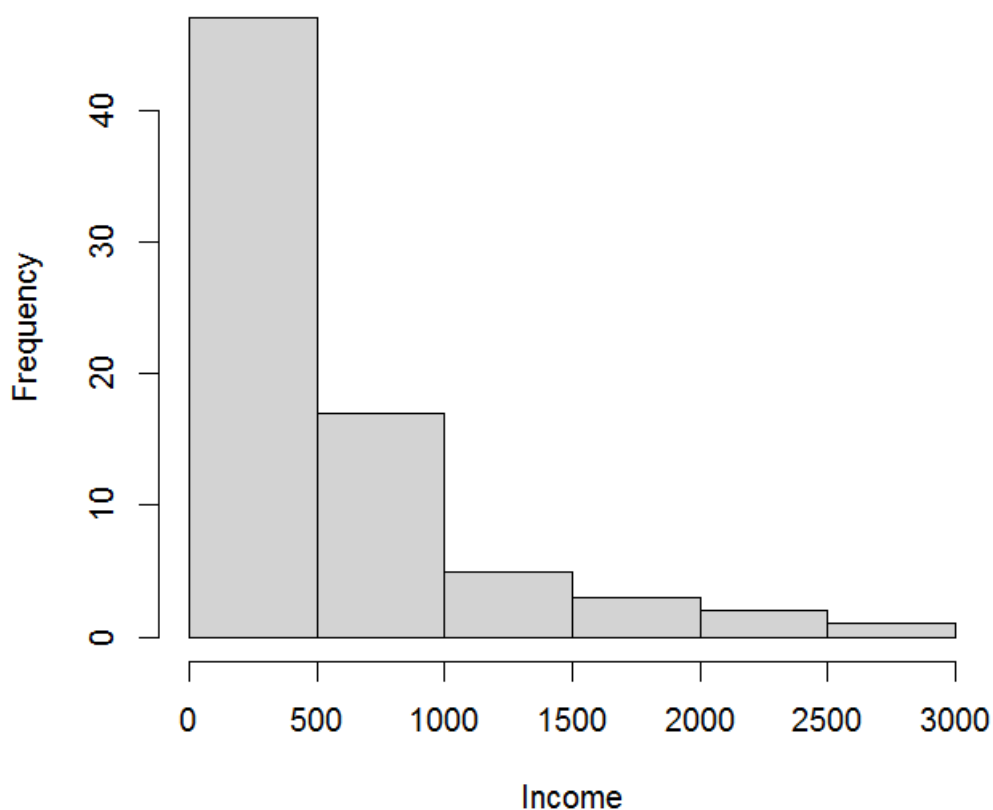
Для начала удалим пустую строку из файла эксель. Будем считать, что общий доход является суммой валового дохода от земледелия и полеводства и дохода от скотоводства и птицы.

В столбце валового дохода заменим пропуски на нули, ни среднее ни дисперсия не изменятся.

Среднее значение дохода – 586.59, дисперсия – 313737.4.

№2

Histogram of Income



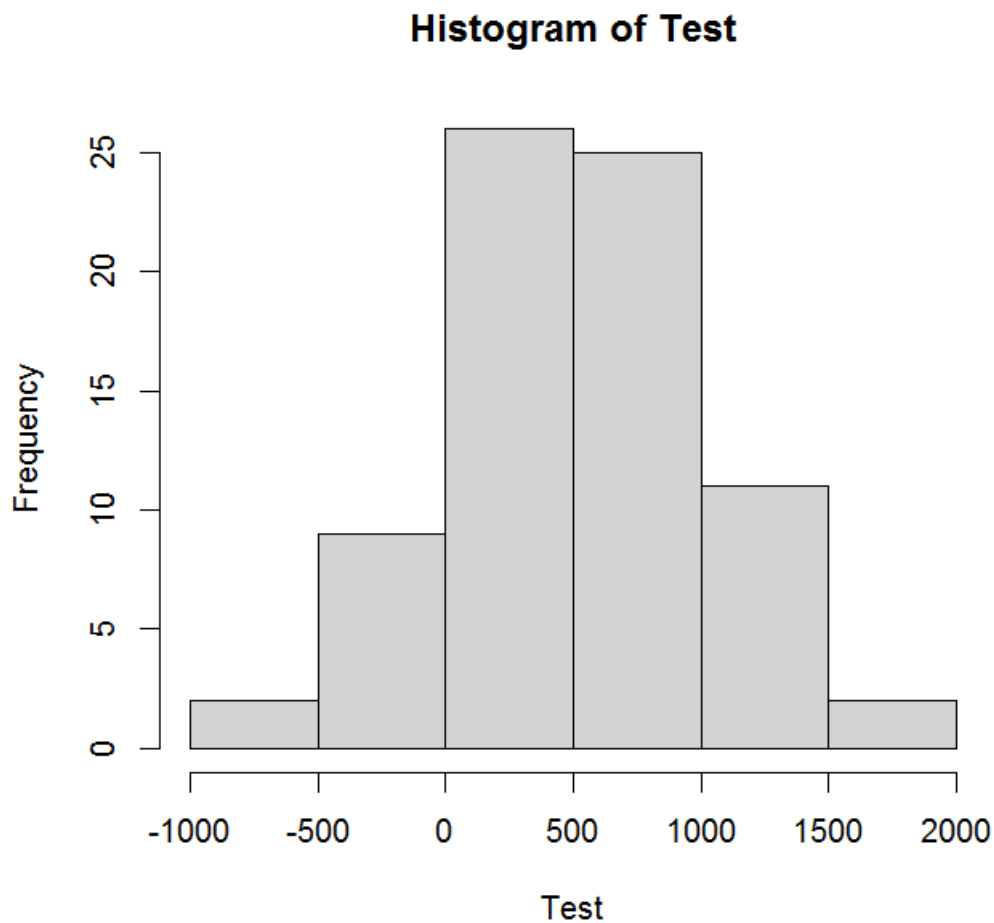
№3

Асимметрия – 2.123, эксцесс – 7.85

Асимметрия характеризует меру скошенности, что заметно на гистограмме – распределение скошено вправо. Поэтому асимметрия больше нуля. В случае, если наблюдается скошенность влево, асимметрия будет отрицательной. Чем больше по модулю асимметрия, тем более значительна скошенность в соответствующую сторону.

Эксцесс характеризует высоту или остроконечность распределения. Чем больше эксцесс, тем выше верхняя точка распределения и наоборот.

Для сравнения сгенерируем случайную выборку с тем же числом элементов и с нашими найденными средним значением и дисперсией.



Асимметрия – 0.123, эксцесс – 2.68

Сравнив частоту наиболее часто встречающихся величин заметим, что у исходной выборки эксцесс действительно больше, и наблюдается правосторонняя асимметрия, в то время как у сгенерированной выборки распределение практически симметрично.

№4

Проводим тесты Шапиро-Уилка и Колмогорова-Смирнова, получая следующие p-value: 1.161e-09 и 0.00075 соответственно. Следовательно, поскольку оба значения меньше 0.05, мы не можем принять гипотезу о нормальности распределения дохода.

№5

За посевную площадь будем считать сумму площадей долевых и арендованных полей.

Распределение дохода не нормально. Критерий Пирсона основывается на нормальности данных, поэтому им мы пользоваться не будем.

Согласно методу Спирмена, корреляция равна 0.812, p-value = 2.2e-16

Согласно методу Кендалла, корреляция равна 0.646, p-value = 7.089e-16

Нулевая гипотеза – равенство корреляции нулю, однако оба значения p-value крайне малы, поэтому мы отвергаем нулевую гипотезу о равенстве корреляции дохода и посевной площади нулю.

№6

Будем считать, что суммарные расходы складываются из расходов на личные и хозяйственные потребности.

Предположим, что исследуемые данные распределены нормально (необходимо для var.test). Получим p-value = 0.001194 < 0.05, значит нужно отвергнуть нулевую гипотезу о равенстве дисперсий. Дисперсии доходов и расходов не равны.

№7

Удалим из вектора стоимости скота ячейку, где нет значения (сам вектор является соответствующим столбцом из экселя).

Воспользуемся следующим преобразованием, предполагая показательность распределения:

$$\begin{aligned}\xi &\sim \exp(\lambda) \\ 2\lambda n\bar{x} &\sim \chi_{2n}^2 \\ \chi_{1-\frac{\alpha}{2}, 2n}^2 &< 2\lambda n\bar{x} < \chi_{\frac{\alpha}{2}, 2n}^2 \\ \frac{2n\bar{x}}{\chi_{\frac{\alpha}{2}, 2n}^2} &< \frac{1}{\lambda} < \frac{2n\bar{x}}{\chi_{1-\frac{\alpha}{2}, 2n}^2}\end{aligned}$$

Нам необходимо найти лишь значения хи-квадрат в критических точках и посчитать среднее по стоимости скота – 382.

Подставив значения получаем доверительный интервал для $1/\lambda$, т.е. для мат. ожидания показательного распределения. Асимптотический доверительный интервал равен (307.98, 486.49).