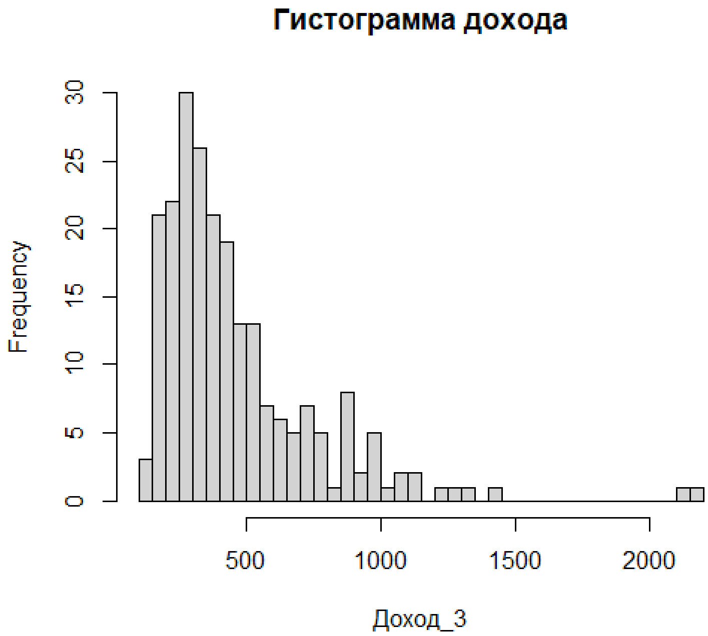
Задание 3

1. Рассчитываем среднее значение и дисперсию дохода с помощью функций mean() и var().

   Среднее значение дохода = 464,242; дисперсия дохода = 90615,324.

2. 

3. Рассчитываем асимметрию и эксцесс с помощью функций skewness() и kurtosis(). Значение асимметрии = 2,289; значение эксцесса = 11,09997.

   Асимметрия больше нуля, следовательно, распределение смещено вправо, имеет длинный хвост справа.

   Эксцесс положителен, что означает, что выбросы в данных интенсивнее, чем для нормального распределения. Эмпирическое распределение является более островершинным относительно эталонного нормального распределения с параметрами $a=\bar{x}$ и $\sigma=S$. И чем больше эксцесс по модулю, тем «аномальнее» высота в ту или иную сторону. В нашем случае эксцесс положительный.

4. Тест на нормальность дохода: используем shapiro.test() и ks.test() (Тест Колмогорова-Смирнова проводим, сравнивая нашу выборку с выборкой из нормальной генеральной совокупности с параметрами $a=\bar{x}$ и $\sigma=S$).

   В первом случае получаем p-value = 2,562e-16, во втором: p-value = 2.1e-05. p-value достаточно маленькие (p-value < 0,05), следовательно, отвергаем нулевую гипотезу о том, что значения дохода распределены нормально, на уровне значимости 5% и выше.

5. Корреляция дохода и посевной площади, её значимость: проводим тесты на корреляцию методами spearman, kendall (метод pearson не используем, так как в предыдущем пункте опровергли гипотезу о нормальности распределения дохода). Получаем p-value < 2.2e-16; p-value < 2.2e-16 соответственно. Значения p-value достаточно маленькие, следовательно, отвергаем нулевую гипотезу, и корреляция значимо отделена от 0.

   Оценки корреляции: 0,7516853, 0,5514624; имеем достаточно высокую корреляцию между доходом и посевной площадью.
6. Гипотеза о равенстве дисперсий дохода и расхода: проводим var.test(). (предположим, что данные распределены нормально). p-value = 0,001524 (p-value < 0,05), следовательно, отвергаем нулевую гипотезу и предполагаем, что истинные дисперсии не равны, на уровне значимости 5% и выше.
7. Доверительный интервал для средней стоимости скота:

   Применяем функцию eexp(), получаем доверительный интервал для параметра показательного распределения. Чтобы найти доверительный интервал для мат. ожидания, которое равно 1/λ, поделим 1 на границы найденного доверительного интервала и поменяем эти значения местами (1/UCL<1/LCL). Получаем: (354; 461).

   Результат совпадает с доверительным интервалом уровня значимости 5%, полученным следующим преобразованием (предполагаем, что стоимость скота имеет показательное распределение):

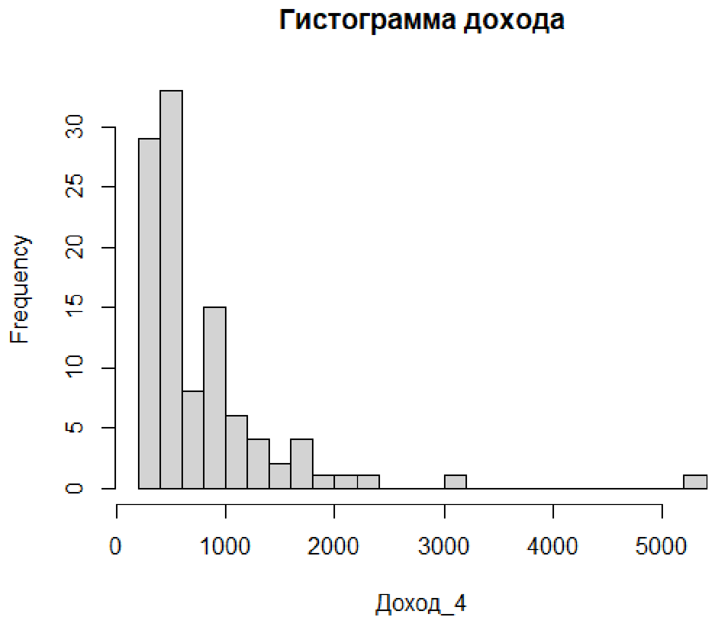
$$\xi \sim exp(\lambda)$$
$$2\lambda n\bar{x} \sim \chi^2_{2n}$$
$$\chi^2_{1-\frac{\alpha}{2},2n} < 2\lambda n\bar{x} < \chi^2_{\frac{\alpha}{2},2n}$$
$$\frac{2n\bar{x}}{\chi^2_{\frac{\alpha}{2},2n}} < \frac{1}{\lambda} < \frac{2n\bar{x}}{\chi^2_{1-\frac{\alpha}{2},2n}}$$

Задание 4

1. Рассчитываем среднее значение и дисперсию дохода с помощью функций mean() и var().

   Среднее значение дохода = 758.927; дисперсия дохода = 430690.803.

2.



3. Рассчитываем асимметрию и эксцесс с помощью функций skewness() и kurtosis(). Значение асимметрии = 3,685; значение эксцесса = 22,578.

   Асимметрия больше нуля, следовательно, распределение смещено вправо, имеет длинный хвост справа.

   Эксцесс положителен, что означает, что выбросы в данных интенсивнее, чем для нормального распределения. Эмпирическое распределение является более островершинным относительно эталонного нормального распределения с параметрами $a=\bar{x}$ и $\sigma=S$. И чем больше эксцесс по модулю, тем «аномальнее» высота в ту или иную сторону. В нашем случае эксцесс положительный.

4. Тест на нормальность дохода: используем shapiro.test() и ks.test() (Тест Колмогорова-Смирнова проводим, сравнивая нашу выборку с выборкой из нормальной генеральной совокупности с параметрами $a=\bar{x}$ и $\sigma=S$).

   В первом случае получаем p-value = 2.323e-14, во втором: p-value = 0.0004466. p-value достаточно маленькие (p-value < 0,05), следовательно, отвергаем нулевую гипотезу о том, что значения дохода распределены нормально, на уровне значимости 5% и выше.

5. Корреляция дохода и посевной площади, её значимость: проводим тесты на корреляцию методами spearman, kendall (метод pearson не используем, так как в предыдущем пункте опровергли гипотезу о нормальности распределения дохода). Получаем p-value = 4.886e-12; p-value = 9.74e-11 соответственно. Значения p-value достаточно маленькие, следовательно, отвергаем нулевую гипотезу, и корреляция значимо отделена от 0.
   Оценки корреляции: 0,6077856; 0,4290236.

   Примечание: посевной площадью считаем сумму площадей под озимую пшеницу и под яровую пшеницу, так как посевной площадью называется часть пахотных земель, занятых посевами сельскохозяйственных культур. Нам известна площадь общая площадь пашни, однако, имеем недостаточно данных, чтобы определить, засевается вся эта общая площадь или нет.

6. Гипотеза о равенстве дисперсий дохода и расхода: проводим var.test() (предположим, что данные распределены нормально). p-value < 2.2e-16 (p-value<0,05), следовательно, отвергаем нулевую гипотезу и предполагаем, что истинные дисперсии не равны на уровне значимости 5% и выше.
7. Доверительный интервал для средней стоимости скота:

   Применяем функцию eexp(), получаем доверительный интервал для параметра показательного распределения. Чтобы найти доверительный интервал для мат. ожидания, которое равно 1/λ, поделим 1 на границы найденного доверительного интервала и поменяем эти значения местами (1/UCL<1/LCL). Получаем: (307; 451).

   Результат совпадает с доверительным интервалом уровня значимости 5%, полученным следующим преобразованием (предполагаем, что стоимость скота имеет показательное распределение):

$$\xi \sim exp(\lambda)$$
$$2\lambda n\bar{x} \sim \chi^2_{2n}$$
$$\chi^2_{1-\frac{\alpha}{2},2n} < 2\lambda n\bar{x} < \chi^2_{\frac{\alpha}{2},2n}$$
$$\frac{2n\bar{x}}{\chi^2_{\frac{\alpha}{2},2n}} < \frac{1}{\lambda} < \frac{2n\bar{x}}{\chi^2_{1-\frac{\alpha}{2},2n}}$$