МИНИСТЕРТСВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования

**«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»**

**(ФГБОУ ВО «КУБГУ»)**

**Факультет компьютерных технологий и прикладной математики**

**Кафедра вычислительных технологий**

**ОТЧЕТ**

**о выполнении лабораторной работы №7**

по дисциплине «Обработка больших данных»

Работу выполнил \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ студент 39/2 группы, Мандыч Д.И.

Работу проверил\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_преподаватель Шиян В.И.

Краснодар 2023

Исходный код: <https://github.com/SnowLukin/BigData/tree/main>

**Набор данных**

Краткая информация о наборе данных:

Набор данных о современных Олимпийских играх, включая все игры от Афин 1896 года до Рио-2016. Эти данные собраны из www.sports-reference.com в мае 2018 года.

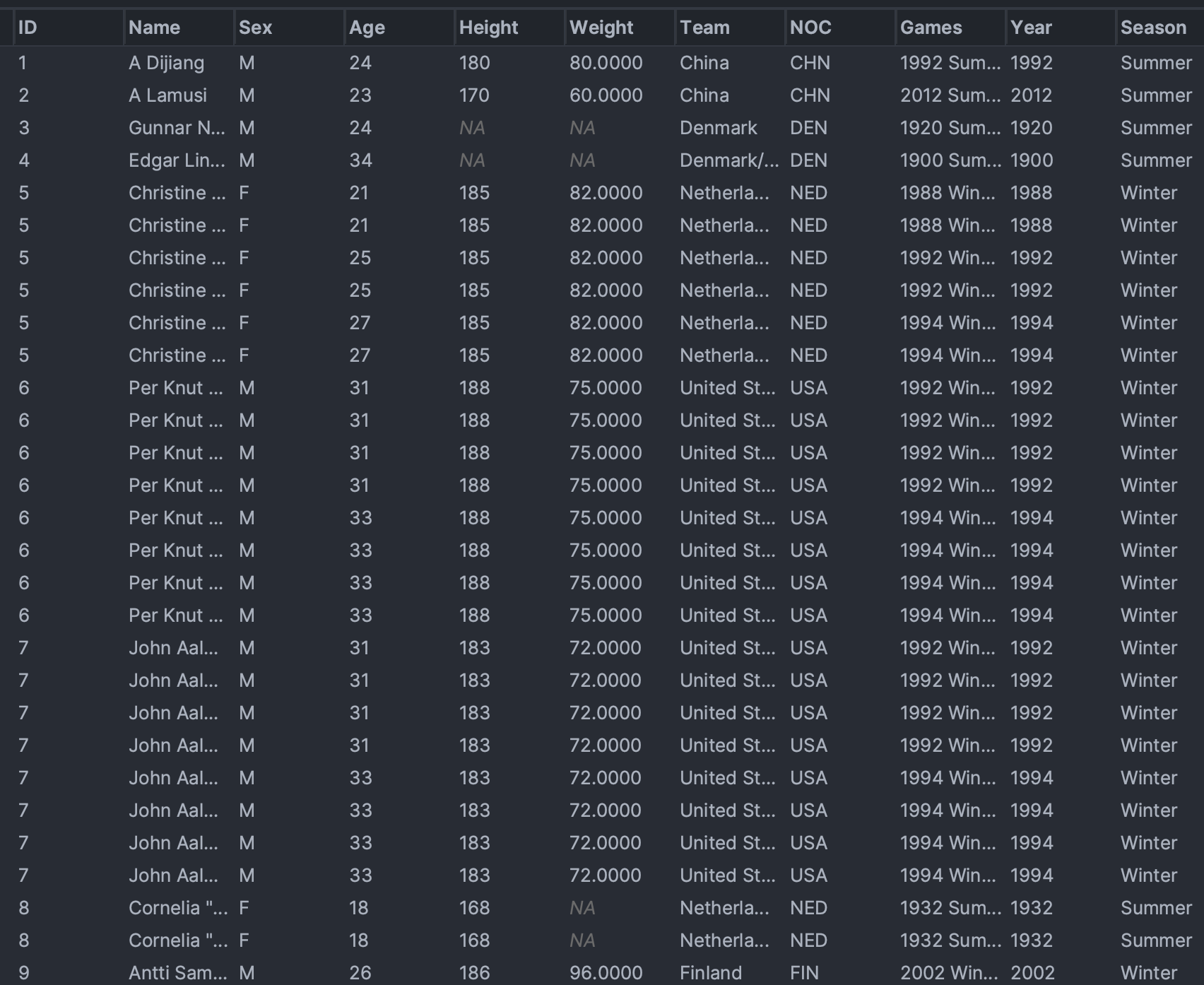


Рисунок 1 – Используемый набор данных

**Анализ гипотез**

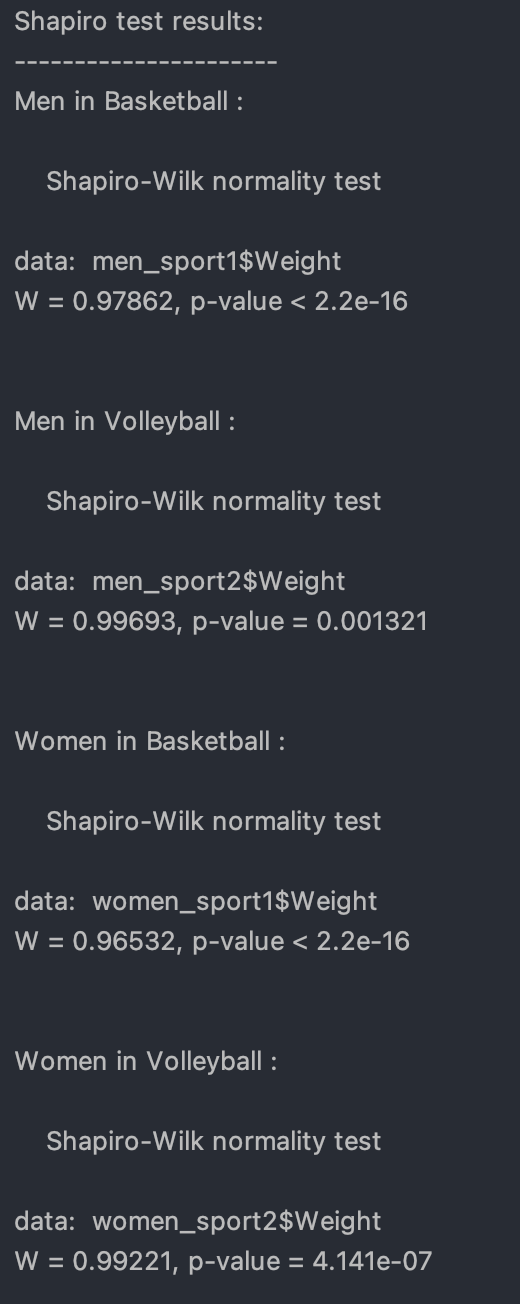


Рисунок 1 – Проверка выборки на нормальность

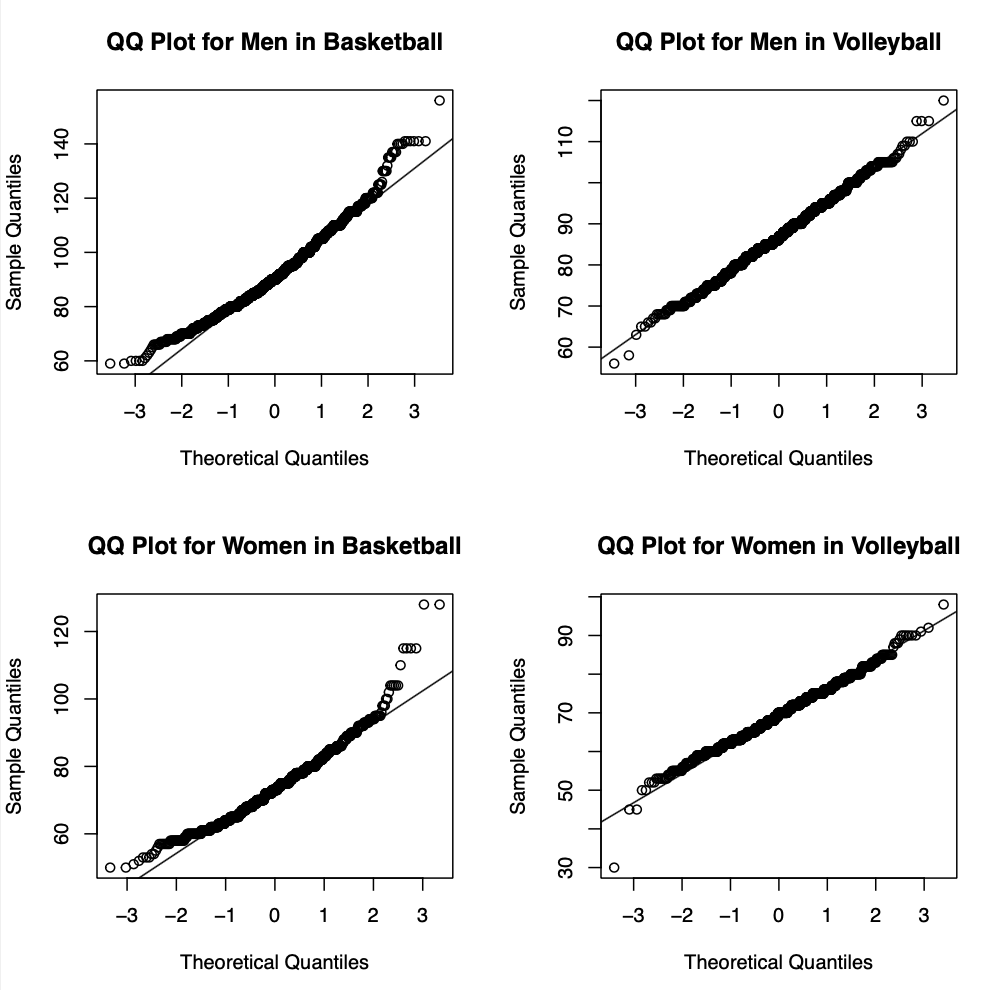


Рисунок 2 – Графическая проверка выборки на нормальность

На основании предоставленной информации о том, что распределение веса мужчин и женщин в баскетболе и волейболе не является нормальным, следует использовать непараметрический тест вместо t-критерия. Тест суммы рангов Уилкоксона (также известный как U-критерий Манна-Уитни) является подходящей альтернативой для сравнения двух независимых выборок, когда данные не распределены нормально.

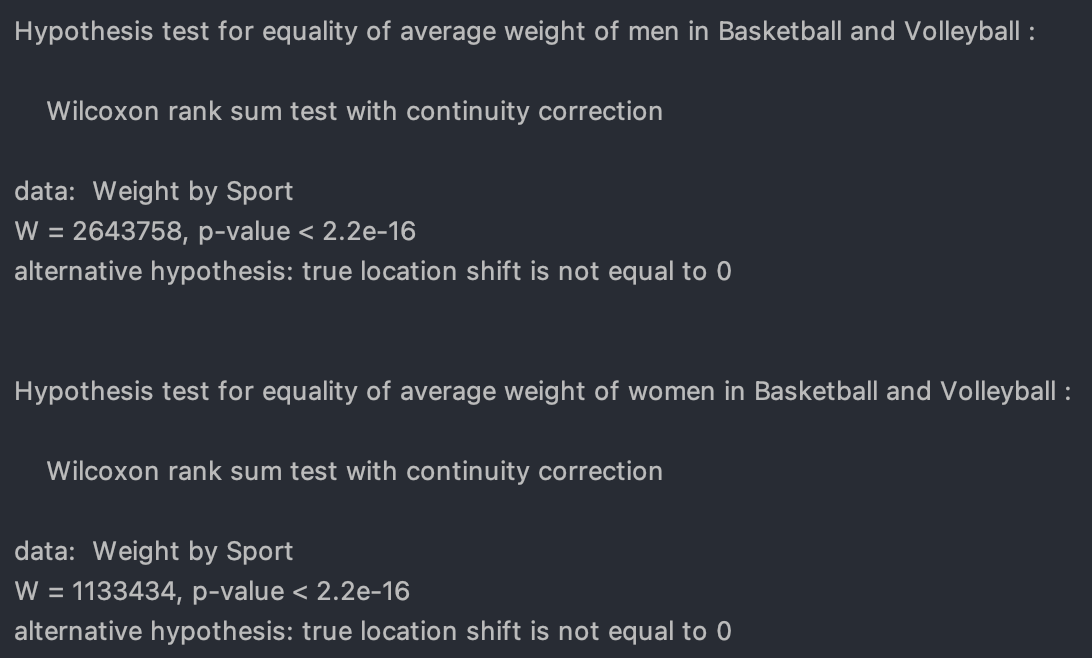


Рисунок 3 – Результат U-критерия Манна-Уитни

На основании предоставленных результатов теста Шапиро сделаем выводы относительно нормальности распределения веса в каждом из четырех подмножеств (мужчины и женщины в баскетболе и волейболе):

Мужчины в баскетболе:

W = 0,97862, p-значение <2,2e-16

Поскольку p-значение меньше 0,05, мы отвергаем нулевую гипотезу и делаем вывод, что распределение веса мужчин в баскетболе не является нормальным.

Мужчины в волейболе:

W = 0,99693, p-значение = 0,001321

Поскольку p-значение меньше 0,05, мы отвергаем нулевую гипотезу и делаем вывод, что распределение веса мужчин в волейболе не является нормальным.

Женщины в баскетболе:

W = 0,96532, значение p <2,2e-16

Поскольку p-значение меньше 0,05, мы отвергаем нулевую гипотезу и делаем вывод, что распределение веса среди женщин в баскетболе не является нормальным.

Женщины в волейболе:

W = 0,99221, значение p = 4,141e-07

Поскольку p-значение меньше 0,05, мы отвергаем нулевую гипотезу и делаем вывод, что распределение веса женщин в волейболе не является нормальным.

Таким образом, согласно результатам теста Шапиро, ни одно из распределений веса для мужчин и женщин в баскетболе и волейболе не является нормальным. Эта информация важна при принятии решения о том, какие статистические тесты использовать для дальнейшего анализа. Поскольку данные не распределяются нормально, следует рассмотреть непараметрические тесты для сравнения средних весов между группами.

Будем использовать критерий суммы рангов Уилкоксона. Критерий суммы рангов Уилкоксона, также известный как критерий Уилкоксона-Манна-Уитни (WMW), является непараметрическим статистическим критерием, используемым для оценки различий между двумя независимыми выборками. Вместо сравнения средних значений, как это делается в параметрических методах, например, t-критерии Стьюдента, критерий Уилкоксона сравнивает медианы выборок.

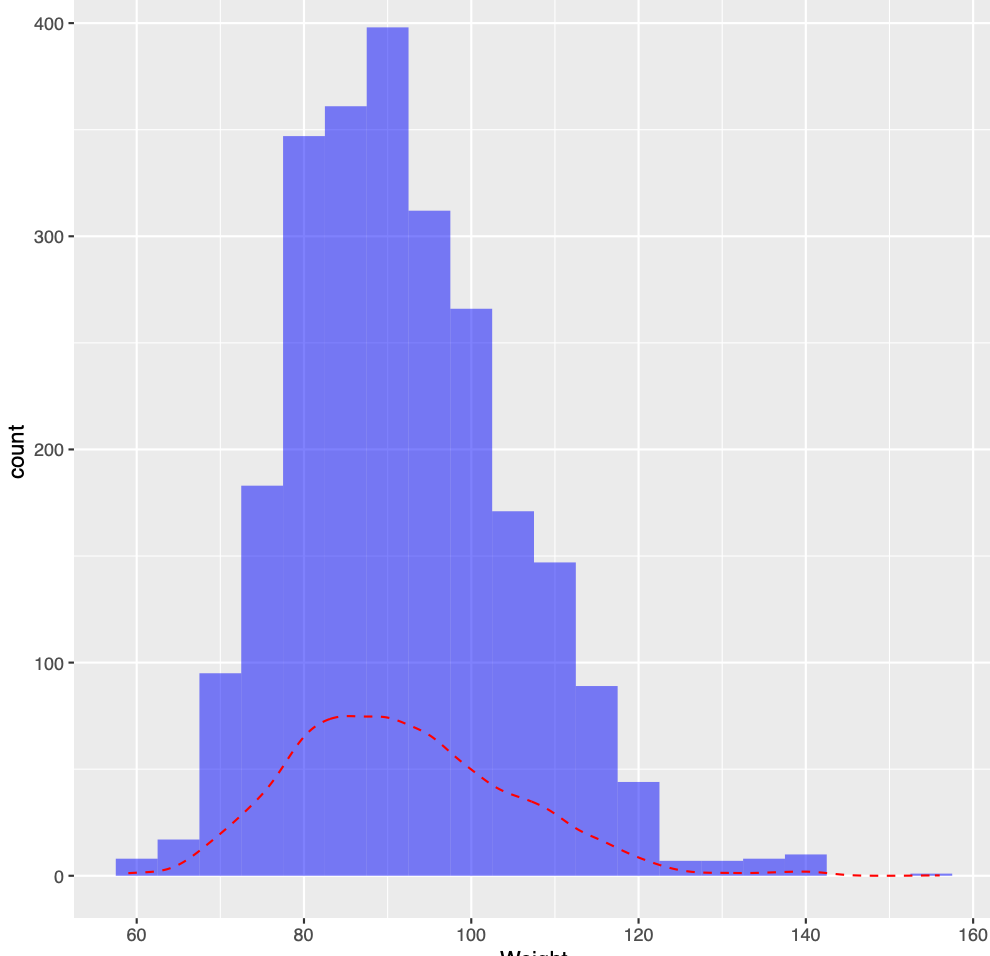


Рисунок 4 – Гистограмма и сглаженная гистограмма веса мужчин баскетболистов

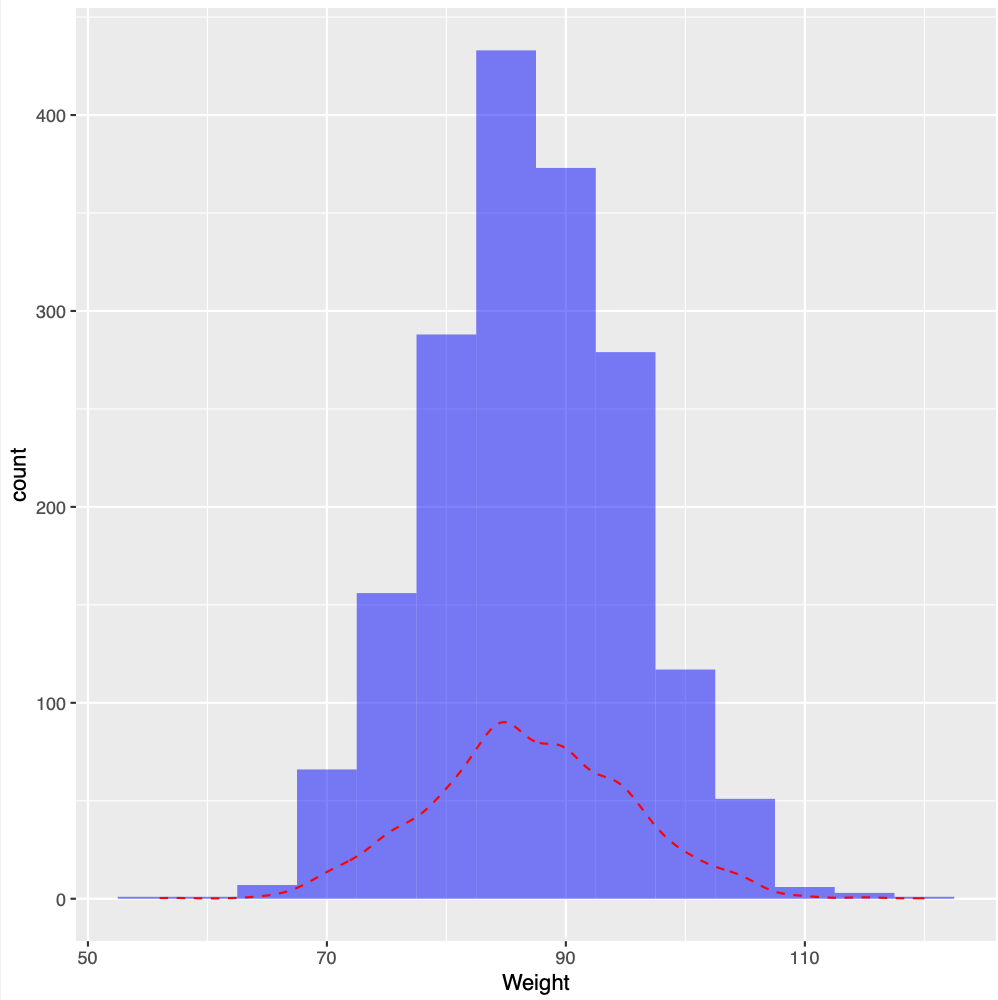
****

Рисунок 5 – Гистограмма и сглаженная гистограмма веса мужчин волейболистов

**A picture containing diagram, screenshot, plot, pixel

Description automatically generated**

Рисунок 6 – Гистограмма и сглаженная гистограмма веса женщин баскетболистов

A picture containing diagram, screenshot, plot, line

Description automatically generated

Рисунок 7 – Гистограмма и сглаженная гистограмма веса мужчин волейболистов

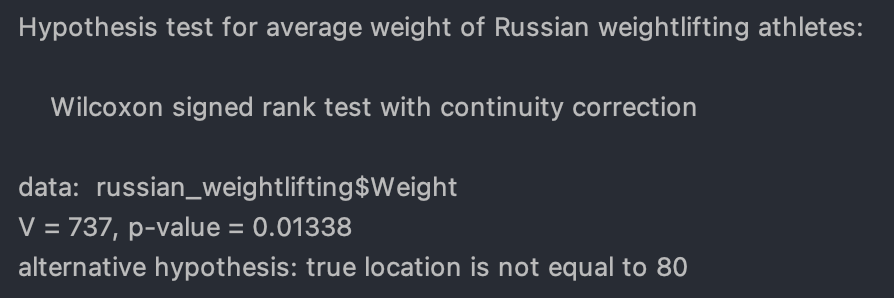


Рисунок 8 – Результат гипотезы о том, что средний вес Российских тяжело атлетов равен 80

**Вывод**

Проверка гипотезы о среднем весе российских тяжелоатлетов:

Критерий знакового ранга Уилкоксона с p-значением 0,01338 (p < 0,05) указывает на то, что мы можем отвергнуть нулевую гипотезу о том, что истинное положение (медиана) распределения веса для российских тяжелоатлетов равно 80. Это говорит о том, что средний вес российских тяжелоатлетов существенно отличается от 80.

Проверка гипотезы о равенстве среднего веса мужчин в баскетболе и волейболе:

Критерий суммы рангов Уилкоксона со значением p < 2,2e-16 (p < 0,05) показывает, что мы можем отвергнуть нулевую гипотезу о том, что истинное смещение местоположения между распределениями веса для мужчин в баскетболе и волейболе равно 0. Это означает, что существует значительная разница в среднем весе мужчин, занимающихся баскетболом и волейболом.

Проверка гипотезы о равенстве среднего веса женщин в баскетболе и волейболе:

Критерий суммы рангов Уилкоксона со значением p < 2,2e-16 (p < 0,05) показывает, что мы можем отвергнуть нулевую гипотезу о том, что истинное смещение местоположения между распределениями веса для женщин в баскетболе и волейболе равно 0. Это означает, что существует значительная разница в среднем весе женщин, занимающихся баскетболом и волейболом.

Таким образом, средний вес российских спортсменов-тяжелоатлетов значительно отличается от 80, и есть существенные различия в среднем весе мужчин и женщин, занимающихся баскетболом и волейболом.