**Работа по Анализу данных**

**Страна – Норвегия**

**Выполнил: Перетокин Никита**

Задача 1.

Определить, одинакова ли среди жителей изучаемой Вами страны частота потребления алкоголя (*alcfreq)* в зависимости размера предприятия *(estsz*)? В данном случае принципиально сравнить две группы: работников самых крупных и самых мелких предприятий.

1. **Сформулируйте задачу исследования (адаптируйте формулировку задания с учетом изучаемой Вами группы)**

Различается ли частота потребления алкоголя работниками самых маленьких и самых больших предприятий в Норвегии.

1. **Сформулируйте содержательную гипотезу:**

Частота потребления алкоголя Норвежскими работниками самых мелких и самых крупных предприятий одинакова.

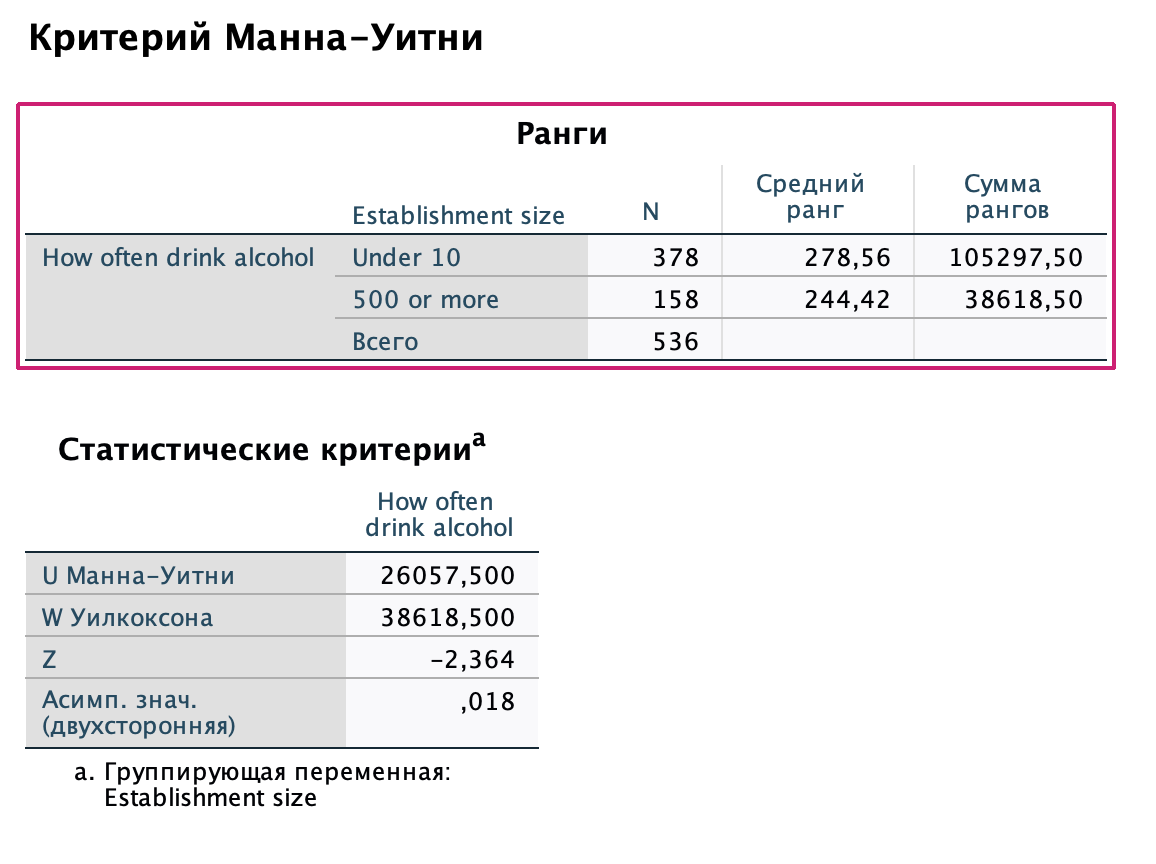
1. **Укажите выбранный̆ для решения метод непараметрической статистики и обоснуйте Ваш выбор:**

Будем использовать тест Манна-Уитни, т.к. необходимо произвести анализ двух независимых выборок, а именно работников самых малых и самых крупных предприятий.

1. **Сформулируйте статистическую гипотезу:**

Средние ранги групп равны; **Альтернативная гипотеза**: Средние ранги групп значительно различаются.

1. **Аргументируйте выбор уровня значимости. Проверьте статистическую гипотезу и сделайте выводы о ее справедливости, подкрепив их всеми необходимыми таблицами из SPSS. Проинтерпретируйте полученные статистические результаты;**

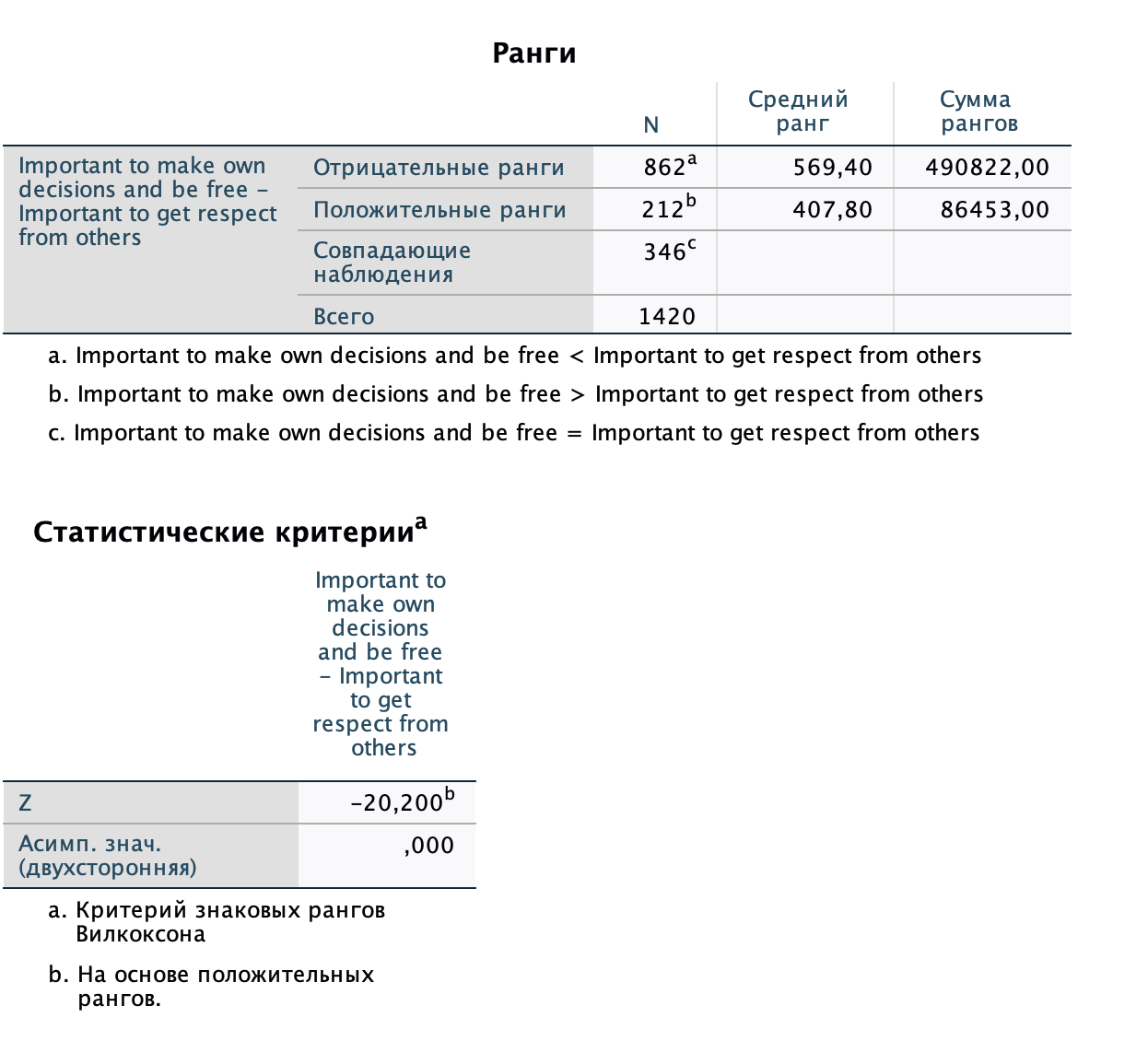


Из таблицы видно, что нулевая статистическая гипотеза о равенстве рангов отвергается на уровне 90% и 95%, но принимается на уровне 99%, значит для первых двух уровней ранги значительно отличается, а для уровня 99% значимых различий нет. Лучше, в качестве доверительного интервала использовать уровень 99%, т.к. выборка большая и асимптотическое значение сравнительно высокое.

1. **Сделайте содержательные выводы (по содержательной задаче и гипотезе)!**

При доверительном интервале 99% можно сделать вывод, что частота потребления алкоголя на самых мелких и на самых крупных предприятиях одинаковы.

**Задача 2.** Определить в одинаковой ли степени представителям изучаемой Вами группы свойственны важность ощущения уважения со стороны других (*iprspot*) и важность самостоятельно принимать решения и быть свободным *(impfree)*?

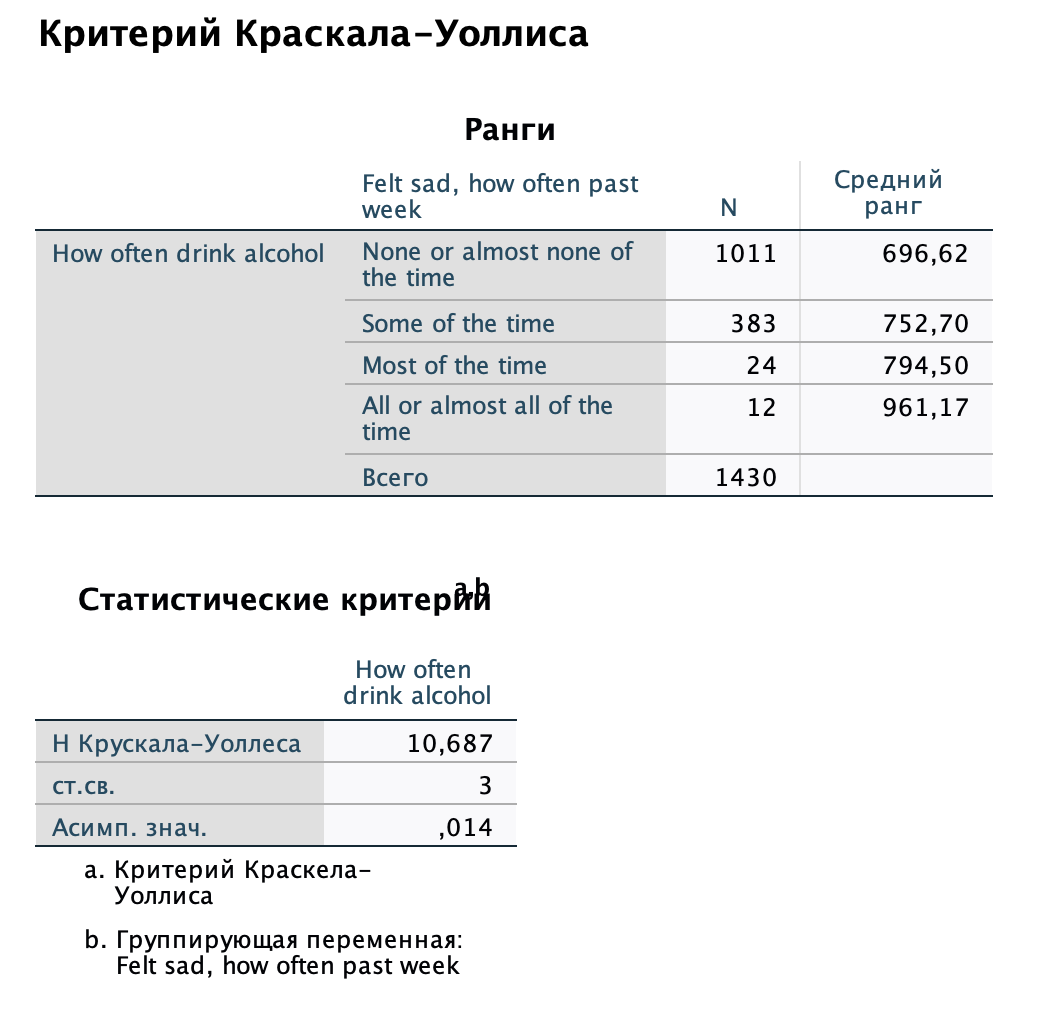
1. **Задача исследования**: В одинаковой ли степени жителям Норвегии важно ощущение уважение со стороны других и возможность самостоятельно принимать решения и быть свободным?
2. **Содержательная Гипотеза:** Жителям Норвегии в одинаковой степени важно ощущение уважение со стороны других и возможность самостоятельно принимать решения и быть свободным
3. **Метод исследования:** Будем использовать тест приписанных знаков рангов Вилкоксона, т.к. нам необходимо сравнить жителей одной стороны, то-есть выборки зависимые
4. **Статистическая гипотеза:** Средние ранги у двух переменных равны; **Альтернативная гипотиза:** Существует значительное различие между средними рангами переменных
5. **Результаты:**

Из таблицы видно, что статистическая гипотеза о равенстве средних рангов отвергается на всех доверительных интервалах, поэтому, на уровне 99% принимается альтернативная гипотеза о существовании существенных различиях между средними рангами переменных. Также из рангов мы видим, что признак 1 имеет более высокий ранг, однако переменная закодирована обратным образом, значит фактор 2 более весомый.

1. **Содержательный вывод:** С доверительной вероятностью 99% можно утверждать, что для жителей Норвегии в разной степени важно ощущение уважение со стороны других и возможность самостоятельно принимать решения и быть свободным. Также можно сделать вывод, что жителям Норвегии более важно самостоятельно принимать решения и быть свободными.

Задача 3.

1. **Задача исследования:** Определить оказывает ли влияние регулярность ощущение грусти на частоту потребления алкоголя жителями Норвегии.
2. **Содержательная гипотеза:** Регулярность ощущения грусти оказывает влияние на частоту потребления алкоголя жителями Норвегии.
3. **Метод исследования:** Будем использовать H-критерий Краскелла-Уоллеса, т.к. нам необходимо выяснить влияние одного фактора, на всю группу, иными словами, у нас несколько независимых выборок, а именно Норвежцев с разной регулярностью грусти.
4. **Статистическая гипотеза:** Средние ранги переменных равны. **Альтернативная гипотеза:** Существует значительная разница между средними рангами переменных.



Для этой группы был выбран доверительный интервал равный 90%, т.к. выборка подгрупп слишком разная по значению.

Таким образом, на уровне 90%, статистическая гипотеза о равенстве средних рангов отвергается, тогда мы можем говорить о зависимости показателей. Также исходя из средних рангов, учитывая обратную кодировку, можно сделать вывод, что последняя группа имеет большее влияние

1. **Содержательный вывод:** На доверительном уровне 90% можно утверждать, что регулярность ощущения грусти влияет на частоту потребления алкоголя Жителями Норвегии. Исходя из средних рангов делаем вывод, что чем чаще житель Норвегии грустит, тем больше его частота потребления алкоголя

**ПРИЛОЖЕНИЕ:**

**Сначала нужно отфильтровать значения по Норвегии:**

USE ALL.

SELECT IF (cntry = 'NO').

EXECUTE.

DATASET ACTIVATE Наборданных1.

**Задание 1: использование теста Манна-Уитни**

NPAR TESTS

/M-W= alcfreq BY estsz(1 5)

/MISSING ANALYSIS.

Здесь в качестве переменных были выбраны значения 1 и 5, как самые крупные и самые малые предприятия

**Задание 2: использование теста приписанных знаков рангов Вилкоксона:**

NPAR TESTS

/WILCOXON=iprspot WITH impfree (PAIRED)

/MISSING ANALYSIS.

**Задание 3: использование теста Краскелла-Уоллиса**

NPAR TESTS

/K-W=alcfreq BY fltsd(1 4)

/MISSING ANALYSIS.

Здесь был выбран диапазон с 1 до 4, не включая не содержательные ответы.