Описательная статистика

Мироненко Ольга

2022-11-08

1. Единицы измерения показателей крови в имеющихся данных не указаны, как и границы их возможных значений, но будем исходить из того, что они могут принимать только положительные значения, поэтому все неположительные значения, обнаруженные при предварительном анализе данных, заменим на пропуски.
2. Я выполню задания в двух вариантах:

– таблицы будут сформированы с помощью готовых функций из пакета gtsummary и “доформатированы” с помощью функций из пакета flextable,

– таблицы будут сформированы “вручную” с помощью функций из пакетов dplyr и purrr и отформатированы с помощью функций из пакета flextable по аналогии с тем, как это было показано в материалах курса.

## Данные по пациентам

В табл.1 представим данные по характеристикам пациентов каждой группы на момент включения в исследование. Сравнение средних значений количественных переменных между группами будет производиться с помощью t-теста, распределений категориальных признаков - с помощью теста Хи-квадрат.

#### Таблица 1 (gtsummary + flextable)

#### Описательная статистика по пациентам

| **Характеристика** | **Группа 1 N = 50** | **Группа 2 N = 50** | **p-value1** |
| --- | --- | --- | --- |
| **Возраст** |  |  | 0.011 |
| N | 50 | 50 |  |
| Среднее (Станд.откл.) | 29.2 (3.8) | 31.3 (3.9) |  |
| Медиана (Q1-Q3) | 29 (27-32) | 32 (29-34) |  |
| Минимум-Максимум | 21-38 | 21-42 |  |
| **Пол, n/N (%)** |  |  | 0.028 |
| Женский | 32/50 (64%) | 21/50 (42%) |  |
| Мужской | 18/50 (36%) | 29/50 (58%) |  |
| **Рост** |  |  | 0.559 |
| N | 50 | 50 |  |
| Среднее (Станд.откл.) | 168.0 (6.2) | 167.4 (5.4) |  |
| Медиана (Q1-Q3) | 169 (164-172) | 167 (164-171) |  |
| Минимум-Максимум | 155-181 | 159-177 |  |
| **Группа крови, n/N (%)** |  |  | 0.340 |
| O (I) | 10/41 (24%) | 15/42 (36%) |  |
| A (II) | 16/41 (39%) | 18/42 (43%) |  |
| B (III) | 11/41 (27%) | 5/42 (12%) |  |
| AB (IV) | 4/41 (9.8%) | 4/42 (9.5%) |  |
| Н.Д. | 9 | 8 |  |
| 1Welch Two Sample t-test; Pearson's Chi-squared test | | | |

#### Таблица 1 (dplyr + purrr + flextable)

#### Описательная статистика по пациентам

| **Характеристика** | **Группа 1 N = 50** | **Группа 2 N = 50** | **p-value1** |
| --- | --- | --- | --- |
| **Возраст** |  |  | 0.011 |
| N | 50 | 50 |  |
| Среднее (Станд.откл.) | 29.2 (3.8) | 31.3 (3.9) |  |
| Медиана (Q1-Q3) | 29 (27-32) | 32 (29-34) |  |
| Минимум-Максимум | 21-38 | 21-42 |  |
| **Пол** |  |  | 0.028 |
| Женский | 32/50 (64%) | 21/50 (42%) |  |
| Мужской | 18/50 (36%) | 29/50 (58%) |  |
| **Рост** |  |  | 0.559 |
| N | 50 | 50 |  |
| Среднее (Станд.откл.) | 168.0 (6.2) | 167.4 (5.4) |  |
| Медиана (Q1-Q3) | 169 (164-172) | 167 (164-171) |  |
| Минимум-Максимум | 155-181 | 159-177 |  |
| **Группа крови** |  |  | 0.340 |
| O (I) | 10/41 (24%) | 15/42 (36%) |  |
| A (II) | 16/41 (39%) | 18/42 (43%) |  |
| B (III) | 11/41 (27%) | 5/42 (12%) |  |
| AB (IV) | 4/41 (10%) | 4/42 (10%) |  |
| Н.Д. | 9 | 8 |  |
| 1Welch Two Sample t-test; Pearson's Chi-squared test | | | |

## Данные по показателям крови по визитам

#### Таблица 2 (gtsummary + flextable)

#### Описательная статистика по показателям крови по визитам

| **Характеристика** | **Визит 1** | | | **Визит 2** | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Группа 1 N = 50** | **Группа 2 N = 50** | **p-value1** | **Группа 1 N = 50** | **Группа 2 N = 50** | **p-value1** |
| **Базофилы** |  |  | 0.043 |  |  | 0.020 |
| N | 47 | 49 |  | 50 | 50 |  |
| Среднее (Станд.откл.) | 0.61 (0.34) | 0.75 (0.34) |  | 0.97 (0.38) | 1.14 (0.36) |  |
| Медиана (Q1-Q3) | 0.56 (0.34-0.84) | 0.72 (0.49-0.95) |  | 0.91 (0.73-1.24) | 1.11 (0.89-1.35) |  |
| Минимум-Максимум | 0.02-1.59 | 0.15-1.72 |  | 0.19-2.00 | 0.38-2.12 |  |
| Н.Д. | 3 | 1 |  |  |  |  |
| **Эозинофилы** |  |  | <0.001 |  |  | 0.001 |
| N | 49 | 47 |  | 50 | 48 |  |
| Среднее (Станд.откл.) | 3.24 (1.70) | 4.57 (2.03) |  | 4.19 (1.75) | 5.48 (2.15) |  |
| Медиана (Q1-Q3) | 3.16 (2.29-4.25) | 4.76 (3.49-6.11) |  | 4.11 (3.20-5.20) | 5.75 (4.25-7.08) |  |
| Минимум-Максимум | 0.27-8.13 | 0.14-8.43 |  | 0.71-9.14 | 0.36-9.45 |  |
| Н.Д. | 1 | 3 |  | 0 | 2 |  |
| **Гемоглобин** |  |  | 0.623 |  |  | 0.623 |
| N | 50 | 50 |  | 50 | 50 |  |
| Среднее (Станд.откл.) | 11.95 (2.10) | 11.77 (1.40) |  | 12.67 (2.10) | 12.49 (1.40) |  |
| Медиана (Q1-Q3) | 11.73 (10.65-13.27) | 11.71 (10.77-13.01) |  | 12.45 (11.37-13.99) | 12.43 (11.49-13.73) |  |
| Минимум-Максимум | 5.35-16.23 | 8.85-15.13 |  | 6.07-16.95 | 9.57-15.85 |  |
| **Эритроциты** |  |  | 0.340 |  |  | 0.340 |
| N | 50 | 50 |  | 50 | 50 |  |
| Среднее (Станд.откл.) | 4.17 (0.67) | 4.04 (0.66) |  | 6.48 (0.67) | 6.36 (0.66) |  |
| Медиана (Q1-Q3) | 4.23 (3.74-4.63) | 4.03 (3.53-4.53) |  | 6.55 (6.06-6.95) | 6.34 (5.84-6.85) |  |
| Минимум-Максимум | 2.82-5.73 | 2.84-5.26 |  | 5.14-8.04 | 5.16-7.58 |  |
| 1Welch Two Sample t-test | | | | | | |

#### Таблица 2 (dplyr + purrr + flextable)

#### Описательная статистика по показателям крови по визитам

| **Характеристика** | **Визит 1** | | | **Визит 2** | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Группа 1 N = 50** | **Группа 2 N = 50** | **p-value1** | **Группа 1 N = 50** | **Группа 2 N = 50** | **p-value1** |
| **Базофилы** |  |  | 0.043 |  |  | 0.020 |
| N | 47 | 49 |  | 50 | 50 |  |
| Среднее (Станд.откл.) | 0.61 (0.34) | 0.75 (0.34) |  | 0.97 (0.38) | 1.14 (0.36) |  |
| Медиана (Q1-Q3) | 0.56 (0.34-0.84) | 0.72 (0.49-0.95) |  | 0.91 (0.73-1.24) | 1.11 (0.89-1.35) |  |
| Минимум-Максимум | 0.02-1.59 | 0.15-1.72 |  | 0.19-2.00 | 0.38-2.12 |  |
| Н.Д. | 3 | 1 |  |  |  |  |
| **Эозинофилы** |  |  | <0.001 |  |  | 0.002 |
| N | 49 | 47 |  | 50 | 48 |  |
| Среднее (Станд.откл.) | 3.24 (1.70) | 4.57 (2.03) |  | 4.19 (1.75) | 5.48 (2.15) |  |
| Медиана (Q1-Q3) | 3.16 (2.29-4.25) | 4.76 (3.49-6.11) |  | 4.11 (3.20-5.20) | 5.75 (4.25-7.08) |  |
| Минимум-Максимум | 0.27-8.13 | 0.14-8.43 |  | 0.71-9.14 | 0.36-9.45 |  |
| Н.Д. | 1 | 3 |  |  | 2 |  |
| **Гемоглобин** |  |  | 0.623 |  |  | 0.623 |
| N | 50 | 50 |  | 50 | 50 |  |
| Среднее (Станд.откл.) | 11.95 (2.10) | 11.77 (1.40) |  | 12.67 (2.10) | 12.49 (1.40) |  |
| Медиана (Q1-Q3) | 11.73 (10.65-13.27) | 11.71 (10.77-13.01) |  | 12.45 (11.37-13.99) | 12.43 (11.49-13.73) |  |
| Минимум-Максимум | 5.35-16.23 | 8.85-15.13 |  | 6.07-16.95 | 9.57-15.85 |  |
| **Эритроциты** |  |  | 0.340 |  |  | 0.340 |
| N | 50 | 50 |  | 50 | 50 |  |
| Среднее (Станд.откл.) | 4.17 (0.67) | 4.04 (0.66) |  | 6.48 (0.67) | 6.36 (0.66) |  |
| Медиана (Q1-Q3) | 4.23 (3.74-4.63) | 4.03 (3.53-4.53) |  | 6.55 (6.06-6.95) | 6.34 (5.84-6.85) |  |
| Минимум-Максимум | 2.82-5.73 | 2.84-5.26 |  | 5.14-8.04 | 5.16-7.58 |  |
| 1Welch Two Sample t-test | | | | | | |

Заметим, что p-values в таблицах 2 выше получились абсолютно идентичными для обоих визитов по тем показателям, где не было пропущенных данных (гемоглобин и эритроциты). Если для каждого наблюдения в выборке рассчитать изменение каждого показателя крови между визитами, то можно заметить (см. табл.3), что по всем показателям, кроме гемоглобина, изменение было абсолютно одинаковым для всех наблюдений без пропусков, для гемоглобина встречается 2 уникальных значения изменения, и то они различаются только четвертым знаком после запятой (что, вероятнее всего, связано со спецификой хранения R-ом десятичных чисел в памяти компьютера, тогда как в действительности, если ту же разницу посчитать в Экселе, то она будет абсолютно одинаковой для всех наблюдений). Иными словами, для всех наблюдений значение на втором визите было получено сдвигом вправо значения на первом визите на одинаковую для каждого показателя крови величину (это верно и для исходных данных, без замены неположительных значений на пропуски) - отсюда и идентичные значения p-values при сравнении групп по средним значениям гемоглобина и эритроцитов внутри каждого визита (разница в средних значениях между группами по каждому показателю не изменилась между визитами). Если бы мы сравнивали средние значения изменений между группами, то они были бы одинаковыми, с нулевой дисперсией, и мы не смогли бы выполнить t-тест на таких данных.

#### Таблица 3

#### Изменение показателей крови между визитами

| **Разница (Визит 2 - Визит 1)** | **Группа 1 N = 50** | **Группа 2 N = 50** |
| --- | --- | --- |
| **Базофилы** |  |  |
| 0.4042 | 47 (100%) | 49 (100%) |
| Н.Д. | 3 | 1 |
| **Эозинофилы** |  |  |
| 1.0148 | 49 (100%) | 47 (100%) |
| Н.Д. | 1 | 3 |
| **Гемоглобин** |  |  |
| 0.720899999999999 | 7 (14%) | 11 (22%) |
| 0.7209 | 43 (86%) | 39 (78%) |
| **Эритроциты** |  |  |
| 2.3155 | 50 (100%) | 50 (100%) |