final

Маслова И

6 ноября 2022 г

# Задание 2

Скачивание

data <- read\_excel('data\_excel.xlsx')

Замена пропущенных значений группы крови, замена запятых и имен переменных

data[data == 'NA'] <- 'Нет данных'   
  
data <- data %>%   
 mutate(across(contains(c("E1", "E2")), function(x) as.numeric(scan(text=x, dec=",", sep="."))))  
  
data <- data %>% rename\_with(function(x) x %>% stri\_replace\_all\_regex(c("\_E1", "\_E2"), c("Визит1", "Визит2"), vectorize\_all = FALSE))

Разделение на визиты

data\_E1 <- data %>% select(!ends\_with("Визит2"))  
data\_E2 <- data %>% select(!ends\_with("Визит1"))

Статистика

statistics <- list( `Количество субъектов` = ~length(.x) %>% as.character(),  
 `Количество (есть данные)` = ~sum(!is.na(.x)) %>% as.character(),  
 `Нет данных` = ~sum(is.na(.x)) %>% as.character(),  
 `Ср. знач.` = ~ifelse(sum(!is.na(.x)) == 0, "Н/П\*", mean(.x, na.rm = TRUE) %>% round(2) %>% as.character()),  
 `Станд. отклон.` = ~ifelse(sum(!is.na(.x)) < 3, "Н/П\*", sd(.x, na.rm = TRUE) %>% round(2) %>% as.character()),  
 `95% ДИ для среднего` = ~sd(.x, na.rm = TRUE) %>% round(2) %>% as.character(),  
 `мин. - макс.` = ~ifelse(sum(!is.na(.x)) == 0, "Н/П\*", paste0(min(.x, na.rm = TRUE) %>% round(2), " - ", max(.x, na.rm = TRUE) %>% round(2)) %>% as.character()),  
 `Медиана` = ~ifelse(sum(!is.na(.x)) == 0, "Н/П\*", median(.x, na.rm = TRUE) %>% round(2) %>% as.character())  
   
 )

Построение таблицы для первого визита

data\_E1 %>%   
 select(`Группа`, where(is.numeric)) %>%  
 group\_by(`Группа`) %>%   
 summarise(across(where(is.numeric), statistics)) %>%   
 pivot\_longer(!`Группа`) %>%  
 separate(name, into = c("Переменная", "Статистика"), sep = "\_") %>%  
 rename(`Значение` = value) %>%  
 flextable() %>%  
 theme\_box() %>%  
 merge\_v(c("Группа", "Переменная")) %>%  
 width(width=1.7)

| **Группа** | **Переменная** | **Статистика** | **Значение** |
| --- | --- | --- | --- |
| Группа 1 | Возраст | Количество субъектов | 50 |
| Количество (есть данные) | 50 |
| Нет данных | 0 |
| Ср. знач. | 29.24 |
| Станд. отклон. | 3.8 |
| 95% ДИ для среднего | 3.8 |
| мин. - макс. | 21 - 38 |
| Медиана | 29 |
| Рост | Количество субъектов | 50 |
| Количество (есть данные) | 50 |
| Нет данных | 0 |
| Ср. знач. | 168.04 |
| Станд. отклон. | 6.16 |
| 95% ДИ для среднего | 6.16 |
| мин. - макс. | 155 - 181 |
| Медиана | 169 |
| БазофилыВизит1 | Количество субъектов | 50 |
| Количество (есть данные) | 50 |
| Нет данных | 0 |
| Ср. знач. | 0.56 |
| Станд. отклон. | 0.38 |
| 95% ДИ для среднего | 0.38 |
| мин. - макс. | -0.22 - 1.59 |
| Медиана | 0.5 |
| ЭозинофилыВизит1 | Количество субъектов | 50 |
| Количество (есть данные) | 50 |
| Нет данных | 0 |
| Ср. знач. | 3.17 |
| Станд. отклон. | 1.75 |
| 95% ДИ для среднего | 1.75 |
| мин. - макс. | -0.3 - 8.13 |
| Медиана | 3.1 |
| ГемоглобинВизит1 | Количество субъектов | 50 |
| Количество (есть данные) | 50 |
| Нет данных | 0 |
| Ср. знач. | 11.95 |
| Станд. отклон. | 2.1 |
| 95% ДИ для среднего | 2.1 |
| мин. - макс. | 5.35 - 16.23 |
| Медиана | 11.73 |
| ЭритроцитыВизит1 | Количество субъектов | 50 |
| Количество (есть данные) | 50 |
| Нет данных | 0 |
| Ср. знач. | 4.17 |
| Станд. отклон. | 0.67 |
| 95% ДИ для среднего | 0.67 |
| мин. - макс. | 2.82 - 5.73 |
| Медиана | 4.23 |
| Группа 2 | Возраст | Количество субъектов | 50 |
| Количество (есть данные) | 50 |
| Нет данных | 0 |
| Ср. знач. | 31.26 |
| Станд. отклон. | 3.94 |
| 95% ДИ для среднего | 3.94 |
| мин. - макс. | 21 - 42 |
| Медиана | 32 |
| Рост | Количество субъектов | 50 |
| Количество (есть данные) | 50 |
| Нет данных | 0 |
| Ср. знач. | 167.36 |
| Станд. отклон. | 5.4 |
| 95% ДИ для среднего | 5.4 |
| мин. - макс. | 159 - 177 |
| Медиана | 167 |
| БазофилыВизит1 | Количество субъектов | 50 |
| Количество (есть данные) | 50 |
| Нет данных | 0 |
| Ср. знач. | 0.74 |
| Станд. отклон. | 0.36 |
| 95% ДИ для среднего | 0.36 |
| мин. - макс. | -0.03 - 1.72 |
| Медиана | 0.71 |
| ЭозинофилыВизит1 | Количество субъектов | 50 |
| Количество (есть данные) | 50 |
| Нет данных | 0 |
| Ср. знач. | 4.24 |
| Станд. отклон. | 2.38 |
| 95% ДИ для среднего | 2.38 |
| мин. - макс. | -1.23 - 8.43 |
| Медиана | 4.64 |
| ГемоглобинВизит1 | Количество субъектов | 50 |
| Количество (есть данные) | 50 |
| Нет данных | 0 |
| Ср. знач. | 11.77 |
| Станд. отклон. | 1.4 |
| 95% ДИ для среднего | 1.4 |
| мин. - макс. | 8.85 - 15.13 |
| Медиана | 11.71 |
| ЭритроцитыВизит1 | Количество субъектов | 50 |
| Количество (есть данные) | 50 |
| Нет данных | 0 |
| Ср. знач. | 4.04 |
| Станд. отклон. | 0.66 |
| 95% ДИ для среднего | 0.66 |
| мин. - макс. | 2.84 - 5.26 |
| Медиана | 4.03 |

Построение таблицы для второго визита

data\_E2 %>%   
 select(`Группа`, where(is.numeric)) %>%  
 group\_by(`Группа`) %>%   
 summarise(across(where(is.numeric), statistics)) %>%   
 pivot\_longer(!`Группа`) %>%  
 separate(name, into = c("Переменная", "Статистика"), sep = "\_") %>%  
 rename(`Значение` = value) %>%  
 flextable() %>%  
 theme\_box() %>%  
 merge\_v(c("Группа", "Переменная")) %>%  
 width(width=1.7)

| **Группа** | **Переменная** | **Статистика** | **Значение** |
| --- | --- | --- | --- |
| Группа 1 | Возраст | Количество субъектов | 50 |
| Количество (есть данные) | 50 |
| Нет данных | 0 |
| Ср. знач. | 29.24 |
| Станд. отклон. | 3.8 |
| 95% ДИ для среднего | 3.8 |
| мин. - макс. | 21 - 38 |
| Медиана | 29 |
| Рост | Количество субъектов | 50 |
| Количество (есть данные) | 50 |
| Нет данных | 0 |
| Ср. знач. | 168.04 |
| Станд. отклон. | 6.16 |
| 95% ДИ для среднего | 6.16 |
| мин. - макс. | 155 - 181 |
| Медиана | 169 |
| БазофилыВизит2 | Количество субъектов | 50 |
| Количество (есть данные) | 50 |
| Нет данных | 0 |
| Ср. знач. | 0.97 |
| Станд. отклон. | 0.38 |
| 95% ДИ для среднего | 0.38 |
| мин. - макс. | 0.19 - 2 |
| Медиана | 0.91 |
| ЭозинофилыВизит2 | Количество субъектов | 50 |
| Количество (есть данные) | 50 |
| Нет данных | 0 |
| Ср. знач. | 4.19 |
| Станд. отклон. | 1.75 |
| 95% ДИ для среднего | 1.75 |
| мин. - макс. | 0.71 - 9.14 |
| Медиана | 4.11 |
| ГемоглобинВизит2 | Количество субъектов | 50 |
| Количество (есть данные) | 50 |
| Нет данных | 0 |
| Ср. знач. | 12.67 |
| Станд. отклон. | 2.1 |
| 95% ДИ для среднего | 2.1 |
| мин. - макс. | 6.07 - 16.95 |
| Медиана | 12.45 |
| ЭритроцитыВизит2 | Количество субъектов | 50 |
| Количество (есть данные) | 50 |
| Нет данных | 0 |
| Ср. знач. | 6.48 |
| Станд. отклон. | 0.67 |
| 95% ДИ для среднего | 0.67 |
| мин. - макс. | 5.14 - 8.04 |
| Медиана | 6.55 |
| Группа 2 | Возраст | Количество субъектов | 50 |
| Количество (есть данные) | 50 |
| Нет данных | 0 |
| Ср. знач. | 31.26 |
| Станд. отклон. | 3.94 |
| 95% ДИ для среднего | 3.94 |
| мин. - макс. | 21 - 42 |
| Медиана | 32 |
| Рост | Количество субъектов | 50 |
| Количество (есть данные) | 50 |
| Нет данных | 0 |
| Ср. знач. | 167.36 |
| Станд. отклон. | 5.4 |
| 95% ДИ для среднего | 5.4 |
| мин. - макс. | 159 - 177 |
| Медиана | 167 |
| БазофилыВизит2 | Количество субъектов | 50 |
| Количество (есть данные) | 50 |
| Нет данных | 0 |
| Ср. знач. | 1.14 |
| Станд. отклон. | 0.36 |
| 95% ДИ для среднего | 0.36 |
| мин. - макс. | 0.38 - 2.12 |
| Медиана | 1.11 |
| ЭозинофилыВизит2 | Количество субъектов | 50 |
| Количество (есть данные) | 50 |
| Нет данных | 0 |
| Ср. знач. | 5.26 |
| Станд. отклон. | 2.38 |
| 95% ДИ для среднего | 2.38 |
| мин. - макс. | -0.21 - 9.45 |
| Медиана | 5.65 |
| ГемоглобинВизит2 | Количество субъектов | 50 |
| Количество (есть данные) | 50 |
| Нет данных | 0 |
| Ср. знач. | 12.49 |
| Станд. отклон. | 1.4 |
| 95% ДИ для среднего | 1.4 |
| мин. - макс. | 9.57 - 15.85 |
| Медиана | 12.43 |
| ЭритроцитыВизит2 | Количество субъектов | 50 |
| Количество (есть данные) | 50 |
| Нет данных | 0 |
| Ср. знач. | 6.36 |
| Станд. отклон. | 0.66 |
| 95% ДИ для среднего | 0.66 |
| мин. - макс. | 5.16 - 7.58 |
| Медиана | 6.34 |

Статистика не нумеречных переменных. Для двух визитов будет одинаковая.

data %>%   
 select(where(is.character)) %>%  
 count(`Группа`, `Пол`, `Группа крови`) %>%  
 group\_by(`Группа`) %>%   
 rename(`Уникальные сочетания` = n) %>%  
 flextable() %>%  
 theme\_box() %>%  
 merge\_v(c("Группа", "Пол")) %>%  
 width(width=1.7)

| **Группа** | **Пол** | **Группа крови** | **Уникальные сочетания** |
| --- | --- | --- | --- |
| Группа 1 | Женский | A (II) | 12 |
| AB (IV) | 2 |
| B (III) | 7 |
| O (I) | 6 |
| Нет данных | 5 |
| Мужской | A (II) | 4 |
| AB (IV) | 2 |
| B (III) | 4 |
| O (I) | 4 |
| Нет данных | 4 |
| Группа 2 | Женский | A (II) | 6 |
| AB (IV) | 1 |
| B (III) | 2 |
| O (I) | 8 |
| Нет данных | 4 |
| Мужской | A (II) | 12 |
| AB (IV) | 3 |
| B (III) | 3 |
| O (I) | 7 |
| Нет данных | 4 |

# Задание 3

Статистические тесты для первого визита и общих нумеречных переменных. Сравнение групп.

p <- wilcox.test(pull(data %>% filter(`Группа` == "Группа 1") %>% select(`БазофилыВизит1`)), pull(data %>% filter(`Группа` == "Группа 2") %>% select(`БазофилыВизит1`)))$p.value  
p <- append(p, wilcox.test(pull(data %>% filter(`Группа` == "Группа 1") %>% select(`Возраст`)), pull(data %>% filter(`Группа` == "Группа 2") %>% select(`Возраст`)))$p.value)  
p <- append(p, wilcox.test(pull(data %>% filter(`Группа` == "Группа 1") %>% select(`Рост`)), pull(data %>% filter(`Группа` == "Группа 2") %>% select(`Рост`)))$p.value)  
p <- append(p, wilcox.test(pull(data %>% filter(`Группа` == "Группа 1") %>% select(`ЭозинофилыВизит1`)), pull(data %>% filter(`Группа` == "Группа 2") %>% select(`ЭозинофилыВизит1`)))$p.value)  
p <- append(p, wilcox.test(pull(data %>% filter(`Группа` == "Группа 1") %>% select(`ГемоглобинВизит1`)), pull(data %>% filter(`Группа` == "Группа 2") %>% select(`ГемоглобинВизит1`)))$p.value)  
p <- append(p, wilcox.test(pull(data %>% filter(`Группа` == "Группа 1") %>% select(`ЭритроцитыВизит1`)), pull(data %>% filter(`Группа` == "Группа 2") %>% select(`ЭритроцитыВизит1`)))$p.value)  
  
as.data.frame(p) %>% flextable() %>% theme\_box() %>% bg(`p` > 0.05, bg = 'orange')

| **p** |
| --- |
| 0.023962747 |
| 0.008403023 |
| 0.413220448 |
| 0.003663798 |
| 0.646637062 |
| 0.302716205 |

Статистические тесты для второго визита. Сравнение групп.

p <- wilcox.test(pull(data %>% filter(`Группа` == "Группа 1") %>% select(`ЭозинофилыВизит2`)), pull(data %>% filter(`Группа` == "Группа 2") %>% select(`ЭозинофилыВизит2`)))$p.value  
p <- append(p, wilcox.test(pull(data %>% filter(`Группа` == "Группа 1") %>% select(`ГемоглобинВизит2`)), pull(data %>% filter(`Группа` == "Группа 2") %>% select(`ГемоглобинВизит2`)))$p.value)  
p <- append(p, wilcox.test(pull(data %>% filter(`Группа` == "Группа 1") %>% select(`ЭритроцитыВизит2`)), pull(data %>% filter(`Группа` == "Группа 2") %>% select(`ЭритроцитыВизит2`)))$p.value)  
p <- append(p, wilcox.test(pull(data %>% filter(`Группа` == "Группа 1") %>% select(`БазофилыВизит2`)), pull(data %>% filter(`Группа` == "Группа 2") %>% select(`БазофилыВизит2`)))$p.value)  
  
as.data.frame(p) %>% flextable() %>% theme\_box() %>% bg(`p` > 0.05, bg = 'orange')

| **p** |
| --- |
| 0.003663798 |
| 0.646637062 |
| 0.302716205 |
| 0.023962747 |