Введение в автоматизацию

Цепелева И.

2022-11-21

library(readxl)  
data <- read\_excel('data\_excel.xlsx')  
summary(data)

## Группа Возраст Пол Рост   
## Length:100 Min. :21.00 Length:100 Min. :155.0   
## Class :character 1st Qu.:28.00 Class :character 1st Qu.:164.0   
## Mode :character Median :30.50 Mode :character Median :168.0   
## Mean :30.25 Mean :167.7   
## 3rd Qu.:33.00 3rd Qu.:171.2   
## Max. :42.00 Max. :181.0   
## Группа крови Базофилы\_E1 Эозинофилы\_E1 Гемоглобин\_E1   
## Length:100 Length:100 Length:100 Length:100   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## Эритроциты\_E1 Базофилы\_E2 Эозинофилы\_E2 Гемоглобин\_E2   
## Length:100 Length:100 Length:100 Length:100   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## Эритроциты\_E2   
## Length:100   
## Class :character   
## Mode :character   
##   
##   
##

str(data)

## tibble [100 x 13] (S3: tbl\_df/tbl/data.frame)  
## $ Группа : chr [1:100] "Группа 1" "Группа 1" "Группа 1" "Группа 1" ...  
## $ Возраст : num [1:100] 31 28 33 26 33 28 27 31 23 29 ...  
## $ Пол : chr [1:100] "Женский" "Женский" "Женский" "Женский" ...  
## $ Рост : num [1:100] 174 157 166 168 170 172 157 174 175 172 ...  
## $ Группа крови : chr [1:100] "A (II)" "A (II)" "NA" "O (I)" ...  
## $ Базофилы\_E1 : chr [1:100] "0,4222" "0,3270" "0,7994" "0,0237" ...  
## $ Эозинофилы\_E1: chr [1:100] "0,6465" "4,9742" "3,3875" "4,5403" ...  
## $ Гемоглобин\_E1: chr [1:100] "10,6842" "9,6169" "10,1628" "10,6391" ...  
## $ Эритроциты\_E1: chr [1:100] "4,2573" "3,8813" "5,0607" "3,8083" ...  
## $ Базофилы\_E2 : chr [1:100] "0,8264" "0,7312" "1,2036" "0,4279" ...  
## $ Эозинофилы\_E2: chr [1:100] "1,6613" "5,9890" "4,4023" "5,5551" ...  
## $ Гемоглобин\_E2: chr [1:100] "11,4051" "10,3378" "10,8837" "11,3600" ...  
## $ Эритроциты\_E2: chr [1:100] "6,5728" "6,1968" "7,3762" "6,1238" ...

library(dplyr)  
library(tidyverse)  
library(flextable)  
library(gtsummary)  
  
data <- data %>%  
 mutate\_at(vars('Базофилы\_E1':'Эритроциты\_E2'), ~ gsub(",", ".", .) %>% as.numeric()) %>%  
 mutate\_at(vars('Базофилы\_E1':'Эритроциты\_E2'), ~ ifelse(. <= 0, NA, .)) %>% #заменяем неположительные значения на NA  
 mutate\_at(vars('Группа', 'Пол'), ~ as.factor(.)) %>%  
 mutate(`Группа крови` = `Группа крови` %>% na\_if('NA')) #заменяем 'NA' на NA

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))),  
 `Медиана` = ~ifelse(sum(!is.na(.x)) == 0, "Н/П\*", median(.x, na.rm = TRUE) %>% round(2) %>% as.character()),  
 `Q1 - Q3` = ~ifelse(sum(!is.na(.x)) == 0, "Н/П\*", paste0(quantile(.x, 0.25, na.rm = TRUE) %>% round(2), " - ", quantile(.x, 0.75, na.rm = TRUE) %>% round(2)))  
)

data %>%   
 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() %>%  
 merge\_v(c("Группа", "Переменная"))

| Группа | Переменная | Статистика | Значение |
| --- | --- | --- | --- |
| Группа 1 | Возраст\_Количество субъектов |  | 50 |
| Возраст\_Количество (есть данные) |  | 50 |
| Возраст\_Нет данных |  | 0 |
| Возраст\_Ср. знач. |  | 29.24 |
| Возраст\_Станд. отклон. |  | 3.8 |
| Возраст\_95% ДИ для среднего |  | 3.8 |
| Возраст\_мин. - макс. |  | 21 - 38 |
| Возраст\_Медиана |  | 29 |
| Возраст\_Q1 - Q3 |  | 27 - 32 |
| Рост\_Количество субъектов |  | 50 |
| Рост\_Количество (есть данные) |  | 50 |
| Рост\_Нет данных |  | 0 |
| Рост\_Ср. знач. |  | 168.04 |
| Рост\_Станд. отклон. |  | 6.16 |
| Рост\_95% ДИ для среднего |  | 6.16 |
| Рост\_мин. - макс. |  | 155 - 181 |
| Рост\_Медиана |  | 169 |
| Рост\_Q1 - Q3 |  | 164 - 172 |
| Базофилы\_E1\_Количество субъектов |  | 50 |
| Базофилы\_E1\_Количество (есть данные) |  | 47 |
| Базофилы\_E1\_Нет данных |  | 3 |
| Базофилы\_E1\_Ср. знач. |  | 0.61 |
| Базофилы\_E1\_Станд. отклон. |  | 0.34 |
| Базофилы\_E1\_95% ДИ для среднего |  | 0.34 |
| Базофилы\_E1\_мин. - макс. |  | 0.02 - 1.59 |
| Базофилы\_E1\_Медиана |  | 0.56 |
| Базофилы\_E1\_Q1 - Q3 |  | 0.34 - 0.84 |
| Эозинофилы\_E1\_Количество субъектов |  | 50 |
| Эозинофилы\_E1\_Количество (есть данные) |  | 49 |
| Эозинофилы\_E1\_Нет данных |  | 1 |
| Эозинофилы\_E1\_Ср. знач. |  | 3.24 |
| Эозинофилы\_E1\_Станд. отклон. |  | 1.7 |
| Эозинофилы\_E1\_95% ДИ для среднего |  | 1.7 |
| Эозинофилы\_E1\_мин. - макс. |  | 0.27 - 8.13 |
| Эозинофилы\_E1\_Медиана |  | 3.16 |
| Эозинофилы\_E1\_Q1 - Q3 |  | 2.29 - 4.25 |
| Гемоглобин\_E1\_Количество субъектов |  | 50 |
| Гемоглобин\_E1\_Количество (есть данные) |  | 50 |
| Гемоглобин\_E1\_Нет данных |  | 0 |
| Гемоглобин\_E1\_Ср. знач. |  | 11.95 |
| Гемоглобин\_E1\_Станд. отклон. |  | 2.1 |
| Гемоглобин\_E1\_95% ДИ для среднего |  | 2.1 |
| Гемоглобин\_E1\_мин. - макс. |  | 5.35 - 16.23 |
| Гемоглобин\_E1\_Медиана |  | 11.73 |
| Гемоглобин\_E1\_Q1 - Q3 |  | 10.65 - 13.27 |
| Эритроциты\_E1\_Количество субъектов |  | 50 |
| Эритроциты\_E1\_Количество (есть данные) |  | 50 |
| Эритроциты\_E1\_Нет данных |  | 0 |
| Эритроциты\_E1\_Ср. знач. |  | 4.17 |
| Эритроциты\_E1\_Станд. отклон. |  | 0.67 |
| Эритроциты\_E1\_95% ДИ для среднего |  | 0.67 |
| Эритроциты\_E1\_мин. - макс. |  | 2.82 - 5.73 |
| Эритроциты\_E1\_Медиана |  | 4.23 |
| Эритроциты\_E1\_Q1 - Q3 |  | 3.74 - 4.63 |
| Базофилы\_E2\_Количество субъектов |  | 50 |
| Базофилы\_E2\_Количество (есть данные) |  | 50 |
| Базофилы\_E2\_Нет данных |  | 0 |
| Базофилы\_E2\_Ср. знач. |  | 0.97 |
| Базофилы\_E2\_Станд. отклон. |  | 0.38 |
| Базофилы\_E2\_95% ДИ для среднего |  | 0.38 |
| Базофилы\_E2\_мин. - макс. |  | 0.19 - 2 |
| Базофилы\_E2\_Медиана |  | 0.91 |
| Базофилы\_E2\_Q1 - Q3 |  | 0.73 - 1.24 |
| Эозинофилы\_E2\_Количество субъектов |  | 50 |
| Эозинофилы\_E2\_Количество (есть данные) |  | 50 |
| Эозинофилы\_E2\_Нет данных |  | 0 |
| Эозинофилы\_E2\_Ср. знач. |  | 4.19 |
| Эозинофилы\_E2\_Станд. отклон. |  | 1.75 |
| Эозинофилы\_E2\_95% ДИ для среднего |  | 1.75 |
| Эозинофилы\_E2\_мин. - макс. |  | 0.71 - 9.14 |
| Эозинофилы\_E2\_Медиана |  | 4.11 |
| Эозинофилы\_E2\_Q1 - Q3 |  | 3.2 - 5.2 |
| Гемоглобин\_E2\_Количество субъектов |  | 50 |
| Гемоглобин\_E2\_Количество (есть данные) |  | 50 |
| Гемоглобин\_E2\_Нет данных |  | 0 |
| Гемоглобин\_E2\_Ср. знач. |  | 12.67 |
| Гемоглобин\_E2\_Станд. отклон. |  | 2.1 |
| Гемоглобин\_E2\_95% ДИ для среднего |  | 2.1 |
| Гемоглобин\_E2\_мин. - макс. |  | 6.07 - 16.95 |
| Гемоглобин\_E2\_Медиана |  | 12.45 |
| Гемоглобин\_E2\_Q1 - Q3 |  | 11.37 - 13.99 |
| Эритроциты\_E2\_Количество субъектов |  | 50 |
| Эритроциты\_E2\_Количество (есть данные) |  | 50 |
| Эритроциты\_E2\_Нет данных |  | 0 |
| Эритроциты\_E2\_Ср. знач. |  | 6.48 |
| Эритроциты\_E2\_Станд. отклон. |  | 0.67 |
| Эритроциты\_E2\_95% ДИ для среднего |  | 0.67 |
| Эритроциты\_E2\_мин. - макс. |  | 5.14 - 8.04 |
| Эритроциты\_E2\_Медиана |  | 6.55 |
| Эритроциты\_E2\_Q1 - Q3 |  | 6.06 - 6.95 |
| Группа 2 | Возраст\_Количество субъектов |  | 50 |
| Возраст\_Количество (есть данные) |  | 50 |
| Возраст\_Нет данных |  | 0 |
| Возраст\_Ср. знач. |  | 31.26 |
| Возраст\_Станд. отклон. |  | 3.94 |
| Возраст\_95% ДИ для среднего |  | 3.94 |
| Возраст\_мин. - макс. |  | 21 - 42 |
| Возраст\_Медиана |  | 32 |
| Возраст\_Q1 - Q3 |  | 29 - 34 |
| Рост\_Количество субъектов |  | 50 |
| Рост\_Количество (есть данные) |  | 50 |
| Рост\_Нет данных |  | 0 |
| Рост\_Ср. знач. |  | 167.36 |
| Рост\_Станд. отклон. |  | 5.4 |
| Рост\_95% ДИ для среднего |  | 5.4 |
| Рост\_мин. - макс. |  | 159 - 177 |
| Рост\_Медиана |  | 167 |
| Рост\_Q1 - Q3 |  | 164 - 171 |
| Базофилы\_E1\_Количество субъектов |  | 50 |
| Базофилы\_E1\_Количество (есть данные) |  | 49 |
| Базофилы\_E1\_Нет данных |  | 1 |
| Базофилы\_E1\_Ср. знач. |  | 0.75 |
| Базофилы\_E1\_Станд. отклон. |  | 0.34 |
| Базофилы\_E1\_95% ДИ для среднего |  | 0.34 |
| Базофилы\_E1\_мин. - макс. |  | 0.15 - 1.72 |
| Базофилы\_E1\_Медиана |  | 0.72 |
| Базофилы\_E1\_Q1 - Q3 |  | 0.49 - 0.95 |
| Эозинофилы\_E1\_Количество субъектов |  | 50 |
| Эозинофилы\_E1\_Количество (есть данные) |  | 47 |
| Эозинофилы\_E1\_Нет данных |  | 3 |
| Эозинофилы\_E1\_Ср. знач. |  | 4.57 |
| Эозинофилы\_E1\_Станд. отклон. |  | 2.03 |
| Эозинофилы\_E1\_95% ДИ для среднего |  | 2.03 |
| Эозинофилы\_E1\_мин. - макс. |  | 0.14 - 8.43 |
| Эозинофилы\_E1\_Медиана |  | 4.76 |
| Эозинофилы\_E1\_Q1 - Q3 |  | 3.49 - 6.11 |
| Гемоглобин\_E1\_Количество субъектов |  | 50 |
| Гемоглобин\_E1\_Количество (есть данные) |  | 50 |
| Гемоглобин\_E1\_Нет данных |  | 0 |
| Гемоглобин\_E1\_Ср. знач. |  | 11.77 |
| Гемоглобин\_E1\_Станд. отклон. |  | 1.4 |
| Гемоглобин\_E1\_95% ДИ для среднего |  | 1.4 |
| Гемоглобин\_E1\_мин. - макс. |  | 8.85 - 15.13 |
| Гемоглобин\_E1\_Медиана |  | 11.71 |
| Гемоглобин\_E1\_Q1 - Q3 |  | 10.77 - 13.01 |
| Эритроциты\_E1\_Количество субъектов |  | 50 |
| Эритроциты\_E1\_Количество (есть данные) |  | 50 |
| Эритроциты\_E1\_Нет данных |  | 0 |
| Эритроциты\_E1\_Ср. знач. |  | 4.04 |
| Эритроциты\_E1\_Станд. отклон. |  | 0.66 |
| Эритроциты\_E1\_95% ДИ для среднего |  | 0.66 |
| Эритроциты\_E1\_мин. - макс. |  | 2.84 - 5.26 |
| Эритроциты\_E1\_Медиана |  | 4.03 |
| Эритроциты\_E1\_Q1 - Q3 |  | 3.53 - 4.53 |
| Базофилы\_E2\_Количество субъектов |  | 50 |
| Базофилы\_E2\_Количество (есть данные) |  | 50 |
| Базофилы\_E2\_Нет данных |  | 0 |
| Базофилы\_E2\_Ср. знач. |  | 1.14 |
| Базофилы\_E2\_Станд. отклон. |  | 0.36 |
| Базофилы\_E2\_95% ДИ для среднего |  | 0.36 |
| Базофилы\_E2\_мин. - макс. |  | 0.38 - 2.12 |
| Базофилы\_E2\_Медиана |  | 1.11 |
| Базофилы\_E2\_Q1 - Q3 |  | 0.89 - 1.35 |
| Эозинофилы\_E2\_Количество субъектов |  | 50 |
| Эозинофилы\_E2\_Количество (есть данные) |  | 48 |
| Эозинофилы\_E2\_Нет данных |  | 2 |
| Эозинофилы\_E2\_Ср. знач. |  | 5.48 |
| Эозинофилы\_E2\_Станд. отклон. |  | 2.15 |
| Эозинофилы\_E2\_95% ДИ для среднего |  | 2.15 |
| Эозинофилы\_E2\_мин. - макс. |  | 0.36 - 9.45 |
| Эозинофилы\_E2\_Медиана |  | 5.75 |
| Эозинофилы\_E2\_Q1 - Q3 |  | 4.25 - 7.08 |
| Гемоглобин\_E2\_Количество субъектов |  | 50 |
| Гемоглобин\_E2\_Количество (есть данные) |  | 50 |
| Гемоглобин\_E2\_Нет данных |  | 0 |
| Гемоглобин\_E2\_Ср. знач. |  | 12.49 |
| Гемоглобин\_E2\_Станд. отклон. |  | 1.4 |
| Гемоглобин\_E2\_95% ДИ для среднего |  | 1.4 |
| Гемоглобин\_E2\_мин. - макс. |  | 9.57 - 15.85 |
| Гемоглобин\_E2\_Медиана |  | 12.43 |
| Гемоглобин\_E2\_Q1 - Q3 |  | 11.49 - 13.73 |
| Эритроциты\_E2\_Количество субъектов |  | 50 |
| Эритроциты\_E2\_Количество (есть данные) |  | 50 |
| Эритроциты\_E2\_Нет данных |  | 0 |
| Эритроциты\_E2\_Ср. знач. |  | 6.36 |
| Эритроциты\_E2\_Станд. отклон. |  | 0.66 |
| Эритроциты\_E2\_95% ДИ для среднего |  | 0.66 |
| Эритроциты\_E2\_мин. - макс. |  | 5.16 - 7.58 |
| Эритроциты\_E2\_Медиана |  | 6.34 |
| Эритроциты\_E2\_Q1 - Q3 |  | 5.84 - 6.85 |

Сформируем статистическую таблицу для категориальных переменных.

data %>%   
 select(`Группа`, where(is.character)) %>%  
 mutate(`Группа крови` = `Группа крови` %>% replace\_na("Нет данных") %>% as.factor()) %>%  
 count(`Группа`, `Группа крови`) %>%  
 group\_by(`Группа`) %>%  
 mutate(`Процент по группе` = (n / sum(n)) %>% round(4) %>% `\*`(100) %>% str\_c("%")) %>%  
 ungroup() %>%  
 mutate(`Процент по выборке` = (n / sum(n)) %>% round(4) %>% `\*`(100) %>% str\_c("%")) %>%  
 flextable() %>%  
 merge\_v('Группа')

| Группа | Группа крови | n | Процент по группе | Процент по выборке |
| --- | --- | --- | --- | --- |
| Группа 1 | A (II) | 16 | 32% | 16% |
| AB (IV) | 4 | 8% | 4% |
| B (III) | 11 | 22% | 11% |
| O (I) | 10 | 20% | 10% |
| Нет данных | 9 | 18% | 9% |
| Группа 2 | A (II) | 18 | 36% | 18% |
| AB (IV) | 4 | 8% | 4% |
| B (III) | 5 | 10% | 5% |
| O (I) | 15 | 30% | 15% |
| Нет данных | 8 | 16% | 8% |

data %>%  
 tbl\_summary(by = "Группа", missing\_text = "Нет данных") %>%   
 add\_p()

| Characteristic | Группа 1, N = 501 | Группа 2, N = 501 | p-value2 |
| --- | --- | --- | --- |
| Возраст | 29.0 (27.0, 32.0) | 32.0 (29.0, 34.0) | 0.008 |
| Пол |  |  | 0.028 |
| Женский | 32 (64%) | 21 (42%) |  |
| Мужской | 18 (36%) | 29 (58%) |  |
| Рост | 169.0 (164.0, 172.0) | 167.0 (164.0, 171.0) | 0.4 |
| Группа крови |  |  | 0.3 |
| A (II) | 16 (39%) | 18 (43%) |  |
| AB (IV) | 4 (9.8%) | 4 (9.5%) |  |
| B (III) | 11 (27%) | 5 (12%) |  |
| O (I) | 10 (24%) | 15 (36%) |  |
| Нет данных | 9 | 8 |  |
| Базофилы\_E1 | 0.56 (0.34, 0.84) | 0.72 (0.49, 0.95) | 0.043 |
| Нет данных | 3 | 1 |  |
| Эозинофилы\_E1 | 3.16 (2.29, 4.25) | 4.76 (3.49, 6.11) | <0.001 |
| Нет данных | 1 | 3 |  |
| Гемоглобин\_E1 | 11.73 (10.65, 13.27) | 11.71 (10.77, 13.01) | 0.6 |
| Эритроциты\_E1 | 4.23 (3.74, 4.63) | 4.03 (3.53, 4.53) | 0.3 |
| Базофилы\_E2 | 0.91 (0.73, 1.24) | 1.11 (0.89, 1.35) | 0.024 |
| Эозинофилы\_E2 | 4.11 (3.20, 5.20) | 5.75 (4.25, 7.08) | <0.001 |
| Нет данных | 0 | 2 |  |
| Гемоглобин\_E2 | 12.45 (11.37, 13.99) | 12.43 (11.49, 13.73) | 0.6 |
| Эритроциты\_E2 | 6.55 (6.06, 6.95) | 6.34 (5.84, 6.85) | 0.3 |
| 1Median (IQR); n (%) | | | |
| 2Wilcoxon rank sum test; Pearson's Chi-squared test; Fisher's exact test; Wilcoxon rank sum exact test | | | |