

Итоговый проект.Хабиров Булат Альбертович

Исследуемые данные

В данном отчете приведена подробная информация о базе данных PISA 2018. База состоит из число наблюдений и число переменных, среди которых число 3747 юношей и 3861 девушек.

Международная программа по оценке образовательных достижений обучающихся PISA (*Programme for International Student Assessment*) является мониторинговым исследованием качества общего образования, которое отвечает на вопрос: «Обладают ли обучающиеся 15-летнего возраста, получающие обязательное общее образование, знаниями и умениями, необходимыми им для полноценного функционирования в современном обществе: решения широкого диапазона задач в различных сферах человеческой деятельности, общения и социальных отношений?» Исследование проводится под эгидой Организации экономического сотрудничества и развития (ОЭСР) трехлетними циклами, начиная с 2000 года.

В исследовании оцениваются компетенции обучающихся по читательской, математической, естественно-научной грамотности, а также дополнительных компетенций, более подробную информацию о которых можно [посмотреть здесь](#).

Данные PISA 2018 изучались с помощью открытого пакета программ RStudio и пакета ggplot (<https://ggplot2.tidyverse.org/>).

```
knitr::opts_chunk$set(echo = T,
                      results = "markup")
pisa_data <- readRDS("Pisa_Russia.rds")
pisa_data <- pisa_data %>%
  filter(!is.na(REPEAT)&!is.na(BELONG)&!is.na(MASTGOAL)&!is.na(WORKMAST)&!is.na(RESILIENCE)&!is.na(COMPETE)&!is.na(GFOFAIL)&!is.na(ATT
LNACT))

sample_data <- pisa_data %>% sample_n(200)

selected_variables <- sample_data %>%
  select(REPEAT, BELONG, MASTGOAL, WORKMAST, RESILIENCE, COMPETE, GFOFAIL, ATTLNACT) %>%
  mutate(
    Mathematics = (sample_data$PV1MATH + sample_data$PV2MATH + sample_data$PV3MATH + sample_data$PV4MATH + sample_data$PV5MATH
+ sample_data$PV6MATH + sample_data$PV7MATH + sample_data$PV8MATH + sample_data$PV9MATH + sample_data$PV10MATH) / 10,
    Reading = (sample_data$PV1READ + sample_data$PV2READ + sample_data$PV3READ + sample_data$PV4READ + sample_data$PV5READ + sa
mple_data$PV6READ + sample_data$PV7READ + sample_data$PV8READ + sample_data$PV9READ + sample_data$PV10READ) / 10,
    Science = (sample_data$PV1SCIE + sample_data$PV2SCIE + sample_data$PV3SCIE + sample_data$PV4SCIE + sample_data$PV5SCIE + sample_
data$PV6SCIE + sample_data$PV7SCIE + sample_data$PV8SCIE + sample_data$PV9SCIE + sample_data$PV10SCIE) / 10,
    Gender = factor(ifelse(sample_data$ST004D01T==1,"female","male"))
  )
selected_variables
```

Описательный анализ.Вариант 1.

```
df <- selected_variables %>%
  select(BELONG, MASTGOAL, WORKMAST, Mathematics)
describe(df)
```

```
##      vars  n mean  sd median trimmed  mad   min   max range
## BELONG    1 200 -0.48 0.79 -0.57 -0.55 0.54 -3.24  2.72  5.96
## MASTGOAL   2 200 -0.22 1.06 -0.43 -0.29 1.16 -2.53  1.85  4.38
## WORKMAST   3 200 -0.32 0.82 -0.26 -0.39 0.76 -2.74  1.82  4.55
## Mathematics 4 200 498.69 72.09 501.78 499.68 74.42 321.30 683.33 362.03
##      skew kurtosis  se
## BELONG    1.13    4.93 0.06
## MASTGOAL   0.36   -0.45 0.08
## WORKMAST   0.77    1.04 0.06
## Mathematics -0.09   -0.42 5.10
```

```
## $BELONG
##
## Shapiro-Wilk normality test
##
## data: x
## W = 0.87777, p-value = 0.0000000001194
##
##
## $MASTGOAL
##
## Shapiro-Wilk normality test
##
## data: x
## W = 0.95301, p-value = 0.000003725
##
##
## $WORKMAST
##
## Shapiro-Wilk normality test
##
## data: x
## W = 0.91683, p-value = 0.00000003432
##
##
## $Mathematics
##
## Shapiro-Wilk normality test
##
## data: x
## W = 0.99446, p-value = 0.6689
```

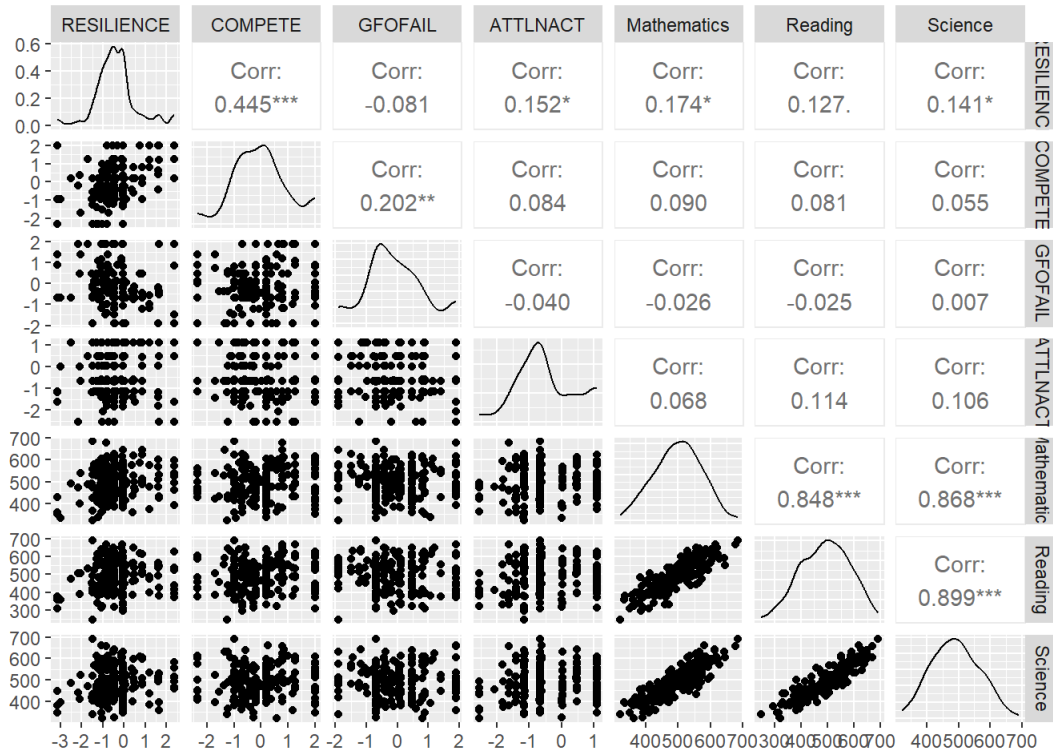
Сравнительный анализ.

```
t_test_result <- t.test(selected_variables[selected_variables$Gender=="female"],$Mathematics,selected_variables[selected_variables$Gender=="male"],$
Mathematics)
```

```
t_test_result
```

```
##
## Welch Two Sample t-test
##
## data: selected_variables[selected_variables$Gender == "female", ]$Mathematics and selected_variables[selected_variables$Gender == "male", ]$Mathematics
## t = 1.2479, df = 196.76, p-value = 0.2135
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -7.379589 32.813923
## sample estimates:
## mean of x mean of y
## 504.9252 492.2080
```

Корреляционный анализ



Выводы

По описательному анализу мы можем сделать вывод о том, что по критерию Шапиро-Уилка переменные BELONG, MASGOAL, WORKMAST из 200 наблюдений - не являются распределенными нормально так как их значение $p\text{-value} < 0.05$

Mathematics может быть принята как нормально распределенная, так как уровень значимости выше 0.05

При сравнительном анализе мы видим по столбчатой диаграмме, то что выборочное среднее по математике мальчиков чуть ниже девочек, возможно, это говорит о том что девочки написали лучше тест по математике.

На основе результатов теста Стьюдента можно сделать вывод, что существует статистически значимое различие в средних значениях математики между девочками и мальчиками. Среднее значение математики у девочек выше, чем у мальчиков, и разница составляет от 1.192029 до 43.961363 в пользу девочек.

На корреляционной матрице мы видим, что более сильная линейная зависимости существует у переменных Mathematics, Reading, Science.