

Zadanie Holík

Marek Holík

```
#system2("open", args="zadanie_Holík.pdf", wait = FALSE)
```

```
data <- read_excel("toughestsport.xlsx")
head(data, 10)
```

```
# A tibble: 10 x 15
```

	SPORT	ENDURANCE	STRENGTH	POWER	SPEED	AGILITY	FLEXIBILITY	NERVE	DURABILITY
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	Boxing	8.63	8.13	8.63	6.38	6.25	4.38	8.88	8.5
2	Ice Hockey	7.25	7.13	7.88	7.75	7.63	4.88	6	8.25
3	Football	5.38	8.63	8.13	7.13	6.38	4.38	7.25	8.5
4	Basketball	7.38	6.25	6.5	7.25	8.13	5.63	4.13	7.75
5	Wrestling	6.63	8.38	7.13	5.13	6.38	7.5	5	6.75
6	Martial A~	5	5.88	7.75	6.38	6	7	6.63	5.88
7	Tennis	7.25	5.13	7.13	6.75	7.75	5.63	3	5
8	Gymnastics	5.38	6.13	6.63	5	6.38	10	7.5	6.88
9	Baseball/~	4.63	5.75	7.63	6.5	6.75	4.75	5.13	5.63
10	Soccer	7.75	4.5	5.13	7.25	8.25	4.75	3.63	6.25

```
# i 6 more variables: `HAND-EYE-CORD` <dbl>, `ANALYTIC APTITUDE` <dbl>,
```

```
# `FIELD CATEGORY` <chr>, `CONTACT CATEGORY` <chr>, TOTAL <dbl>, RANK <dbl>
```

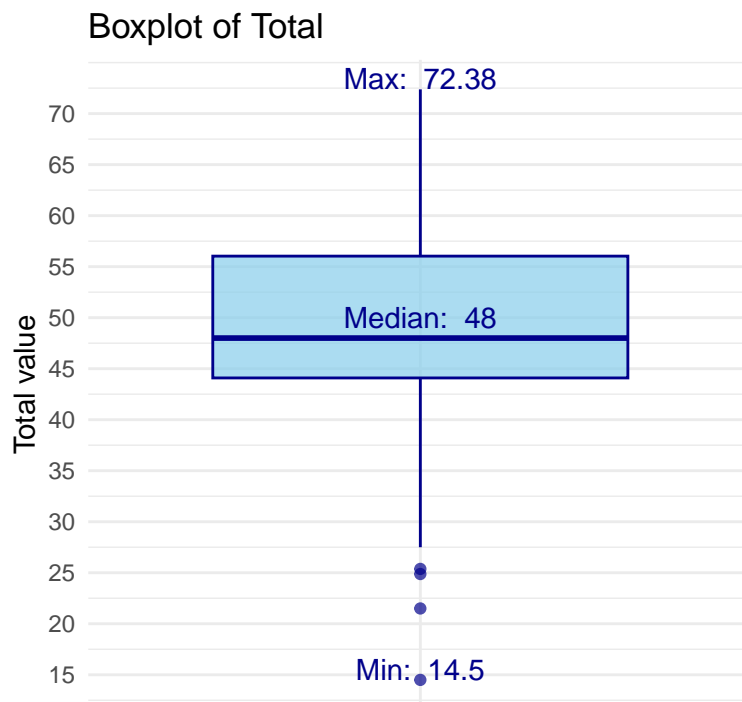
Ako projekt na štatistické analyzovanie som si vybral dataset o športoch, ktorý obsahuje zoznam 60 športov a ich fyziologické nároky. Dataset obsahuje 14 premenných, kde 10 z ich obsahujú číselné hodnoty reprezentujúce: vytrvalosť, silu, explozivitu, rýchlosť, obratnosť, flexibilitu, nervovú odolnosť, výdrž silového zaťaženia, kordináciu a analytické schopnosti. 2 premenné obsahujú kategóriu športu a jeho kontaktnosť. Posledné premenné obsahujú celkové hodnotenie športu vypočítané súčtom všetkých predchádzajúcich hodnôt a výslednú hodnotu v rebríčku.

1. Využitím popisnej štatistiky charakterizujte premennú Total. Výsledky zobrazte graficky a popíšte aj použité grafy. Ktorý Field category je v súbore najviac zastúpený?

```

ggplot(data, aes(x = " ", y = TOTAL)) +
  geom_boxplot(fill = "skyblue", color = "darkblue", alpha = 0.7) +
  annotate("text", x = 1, y = min(data$TOTAL), label = paste("Min: ", round(min(data$TOTAL), 1))) +
  annotate("text", x = 1, y = median(data$TOTAL)+1, label = paste("Median: ", round(median(data$TOTAL), 1))) +
  annotate("text", x = 1, y = max(data$TOTAL), label = paste("Max: ", round(max(data$TOTAL), 1))) +
  scale_y_continuous(breaks = seq(0, max(data$TOTAL), by = 5)) +
  labs(title = "Boxplot of Total", x = " ",
       y = "Total value") +
  theme_minimal() +
  theme(axis.text.x = element_blank(), axis.ticks.x = element_blank())

```



```

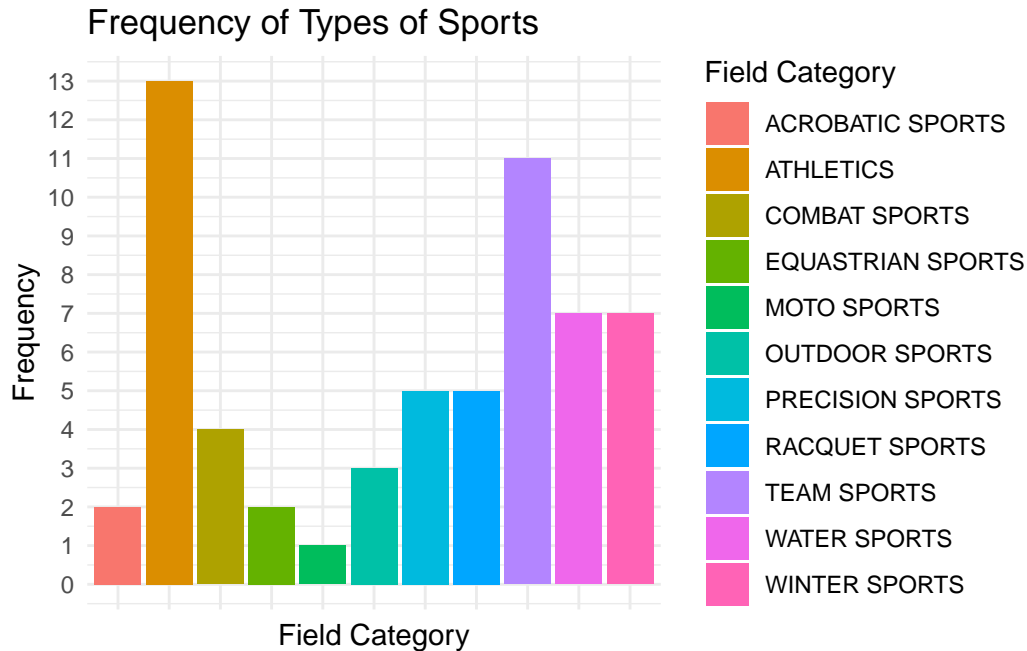
category_counts <- table(data$"FIELD CATEGORY")

plot_data <- data.frame(Field_Category = names(category_counts), Frequency = as.numeric(category_counts))

ggplot(plot_data, aes(x = Field_Category, y = Frequency, fill = Field_Category)) +
  geom_bar(stat = "identity", ) +
  labs(title = "Frequency of Types of Sports", x = "Field Category", y = "Frequency", fill = "Field Category")

```

```
scale_y_continuous(breaks = seq(0, max(plot_data$Frequency), by = 1)) +
theme_minimal() +
theme(axis.text.x = element_blank(), axis.ticks.x = element_blank())
```



2. Vhodným testom overte na hladine významnosti 0.1, či je Analytic aptitude najviac 4. Porovnajte Analytic aptitude, vzhľadom na to, či ide o kontaktný alebo bezkontaktný šport. Je tento rozdiel štatisticky významný na hladine významnosti 0.01?

```
t_test_result <- t.test(data$"ANALYTIC APTITUDE"[data$"CONTACT CATEGORY" == "YES"], mu = 4)
```

3. Existuje štatisticky významný rozdiel v premennej Total vzhľadom na Field category (berme do úvahy len team sports, winter sports a precision sports) alebo vzhľadom na to či je Durability pod 5 alebo 5 a viac? Overte vhodným testom a aj graficky. Ak rozdiel existuje, tak medzi ktorými? Existuje aj interakcia medzi premennými?

```
# Subset data for relevant categories
subset_data <- subset(data, data$"FIELD CATEGORY" %in% c("TEAM SPORTS", "WINTER SPORTS", "PRECISION SPORTS"))
head(subset_data, 10)
```

```
# A tibble: 10 x 15
```

```
SPORT      ENDURANCE STRENGTH POWER SPEED AGILITY FLEXIBILITY NERVE DURABILITY
```

```

      <chr>      <dbl>   <dbl> <dbl> <dbl>   <dbl>      <dbl> <dbl>      <dbl>
1 Ice Hockey    7.25    7.13 7.88 7.75    7.63      4.88 6        8.25
2 Football     5.38    8.63 8.13 7.13    6.38      4.38 7.25     8.5
3 Basketball    7.38    6.25 6.5 7.25    8.13      5.63 4.13     7.75
4 Soccer        7.75    4.5 5.13 7.25    8.25      4.75 3.63     6.25
5 Skiing: A~    5.13    5.25 6    7.38    6.13      5.63 8.38     6
6 Water Polo    7.88    6.63 6.88 5.38    6.38      5    4.25     6.38
7 Rugby         6.75    7    6.38 5.88    6        4.13 6.5      7.88
8 Lacrosse      6.63    5.13 5.75 7        6.63      4.75 4.38     6.13
9 Field Hoc~    6.75    4.5 5.38 6        5.75      4.63 3.75     5
10 Speed Ska~   7.63    7.25 7.38 8.88    4        4.25 4.5      4.63
# i 6 more variables: `HAND-EYE-CORD` <dbl>, `ANALYTIC APTITUDE` <dbl>,
#   `FIELD CATEGORY` <chr>, `CONTACT CATEGORY` <chr>, TOTAL <dbl>, RANK <dbl>

```

```
# ANOVA test
```

```
anova_result <- aov(subset_data$TOTAL ~ subset_data$`FIELD CATEGORY` * subset_data$DURABILITY)
```

```
summary(anova_result)
```

```

              Df Sum Sq Mean Sq F value
subset_data$`FIELD CATEGORY`      2   3796   1898.0 113.809
subset_data$DURABILITY             1    816    815.6  48.905
subset_data$`FIELD CATEGORY`:subset_data$DURABILITY  2     33     16.4   0.985
Residuals                        17    284     16.7

              Pr(>F)
subset_data$`FIELD CATEGORY` 1.43e-10 ***
subset_data$DURABILITY       2.17e-06 ***
subset_data$`FIELD CATEGORY`:subset_data$DURABILITY 0.394
Residuals
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

4. Je medzi Endurance a Nerve bezkontaktných športov štatisticky významný vzťah? Ak áno, popíšte slovne ako sa prejavuje a popíšte ho vhodným modelom. Vhodnosť modelu overte.