World Happiness Report

X Æ A-Xii

6.12.2021.

Motivacija i opis problema

World Happiness Report je publikacija Mreže rješenja za održivi razvoj Ujedinjenih naroda koja sadrži podatke o osjećaju sreće pojedinih nacija. Podatci su dobiveni kroz ankete koje provode Gallup i Lloyd's Register Foundation. Prvi je izvještaj objavljen 2012. godine, a od 2016. se objavljuje na Međunarodni dan sreće 20. ožujka.

Učitavanje podataka o svjetskom bogatstvu 2021. godine

```
wealth_data <- read_excel("E:/FER/Statistička analiza podataka/Projekt/files/credit_suisse_global_wealth
```

```
## New names:
## * `` -> ...6
## * `` -> ...7
## * `` -> ...8
## * `` -> ...9
```

```
dim(wealth_data)
```

```
## [1] 169  10
```

```
head(wealth_data)
```

```
## # A tibble: 6 x 10
##   `Country name` `Adults (thousands)` `Mean wealth per adu~ `Median wealth per ~
##   <chr>                         <dbl>                 <dbl>                 <dbl>
## 1 <NA>                             NA                    NA                    NA
## 2 Afghanistan                   18356                  1744                   734
## 3 Albania                        2187                 30524                 15363
## 4 Algeria                       27620                  8871                  2302
## 5 Angola                        14339                  3529                  1131
## 6 Argentina                     30799                  7224                  2157
## # ... with 6 more variables:
## #   Distribution of adults (%) by wealth range (USD) <chr>, ...6 <chr>,
## #   ...7 <chr>, ...8 <chr>, ...9 <chr>, Gini (%) <dbl>
```

Učitavanje podataka o globalnoj sreći 2020. godine

You can also embed plots, for example:

```
whr2020_data <- read_excel("E:/FER/Statistička analiza podataka/Projekt/files/WHR_2020.xlsx")

dim(whr2020_data)
```

```
## [1] 153   9
```

```
head(whr2020_data)
```

```
## # A tibble: 6 x 9
##   `Country name` `Regional indicator` `Ladder score` `Logged GDP per capita`
##   <chr>          <chr>                         <dbl>                   <dbl>
## 1 Finland        Western Europe                 7.81                    10.6
## 2 Denmark        Western Europe                 7.65                    10.8
## 3 Switzerland    Western Europe                 7.56                    11.0
## 4 Iceland        Western Europe                 7.50                    10.8
## 5 Norway         Western Europe                 7.49                    11.1
## 6 Netherlands    Western Europe                 7.45                    10.8
## # ... with 5 more variables: Social support <dbl>,
## #   Healthy life expectancy <dbl>, Freedom to make life choices <dbl>,
## #   Generosity <dbl>, Perceptions of corruption <dbl>
```

Učitavanje podataka o globalnoj sreći 2021. godine

You can also embed plots, for example:

```
whr2021_data <- read_excel("E:/FER/Statistička analiza podataka/Projekt/files/WHR_2021.xlsx")

dim(whr2021_data)
```

```
## [1] 149  11
```

```
head(whr2021_data)
```

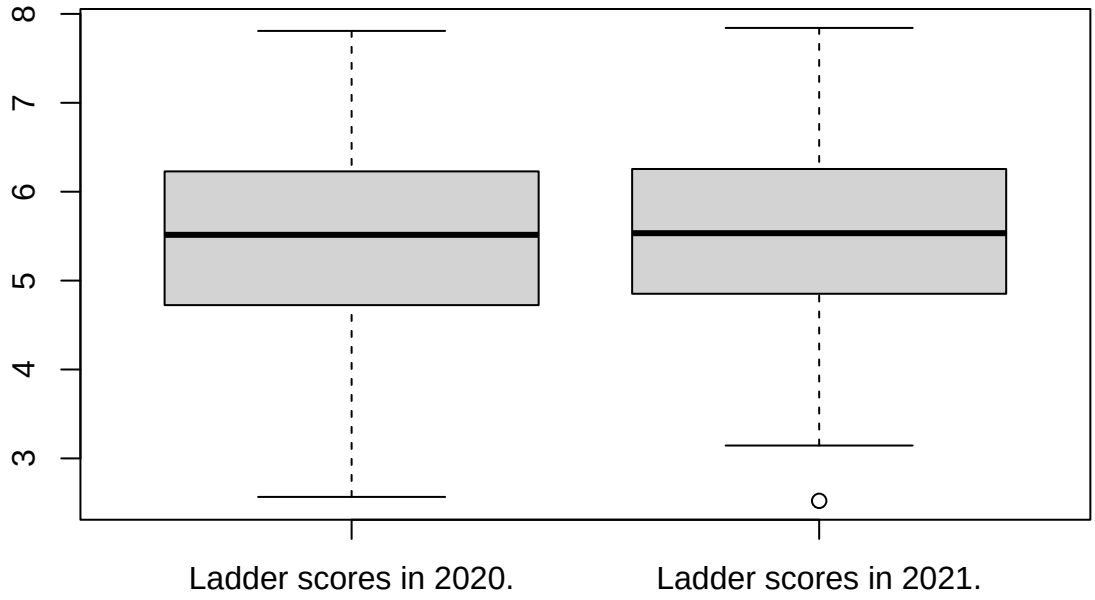
```
## # A tibble: 6 x 11
##   `Country name` `Regional indicator` `Ladder score` `Logged GDP per capita`
##   <chr>          <chr>                         <dbl>                   <dbl>
## 1 Finland        Western Europe                 7.84                    10.8
## 2 Denmark        Western Europe                 7.62                    10.9
## 3 Switzerland    Western Europe                 7.57                    11.1
## 4 Iceland        Western Europe                 7.55                    10.9
## 5 Netherlands    Western Europe                 7.46                    10.9
## 6 Norway         Western Europe                 7.39                    11.1
## # ... with 7 more variables: Social support <dbl>,
## #   Healthy life expectancy <dbl>, Freedom to make life choices <dbl>,
## #   Generosity <dbl>, Perceptions of corruption <dbl>, Income Gini <dbl>,
## #   Wealth Gini <dbl>
```

Je li razina sreće veća u 2020. ili 2021. godini?

Ovo pitanje ćemo provjeravati uparenim t-testom. Podaci koje koristimo su razlike rezultata WHR-a u 2021. i 2020. godini za iste države.

Prvo ćemo napraviti dva boxplota kako bi vizualizirali podatke za pojedinu godinu.

```
boxplot(whr2020_data$`Ladder score`, whr2021_data$`Ladder score`, names = c("Ladder scores in 2020.", "
```



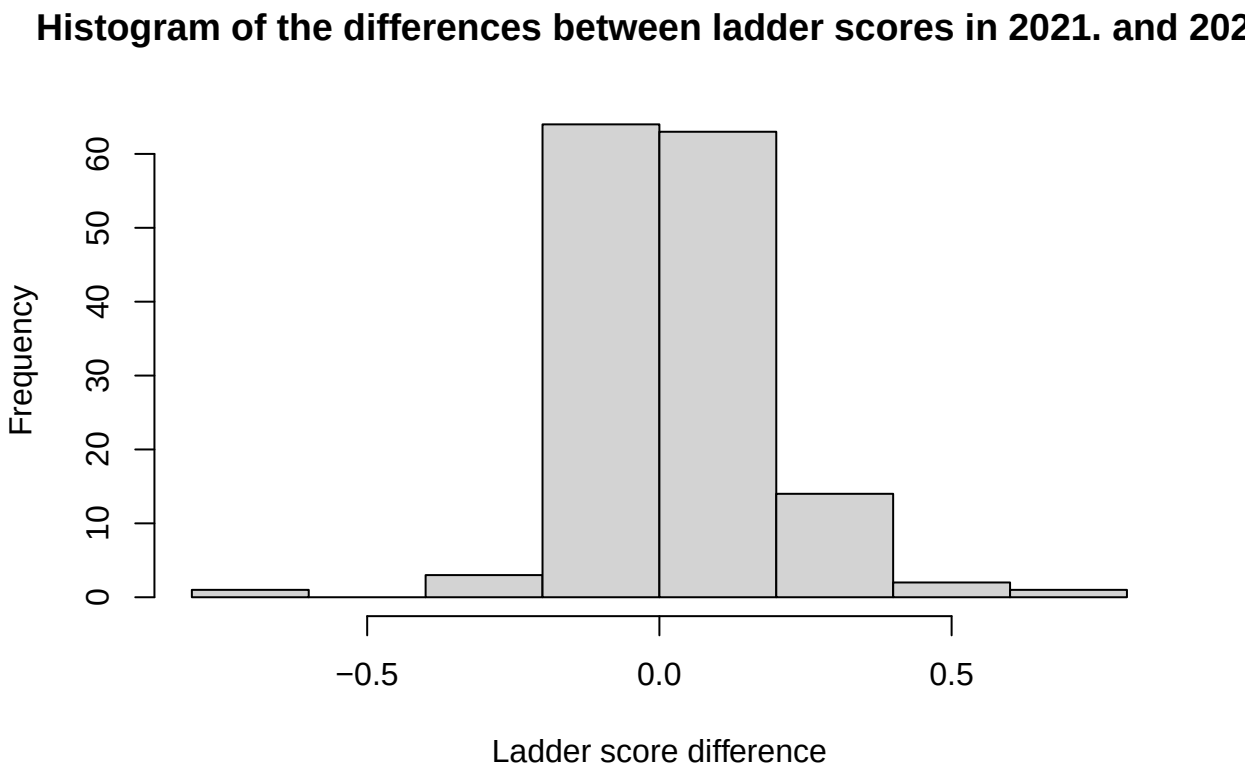
Vidimo da su srednje vrijednosti rezultata za obje godine otprilike jednake.

Sada ćemo prikazati razlike rezultata pomoću histograma kako bismo se uvjerili u normalnost podataka, budući da je to uvjet za provođenje uparenog t-testa.

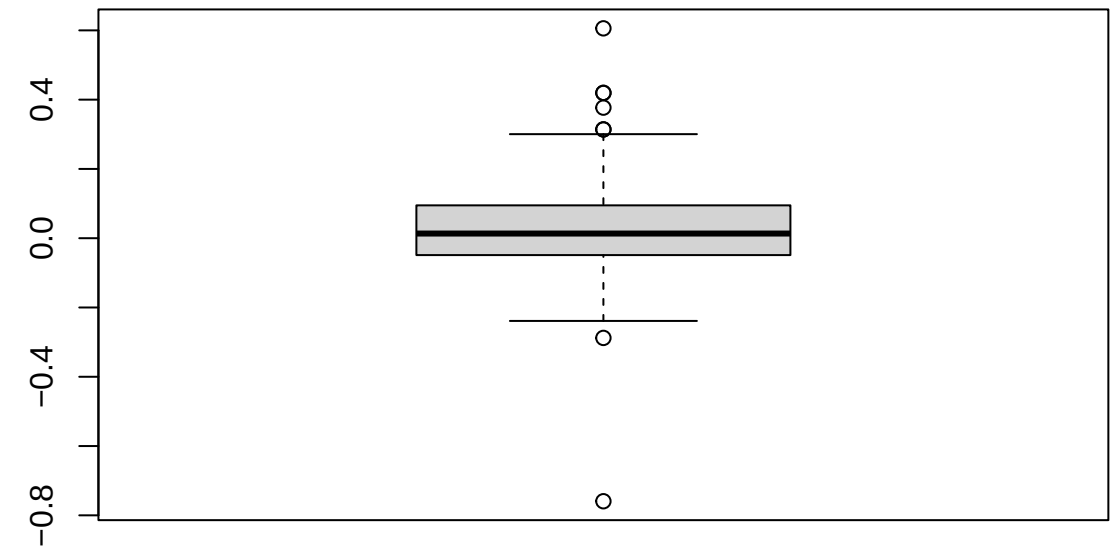
Također ćemo ih prikazati pomoću boxplota, da lakše uočimo stršeće vrijednosti.

```
whr_merged = merge(whr2021_data, whr2020_data, by="Country name")
ladderScore_differences = whr_merged$`Ladder score.x`- whr_merged$`Ladder score.y`

hist(ladderScore_differences, xlab="Ladder score difference", main="Histogram of the differences betwee
```




```
boxplot(ladderScore_differences, main="Boxplot of the differences between ladder scores in 2021. and 20
```



Mogli bismo otprilike reći da podaci jesu normalno distribuirani, no da su ipak više zbijeni oko sredine.

S ciljem povećavanja normalnosti, izbacit ćemo stršeće vrijednosti te ponovno prikazati histogram i boxplot dobivenih podataka.

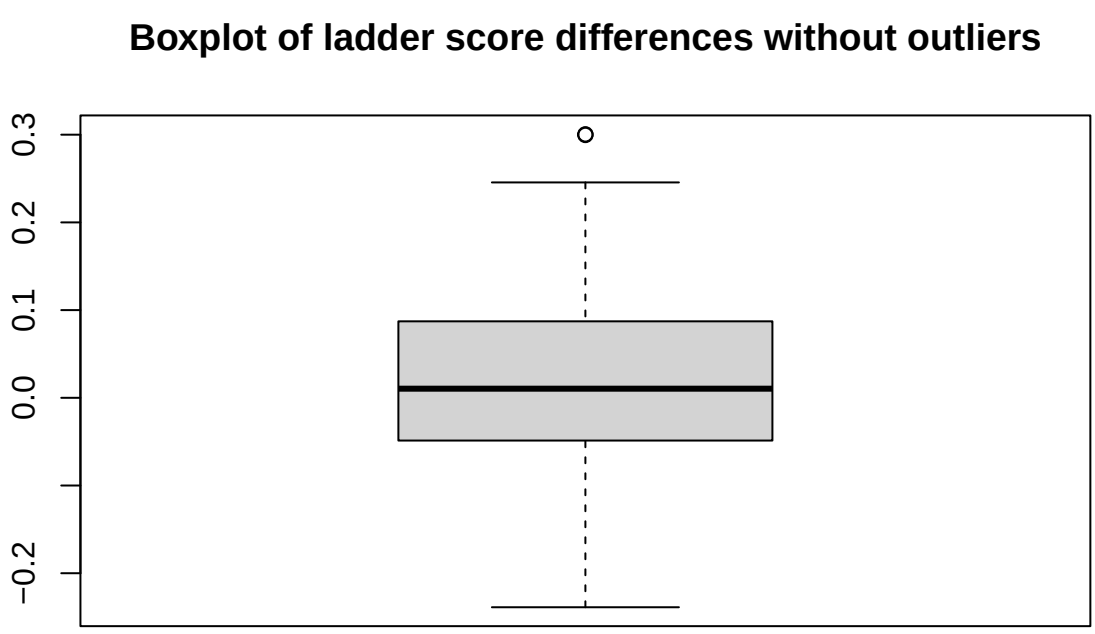
```
ladderScore_differences_no_outliers = ladderScore_differences[!ladderScore_differences %in% boxplot.sta

length(ladderScore_differences) - length(ladderScore_differences_no_outliers)
```

```
## [1] 8
```

```
hist(ladderScore_differences_no_outliers, xlab="Ladder score difference", main="Histogram of ladder sco
```

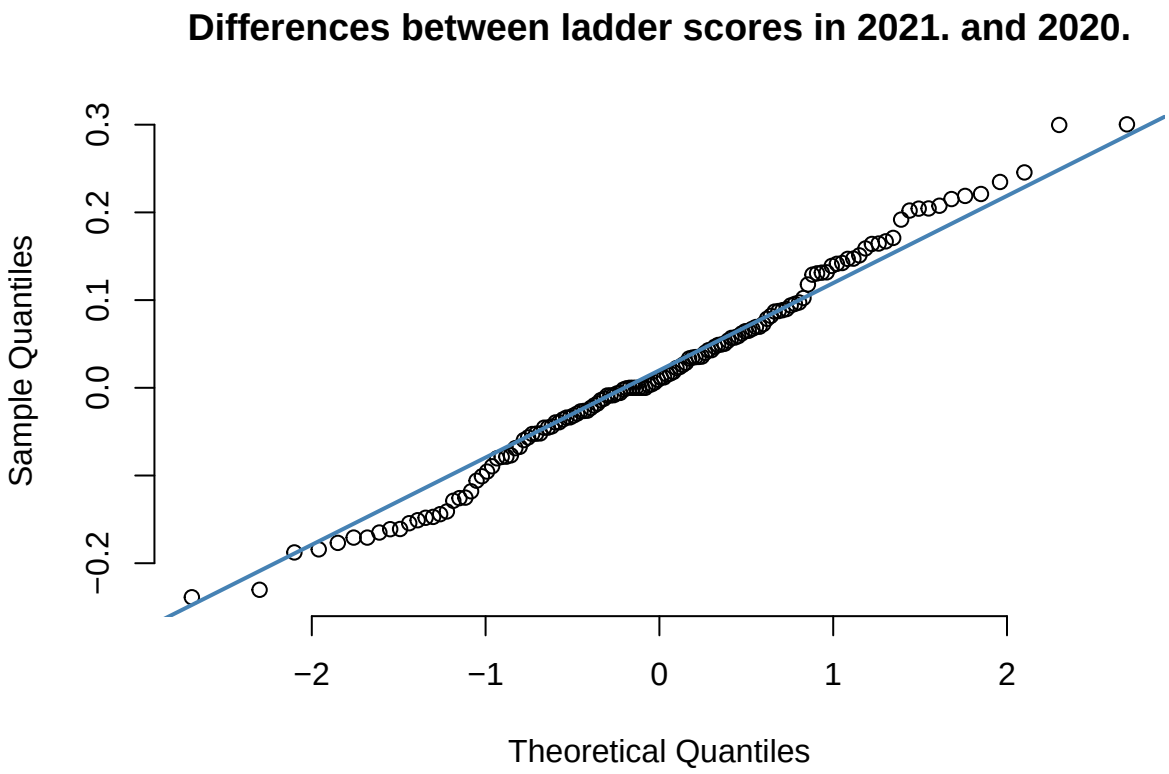

```
boxplot(ladderScore_differences_no_outliers, main="Boxplot of ladder score differences without outliers
```



Uočavamo da su nam sadašnji podaci distribuirani puno više normalno, nego što su prije bili, a izbacili smo samo 8 vrijednosti.

Da bismo se uvjerili u normalnost podataka, možemo je provjeriti i pomoću qq-plota.

```
qqnorm(ladderScore_differences_no_outliers, pch = 1, frame = FALSE, main="Differences between ladder sc
qqline(ladderScore_differences_no_outliers, col = "steelblue", lwd = 2)
```



Kao i histogram, qq-plot nam upućuje na normalnost podataka. Jeini podaci koji se ne ravnaju savršeno po normalnoj distribuciji su oni rubni.

Još da budemo sasvim sigurni u normalnost naših podataka, provest ćemo Kolmogorov-Smirnovljev test. Hipoteze su nam sljedeće:

$$H_0 : \text{podaci su normalno distribuirani}$$
$$H_1 : \text{podaci nisu normalno distribuirani}$$

```
ks.test(ladderScore_differences_no_outliers, "pnorm", mean(ladderScore_differences_no_outliers), sd(lad
```

```
## Warning in ks.test(ladderScore_differences_no_outliers, "pnorm",
## mean(ladderScore_differences_no_outliers), : ties should not be present for the
## Kolmogorov-Smirnov test
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  ladderScore_differences_no_outliers
## D = 0.04168, p-value = 0.9681
## alternative hypothesis: two-sided
```

Budući da je p-vrijednost znatno veća od 0.05, ne odbijamo nul hipotezu o normalnosti podataka te možemo krenuti s obostranim t-testom.

Hipoteze nam glase ovako:

$$H_0 : \mu_{2021} = \mu_{2020}$$
$$H_1 : \mu_{2021} \neq \mu_{2020}$$

```
t.test(whr_merged$`Ladder score.x`, whr_merged$`Ladder score.y`, paired=TRUE, alternative="two.sided",c
```

```
##
##  Paired t-test
##
## data:  whr_merged$`Ladder score.x` and whr_merged$`Ladder score.y`
## t = 2.0746, df = 147, p-value = 0.03977
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.001230481 0.050664122
## sample estimates:
## mean of the differences
##               0.0259473
```

Budući da smo dobili p-vrijednost manju od 0.05, odbijamo hipotezu o jednakosti rezultata WHR-a u 2020. i 2021. godini u korist alternativne hipoteze.

Provest ćemo još jedan t-test, no ovaj put jednostrani sa sljedećim hipotezama:

$$H_0 : \mu_{2021} <= \mu_{2020}$$
$$H_1 : \mu_{2021} > \mu_{2020}$$

```
t.test(whr_merged$`Ladder score.x`, whr_merged$`Ladder score.y`, paired=TRUE, alternative="greater",con
```

```
##
##  Paired t-test
##
## data:  whr_merged$`Ladder score.x` and whr_merged$`Ladder score.y`
## t = 2.0746, df = 147, p-value = 0.01988
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.005244588         Inf
## sample estimates:
## mean of the differences
##               0.0259473
```

Zbog p-vrijednosti manje od 0.05, odbijamo nul hipotezu i prihvaćemo alternativnu, odnosno da je razina sreće veća u 2021. nego što je bila u 2020. godini.