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>
```

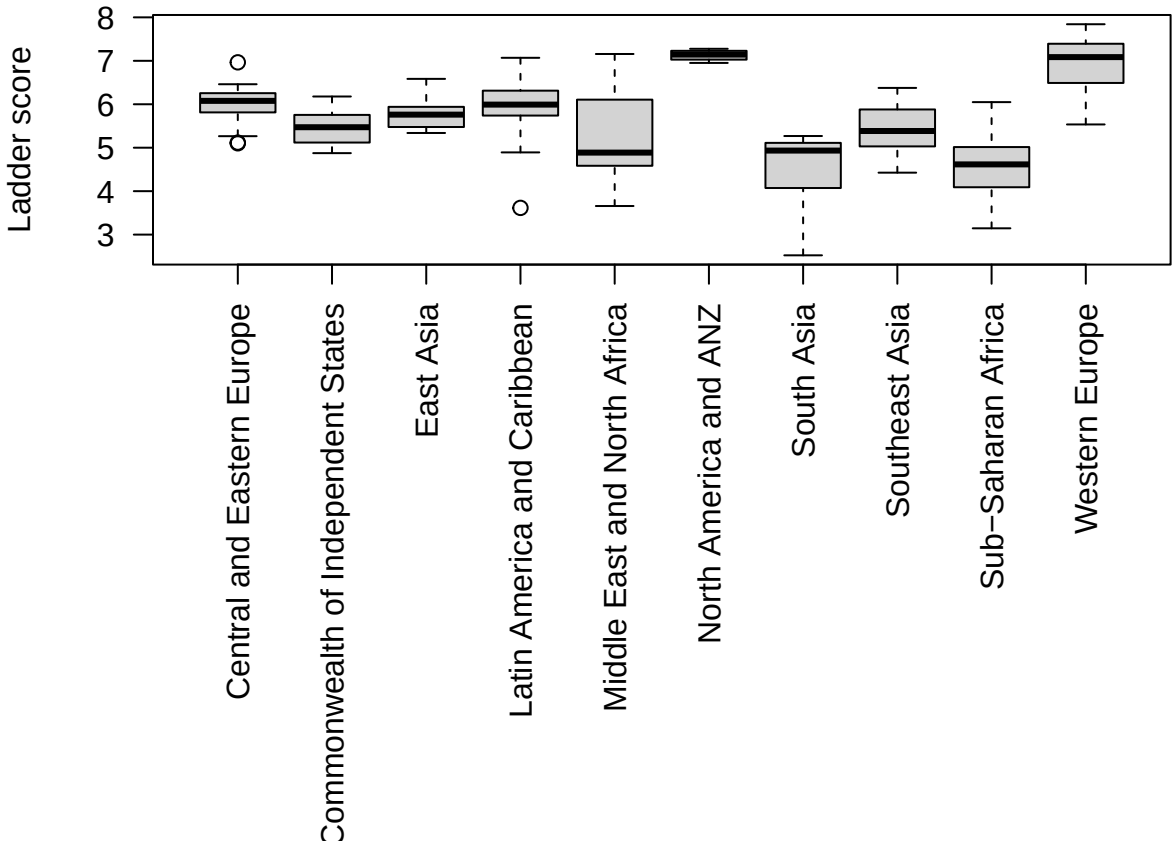
Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

Postoje li razlike u iskazanoj sreći medu različitim regijama?

Na ovo pitanje ćemo odgovoriti korištenjem jednofaktorskom ANOVA metodom.

U sljedećem isječku ćemo prikazati box plot dijagrame sreće po pojedinim regijama.

```
par(mar=c(15,5,1,1))
boxplot(`Ladder score`~`Regional indicator`,data = whr2021_data, las = 2, xlab = "" )
```



Boxplot nas upućuje da postoje razlike u iskazanim srećama po regijama. To ćemo potvrditi ANOVA metodom.

Uvjeti za ANOVA-u su normalnost i nezavisnost podataka, te homogenost varijanci među regijama. Nezavisnost podataka možemo pretpostaviti. Normalnost podataka po regijama ćemo provjeriti s Kolmogorov-Smirnovim testom.Hipoteze su nam sljedeće:

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

te

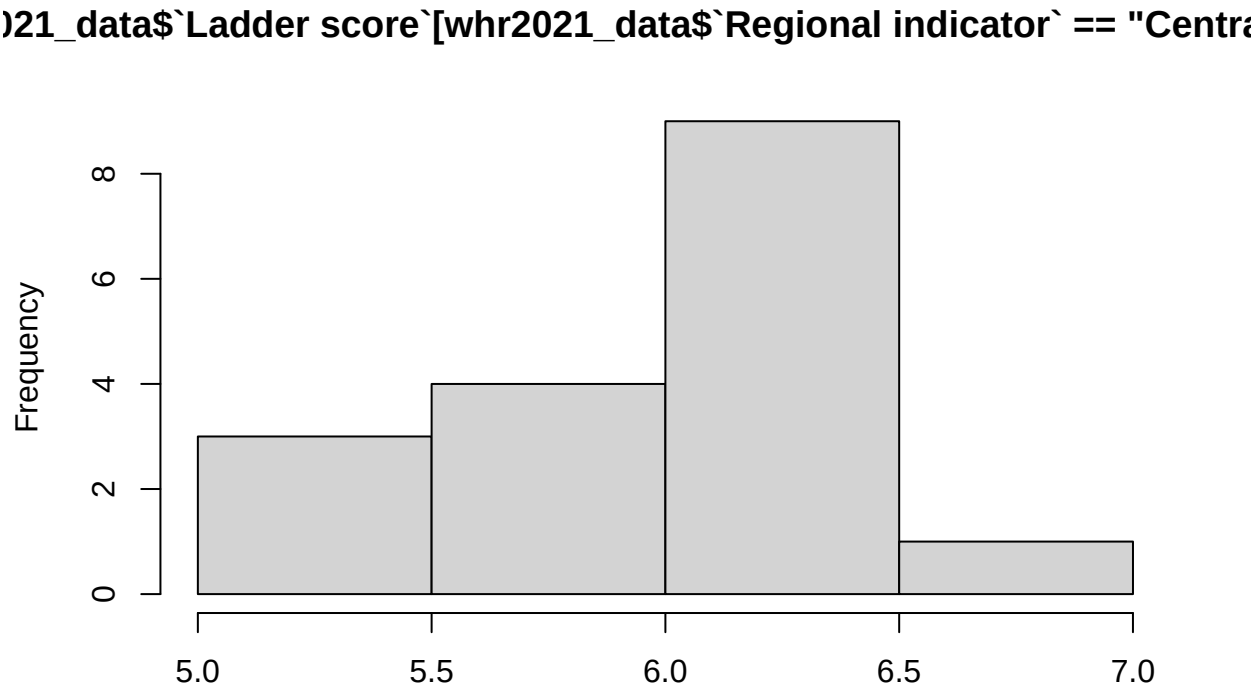
$$\alpha = 0.05$$

```
ks.test(whr2021_data$`Ladder score`[whr2021_data$`Regional indicator`=='Central and Eastern Europe'], "
```

```
##
##  One-sample Kolmogorov-Smirnov test
```

```
##
## data:  whr2021_data$‘Ladder score‘[whr2021_data$‘Regional indicator‘ == "Central and Eastern Europe"]
## D = 0.15266, p-value = 0.7689
## alternative hypothesis: two-sided
```

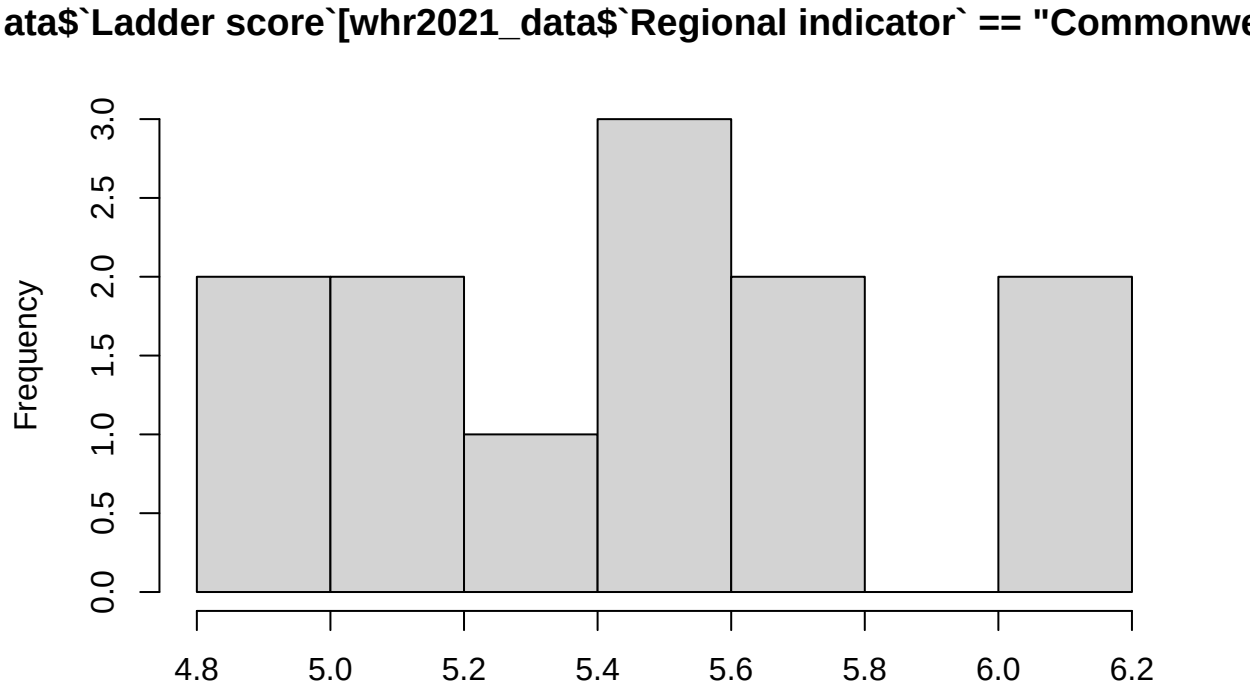
```
hist(whr2021_data$`Ladder score`[whr2021_data$`Regional indicator`=='Central and Eastern Europe'])
```




```
ks.test(whr2021_data$`Ladder score`[whr2021_data$`Regional indicator`=='Commonwealth of Independent Sta
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  whr2021_data$‘Ladder score‘[whr2021_data$‘Regional indicator‘ == "Commonwealth of Independent
## D = 0.1077, p-value = 0.9962
## alternative hypothesis: two-sided
```

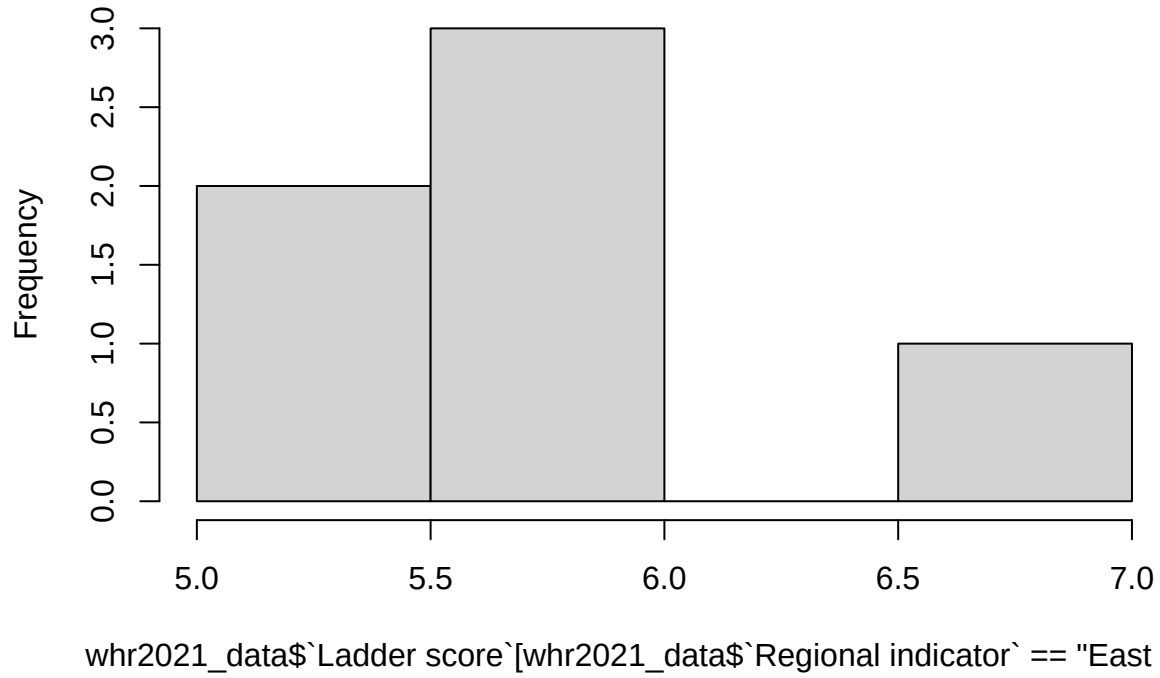
```
hist(whr2021_data$`Ladder score`[whr2021_data$`Regional indicator`=='Commonwealth of Independent States
```




```
ks.test(whr2021_data$`Ladder score`[whr2021_data$`Regional indicator`=='East Asia'], "pnorm", mean(whr2
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  whr2021_data$‘Ladder score‘[whr2021_data$‘Regional indicator‘ == "East Asia"]
## D = 0.21724, p-value = 0.8868
## alternative hypothesis: two-sided
```

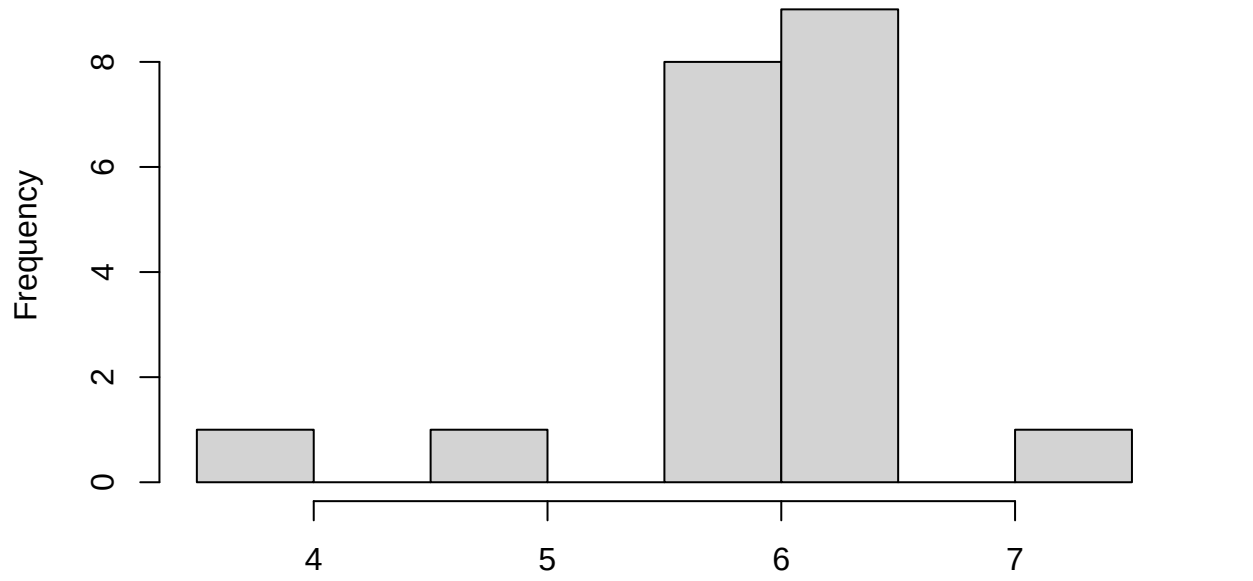
```
hist(whr2021_data$`Ladder score`[whr2021_data$`Regional indicator`=='East Asia'])
```




```
ks.test(whr2021_data$`Ladder score`[whr2021_data$`Regional indicator`=='Latin America and Caribbean'], '
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  whr2021_data$‘Ladder score‘[whr2021_data$‘Regional indicator‘ == "Latin America and Caribbean'
## D = 0.20631, p-value = 0.3171
## alternative hypothesis: two-sided
```

```
hist(whr2021_data$`Ladder score`[whr2021_data$`Regional indicator`=='Latin America and Caribbean'])
```



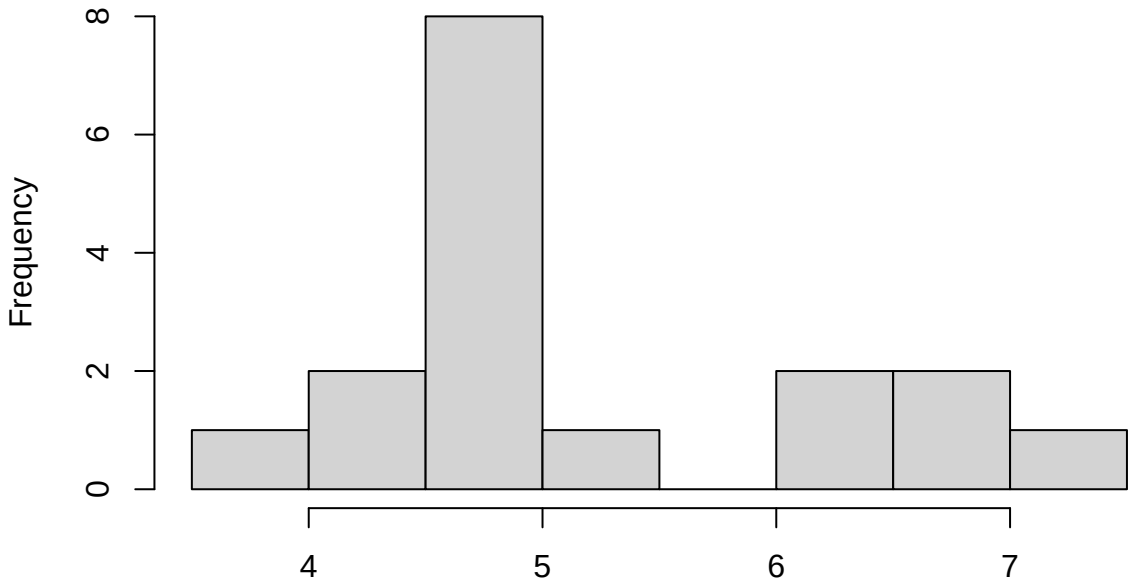
2021_data$`Ladder score`[whr2021_data$`Regional indicator` == "Latin America and Ca

```
ks.test(whr2021_data$`Ladder score`[whr2021_data$`Regional indicator`=='Middle East and North Africa'],
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  whr2021_data$‘Ladder score‘[whr2021_data$‘Regional indicator‘ == "Middle East and North Africa
## D = 0.25437, p-value = 0.186
## alternative hypothesis: two-sided
```

```
hist(whr2021_data$`Ladder score`[whr2021_data$`Regional indicator`=='Middle East and North Africa'])
```

21_data$`Ladder score`[whr2021_data$`Regional indicator` == "Middle

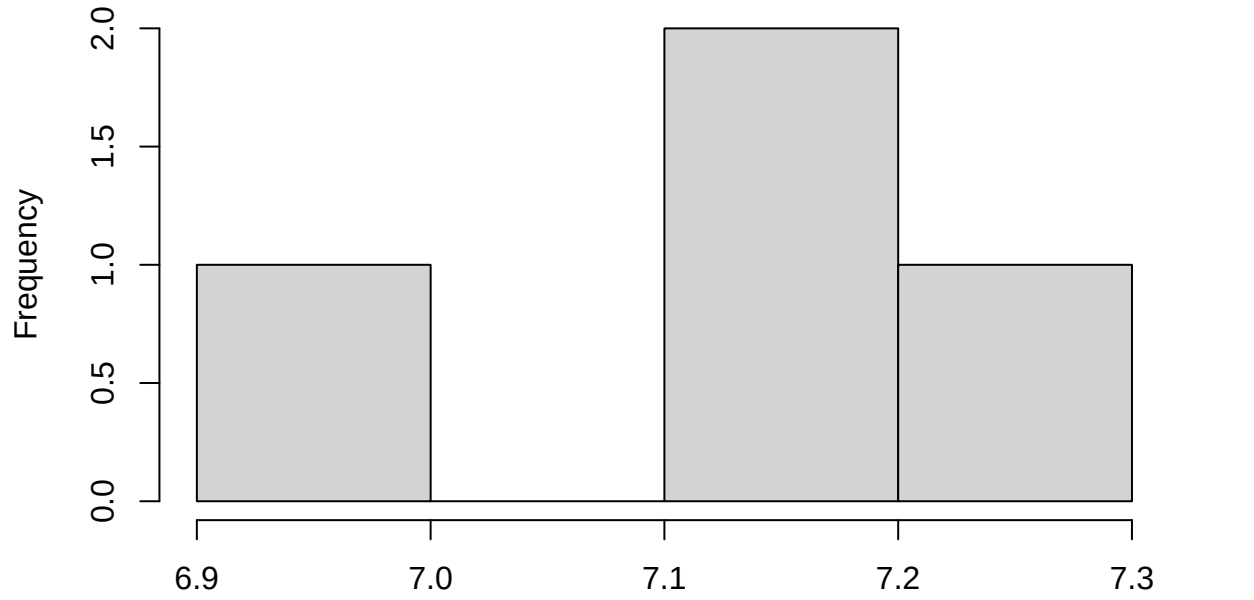


2021_data$`Ladder score`[whr2021_data$`Regional indicator` == "Middle East and Nort

```
ks.test(whr2021_data$`Ladder score`[whr2021_data$`Regional indicator`=='North America and ANZ'], "pnorm
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  whr2021_data$‘Ladder score‘[whr2021_data$‘Regional indicator‘ == "North America and ANZ"]
## D = 0.17678, p-value = 0.9972
## alternative hypothesis: two-sided
```

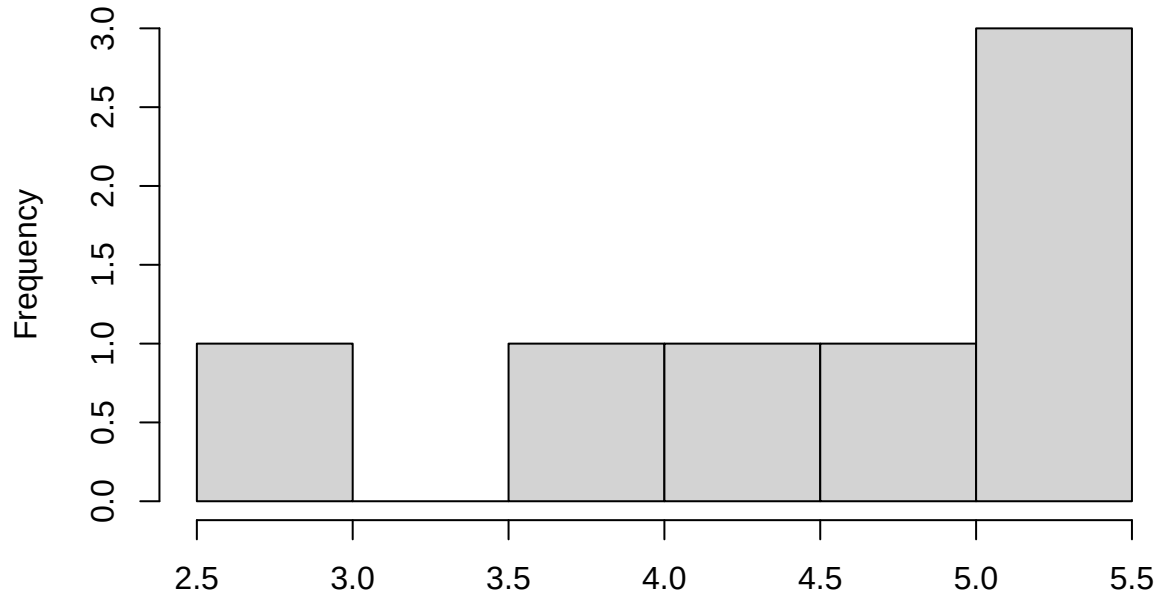
```
hist(whr2021_data$`Ladder score`[whr2021_data$`Regional indicator`=='North America and ANZ'])
```




```
ks.test(whr2021_data$`Ladder score`[whr2021_data$`Regional indicator`=='South Asia'], "pnorm", mean(whr
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  whr2021_data$‘Ladder score‘[whr2021_data$‘Regional indicator‘ == "South Asia"]
## D = 0.26133, p-value = 0.6354
## alternative hypothesis: two-sided
```

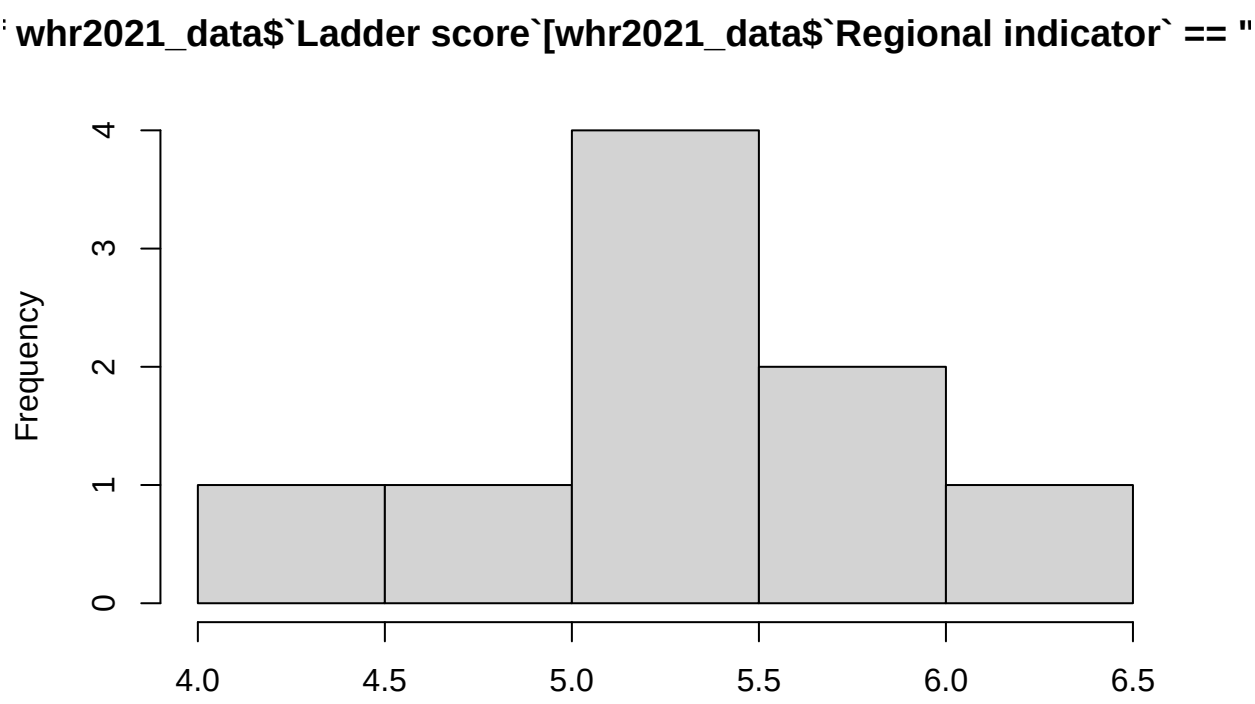
```
hist(whr2021_data$`Ladder score`[whr2021_data$`Regional indicator`=='South Asia'])
```




```
ks.test(whr2021_data$`Ladder score`[whr2021_data$`Regional indicator`=='Southeast Asia'], "pnorm", mean
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  whr2021_data$‘Ladder score‘[whr2021_data$‘Regional indicator‘ == "Southeast Asia"]
## D = 0.16447, p-value = 0.9367
## alternative hypothesis: two-sided
```

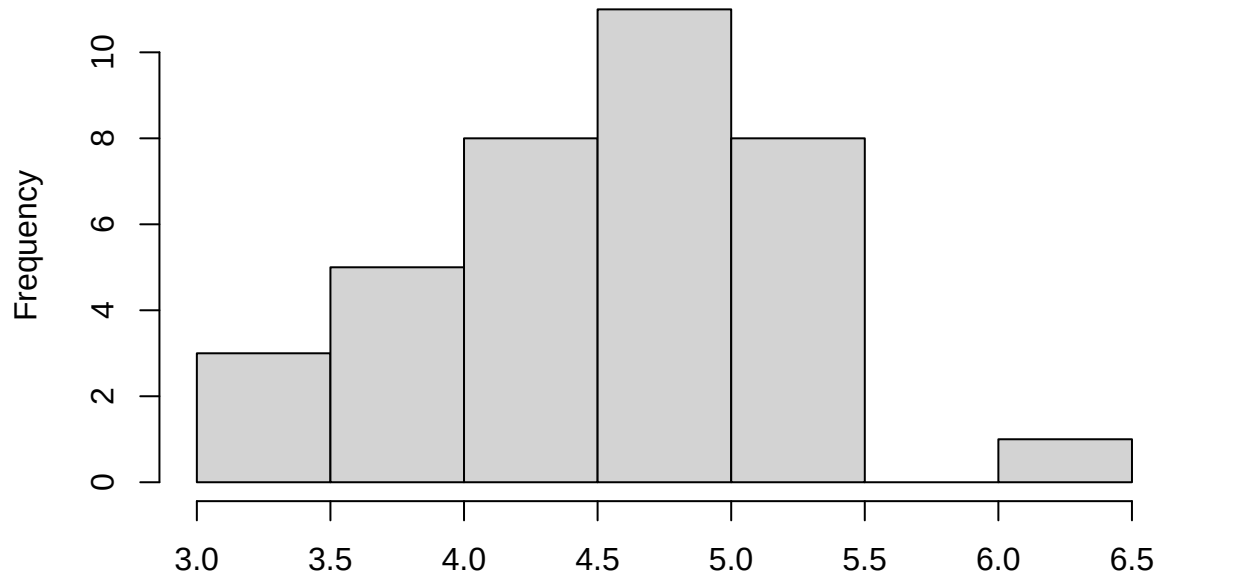
```
hist(whr2021_data$`Ladder score`[whr2021_data$`Regional indicator`=='Southeast Asia'])
```




```
ks.test(whr2021_data$`Ladder score`[whr2021_data$`Regional indicator`=='Sub-Saharan Africa'], "pnorm", m
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  whr2021_data$‘Ladder score‘[whr2021_data$‘Regional indicator‘ == "Sub-Saharan Africa"]
## D = 0.1039, p-value = 0.7942
## alternative hypothesis: two-sided
```

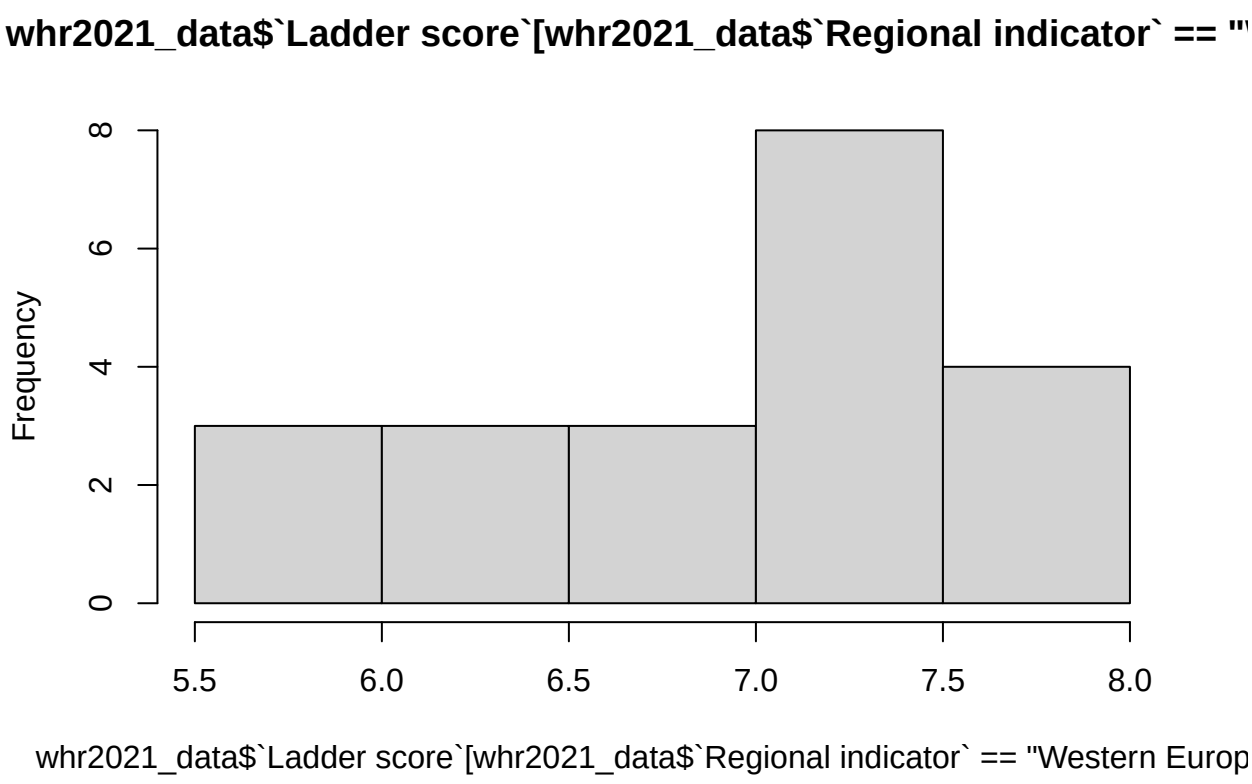
```
hist(whr2021_data$`Ladder score`[whr2021_data$`Regional indicator`=='Sub-Saharan Africa'])
```




```
ks.test(whr2021_data$`Ladder score`[whr2021_data$`Regional indicator`=='Western Europe'], "pnorm", mean
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  whr2021_data$‘Ladder score‘[whr2021_data$‘Regional indicator‘ == "Western Europe"]
## D = 0.16103, p-value = 0.5918
## alternative hypothesis: two-sided
```

```
hist(whr2021_data$`Ladder score`[whr2021_data$`Regional indicator`=='Western Europe'])
```



P-vrijednosti na svim testovima su nam veće od kritične vrijednosti te ne odbijamo nul hipotezu.

Sada trebamo analizirati homogenost varijanci regija što ćemo napraviti s Bartlettovim testom. Hipoteze su nam sljedeće:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$$
$$H_1 : \neg H_0.$$

te

$$\alpha = 0.05$$

```
bartlett.test(whr2021_data$`Ladder score` ~ whr2021_data$`Regional indicator`)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  whr2021_data$‘Ladder score‘ by whr2021_data$‘Regional indicator‘
## Bartlett’s K-squared = 21.976, df = 9, p-value = 0.008955
```

P-vrijednost je manja od kritične vrijednosti tako da ne odbacujemo nul hipotezu.

Sada možemo napraviti jednofaktorsku ANOVA-u. Hipoteze su nam sljedeće:

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$$
$$H_1 : \neg H_0.$$

te

$$\alpha = 0.05k = 10$$

```r
luck = aov(whr2021_data$`Ladder score` ~ whr2021_data$`Regional indicator`)
summary(luck)
```

```
##                                    Df Sum Sq Mean Sq F value Pr(>F)
## whr2021_data$‘Regional indicator‘   9 106.05  11.783   25.34 <2e-16 ***
## Residuals                         139  64.64   0.465
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
```

Kao što je sugerirano u grafu s početka, ANOVA potvrđuje da postoji razlika u iskazanoj sreći među regijama.