

Analiza antropometrijskih podataka američkih vojnih snaga

Alen Kurtović, Marin Matjanec, Ivan Rep, Josip Šestak

11/26/2020

Contents

1	Uvod	2
2	Deskriptivna analiza nekih općenitih značajki	2
2.1	Dob vojnika	2
2.2	Spol vojnika	4
2.3	Etnicitet vojnika	4
3	Deskriptivna analiza antropometrijskih značajki	5
3.1	Visina vojnika	5
3.2	Težina vojnika	7
3.3	Raspon ruku vojnika	9
4	ANOVA	10
4.1	Opseg glave vojnika	10
5	Testiranje hipoteza	13
5.1	Je li raspon ruku muškaraca u različitim kampovima jednak?	13
5.2	Je li raspon ruku žena u različitim kampovima jednak?	14
5.3	Utječe li rasa na pripadnost grani vojske?	16
5.4	Imaju li dešnjaci veći opseg desnog bicepsa od ljevaka?	18
6	Linearna regresija	22
6.1	Postoji li veza između nekih varijabli i kilaže vojnika?	22
6.2	Usporedba više parametara i tjelesne mase	26
7	Koje mjere kaciga biste preporučili američkoj vojsci i koliko različitih veličina?	29
7.1	Vizualizacija i numeričke karakteristike	29
7.2	Prijedlozi veličina kaciga	33

1 Uvod

Za ovu analizu koristiti ćemo skup podataka ANSUR II. Ovaj skup podataka opisuje razne mjere ljudskog tijela i pojedinih njegovih segmenata. Uz neke uobičajene mjere kao što su visina, težina, dob i indeks tjelesne mase(BMI), pojavljuju se i manje poznate veličine. Neke od njih su npr. opseg prsa, opseg gležnja, širina uha ili visina od poda do ručnog zgloba. Osim primjena u medicini, ove mjere koriste se u dizajnu proizvoda, ergonomiji, forenzici i antropologiji.

Konkretno, u skupu se nalaze podaci za 6068 osoba sakupljenih od 2010. do 2012. godine u raznim američkim vojničkim kampovima. Podaci sadrže 107 značajki koje se mogu podijeliti u općenite značajke i antropometrijske značajke.

2 Deskriptivna analiza nekih općenitih značajki

2.1 Dob vojnika

Jedna općenita značajka koja nas može zanimati je dob vojnika. Prvo što se može provjeriti je ispravnost unesenih podataka. Sljedeći odsječak prikazuje tip varijable i koliko vrijednosti varijable “Age” nije dostupno:

```
age <- data$Age

# Tip varijable
cat("Type of variable: ",class(age) )

## Type of variable:  numeric

# Broj redaka s vrijednosti NA
countNaValues <- function( values ) {
  cat("Number of NA values: ",sum(is.na(values)), "\n") }

countNaValues(age)

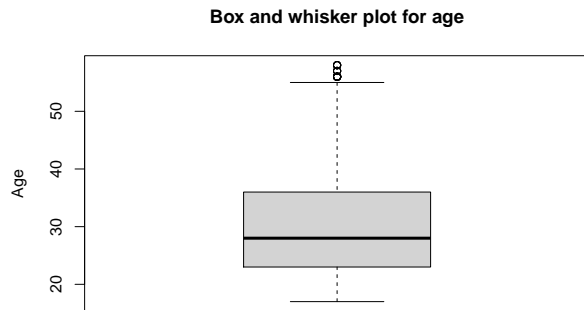
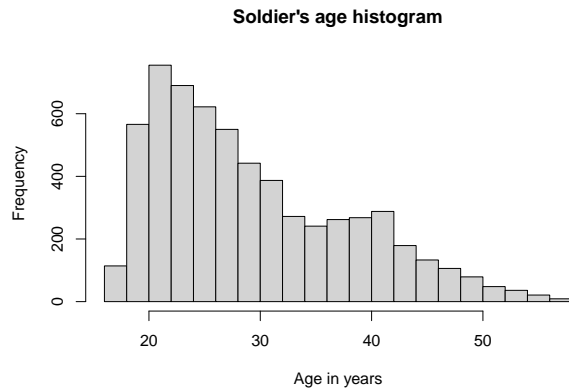
## Number of NA values:  0
```

Vizualizacija dobi

Razumno je razmišljati da u vojsci prevladavaju mladi ljudi, a da je znatno manje zrelih i starijih ljudi zbog naglašene potrebne fizičke spremnosti. Korisno je vizualizirati ovakve podatke, pa se može pogledati kako izgleda histogram i pravokutni dijagram:

```
# Histogram
hist(age,
      breaks=15,
      main="Soldier's age histogram",
      xlab="Age in years",
      ylab='Frequency' )

# Pravokutni dijagram
boxplot(age,
        main="Box and whisker plot for age",
        ylab="Age" )
```



Iz histograma možemo tvrditi da je stvarna situacija onakva kakvu smo pretpostavili i vidimo da je distribucija desno zakrivljena.

Na gornjem rubu možemo uočiti nekoliko vrijednosti koje su izvan izdanaka. Njih nazivamo outlieri i one ukazuju da je moguće da je došlo do neke pogreške prilikom prikupljanja. U ovom kontekstu ima smisla da te vrijednosti nisu krive jer je razumno da postoji manji broj starijih ljudi koji obnašaju neke visoke činove u vojsci. Zbog toga ih nismo izbacivali iz skupa.

Mjere centra

Možemo odrediti i neke mjere centra za dob vojnika. Pošto je distribucija zakrivljena, dobre mjere centra su medijan i podrezana aritmetička sredina. Njihove vrijednosti i vrijednosti ostalih mjera centra glase:

Koeficijent asimetrije je pozitivan i možemo uočiti da su ostali podaci poredani od najmanjeg prema najvećem na sljedeći način: mod, medijan, podrezana aritmetička sredina, aritmetička sredina. Ovakve vrijednosti i odnosi se i očekuju upravo zbog desne zakrivljenosti.

Mjere rasipanja

Sljedeće pitanje vezano za ovu značajku koje se može promatrati jest rasipanje. Iz histograma se može dobiti privid u izgled distribucije i utvrditi da je distribucija desno zakrivljena. U ovom slučaju, najbolja mjera rasipanja je IQR zbog svoje robusnosti:

```
# Rang
cat("Range: ",max(age)-min(age),"\n")
```

```
## Range: 41
```

```
# IQR
cat("IQR: ",IQR(age),"\n")
```

```
## IQR: 13
```

```
# Varijanca
cat("Variance: ",var(age),"\n")
```

```
## Variance: 75.24954
```

```
# Standardna devijacija
cat("Standard deviation: ",sd(age),"\n")
```

```
## Standard deviation: 8.67465
```

```
# Koeficijent varijacije
cat("Variation coefficient: ",sd(age)/mean(age),"\n")
```

```
## Variation coefficient: 0.2914751
```

Iz ranga se može zaključiti da je razlika između najmlađeg i najstarijeg vojnika 41 godina, a *IQR* daje predodžbu gdje se nalazi 50% podataka. Varijanca i standardna devijacija koje su ovdje izračunate zapravo su nepristrane procjene za varijancu i standardnu devijaciju.

2.2 Spol vojnika

Zanimanje vojnika se smatra muškim zanimanjem pa se i to može provjeriti u ovom skupu podataka. Prije svega potrebno je provjeriti ispravnost podataka. Varijabla spola može poprimiti samo dvije vrijednosti ili biti nedostupna:

```
gender <- data$Gender
```

```
# Dostupnost podataka  
countNaValues(gender)
```

```
## Number of NA values: 0
```

```
# Tip variable  
cat("Type of variable: ",class(gender),"\n")
```

```
## Type of variable: character
```

```
# Pretvorba podatka iz character u factor  
gender <- factor(gender)  
cat("Type of variable after change; ",class(gender),"\n")
```

```
## Type of variable after change; factor
```

```
# Koliko ima žena u skupu  
females <- length(which(gender=="Female"))  
cat("Number of females: ",females,"\n")
```

```
## Number of females: 1986
```

```
# Koliko ima muškaraca u skupu  
males <- length(which(gender=="Male"))  
cat("Number of males: ",males,"\n")
```

```
## Number of males: 4082
```

```
# Je li broj muškaraca i žena jednak ukupnom broju ljudi u skupu  
all <- length(gender)  
cat("Number of females and males equal to the whole set: ",males + females == all,"\n")
```

```
## Number of females and males equal to the whole set: TRUE
```

2.3 Etnicitet vojnika

Varijabla koja predstavlja narodnost vojnika je etnicitet. Kao i inače, potrebno je provjeriti dostupnost podataka i njihov tip:

```
## Number of NA values: 4647
```

```
## Amount of non NA values: 1421
```

S ovom varijablom postoji problem, a to je da ima 4647 nedostupnih podataka. To je velik dio skupa podataka, pa se ne može provoditi deskriptivna analiza nad ovom varijablom.

3 Deskriptivna analiza antropometrijskih značajki

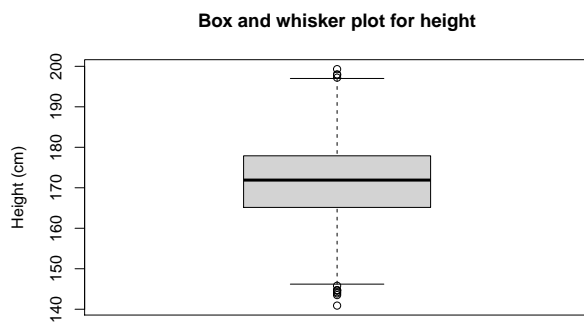
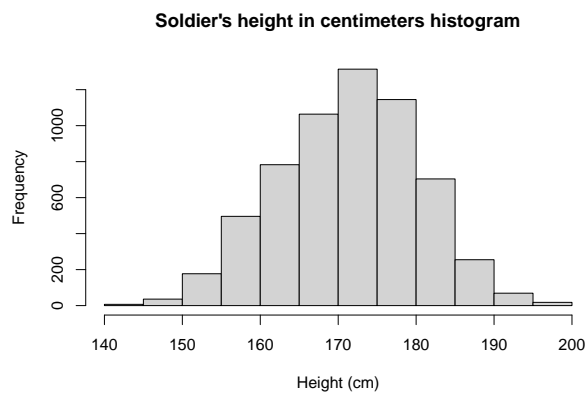
3.1 Visina vojnika

Jedna od osnovnih antropometrijskih značajki je visina. Prvo se provjerava ispravnost i pretvaraju podaci u centimetre:

```
## Number of NA values: 0
## Type of variable: numeric
## Height after conversion to cm: 156 166.5 171.1 166 157.2 170.7
```

Vizualizacija visine

Poslije ovoga mogu se vizualizirati podaci histogramom:



```
# Kvantili
cat("Quantiles:\n")
```

```
## Quantiles:
```

```
quantile(heights)
```

```
##      0%      25%      50%      75%     100%
## 140.900 165.175 171.900 177.900 199.300
```

Može se uočiti da je distribucija nalik normalnoj. Postoji i nekoliko ekstremnih vrijednosti. Pravokutni dijagram pokazuje i prisutnost mogućih outliera.

Mjere centra visine

Pošto distribucija nije zakrivljena očekuje se da će medijan, srednja vrijednost i mod biti podjednakih vrijednosti. Mogu se provjeriti te mjere za ovu varijablu.

```
## Mean: 171.4406
## Median: 171.9
## Mod: 174
```

Izračunate vrijednosti su blizu kao što se i očekuje.

Mjere rasipanja visine

Također, zbog simetričnosti distribucije dobru predodžbu varijabilnosti distribucije predstavlja standardna devijacija. Uz to može se i izračunati rang, IQR, varijanca i koeficijent varijacije.

```
## Range: 58.4
## IQR: 12.725
## Variance: 81.02761
## Standard deviation: 9.001534
## Variation coefficient: 0.05250526
```

Usporedba spola i visine

Ono što se nije uzimalo u obzir dosad je spol, a to je značajka koja gotovo sigurno utječe na distribuciju. Uzme li se to u obzir i može se istražiti kako se ponaša visina muškaraca i visina žena:

```
# Visine svih muškaraca u centimetrima
maleHeights <- data[data$Gender=="Male",]$stature/10

# Visine svih žena u centimetrima
femaleHeights <- data[data$Gender=="Female",]$stature/10
```

Ova varijabla može se grafički prikazati i mogu se odrediti njene numeričke karakteristike:

```
# RGB polje za plavu boju
rgbArr <- col2rgb(blue)/255
col1 <- rgb(rgbArr[1], rgbArr[2], rgbArr[3], 0.5)

# Histogram - visina i spol muškaraca
hist(maleHeights,
     breaks=20,
     main="Soldier's height in centimeters and gender histogram",
     xlab="Height (cm)",
     ylab='Frequency',
     col=col1)

# RGB polje za smeđu boju
rgbArr2 <- col2rgb(brown)/255
col2 <- rgb(rgbArr2[1], rgbArr2[2], rgbArr2[3], 0.5)

# Histogram - visina i spol žena
hist(femaleHeights,
     breaks=15,
     col=col2,
     add=T)      # Parametar koji postavlja drugi histogram na plot prvoga

# Legenda za spol - histogram
legend("topleft",
     inset=.05,
     title="Gender",
     legend=c("Male","Female"),
     fill=c(col1,col2))

# Pravokutni dijagram - visina i spol
boxplot(data$stature/10 ~ Gender,
     data=data,
     xlab = "Gender ",
     ylab = "Height (cm)",
     main = "Soldier's height in centimeters and gender box plot",
```

```

col = c(brown, blue))

# Legenda za spol - pravokutni dijagram
legend("topleft",
      inset=.05,
      title="Gender",
      legend=c("Male","Female"),
      fill=c(blue, brown) )

# Numeričke karakteristike - muška visina
cat( "Mean (male): ", mean(maleHeights),"\n")

## Mean (male):  175.6215

cat( "Standard deviation (male): ",sd(maleHeights),"\n")

## Standard deviation (male):  6.855079

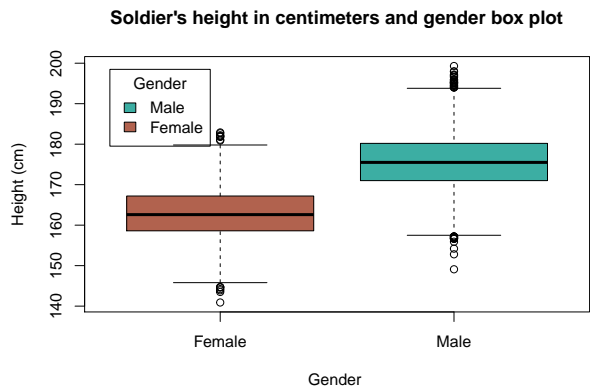
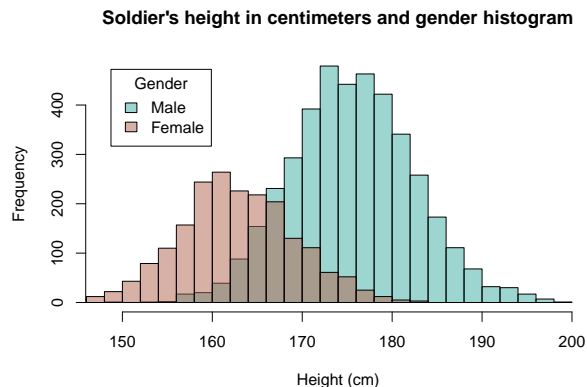
# Numeričke karakteristike - ženska visina
cat( "Mean (female): ", mean(femaleHeights),"\n")

## Mean (female):  162.8473

cat( "Standard deviation (female): ",sd(femaleHeights),"\n")

## Standard deviation (female):  6.419951

```



Može se vidjeti da su se rezultati promijenili podjelom u pojedini spol. Možemo vidjeti da je prosjek visine muškaraca veći od prosjeka visine žena, dok je varijabilnost obje grupe podjednaka. Važno je spomenuti da varijabla spola ne mora nužno uzrokovati ovu promjenu. Postoji mogućnost da neka skrivena varijabla zapravo uzrokuje ovu promjenu.

3.2 Težina vojnika

Druga antropometrijska značajka koja se može obraditi je težina. Naputak na web stranici ovog dataseta kaže da su podaci u hektogramima unatoč nazivu. Potrebno je provjeriti ispravnost podataka i pretvoriti podatke u kilograme.

```

## Number of NA values:  0

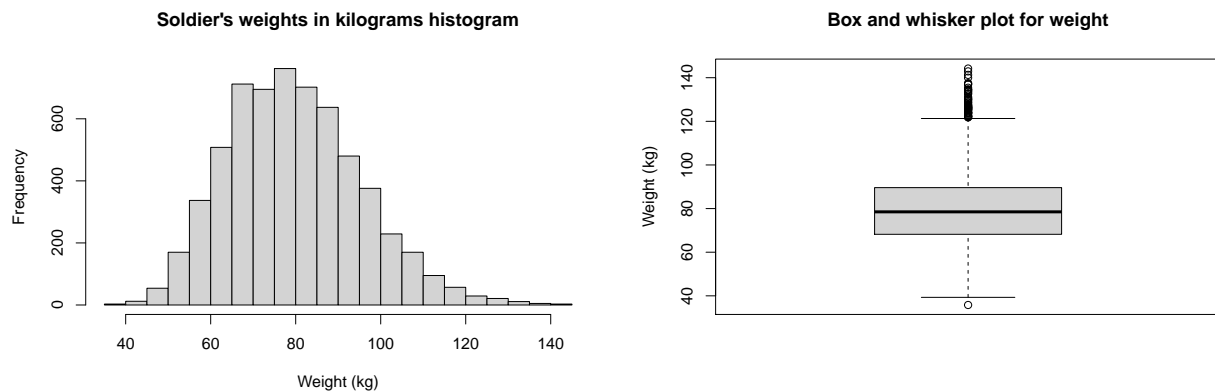
## Type of variable:  numeric

## Weight after conversion to kg:  65.7 53.4 66.3 78.2 88.6 83.7

```

Vizualizacija težine

Poslije ovoga možemo vizualizirati podatke histogramom i pravokutnim dijagramom:



Može se uočiti da je distribucija nalik normalnoj.

Mjere centra težine

Kao i u slučaju s visinom, distribucija je simetrična pa se očekuje da će medijan, srednja vrijednost i mod biti podjednake vrijednosti.

```
## Mean: 79.70939
```

```
## Median: 78.5
```

```
## Mod: 79.1
```

Izračunate vrijednosti su podjednake iznosa.

Mjere rasipanja težine

Mogu se i izračunati mjere varijabilnosti za ovu varijablu npr. rang, IQR, varijanca, standardna devijaciju i koeficijent varijacije.

```
## Range: 108.4
```

```
## IQR: 21.4
```

```
## Variance: 245.0107
```

```
## Standard deviation: 15.65282
```

```
## Variation coefficient: 0.1963735
```

Usporedba spola i težine

Jednako kao u slučaju s visinom, ovdje nije uzet u obzir spol vojnika. Kada se to uzme u obzir podaci izgledaju drukčije, a i numeričke karakteristike se mijenjaju.

```
## Mean (male): 85.52396
```

```
## Standard deviation (male): 14.2207
```

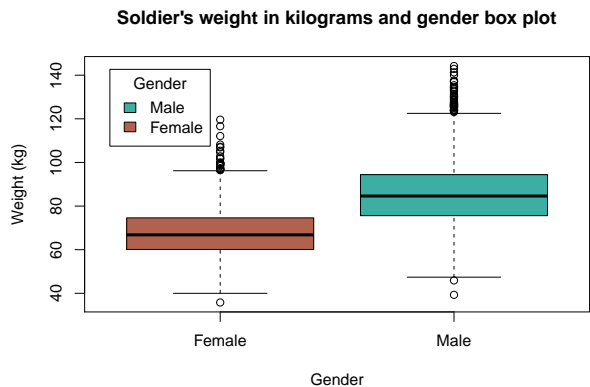
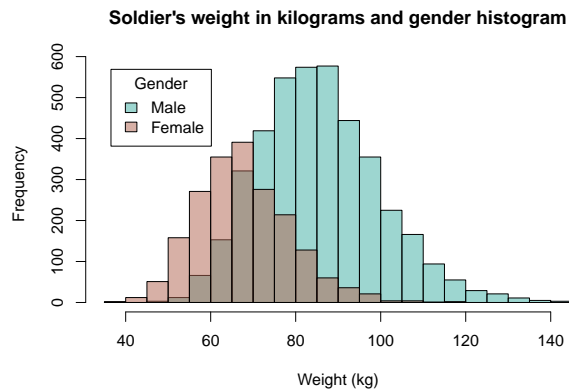
```
## IQR (male): 18.8
```

```
## Mean (female): 67.75821
```

```
## Standard deviation (female): 10.98465
```



```
## IQR (female): 14.475
```



Kao i u slučaju s visinom može se uočiti da je došlo do promjena u numeričkim karakteristikama i izgledu podataka. Žene imaju manju prosječnu kilažu, dok muškarci imaju veću prosječnu kilažu, a varijabilnost muške kilaže je nešto veća. To se može i potvrditi pomoću interkvartilnog ranga koji je također veći kod muškaraca.

3.3 Raspon ruku vojnika

U skupu podataka se nalaze vojnici iz nekoliko vojnih kampova u SAD-u. Može se provjeriti koji su to uopće kampovi, kako su zastupljeni i vizualizirati podatke:

```
# Provjera dostupnosti svih podataka
countNaValues(data$Installation)
```

```
## Number of NA values: 0
```

```
cat("Type of variable: ",class(data$Installation),"\n")
```

```
## Type of variable: character
```

```
cat("Camps:\n")
```

```
## Camps:
```

```
unique(data$Installation)
```

```
## [1] "Fort Hood"      "Fort Bliss"     "Camp Atterbury" "Fort Drum"
## [5] "Fort McCoy"     "Fort Lee"       "Fort Stewart"   "Fort Bragg"
## [9] "Fort Gordon"    "Fort Huachuca" "Camp Shelby"    "Fort Rucker"
```

```
# Pretvorba u factor
```

```
installations <- factor(data$Installation)
```

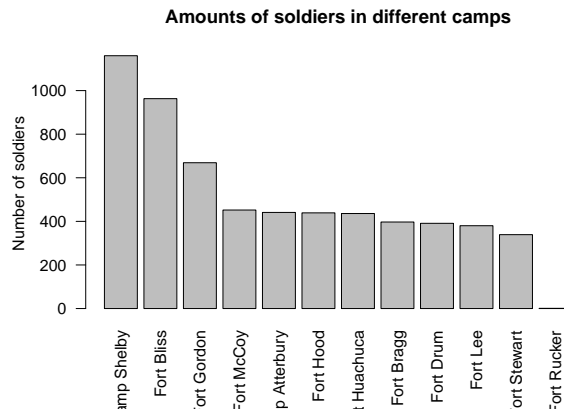
```
cat("Type of variable after conversion: ",class(installations),"\n")
```

```
## Type of variable after conversion: factor
```

```
camps <- table(sort(installations))
```

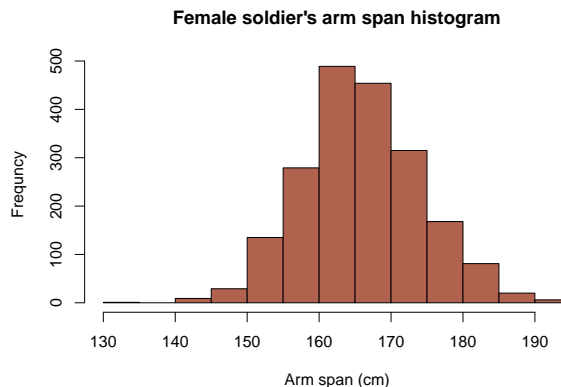
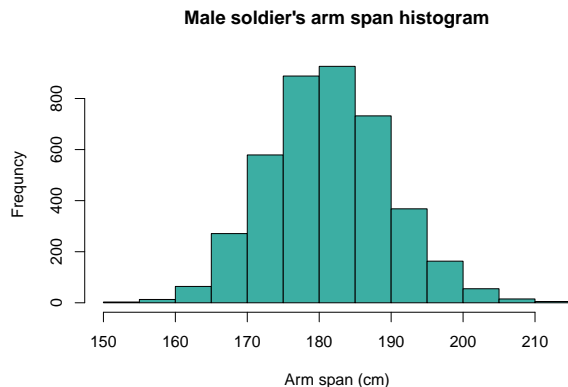
```
# Stupičasti dijagram
```

```
barplot( sort(camps, decreasing=TRUE),
  main="Amounts of soldiers in different camps",
  ylab="Number of soldiers",
  las=2)
```



Zbog malog broja testiranih ljudi iz kampa Fort Rucker nad njima nećemo testirati hipoteze.

Kod dizajna proizvoda mogao bi se koristiti raspon ruku vojnika. To možemo iskoristiti da npr. odredimo dužine rukava na uniformama za veličinu muške ili ženske uniforme. Može se pogledati kako se ponaša ova varijabla.



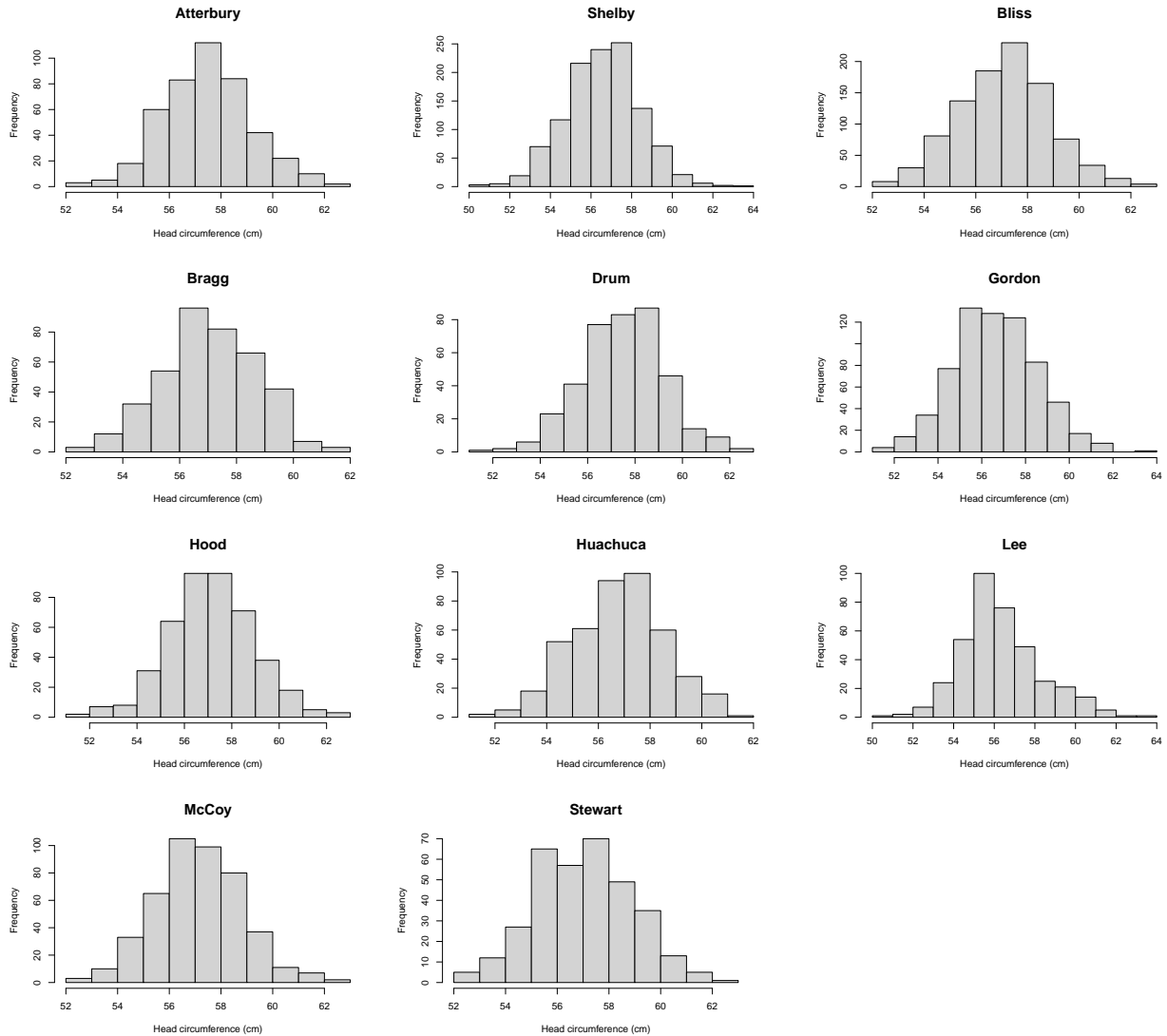
Zatim se može provjeriti postoji li neki kamp koji odstupa. Za njega je onda potrebno napraviti uniforme drukčijih veličina od ostalih kampova npr. može se ispitati razlikuje li se raspon ruku između kampova Camp Shelby i Fort Bliss.

4 ANOVA

4.1 Opseg glave vojnika

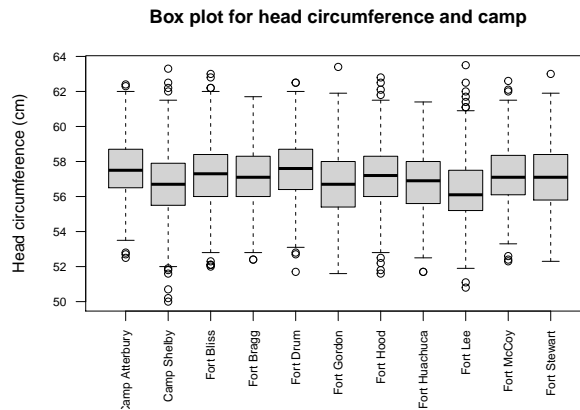
Može nas zanimati i je li raspršenje varijable opsega glave jednako u svim kampovima. Ova varijabla bi se mogla koristiti pri dizajnu kaciga za američku vojsku. Za ovo se može koristiti analiza varijance. Pretpostavke koje je potrebno zadovoljiti za ovo su: nezavisnost pojedinih podataka u uzorcima, normalna razdioba podataka i homogenost varijanci među populacijama. Nezavisnost podataka može se tvrditi jer potječu iz različitih kampova. U nastavku su prikazani histogrami kako bi se utvrdila normalnost podataka pa zatim pravokutni dijagram. Zbog vrlo malog broja ispitanih u kampu Fort Rucker izbaciti ćemo ga iz skupa podataka za ovaj test.

Vizualizacija opsega glave



```
headCircNoRucker <- data[data$Installation!="Fort Rucker",]$headcircumference/10
installationNoRucker <- factor(data[data$Installation!="Fort Rucker",]$Installation)

boxplot(headCircNoRucker ~ installationNoRucker,
        main = "Box plot for head circumference and camp",
        ylab = "Head circumference (cm)",
        xlab = "", las = 2, cex.axis = 0.71)
```



Homogenost varijance

Postave se hipoteze za jednakost varijanci:

- 1) $H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_{11}^2$
- 2) $H_1 : \neg H_0$
- 3) $\alpha = 0.05$

Jednakost varijanci može se testirati pomoću Bartlettovog testa.

```
bartlett.test(headCircNoRucker ~ installationNoRucker)
```

```
##
## Bartlett test of homogeneity of variances
##
## data: headCircNoRucker by installationNoRucker
## Bartlett's K-squared = 22.295, df = 10, p-value = 0.01367
```

P-vrijednost je vrlo mala, ali skup podataka je velik, na dijagramu raspršenja varijance se ne razlikuju puno i ANOVA je vrlo robusna metoda na svoje pretpostavke pa se test može nastaviti.

ANOVA test

ANOVA hipoteze:

- 1) $H_0 : \mu_1 = \mu_2 = \dots = \mu_{11}$
- 2) $H_1 : \text{barem jedan } \mu_k \text{ se razlikuje od ostalih}$
 $k \in \{1, \dots, 11\}$
- 3) $\alpha = 0.05$

```
summary(aov(headCircNoRucker ~ installationNoRucker))
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## installationNoRucker    10      669    66.94  20.65 <2e-16 ***
## Residuals              6056    19627     3.24
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Mala p-vrijednost pokazuje da se odbija hipoteza H_0 u korist hipoteze H_1 .

5 Testiranje hipoteza

5.1 Je li raspon ruku muškaraca u različitim kampovima jednak?

- 1) $H_0 \dots$ Raspon ruku muškaraca u kampu Fort Shelby jednak je onome u kampu Fort Bliss
$$\mu_1 = \mu_2, d_0 = 0$$
- 2) $H_1 \dots$ Raspon ruku muškaraca u kampu Fort Shelby nije jednak onome u kampu Fort Bliss
$$\mu_1 \neq \mu_2, d_0 \neq 0$$
- 3) $\alpha = 0.05$

Za provesti ovaj test potrebne su numeričke karakteristike ovih varijabli:

```
campShelbyMaleSpan <- data[ data$Installation=="Camp Shelby" & data$Gender=="Male", ]$span/10
fortBlissMaleSpan <- data[ data$Installation=="Fort Bliss" & data$Gender=="Male", ]$span/10

# Aritmetičke sredine
cat("Mean (Camp Shelby): ",mean(campShelbyMaleSpan),"\n")

## Mean (Camp Shelby): 179.5016
cat("Mean: (Fort Bliss)",mean(fortBlissMaleSpan),"\n")

## Mean: (Fort Bliss) 182.7469

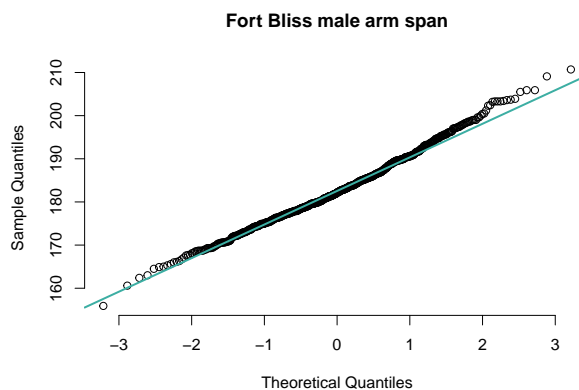
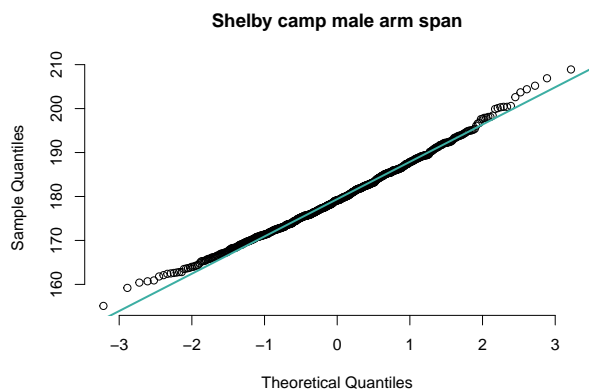
# Standardne devijacije
cat("Standard deviation (Camp Shelby): ",sd(campShelbyMaleSpan),"\n")

## Standard deviation (Camp Shelby): 8.25697
cat("Standard deviation (Fort Bliss): ",sd(fortBlissMaleSpan),"\n")

## Standard deviation (Fort Bliss): 8.129368

# QQ plot muški raspon - Camp Shelby
qqnorm(campShelbyMaleSpan, pch = 1, frame = FALSE,main='Shelby camp male arm span')
qqline(campShelbyMaleSpan, col = blue, lwd = 2)

# QQ plot muški raspon - Fort Bliss
qqnorm(fortBlissMaleSpan, pch = 1, frame = FALSE,main='Fort Bliss male arm span')
qqline(fortBlissMaleSpan, col = blue, lwd = 2)
```



Iz ispisa se vidi da postoje pokazatelji da nisu jednaki, no svejedno se može provesti test. Može se uočiti da su standardne devijacije podjednake. Nezavisnost ovih uzoraka može se pretpostaviti jer potječu iz različitih

kampova, a normalnost se može pretpostaviti jer se podaci poklapaju s QQ plotom. Još se može provesti test o varijancama da bi provjerili njihovu jednakost:

1) $H_0 \dots$ Varijance distribucija se ne razlikuju

$$\sigma_1 = \sigma_2$$

2) $H_1 \dots$ Varijance distribucija se razlikuju

$$\sigma_1 \neq \sigma_2$$

3) $\alpha = 0.05$

```
var.test(campShelbyMaleSpan, fortBlissMaleSpan)
```

```
##
## F test to compare two variances
##
## data: campShelbyMaleSpan and fortBlissMaleSpan
## F = 1.0316, num df = 771, denom df = 765, p-value = 0.6662
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.8954534 1.1884900
## sample estimates:
## ratio of variances
##      1.031639
```

Iz ovog rezultata može se zaključiti da se ne odbacuje hipoteza H_0 koja glasi da su varijance jednake jer je p-vrijednost vrlo velika.

Sada se može nastaviti s početnim testom:

```
t.test(campShelbyMaleSpan, fortBlissMaleSpan, alt = "two.sided", var.equal = TRUE)
```

```
##
## Two Sample t-test
##
## data: campShelbyMaleSpan and fortBlissMaleSpan
## t = -7.7665, df = 1536, p-value = 1.465e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.064954 -2.425671
## sample estimates:
## mean of x mean of y
##  179.5016  182.7469
```

Iz ovog ispisa može se zaključiti da je p-vrijednost izrazito mala. Zbog toga je potrebno odbaciti hipotezu H_0 u korist H_1 i reći da na ovoj razini značajnosti sredine ovih populacija nisu jednake.

5.2 Je li raspon ruku žena u različitim kampovima jednak?

Ovaj test može se provesti i u kampovima Fort Lee i Fort McCoy na rasponu ruku žena.

```
## Mean (Fort McCoy):  166.8252
## Mean (Fort Lee):   166.7297
## Standard deviation (Fort McCoy):  7.728267
## Standard deviation (Fort Lee):    8.504761
```

Vidi se da su aritmetičke sredine vrlo blizu, no varijance se u manjoj mjeri razlikuju. Provede li se test o jednakosti varijanci može se saznati nešto više o tome. Postave se hipoteze:

1) $H_0 \dots$ *Varijance distribucija se ne razlikuju*

$$\sigma_1 = \sigma_2$$

2) $H_1 \dots$ *Varijance distribucija se razlikuju*

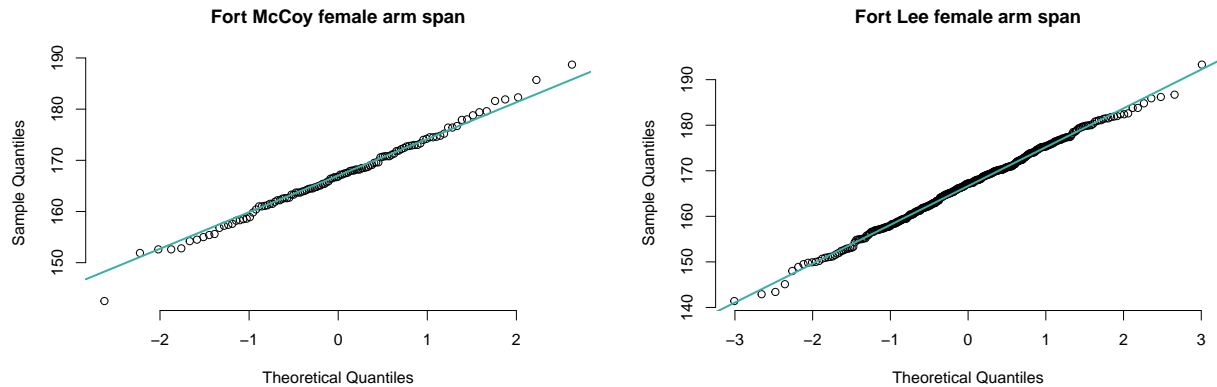
$$\sigma_1 \neq \sigma_2$$

3) $\alpha = 0.05$

```
##
## F test to compare two variances
##
## data: fortMcCoyFemaleSpan and fortLeeFemaleSpan
## F = 0.82573, num df = 114, denom df = 379, p-value = 0.2236
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.6208129 1.1249937
## sample estimates:
## ratio of variances
##      0.8257337
```

Iz ispisa se može zaključiti da se na zadanoj razini značajnosti ne može odbiti hipoteza H_0 da su varijance populacija jednake.

Zatim se provodi isti test kao i u slučaju s muškarcima iz kampova Fort Bliss i Camp Shelby. Pretpostavka nezavisnosti vrijedi zbog različitosti kampova pa se još provjerava vrijedi li pretpostavka normalnosti.



Ovo potvrđuje i pretpostavku normalnosti pa se postave hipoteze i odredi razina značajnosti:

1) $H_0 \dots$ *Raspon ruku žena u kampu Fort McCoy jednak je rasponu ruku žena u kampu Fort Lee*

$$\mu_1 = \mu_2, d_0 = 0$$

2) $H_1 \dots$ *Raspon ruku žena u kampu Fort McCoy nije jednak rasponu ruku žena u kampu Fort Lee*

$$\mu_1 \neq \mu_2, d_0 \neq 0$$

3) $\alpha = 0.05$

```
##
## Two Sample t-test
##
## data: fortMccoyFemaleSpan and fortLeeFemaleSpan
## t = 0.10768, df = 493, p-value = 0.9143
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -1.646760 1.837721
## sample estimates:
## mean of x mean of y
## 166.8252 166.7297
```

Na zadanoj razini značajnosti ne odbija se hipoteza H_0 koja kaže da je raspon ruku žena u kampu Fort McCoy jednak rasponu ruku žena u kampu Fort Lee što potvrđuje vrlo velika p-vrijednost.

5.3 Utječe li rasa na pripadnost grani vojske?

U američkim vojnim snagama je prisutan širok raspon rasa zbog mnogih seoba koje su se dogodile na tom području. Osim bijelaca i crnaca česti su latinoamerikanci i azijati. Može se provjeriti koje su uopće rase i grane vojske sadržane u skupu podataka:

```

races <- as.factor(data$DODRace)
branches <- as.factor(data$Branch)

cat("Race codes: ",levels(races),"\n")

## Race codes:  1 2 3 4 5 6 8

cat("Branches: ",levels(branches),"\n")

## Branches:  Combat Arms Combat Service Support Combat Support

raceSums <- table(races)
branchSums <- table(branches)

contingencyTable <- table(races, branches)
contingencyTable
```

```
##      branches
## races Combat Arms Combat Service Support Combat Support
## 1      1146      1911      735
## 2      173      745      380
## 3      189      365      125
## 4       44       99       45
## 5       16       25        8
## 6       13       27       19
## 8        1        2        0
```

U gornjem ispisu vide se svi mogući kodove za rase, sve moguće grane vojske i kontingencijska tablica za te dvije varijable.

Još je potrebno provjeriti frekvencije razreda. Svi razredi moraju imati frekvenciju veću ili jednaku 5 što ovdje nije zadovoljeno. Pripadnike iz rase 8 ćemo dodati u rasu 6:

```

newRowSix <- contingencyTable[ races="8", ] + contingencyTable[ races="6", ]

contingencyTable <- contingencyTable[!(row.names(contingencyTable) %in% c('8', '6')), ]
contingencyTable <- rbind(contingencyTable, "6"=newRowSix)
```



```
contingencyTable
```

```
##      Combat Arms Combat Service Support Combat Support
## 1      1146              1911              735
## 2      173              745              380
## 3      189              365              125
## 4       44              99              45
## 5       16              25              8
## 6       14              29              19
```

Postave se hipoteze i razina značajnosti:

- 1) $H_0 \dots$ *Pripadnost određenoj rasi i pripadnost grani vojske su nezavisna obilježja*
$$o_{ij} = e_{ij}, \quad \forall i \in \{1, \dots, 6\}, j \in \{1, \dots, 3\}$$
- 2) $H_1 \dots$ *Pripadnost određenoj rasi i pripadnost grani vojske nisu nezavisna obilježja*
$$o_{ij} \neq e_{ij}, \quad \exists i \in \{1, \dots, 6\}, j \in \{1, \dots, 3\}$$
- 3) $\alpha = 0.05$

Sad se može provesti χ^2 test:

```
chisq.test(contingencyTable)
```

```
##
## Pearson's Chi-squared test
##
## data:  contingencyTable
## X-squared = 168.69, df = 10, p-value < 2.2e-16
```

Iz vrlo male p-vrijednosti može se zaključiti da se hipoteza H_0 odbija u korist hipoteze H_1 .

5.4 Imaju li dešnjaci veći opseg desnog bicepsa od ljevaka?

Logično je da vojnici dešnjaci imaju veći desni biceps jer češće koriste tu ruku. Prvo se provjerava postojanje nedostupnih vrijednosti i filtriraju podaci.

```
males <- data[data$Gender=="Male",]
```

```
rightRBicep <- males[males$WritingPreference=="Right hand",]$bicepscircumferenceflexed/10  
cat("Right handed:\n")
```

```
## Right handed:
```

```
countNaValues(rightRBicep)
```

```
## Number of NA values: 0
```

```
leftRBicep <- males[males$WritingPreference=="Left hand",]$bicepscircumferenceflexed/10  
cat("Left handed:\n")
```

```
## Left handed:
```

```
countNaValues(leftRBicep)
```

```
## Number of NA values: 0
```

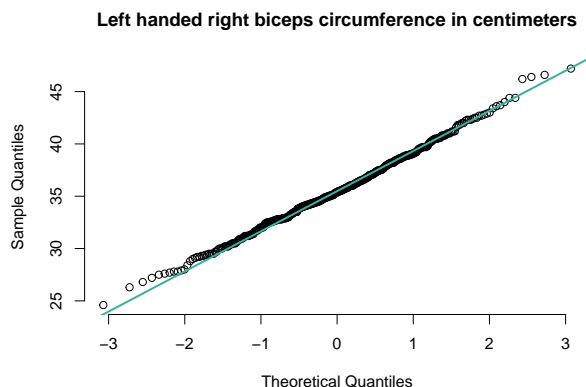
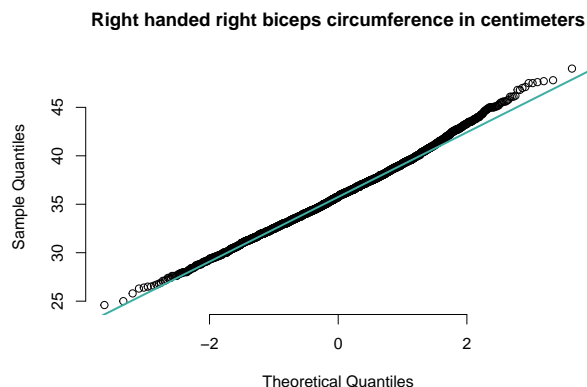
```
ambiRBicep <-males[males$WritingPreference=="Either hand (No preference)",]$bicepscircumferenceflexed/10  
cat("Ambidextrous:\n")
```

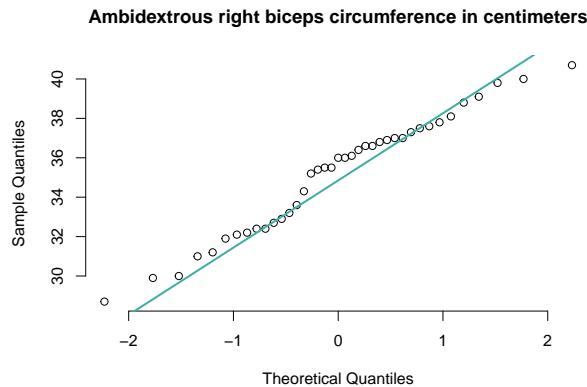
```
## Ambidextrous:
```

```
countNaValues(ambiRBicep)
```

```
## Number of NA values: 0
```

Zatim se grafički prikazu podaci da se vidi ravnaaju li se distribucije prema normalnoj razdiobi.





Za prve dvije distribucije može se pretpostaviti normalnost, no vidi se da treća odstupa. Nezavisnost između svih grupa se pretpostavlja jer se radi o različitim grupama ljevaka, desnjaka i ambidekstretnih vojnika. Mogu se izračunati mjere centra i raspršenja:

```
## Mean (Right Handed) : 35.85597
## Mean (Left Handed) : 35.54077
## Mean (Ambidextrous) : 35.18462
## Standard deviation (Right Handed) : 3.433578
## Standard deviation (Left Handed) : 3.693536
## Standard deviation (Ambidextrous) : 3.025987
```

Postoje pokazatelji da opseg bicepsa desnjaka premašuje opseg bicepsa ljevaka, pa i ambidekstretnih vojnika. Zatim se može provesti test za grupe desnjaka i ljevaka. Prvo je potrebno provjeriti jednakost varijanci pomoću F testa koji smijemo koristiti zbog pretpostavki normalnosti i nezavisnosti:

1) $H_0 \dots$ *Varijance su jednake*

$$\sigma_1 = \sigma_2$$

2) $H_1 \dots$ *Varijance se razlikuju*

$$\sigma_1 \neq \sigma_2$$

3) $\alpha = 0.05$

```
##
## F test to compare two variances
##
## data: rightRBicep and leftRBicep
## F = 0.86419, num df = 3576, denom df = 465, p-value = 0.0315
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.7509722 0.9872488
## sample estimates:
## ratio of variances
##      0.8641899
```

Na razini značajnosti od $\alpha = 0.05$ odbija se hipoteza H_0 o jednakosti varijanca u korist alternative H_1 .

Iz ovoga se može zaključiti da je prikladan T test s nejednakim varijancama. Postave se hipoteze i provede test:

1) $H_0 \dots$ *Opseg desnog bicepsa dešnjaka je jednak opsegu desnog bicepsa ljevaka*

$$\mu_1 = \mu_2, \quad d_0 = 0$$

2) $H_1 \dots$ *Opseg desnog bicepsa dešnjaka je veći od desnog bicepsa ljevaka*

$$\mu_1 > \mu_2, \quad d_0 > 0$$

$$3) \quad \alpha = 0.05$$

```
##
## Welch Two Sample t-test
##
## data: rightRBicep and leftRBicep
## t = 1.7465, df = 574.65, p-value = 0.04063
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.01786278      Inf
## sample estimates:
## mean of x mean of y
##  35.85597  35.54077
```

Odbija se hipoteza H_0 jer je p-vrijednost manja od razine značajnosti.

Može se i testirati jednakost sredina dešnjaka i ambidekstretnih vojnika. Postave se hipoteze i provjeri jednakost varijanci:

1) $H_0 \dots$ *Varijance su jednake*

$$\sigma_1 = \sigma_2$$

2) $H_1 \dots$ *Varijance se razlikuju*

$$\sigma_1 \neq \sigma_2$$

$$3) \quad \alpha = 0.05$$

```
##
## F test to compare two variances
##
## data: rightRBicep and ambiRBicep
## F = 1.2875, num df = 3576, denom df = 38, p-value = 0.3294
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.7737646 1.9333947
## sample estimates:
## ratio of variances
##          1.287537
```

P-vrijednost je znatno veća od razine značajnosti α , pa se hipoteza H_0 ne odbija.

Hipoteze za jednakost sredina ovih dviju populacija glase:

1) $H_0 \dots$ *Opseg desnog bicepsa dešnjaka je jednak opsegu desnog bicepsa ambidekstretnih vojnika*

$$\mu_1 = \mu_2, \quad d_0 = 0$$

2) $H_1 \dots$ *Opseg desnog bicepsa dešnjaka je veći od desnog bicepsa ambidekstretnih vojnika*

$$\mu_1 > \mu_2, \quad d_0 > 0$$

$$3) \quad \alpha = 0.05$$

```
##
## Two Sample t-test
##
## data:  rightRBicep and ambiRBicep
## t = 1.2159, df = 3614, p-value = 0.1121
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -0.2370891      Inf
## sample estimates:
## mean of x mean of y
##  35.85597  35.18462
```

P-vrijednost je veća od zadane razine značajnosti, pa se hipoteza H_0 ne odbija.

6 Linearna regresija

6.1 Postoji li veza između nekih varijabli i kilaže vojnika?

6.1.1 Usporedba tjelesne mase i širine struka

Poznato je da se povećanjem tjelesne mase skupljaju masne naslage i to često u predjelu trbuha. Možemo provjeriti na koji način su u ovom skupu podataka povezane ove dvije varijable:

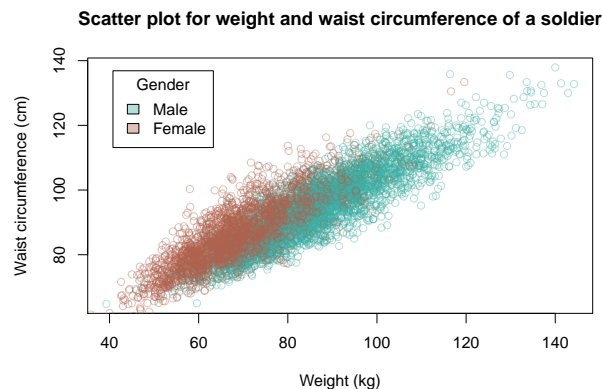
```
waist <- data$waistcircumference/10
maleWaist <- ( data[data$Gender=="Male", ] )$waistcircumference/10
femaleWaist <- ( data[data$Gender=="Female", ] )$waistcircumference/10

# Smanjivanje prozirnosti da se točke bolje vide
col1 <- "#3CAEA365"
col2 <- "#B1624E65"

# Stvaranje plota i ubacivanje točaka o muškarcima
plot(maleWeights,
     maleWaist,
     col=c(col1),
     main="Scatter plot for weight and waist circumference of a soldier",
     xlab="Weight (kg)",
     ylab="Waist circumference (cm)")

# Dodavanje točaka za žene
points(femaleWeights,
       femaleWaist,
       col=c(col2))

# Legenda za spol
legend("topleft",
      inset=.05,
      title="Gender",
      legend=c("Male", "Female"),
      fill=c(col1, col2))
```



Već iz dijagrama raspršenja može se naslutiti nekakva linearnu povezanost ovih dviju varijabli. Podaci se još mogu prikazati i uz pravac linearne regresije nakon podjele po spolu:

```
# Model linearne regresije ( y ~ x )
reg1 <- lm( femaleWaist ~ femaleWeights, data = data )
```

```

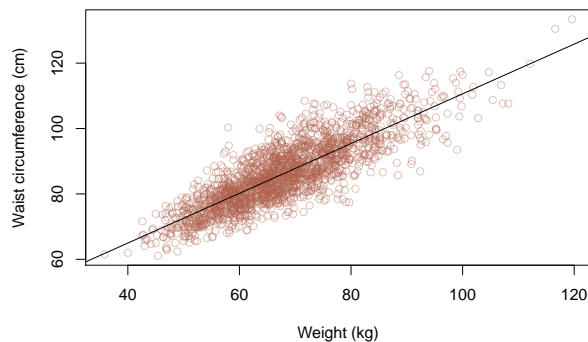
reg2 <- lm( maleWaist ~ maleWeights, data = data )

# Dodavanje točaka za žene
plot(femaleWeights,
     femaleWaist,
     col=c(col2),
     main="Linear regression plot for weight and waist circumference of a female soldier",
     xlab="Weight (kg)",
     ylab="Waist circumference (cm)")
# Pravac regresije za žene
abline(reg1)

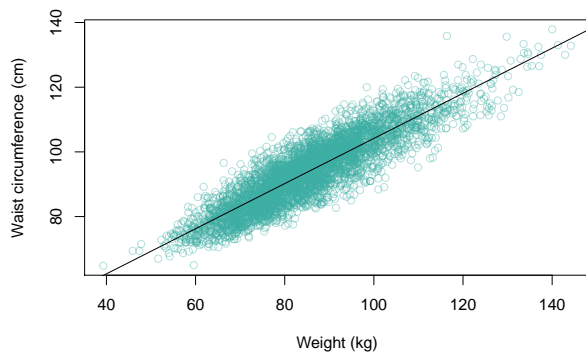
# Dodavanje točaka za muškarce
plot(maleWeights,
     maleWaist,
     col=c(col1),
     main="Linear regression plot for height and weight of a male soldier",
     xlab="Weight (kg)",
     ylab="Waist circumference (cm)")
# Pravac regresije za muškarce
abline(reg2)

```

Linear regression plot for weight and waist circumference of a female soldier



Linear regression plot for height and weight of a male soldier



6.1.2 Provjere pretpostavki linearne regresije između visine i težine

Ovdje su prešućene pretpostavke modela linearne regresije, a to su homogenost varijance i ravnanje reziduala po normalnoj razdiobi: $\mathcal{N}(0, \sigma^2)$. Ove pretpostavke mogu se provjeriti raznim vizualizacijama kao što su histogrami i QQ plotovi reziduala:

```

# Histogram reziduala - žene
hist(reg1$residuals,
     breaks = 15,
     main="Residual histogram for weight and waist circumference of a female soldier",
     xlab = "Waist circumference residuals (cm)",
     col = brown)

# Scatter plot reziduala - žene
plot(femaleWaist, rstandard(reg1),
     main="Residual scatter plot for weight and waist circumference of a female soldier",
     ylab = "Waist circumference residuals (cm)",
     xlab = "Waist circumference (cm)",

```

```

col = brown)

# QQ plot - žene
qqnorm(rstandard(reg1),
      main="Residual QQ plot for weight and waist circumference of a female soldier")

qqline(rstandard(reg1), col = blue, lwd = 2)

# Histogram reziduala - muškarci
hist(reg2$residuals,
     breaks = 15,
     main="Residual histogram for weight and waist circumference of a male soldier",
     xlab = "Waist circumference residuals (cm)",
     col = blue)

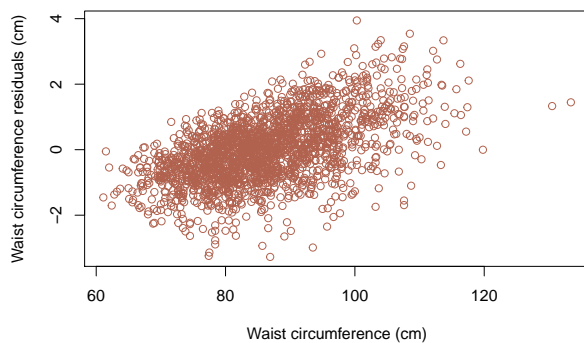
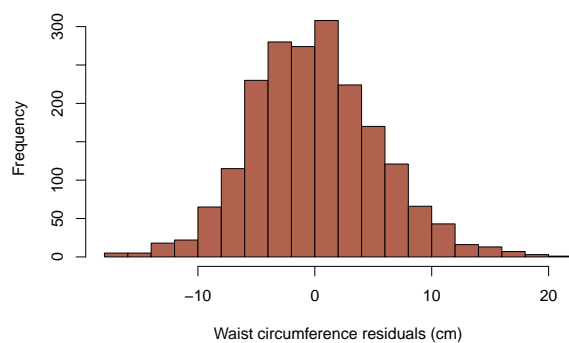
# Scatter plot reziduala - muškarci
plot(maleWaist, rstandard(reg2),
     main="Residual scatter plot for weight and waist circumference of a male soldier",
     ylab = "Waist circumference residuals (cm)",
     xlab = "Waist circumference (cm)",
     col = blue)

# QQ plot - muškarci
qqnorm(rstandard(reg2),
      main="Residual QQ plot for weight and waist circumference of a male soldier")

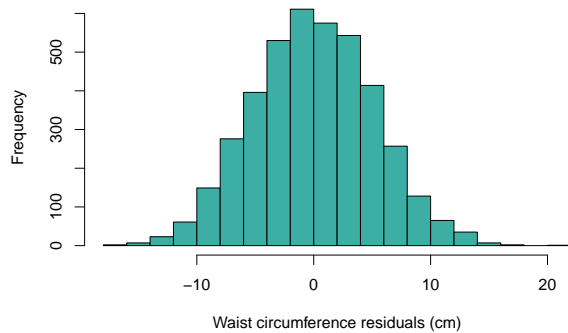
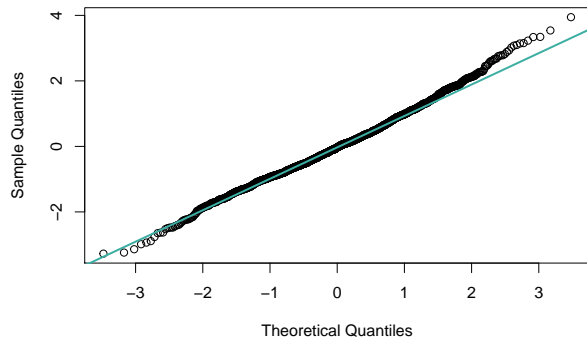
qqline(rstandard(reg1), col = blue, lwd=2)

```

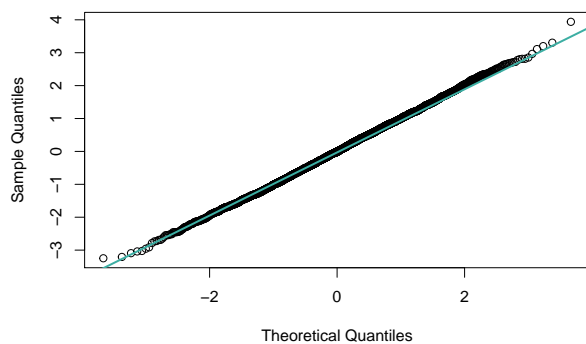
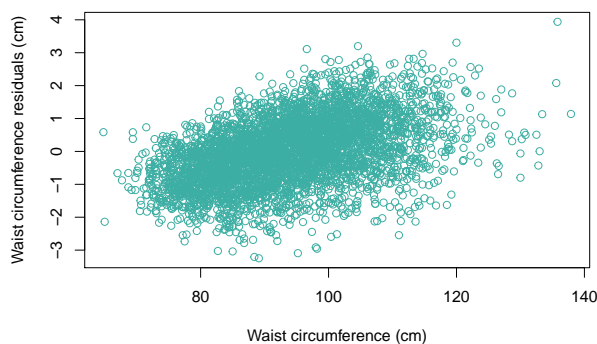
Residual histogram for weight and waist circumference of a female soldier Residual scatter plot for weight and waist circumference of a female soldier



Residual QQ plot for weight and waist circumference of a female sold Residual histogram for weight and waist circumference of a male sold



Residual scatter plot for weight and waist circumference of a male sold Residual QQ plot for weight and waist circumference of a male sold



Ovi grafovi potvrđuju prethodno spomenute pretpostavke. Iz QQ plota može se vidjeti da se reziduali ravnaaju po normalnoj razdiobi s parametrima 0 i σ^2 , a dijagram raspršenja ukazuje da je standardna devijacija podjednaka. Pomoću funkcije summary mogu se saznati još mnoge stvari o modelu. U prvom retku ispisuje se nešto više o raspodjeli reziduala. Dio s koeficijentima daje b_0 i b_1 parametre iz procijenjenog modela i provodi testove s hipotezama:

- 1) $H_0 : \beta_i = 0$
 - 2) $H_1 : \beta_i \neq 0$
- $i \in \{0, 1\}$

pomoću t-testa nad modelom.

```
cat("Tests and parameters for female weight and waist circumference regression:\n")
```

```
## Tests and parameters for female weight and waist circumference regression:
```

```
summary(reg1)
```

```
##
```

```
## Call:
```

```
## lm(formula = femaleWaist ~ femaleWeights, data = data)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
```

```
## -17.9308  -3.7157  -0.1588   3.3797  21.6307
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) 34.55983 0.76938 44.92 <2e-16 ***
## femaleWeights 0.76051 0.01121 67.85 <2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.486 on 1984 degrees of freedom
## Multiple R-squared: 0.6988, Adjusted R-squared: 0.6987
## F-statistic: 4604 on 1 and 1984 DF, p-value: < 2.2e-16
```

```
cat("Correlation coefficient: ",cor(femaleWeights, femaleWaist))
```

```
## Correlation coefficient: 0.8359622
```

```
cat("\n\n")
```

```
cat("Tests and parameters for male weight and waist circumference regression:\n")
```

```
## Tests and parameters for male weight and waist circumference regression:
```

```
summary(reg2)
```

```
##
## Call:
## lm(formula = maleWaist ~ maleWeights, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.663  -3.529  -0.036   3.520  20.199
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.384462   0.489757   70.21  <2e-16 ***
## maleWeights  0.697738   0.005649  123.52  <2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.132 on 4080 degrees of freedom
## Multiple R-squared: 0.789, Adjusted R-squared: 0.7889
## F-statistic: 1.526e+04 on 1 and 4080 DF, p-value: < 2.2e-16
```

```
cat("Correlation coefficient: ",cor(maleWeights, maleWaist))
```

```
## Correlation coefficient: 0.8882537
```

Može se vidjeti da se za sva tri testa vezana za parametre modela, hipoteza H_0 odbija u korist hipoteze H_1 što govori da parametri nisu jednaki nula. Korelacijski koeficijenti pokazuju da modeli u nekoj mjeri objašnjavaju tjelesnu masu, no poželjno je da su većeg iznosa.

6.2 Usporedba više parametara i tjelesne mase

Možemo sada pokušati objasniti tjelesnu masu pomoću nekih drugih značajki. Varijable koje bi mogle objašnjavati tjelesnu masu su visina, opseg struka, opseg zapešća, opseg bedra, opseg lista i opseg vrata. Za ovo se može koristiti model višestruke regresije, no potrebno je paziti na međusobnu koreliranost regresora. Ovo se lako provjeri pomoću korelacijske matrice:

```
height <- data$stature/10
waist <- data$waistcircumference/10
wrist <- data$wristcircumference/10
```

```
ankle <- data$anklecircumference/10
thigh <- data$thighcircumference/10
calf <- data$calfcircumference/10
neck <- data$neckcircumference/10

cat("Correlation matrix of regressors:\n")
```

```
## Correlation matrix of regressors:
cor(cbind(height, waist, ankle, thigh, calf, neck))
```

```
##          height      waist      ankle      thigh      calf      neck
## height  1.0000000  0.3687336  0.5170262  0.2524165  0.4002015  0.6311859
## waist   0.3687336  1.0000000  0.5693055  0.7830274  0.6953047  0.6762682
## ankle   0.5170262  0.5693055  1.0000000  0.6016520  0.8178611  0.5803084
## thigh   0.2524165  0.7830274  0.6016520  1.0000000  0.7771693  0.4540465
## calf    0.4002015  0.6953047  0.8178611  0.7771693  1.0000000  0.5796284
## neck    0.6311859  0.6762682  0.5803084  0.4540465  0.5796284  1.0000000
```

U ovom ispisu neki od regresora imaju relativno velik koeficijent korelacije. Mogu se provjeriti ove vrijednosti za npr. visinu, opseg struka i opseg gležnja.

```
##          height      waist      ankle
## height  1.0000000  0.3687336  0.5170262
## waist   0.3687336  1.0000000  0.5693055
## ankle   0.5170262  0.5693055  1.0000000
```

Pošto muškarci imaju više testosterona očekuje se da imaju jaču tjelesnu građu. Ova nominalna varijabla može se uključiti u model višestruke regresiju. Provjeriti ćemo dijagramom raspršenja utječe li spol na tjelesnu masu i testirati hipoteze $\beta_i = 0$, $i \in \{0, 1, 2, 3, 4\}$ s alternativom $\beta_i \neq 0$ o ovom modelu.



Iz ovog ispisa vidimo da je p-vrijednost za sve testove izrazito mala, što potvrđuje odbijanje hipoteze H_0 u korist alternative H_1 . Možda opseg gležnja i ne utječe toliko na tjelesnu masu vojnika. Moguće je izbaciti ga iz modela i usporediti rezultate.

```
##
## Call:
## lm(formula = weights ~ height + waist + Gender_Male, data = dummyGenderData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.6717  -3.4263  -0.1466   3.2147  20.3390
##
```

```
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.175e+02  1.585e+00  -74.12  <2e-16 ***
## height      6.142e-01  9.869e-03   62.24  <2e-16 ***
## waist       9.897e-01  6.137e-03  161.26  <2e-16 ***
## Gender_Male  2.035e+00  1.862e-01   10.93  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.043 on 6064 degrees of freedom
## Multiple R-squared:  0.8963, Adjusted R-squared:  0.8962
## F-statistic: 1.747e+04 on 3 and 6064 DF,  p-value: < 2.2e-16
```

Vrijednost prilagođenog koeficijenta determinacije R^2 i R_{adj}^2 se smanjila za oko 0.03 što pokazuje da opseg gležnja ipak ima neki utjecaj tjelesnu masu pa ga ne moramo izbacivati iz modela.

7 Koje mjere kaciga biste preporučili američkoj vojsci i koliko različitih veličina?

Kako bi se moglo reći nešto više o veličini kaciga za američku vojsku, potrebno je provjeriti npr. varijablu opsega, širine, dužine glave i udaljenosti od vrha glave do uha. Pošto su muškarci jače tjelesne građe može se pogledati utjecaj spola na ove varijable.

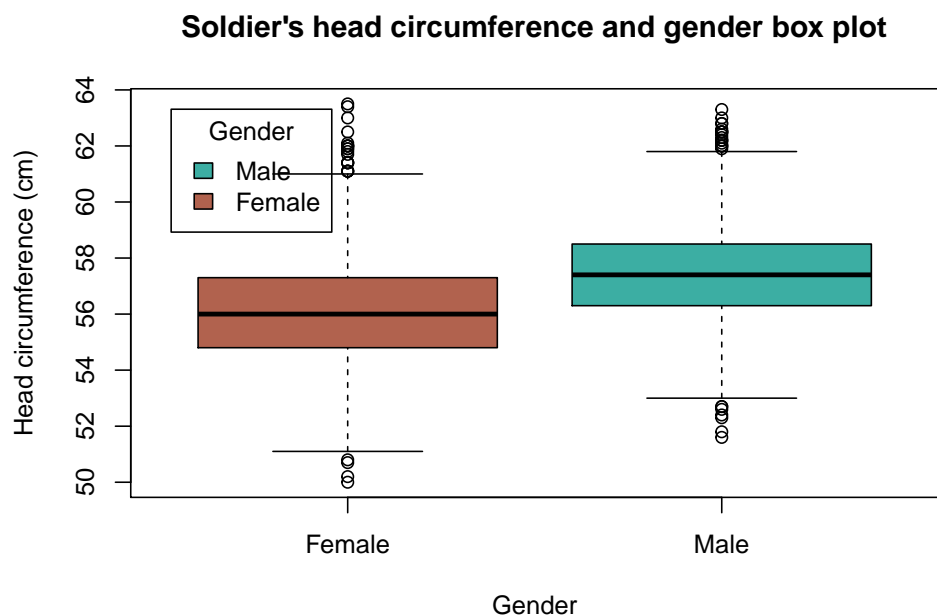
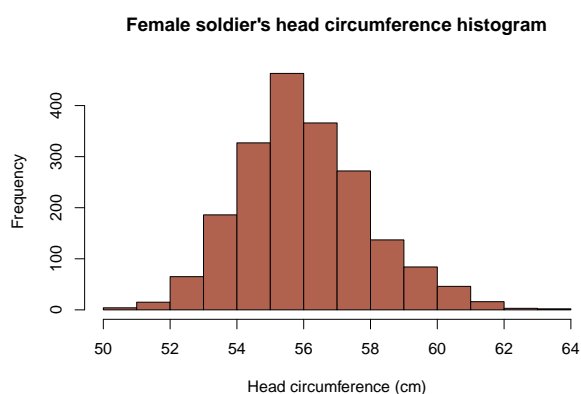
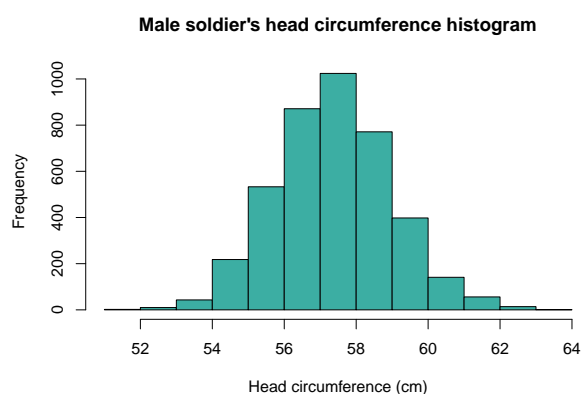
7.1 Vizualizacija i numeričke karakteristike

Male head circumference:

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	51.60	56.30	57.40	57.44	58.50	63.30

Female head circumference:

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	50.00	54.80	56.00	56.11	57.30	63.50

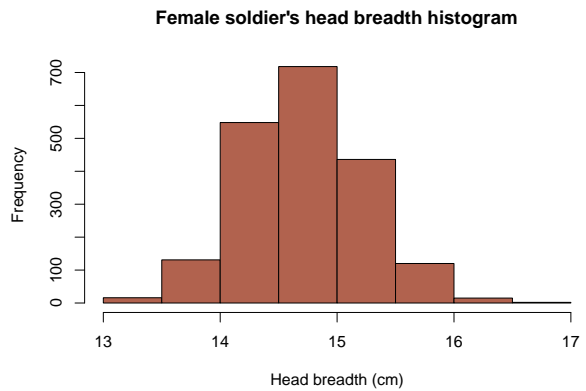


Male head breadth:

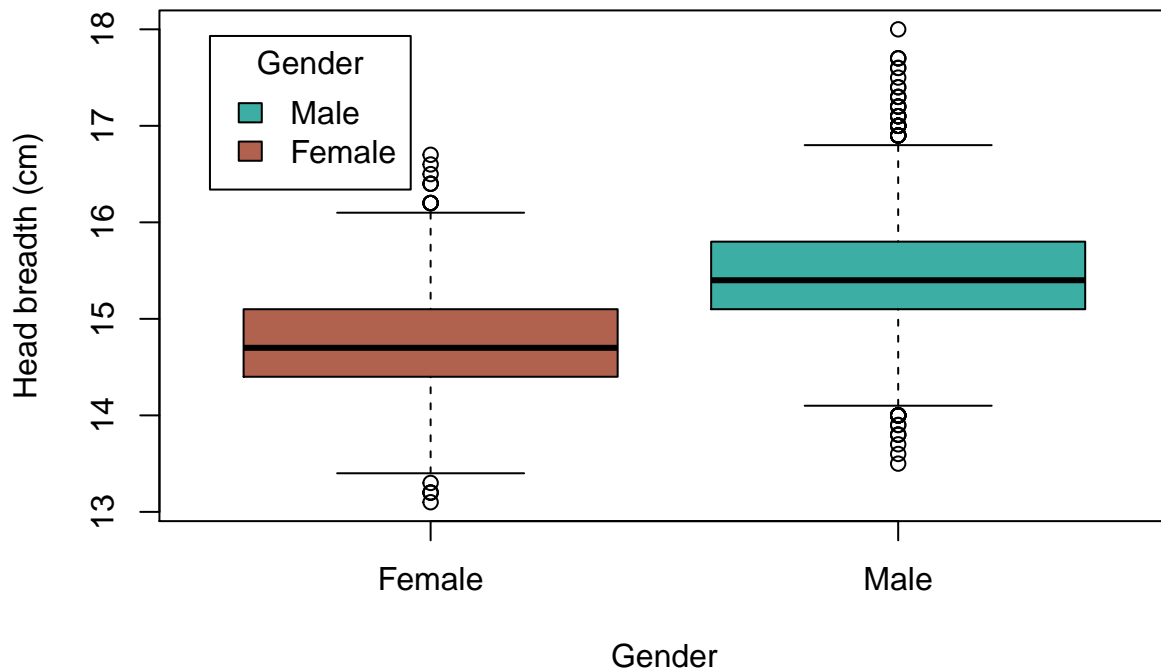
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	13.50	15.10	15.40	15.43	15.80	18.00

Female head breadth:

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	13.10	14.40	14.70	14.78	15.10	16.70



Soldier's head breadth and gender box plot

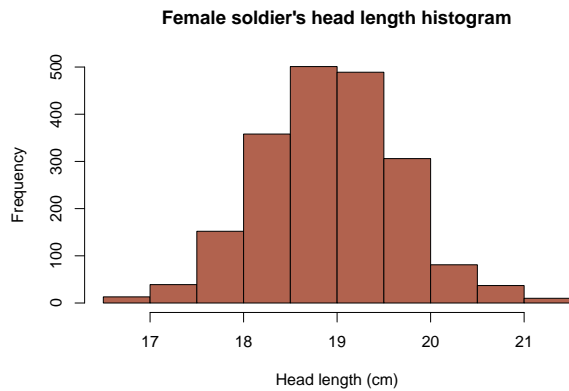
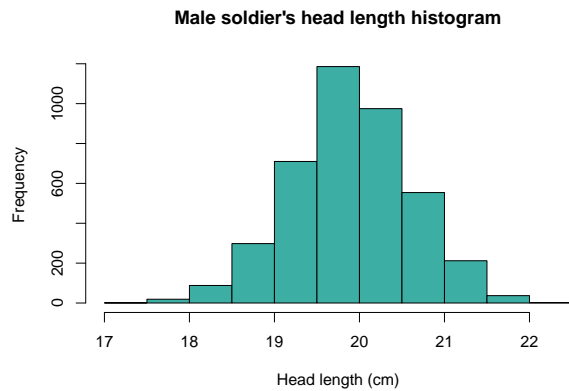


Male head length:

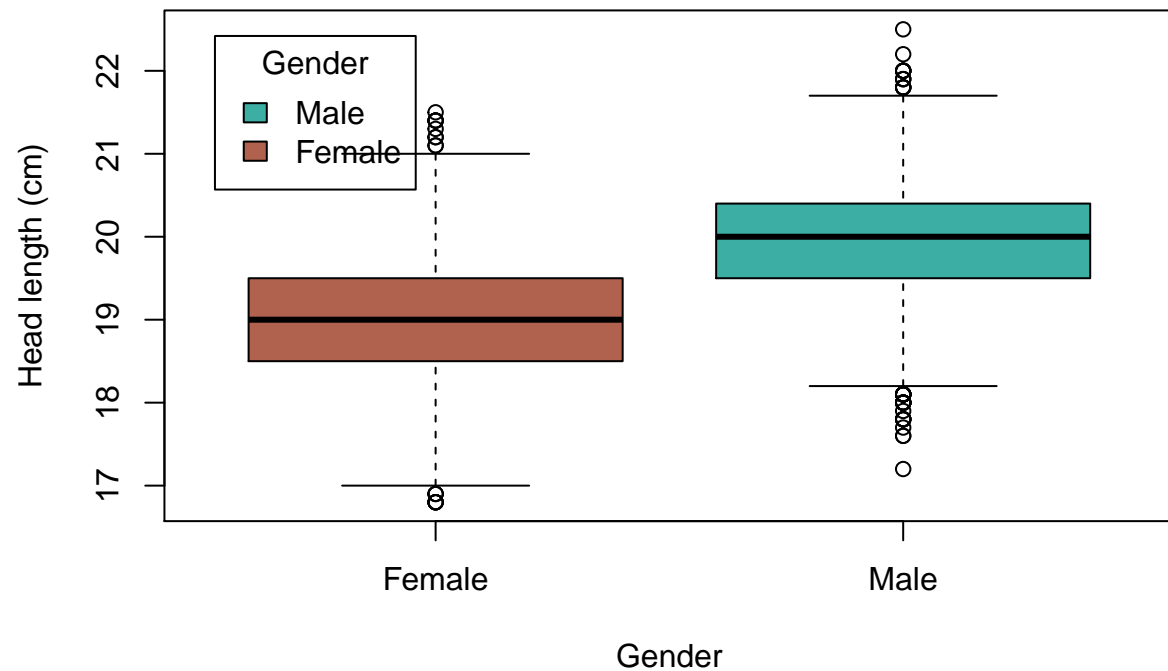
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	17.20	19.50	20.00	19.95	20.40	22.50

Female head length:

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	16.80	18.50	19.00	18.98	19.50	21.50



Soldier's head length and gender box plot



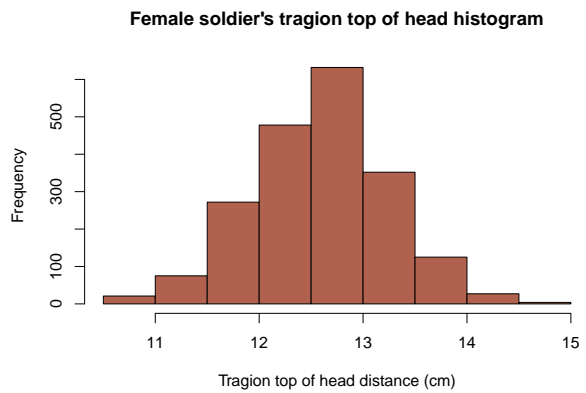
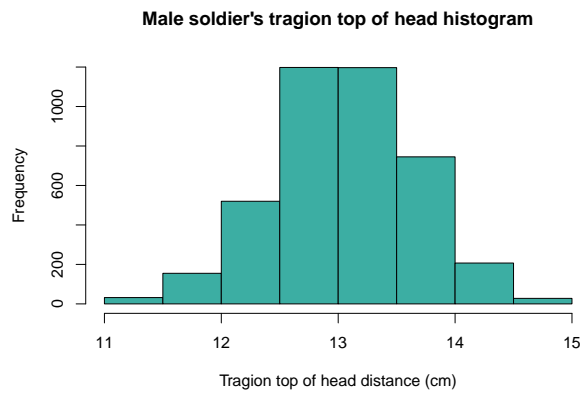
Male tragon top of head distance:

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	11.00	12.70	13.10	13.11	13.50	15.00

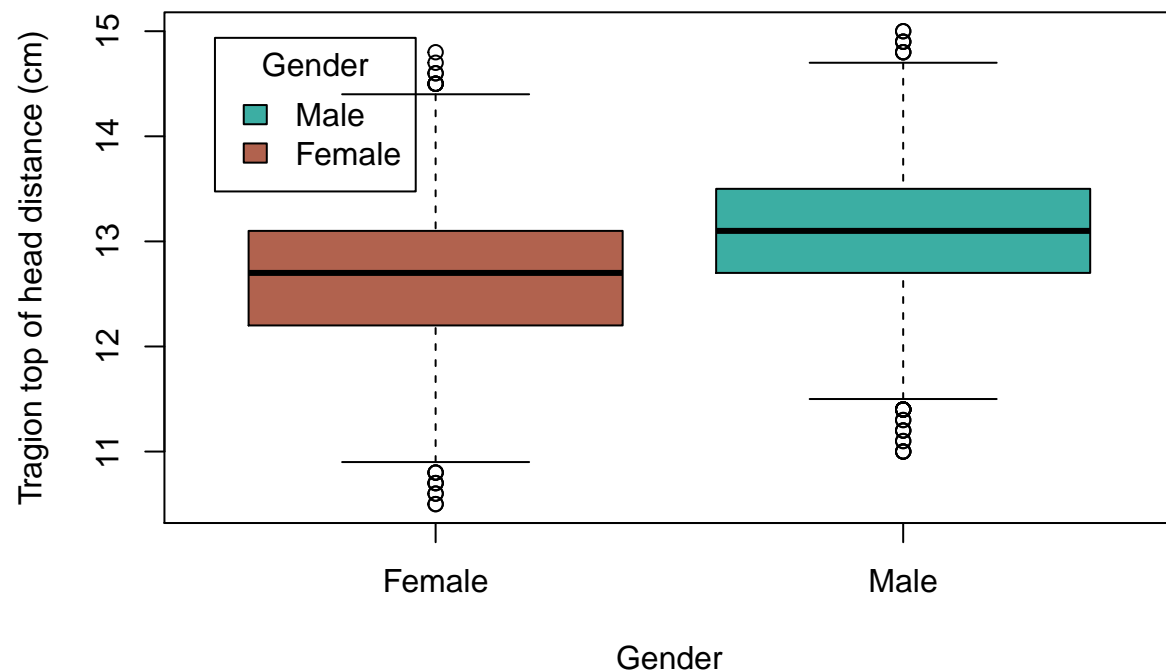
Female tragon top of head distance:

n

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	10.50	12.20	12.70	12.65	13.10	14.80



Soldier's tragon top of head and gender box plot



7.2 Prijedlozi veličina kaciga

Veličine kaciga				
Oznaka veličine	Opseg kacige (cm)	Širina kacige (cm)	Dužina kacige (cm)	Dubina kacige (cm)
S	55	15	18.5	12
M	58	16	20	13
L	61	17	21.5	14
XL	64	18.5	23	15

Uzevši u obzir numeričke karakteristike oba spola, odlučili smo se za 4 veličine kacige. Veličine koje smo odabrali su: S, M, L i XL. U tablici su navedene sve mjere koje bi pojedine veličine imale. Čini nam se da bi na ovaj način svatko dobio kacigu koja bi mu bila udobna za nositi, ali da opet nije potrebno napraviti puno različitih veličina. Ove veličine su univerzalne i namijenjene za oba spola. Muškarci bi uglavnom nosili M ili L kacige, a većina žena S ili M kacige.