

SAP - Projekt

Case study *Antropometrijski podaci američkih vojnih snaga*

Marija Dragošević, Petra Ilić, Petra Jović, Bartol Rod

15. January, 2021.

Sadržaj

1	Uvod	3
1.1	Prikaz podataka	3
2	Analiza vojnih kampova	6
2.1	Osnovne metode deskriptivne statistike	6
2.2	ANOVA	7
2.3	Logistička regresija	13
3	Predviđanje kilaže vojnika	14
3.1	Jednostavna regresija	14
3.1.1	Normalnost reziduala i homogenost varijance	20
3.1.2	Ocjena kvalitete linearног modela i statističko zaključivanje o procijenjenom modelu	29
3.1.3	Korelacijski koeficijenti u linearном modelu	31
3.2	Višestruka regresija	32
3.2.1	Provjera normalnosti reziduala	34
3.2.2	Uključivanje kategorijskih varijabli	36
3.2.3	Zaključak	37
4	Opseg bicepsa	38
4.1	Prikaz raspodjele vojnika ovisno o preferenciji ruke i spolu	38
4.2	Test o dvije proporcije: dva uzorka	40
4.2.1	Udjeli ljevaka ovisno o spolu	40
4.3	Opseg bicepsa ovisno o preferenciji ruke	41
4.3.1	Opseg desnog bicepsa dešnjaka i ljevaka	41
4.3.1.1	Provjera normalnosti podataka	41
4.3.1.2	Test o jednakosti varijanci	48
4.3.1.3	t-test uz pretpostavku jednakosti varijanci	48
4.3.2	Opseg desnog bicepsa dešnjaka i ambidekstrenih vojnika	48
4.3.2.1	Test o jednakosti varijanci	48
4.3.2.2	t-test uz pretpostavku jednakosti varijanci	49
5	Mjere kaciga	50
5.1	Klase podataka	50
5.2	Summary	50
5.3	Mjere centralne tendencije i mjere rasipanja	50
5.4	Vizualizacija podataka	50
5.5	Prikaz veličina kaciga	56
5.6	Histogrami s custom breakpointovima po veličinama kaciga	62
5.7	Brojevi vojnika kojima treba određena veličina kacige	66
5.8	Postotci udjela veličina kaciga u vojski	67

1 Uvod

U ovom projektu bavit će se analizom antropometrijskih podatke američkih vojnih snaga (ANSUR_II). Podatci se sastoje od 6068 vojnika i 107 značajki sakupljenih od 2010 do 2012 u raznim američkim vojnim kampovima.

1.1 Prikaz podataka

abdominalextensiondepthsitting	231	194	183
acromialheight	1282	1379	1369
acromionradialelength	301	320	329
anklecircumference	204	207	233
axillaheight	1180	1292	1271
balloffootcircumference	222	225	237
balloffootlength	177	178	196
biacromialbreadth	373	372	397
bicepscircumferenceflexed	315	272	300
bicristalbreadth	263	250	276
bideltoidbreadth	466	430	450
bimalleolarbreadth	65	64	69
bitragionchinarc	338	294	309
bitragionsubmandibulararc	301	270	270
bizygomaticbreadth	141	126	128
buttockcircumference	1011	893	987
buttockdepth	223	186	204
buttockheight	836	900	861
buttockkneelength	587	583	583
buttoockpopliteallength	476	483	466
calfcircumference	360	350	384
cervicaleheight	1336	1440	1451
chestbreadth	274	261	287
chestcircumference	922	839	874
chestdepth	245	206	223
chestheight	1095	1234	1226
crotchheight	759	835	821
crotchlenthomphalion	557	549	643
crotchlenthposterioromphalion	310	329	374
earbreadth	35	32	36
earlength	65	60	65
earprotrusion	16	23	26
elbowrestheight	220	208	204
eyeheightssitting	713	726	790
footbreadthhorizontal	91	91	100
footlength	246	249	265
forearmcenterofgriplength	316	341	343
forearmcircumferenceflexed	265	247	262
forearmforearmbreadth	517	468	488
forearmhandlength	432	463	469
functionalleglength	1028	1117	1060
handbreadth	75	78	84
handcircumference	182	187	198

(continued)

handlength	184	189	195
headbreadth	141	138	146
headcircumference	548	535	588
headlength	191	180	207
heelanklecircumference	314	307	331
heelbreadth	69	60	70
hipbreadth	345	315	356
hipbreadthsitting	388	335	399
iliocristaleheight	966	1048	1043
interpillarybreadth	645	595	655
interscyei	363	340	345
interscyeii	399	375	399
kneeheightmidpatella	435	483	470
kneeheightsitting	496	532	530
lateralfemoralepicondyleheight	447	492	469
lateralmalleolusheight	55	69	64
lowerthighcircumference	404	334	401
mentonsellionlength	118	115	135
neckcircumference	335	302	325
neckcircumferencebase	368	345	369
overheadfingertipreachsitting	1268	1389	1414
palmlength	113	110	122
poplitealheight	362	426	398
radialestyliionlength	235	259	258
shouldercircumference	1062	1014	1049
shoulderelbowlength	327	346	362
shoulderlength	148	142	164
sittingheight	803	835	904
sleevelengthspinewrst	809	810	855
sleeveoutseam	513	575	568
span	1647	1751	1779
stature	1560	1665	1711
suprasternaleheight	1280	1372	1383
tenthribheight	1013	1107	1089
thighcircumference	622	524	577
thighclearance	174	152	164
thumbtipreach	736	771	814
tibialheight	430	475	458
tragiontopofhead	110	125	129
trochanterionheight	844	901	882
verticaltrunkcircumferenceusa	1488	1470	1542
waistbacklength	406	422	419
waistbreadth	295	254	269
waistcircumference	850	708	727
waistdepth	217	168	159
waistfrontlengthsitting	345	329	367
waistheightomphalion	942	1032	1035
weightkg	657	534	663
wristcircumference	152	155	162

(continued)

wristheight	756	815	799
Gender	Female	Female	Female
Installation	Fort Hood	Fort Hood	Fort Hood
Component	Regular Army	Regular Army	Regular Army
Branch	Combat Support	Combat Service Support	Combat Service Support
Age	26	21	23
Heightin	61	64	68
Weightlbs	142	120	147
WritingPreference	Right hand	Right hand	Right hand

2 Analiza vojnih kampova

2.1 Osnovne metode deskriptivne statistike

```
## # A tibble: 12 x 2
## # Groups:   Installation [12]
##   Installation     n
##   <chr>        <int>
## 1 Camp Atterbury    441
## 2 Camp Shelby      1160
## 3 Fort Bliss       963
## 4 Fort Bragg       397
## 5 Fort Drum        391
## 6 Fort Gordon      669
## 7 Fort Hood        439
## 8 Fort Huachuca    436
## 9 Fort Lee          380
## 10 Fort McCoy       452
## 11 Fort Rucker       1
## 12 Fort Stewart     339
```

Prebrojavanjem osoba po kampovima možemo vidjeti da se u kampu "Fort Rucker" nalazi samo jedna osoba, stoga taj kamp izbacujemo iz analize.

Za analizu smo nasumično odabrali sljedeće atribute:

- opseg prsa (chestcircumference)
- duljinu ruke (handlength)
- širina bokova (hipbreadth)
- službu (Branch)
- spol (Gender)

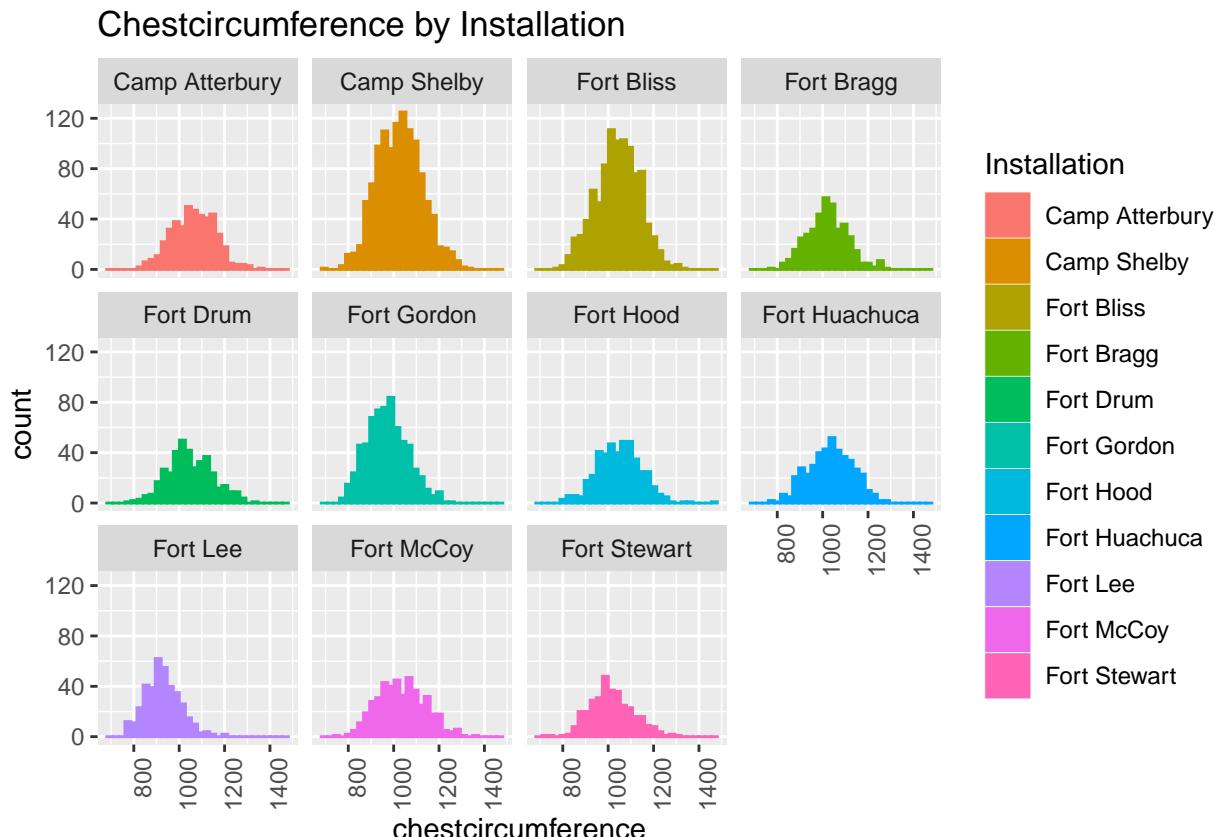
```
##   chestcircumference   handlength   hipbreadth   Gender
##   Min. : 695           Min. :145.0     Min. :264.0   Length:6067
##   1st Qu.: 949         1st Qu.:181.0     1st Qu.:332.0   Class  :character
##   Median :1021         Median :190.0     Median :347.0   Mode   :character
##   Mean   :1022         Mean   :189.3     Mean   :348.4
##   3rd Qu.:1092         3rd Qu.:197.0     3rd Qu.:364.0
##   Max.  :1469         Max.  :239.0     Max.  :473.0
##   Installation       Branch
##   Length:6067         Length:6067
##   Class  :character   Class  :character
##   Mode   :character   Mode   :character
##
##
```

Iz izračuna srednje vrijednosti, medijana i standardne devijacije možemo naslutiti razliku u podatcima razvrstanim po kampovima. Uočavamo kako je i srednja vrijednost približno jednaka medijanu što upućuje na normalnost podataka. Razvrstavanjem po spolu u vojnim kampovima veći je udio muškaraca, iznimka je "Fort Lee" u kojem su samo žene. Također u kampu "Fort Lee" vidimo kako su i niže vrijednosti srednjih vrijednosti podataka nego u ostalim kampovima, no standardna devijacija ne odstupa.

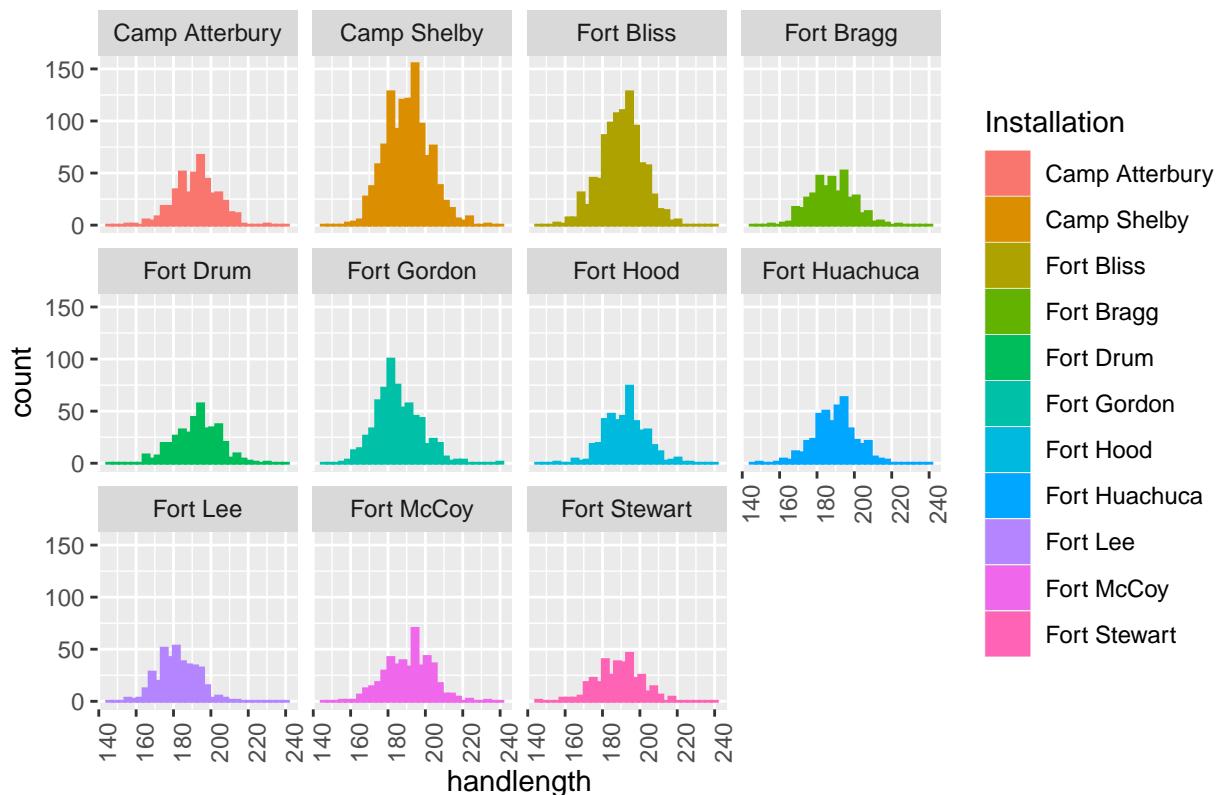
2.2 ANOVA

```
##      Installation chestcircumference.p handlength.p    hipbreadth
## 1      Ungrouped   4.664006e-05 5.729598e-09 1.073183e-18
## 2      Fort Hood   7.832516e-01 3.348776e-02 1.456905e-05
## 3      Fort Bliss  1.792532e-01 8.196402e-05 2.577246e-02
## 4 Camp Atterbury  6.656600e-01 9.231877e-02 8.424878e-03
## 5      Fort Drum   4.796294e-03 2.378153e-02 3.905721e-02
## 6      Fort McCoy  1.274228e-01 1.569588e-06 1.074329e-01
## 7      Fort Lee    5.751931e-02 6.642832e-02 5.422562e-01
## 8      Fort Stewart 9.520247e-03 9.385869e-02 4.391390e-04
## 9      Fort Bragg  7.155254e-01 1.268504e-01 5.580023e-05
## 10     Fort Gordon 9.824675e-02 3.498124e-08 9.352386e-04
## 11     Fort Huachuca 4.262501e-01 1.010728e-01 5.084946e-02
## 12     Camp Shelby  4.069465e-02 1.958891e-03 1.013141e-06
```

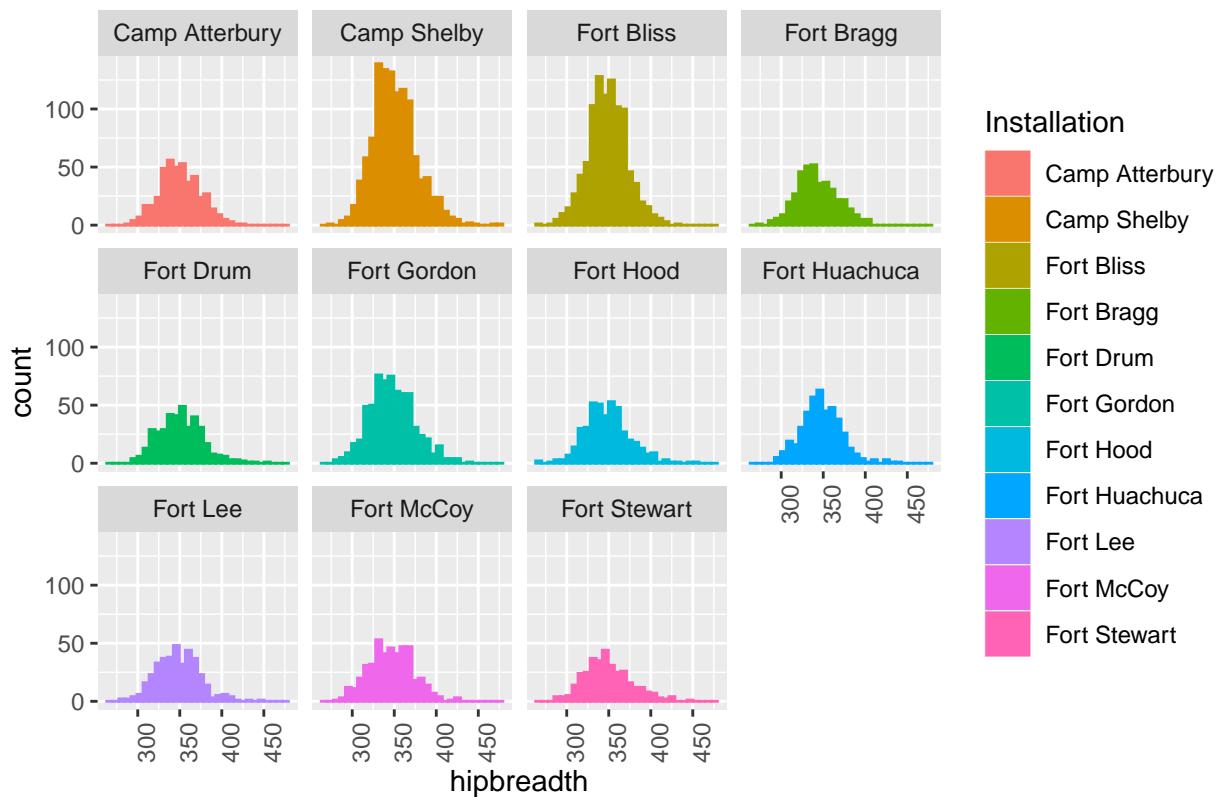
Provodenjem Kolmogorov-Smirnovog testa možemo zaključiti kako se neki podatci ponašaju veoma normalno. Također znatna razlika u p-vrijednosti negrupiranih podataka također može ukazati na postojanje razlike u grupama.



Handlength by Installation

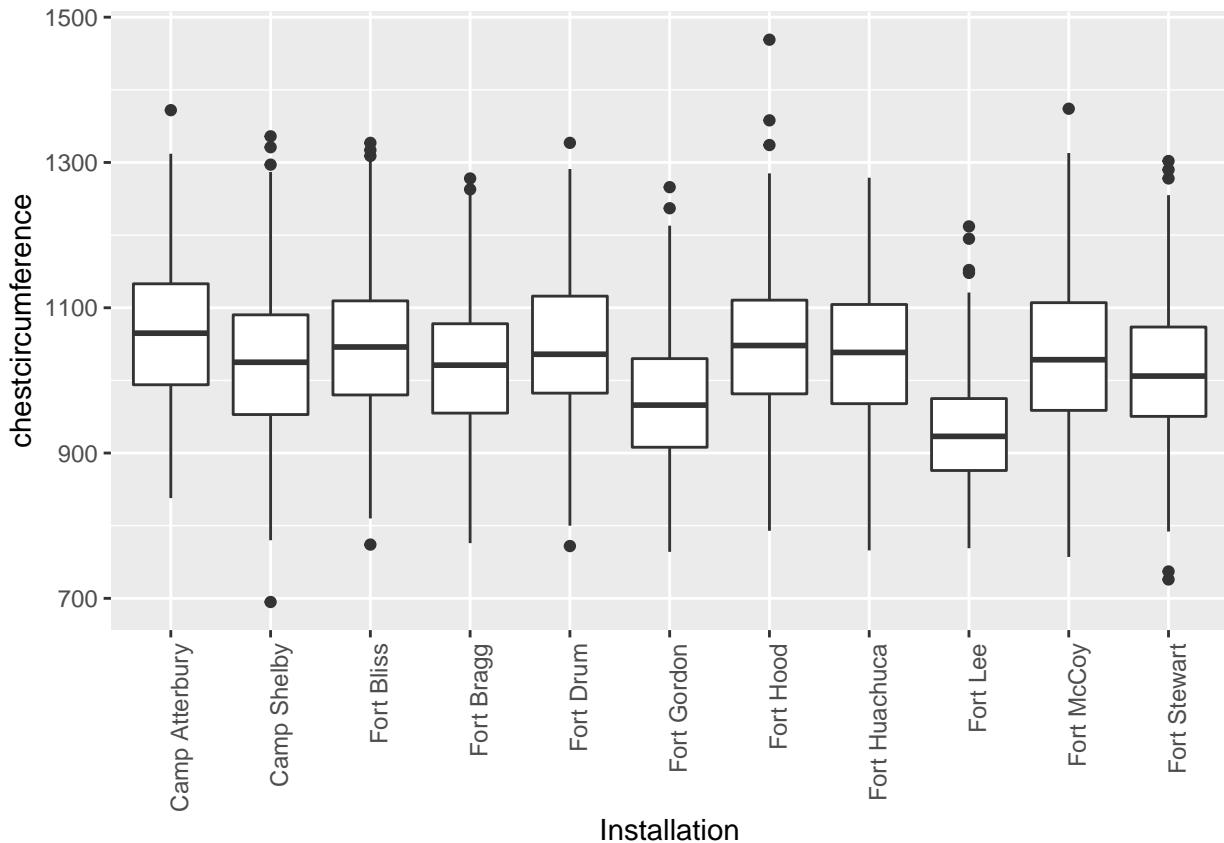


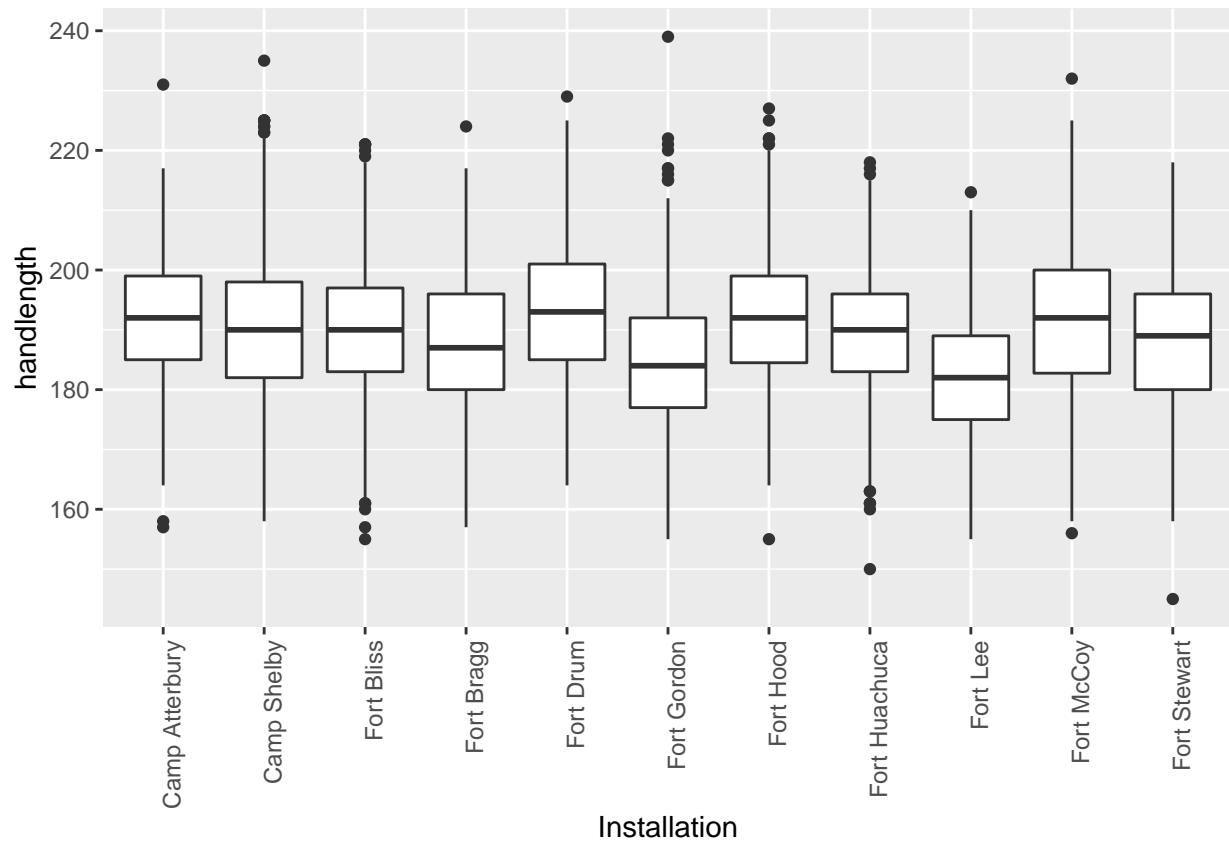
Hipbreadth by Installation

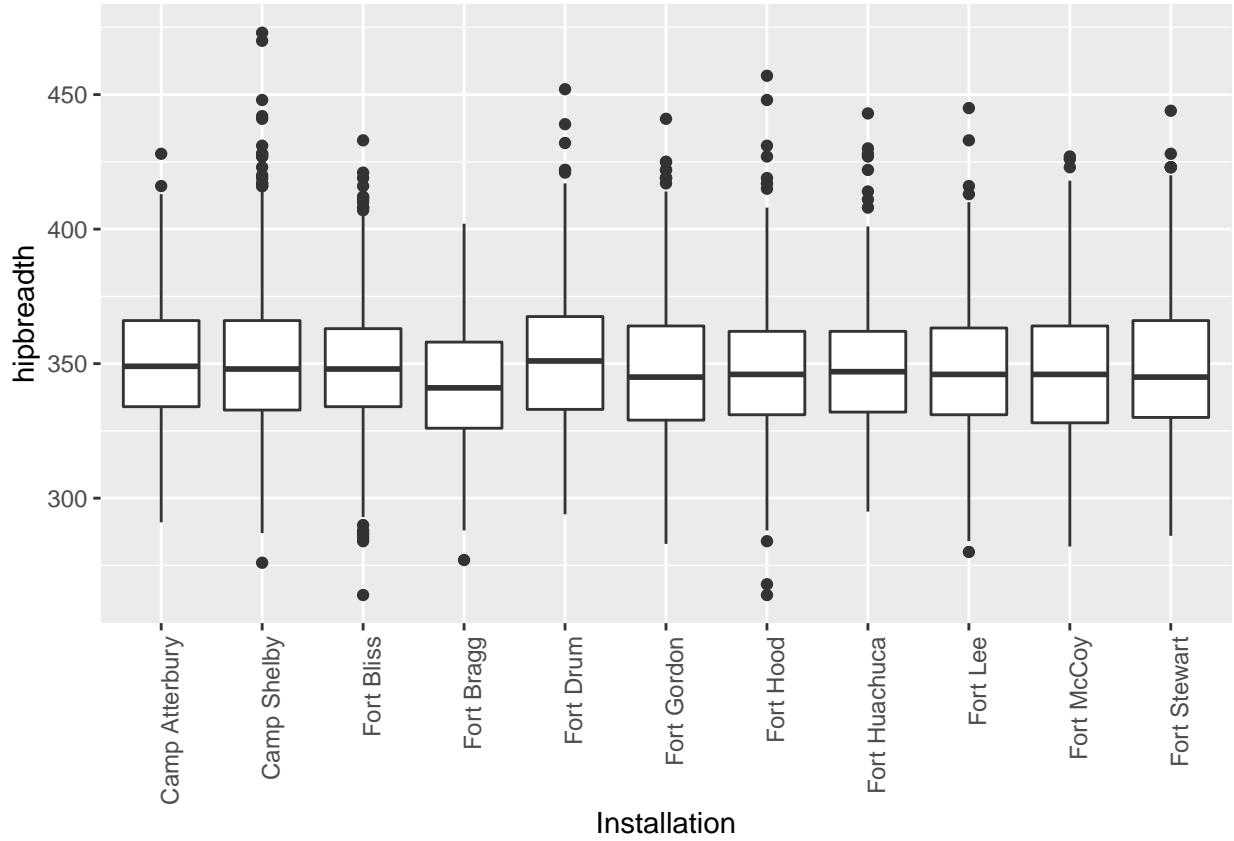


Testiranjem na normalnost te iz grafičkog prikaza značajki kampova možemo naslutiti kako su podaci u kampovima normalno raspoređeni.

```
##  
##  Bartlett test of homogeneity of variances  
##  
## data: df.camps$chestcircumference by df.camps$Installation  
## Bartlett's K-squared = 52.789, df = 10, p-value = 8.15e-08  
  
##  
##  Bartlett test of homogeneity of variances  
##  
## data: df.camps$handlength by df.camps$Installation  
## Bartlett's K-squared = 17.941, df = 10, p-value = 0.05596  
  
##  
##  Bartlett test of homogeneity of variances  
##  
## data: df.camps$hipbreadth by df.camps$Installation  
## Bartlett's K-squared = 43.986, df = 10, p-value = 3.311e-06
```







Grafički prikaz opsega prsa i duljinu ruke sugerira na jasnu razliku u podacima među kampvima, no kod prikaza opsega kuka to ne možemo vidjeti.

```
##           Df   Sum Sq Mean Sq F value Pr(>F)
## df.camps$Installation 10 704592 704592 78.38 <2e-16 ***
## Residuals             6056 54441338     8990
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##           Df   Sum Sq Mean Sq F value Pr(>F)
## df.camps$Installation 10 44991    4499   35.95 <2e-16 ***
## Residuals             6056 757844     125
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##           Df   Sum Sq Mean Sq F value   Pr(>F)
## df.camps$Installation 10 28010  2801.0   4.401 3.46e-06 ***
## Residuals             6056 3854373   636.5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Provodenjem ANOVA testa po pojedinim značajkama uočavamo kako postoji značajna razlika u odabranim antropometrijskim podacima razvrstanih po kampovima.

2.3 Logistička regresija

Želimo predvidjeti hoće li neki odabrani vojnik po svojim podatcima pripasti nekom kampu ili ne.

3 Predviđanje kilaže vojnika

Kako bismo mogli predvidjeti kilažu vojnika, ispitat ćemo različite varijable koje bi mogle utjecati na nju:

- visina
- opseg struka
- opseg vrata
- duljina prednjeg struka u sjedećem položaju
- opseg bedra

3.1 Jednostavna regresija

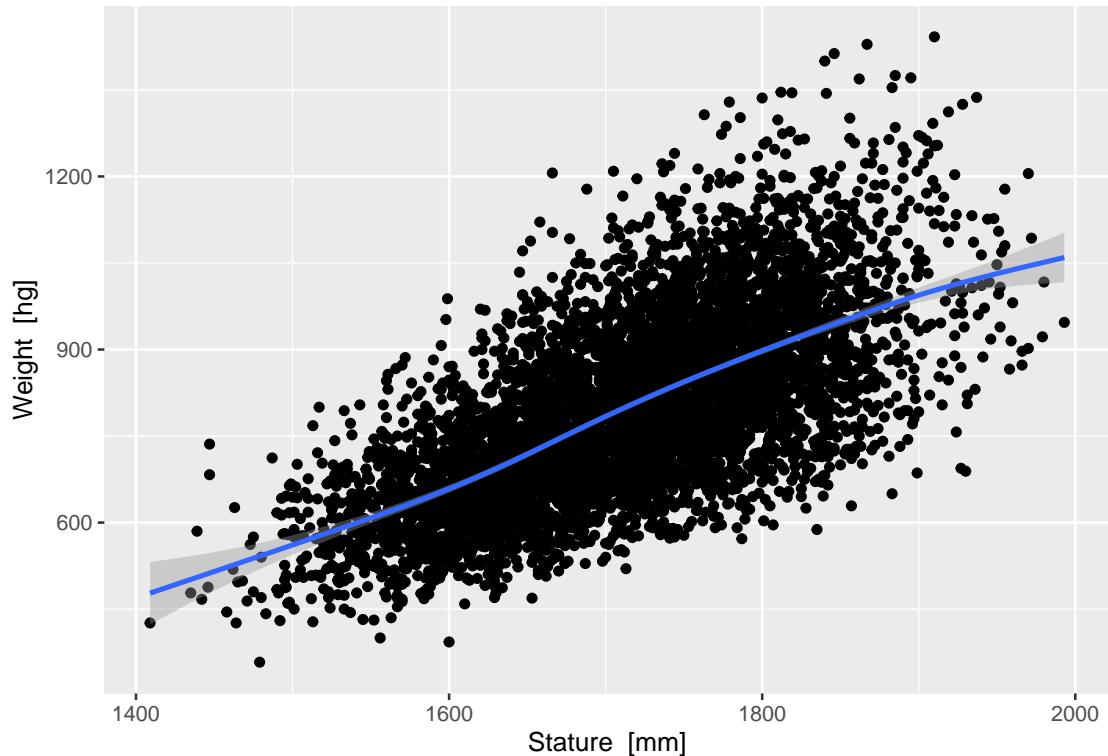


Figure 1: Utjecaj visine na kilažu

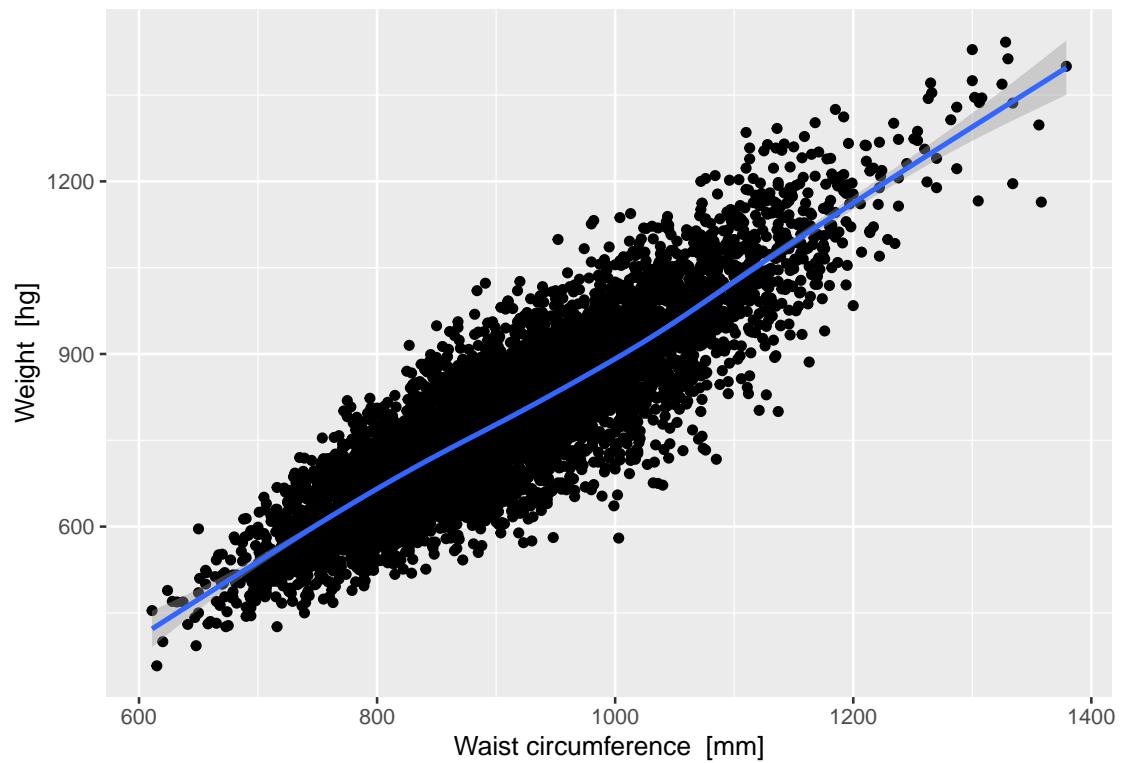


Figure 2: Utjecaj opsega struka na kilažu

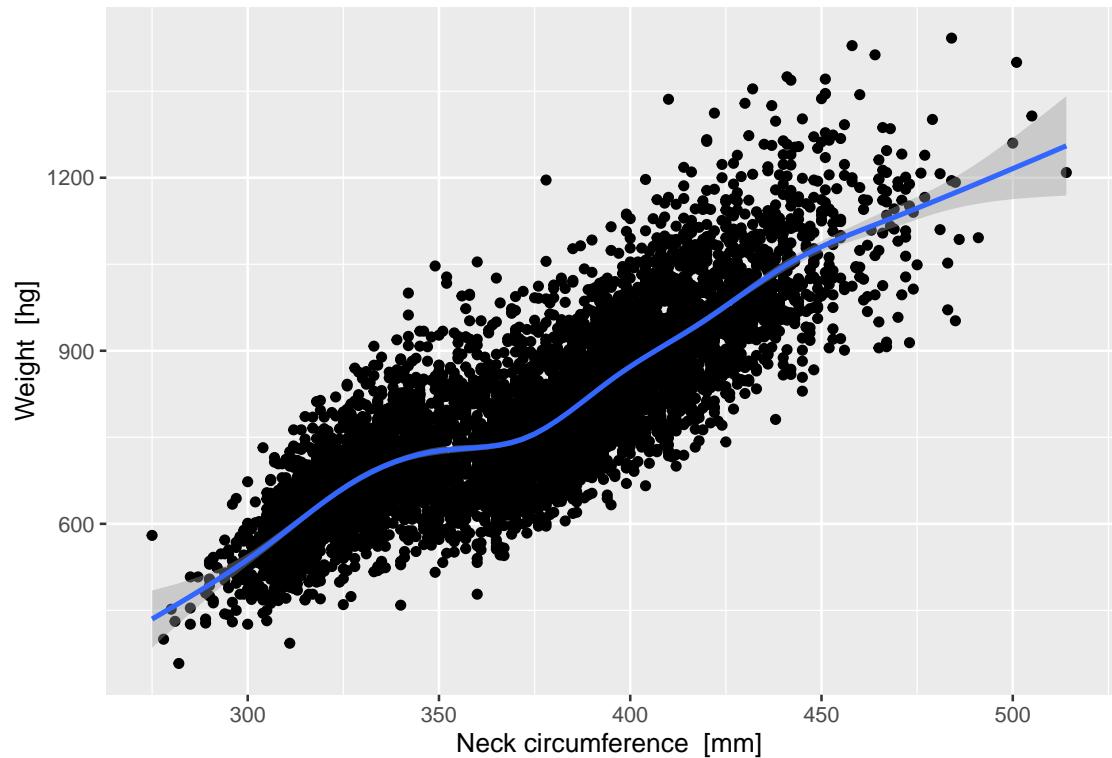


Figure 3: Utjecaj opsega vrata na kilažu

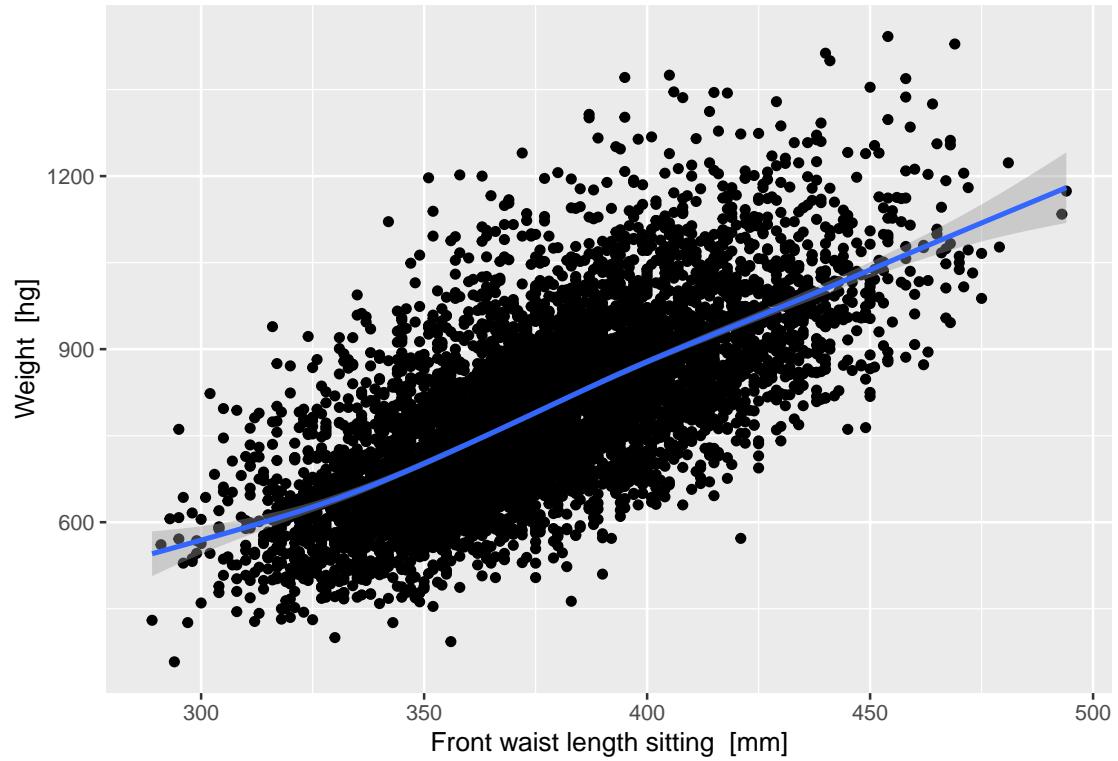


Figure 4: Utjecaj duljine prednjeg struka u sjedecem položaju na kilažu

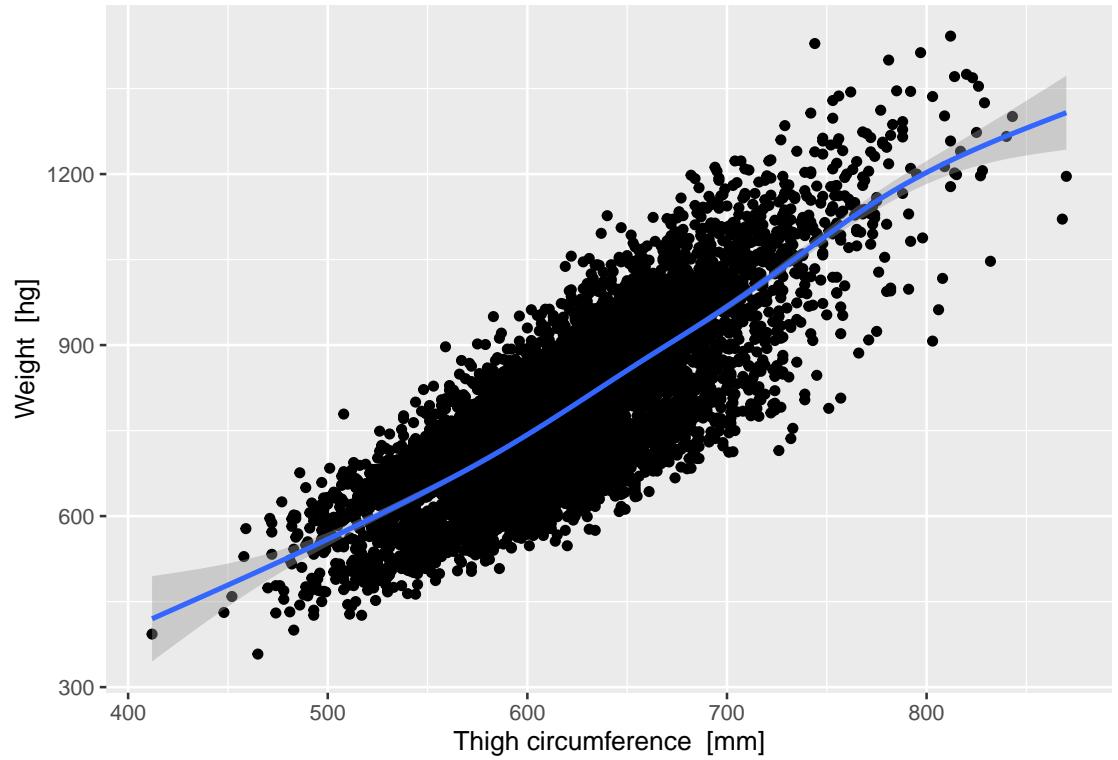


Figure 5: Utjecaj opsega bedra na kilažu

Iz grafičkih prikaza vidi se da sve navedene varijable imaju pozitivan utjecaj na izlaznu varijablu kilaže. Najizraženiji je utjecaj opsega struka.

Kako bismo mogli ispitati pojedinačni utjecaj varijabli, procijenit ćemo model jednostavne regresije u kojem su navedenih pet varijabli (svaka u zasebnom modelu) nezavisne, a kilaža zavisna varijabla.

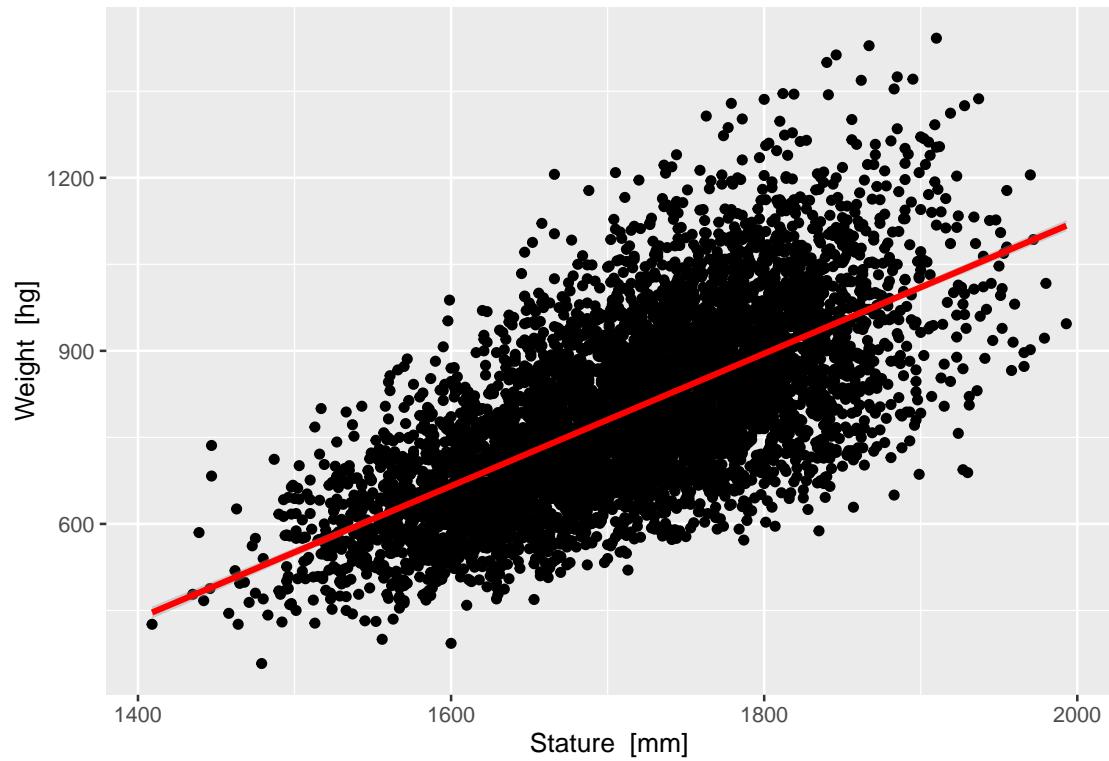


Figure 6: Linearni model kilaže i visine

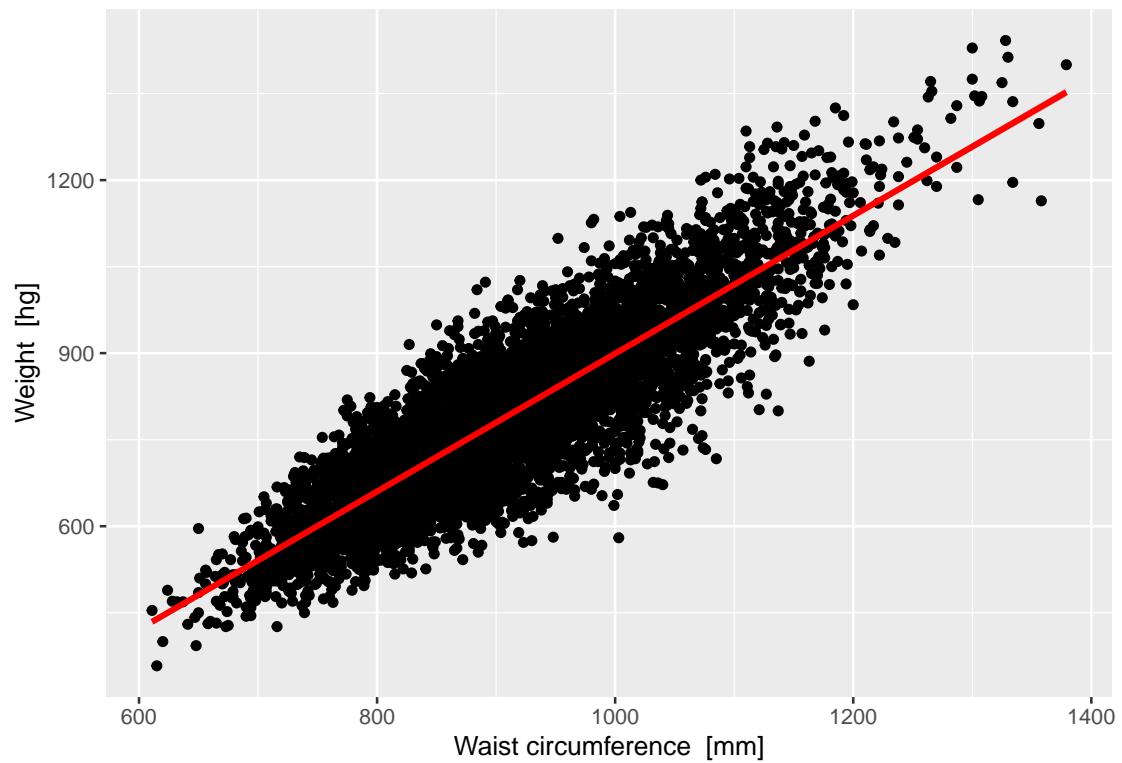


Figure 7: Linearni model kilaže i opsega struka

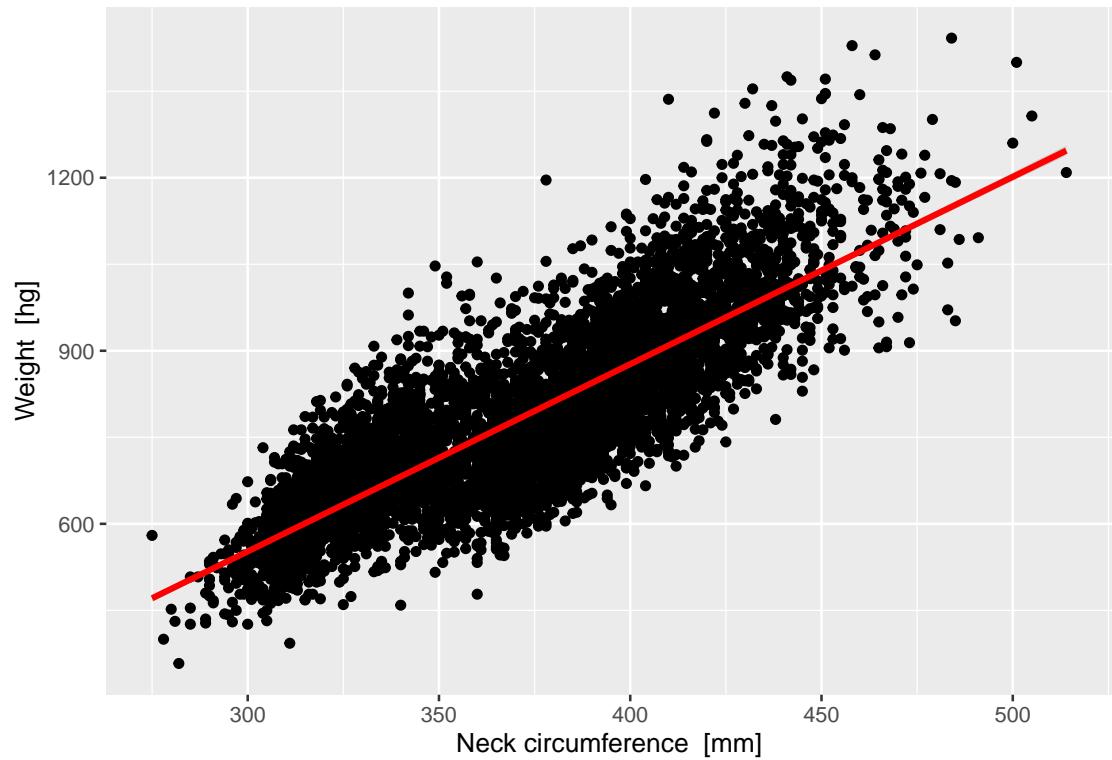


Figure 8: Linearni model kilaže i opsega vrata

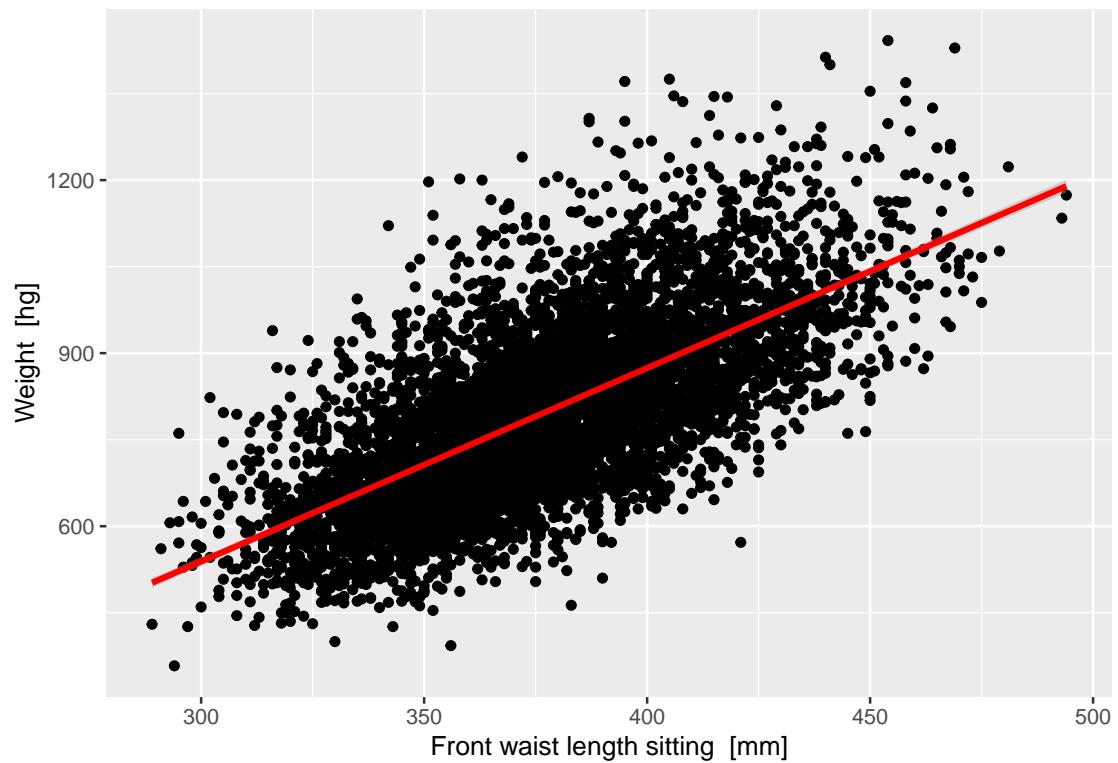


Figure 9: Linearni model kilaže i duljine prednjeg struka u sjedecem položaju

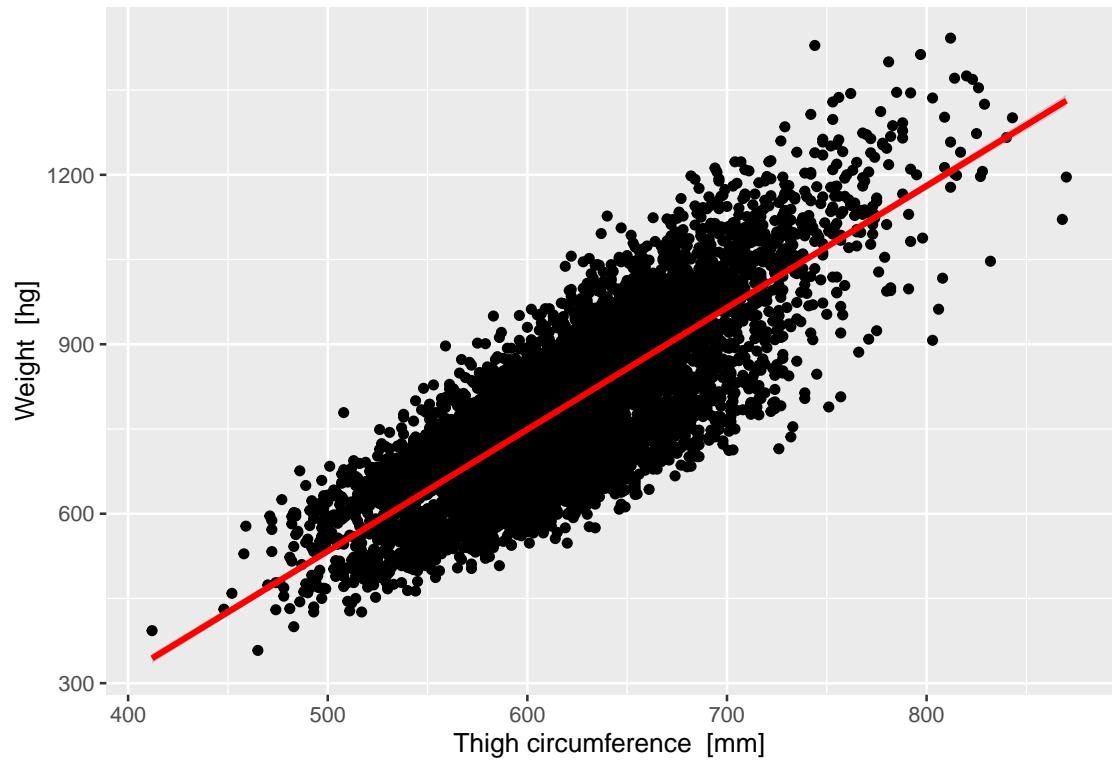


Figure 10: Linearni model kilaže i opsega bedra

Nagibi pravaca linearne regresije potvrđuju tvrdnje o izraženom utjecaju razmatranih varijabli na kilažu vojnika. Kako bi se dobiveni modeli analizirali i usporedili, potrebno je provjeriti da pretpostavke modela o regresorima i rezidualima nisu narušene. Mora vrijediti normalnost reziduala i homogenost varijance i regresori u višestrukoj regresiji ne smiju biti međusobno jako korelirani.

3.1.1 Normalnost reziduala i homogenost varijance

Normalnost reziduala provjerit ćemo grafički te statistički pomoću Lillieforsove inačice Kolmogorov-Smirnovljevog testa.

Visina:

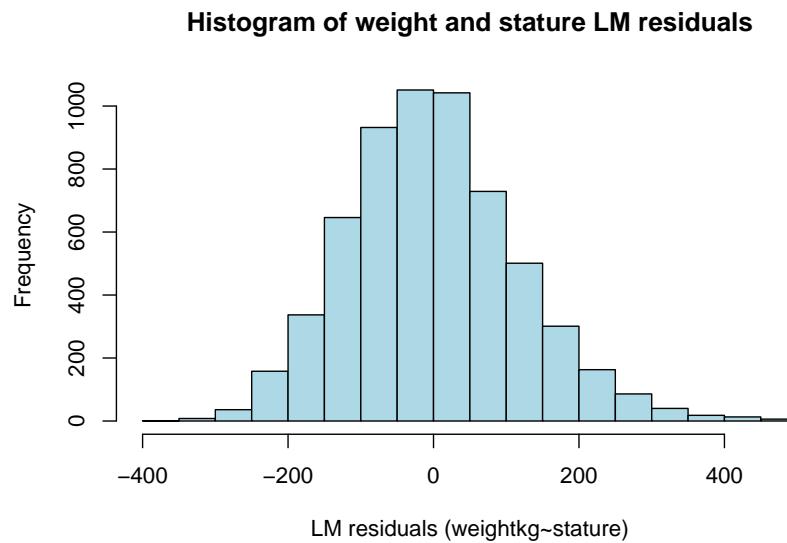


Figure 11: Histogram reziduala modela kilaže i visine

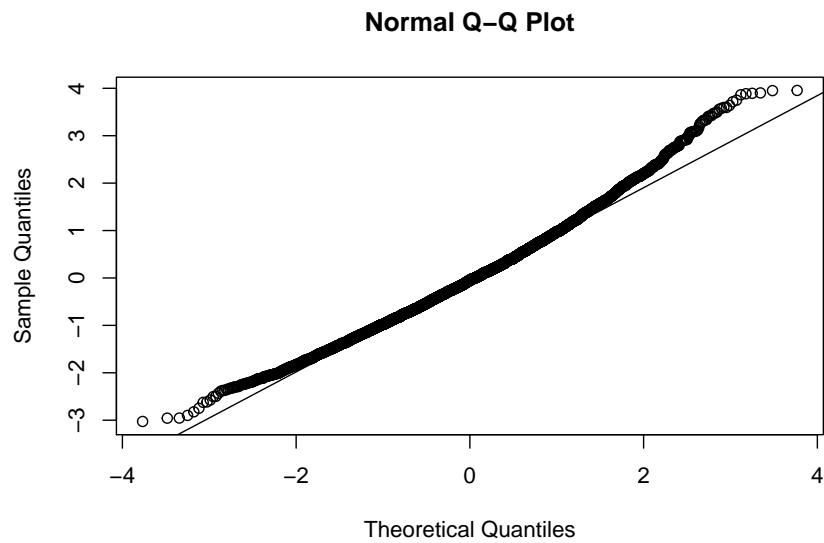


Figure 12: Prikaz g-g plota reziduala modela kilaže i visine s linijom normalne distribucije

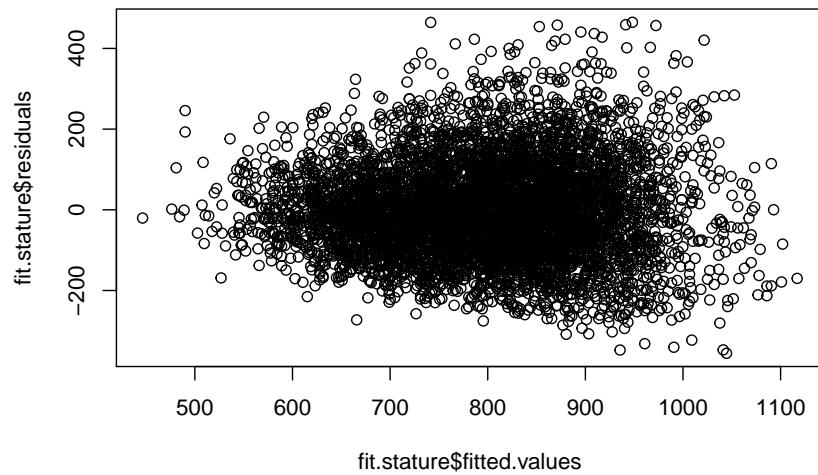


Figure 13: Reziduali modela kilaže i visine u ovisnosti o procjenama modela

```
##  
## Lilliefors (Kolmogorov-Smirnov) normality test  
##  
## data: rstandard(fit.stature)  
## D = 0.032889, p-value < 2.2e-16
```

Opseg struka:

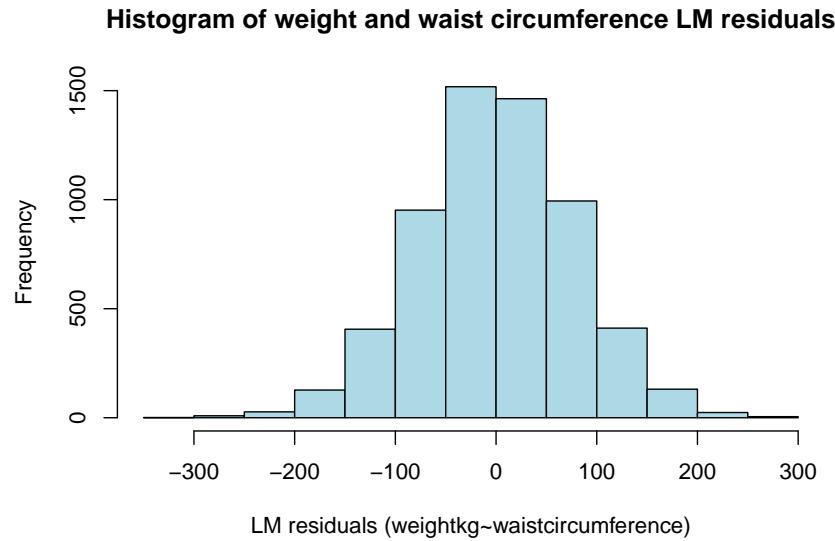


Figure 14: Histogram reziduala modela kilaže i opsega struka

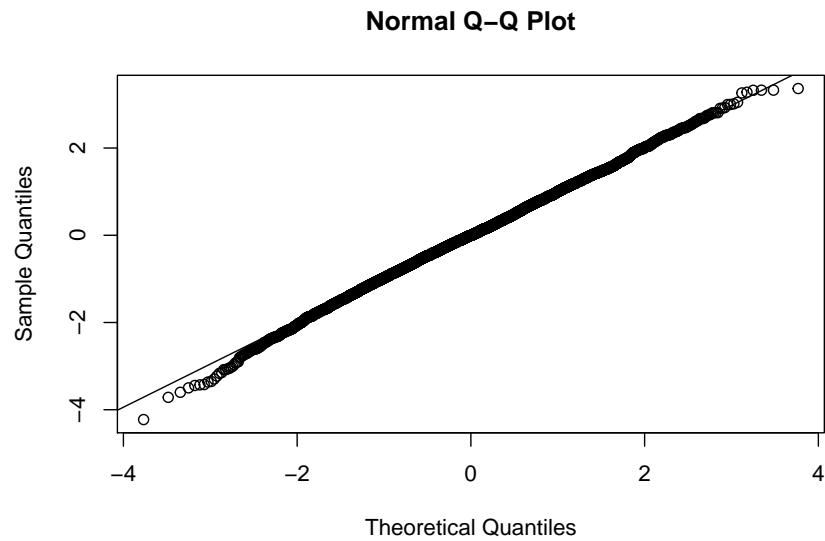


Figure 15: Prikaz g-g plota reziduala modela kilaže i opsega struka s linijom normalne distribucije

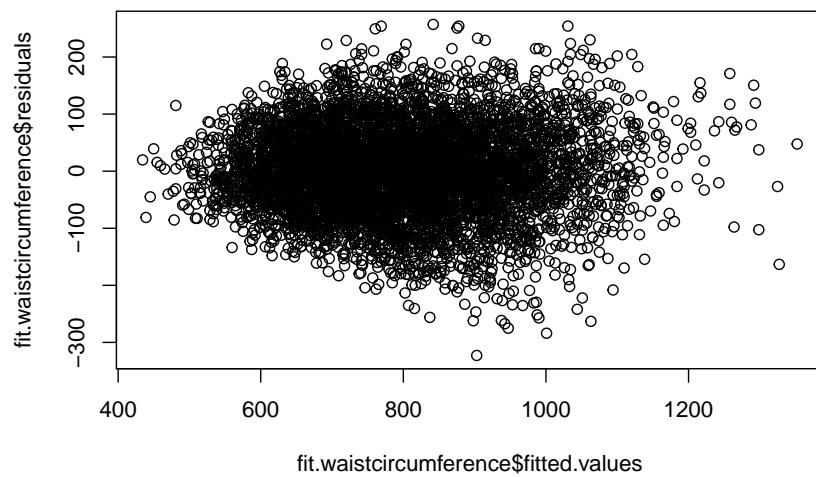


Figure 16: Reziduali modela kilaže i opsega struka u ovisnosti o procjenama modela

```
##  
## Lilliefors (Kolmogorov-Smirnov) normality test  
##  
## data: rstandard(fit.waistcircumference)  
## D = 0.011918, p-value = 0.0486
```

Opseg vrata:

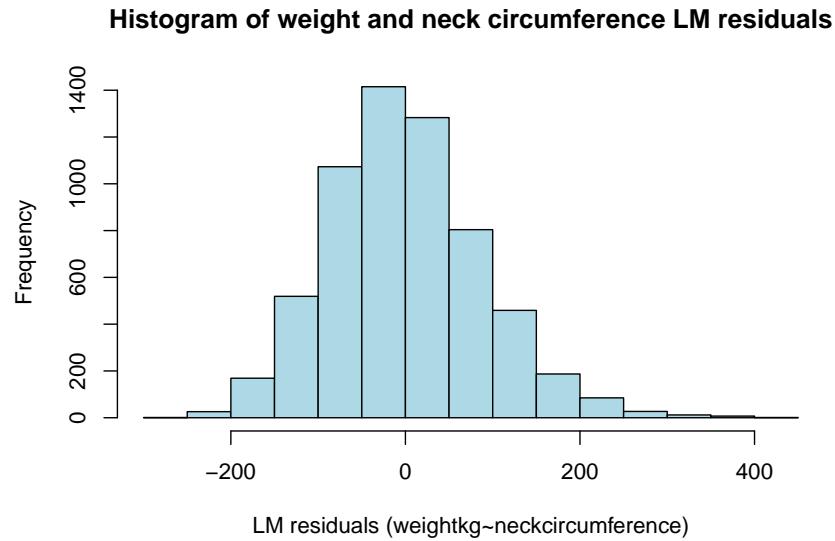


Figure 17: Histogram reziduala modela kilaže i opsega vrata

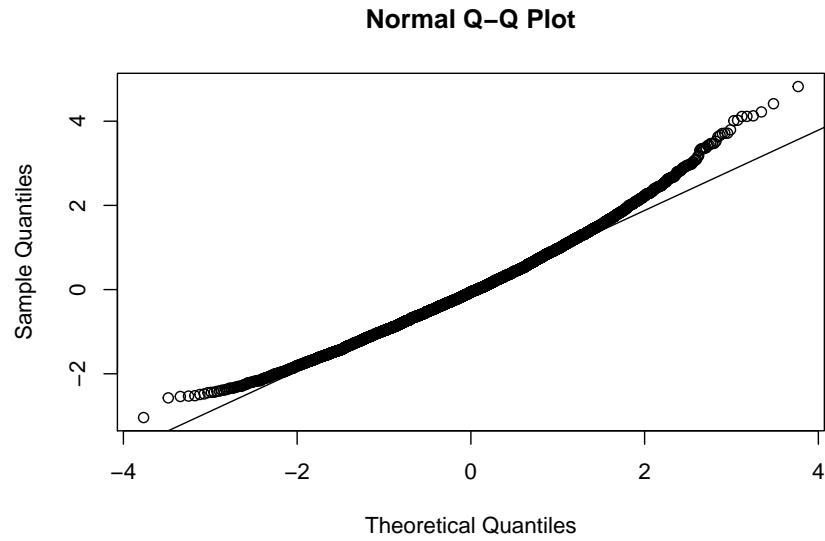


Figure 18: Prikaz g-g plota reziduala modela kilaže i opsega vrata s linijom normalne distribucije

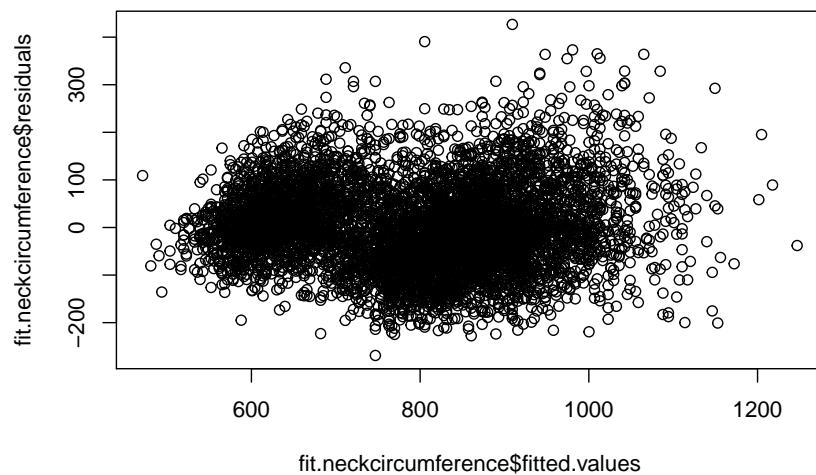


Figure 19: Reziduali modela kilaže i opsega vrata u ovisnosti o procjenama modela

```
##  
## Lilliefors (Kolmogorov-Smirnov) normality test  
##  
## data: rstandard(fit.neckcircumference)  
## D = 0.031255, p-value = 9.693e-15
```

Duljina prednjeg struka u sjedećem položaju:

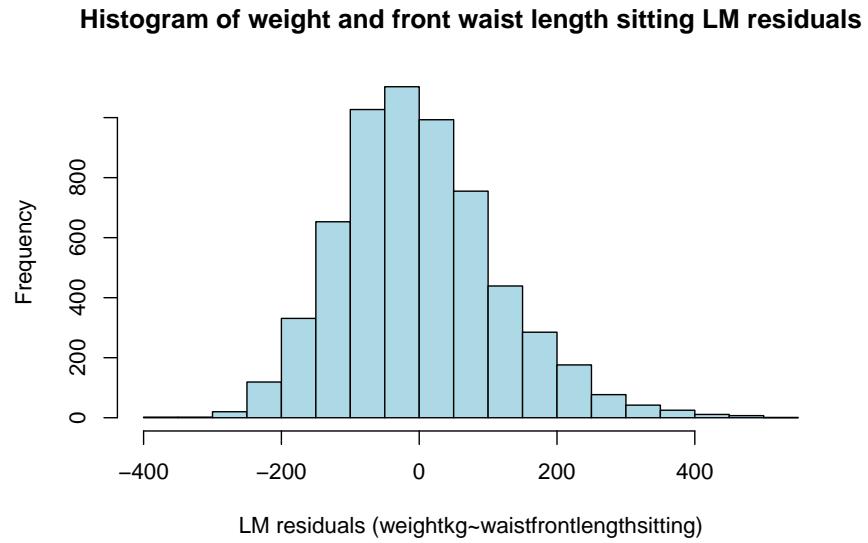


Figure 20: Histogram reziduala modela kilaže i duljine prednjeg struka u sjedecem položaju

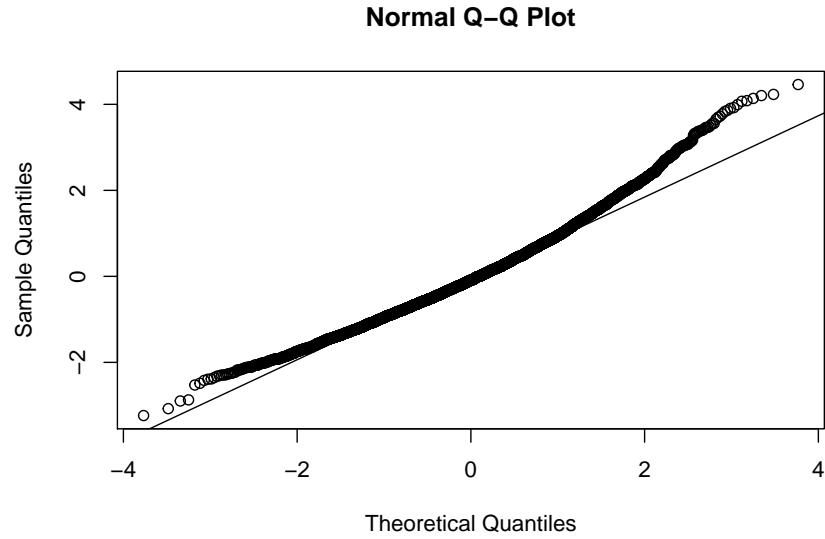


Figure 21: Prikaz g-g plota reziduala modela kilaže i duljine prednjeg struka u sjedecem položaju s linijom normalne distribucije

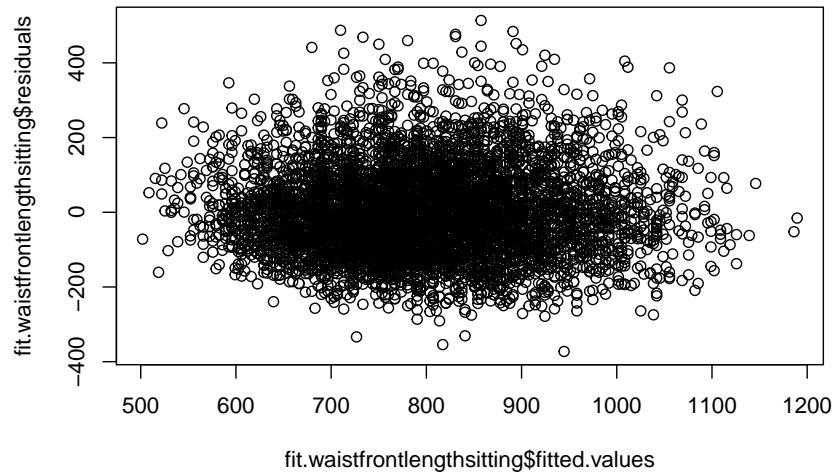


Figure 22: Reziduali modela kilaže i duljine prednjeg struka u sjedecem položaju u ovisnosti o procjenama modela

```
##  
## Lilliefors (Kolmogorov-Smirnov) normality test  
##  
## data: rstandard(fit.waistfrontlengthsitting)  
## D = 0.043958, p-value < 2.2e-16
```

Opseg bedra:

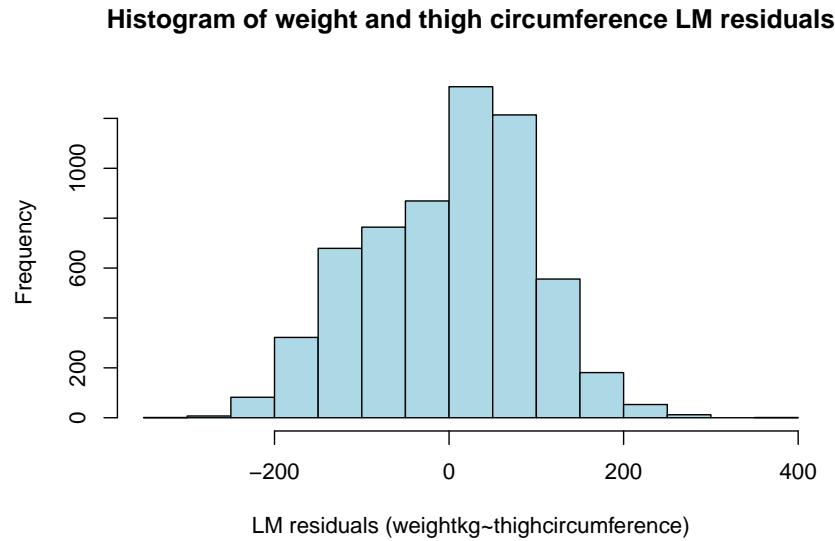


Figure 23: Histogram reziduala modela kilaže i opsega bedra

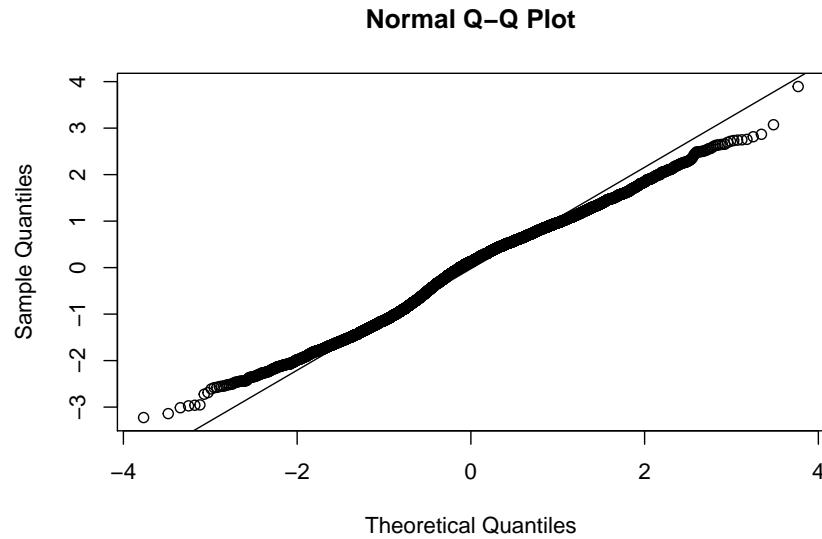


Figure 24: Prikaz g-g plota reziduala modela kilaže i opsega bedra s linijom normalne distribucije

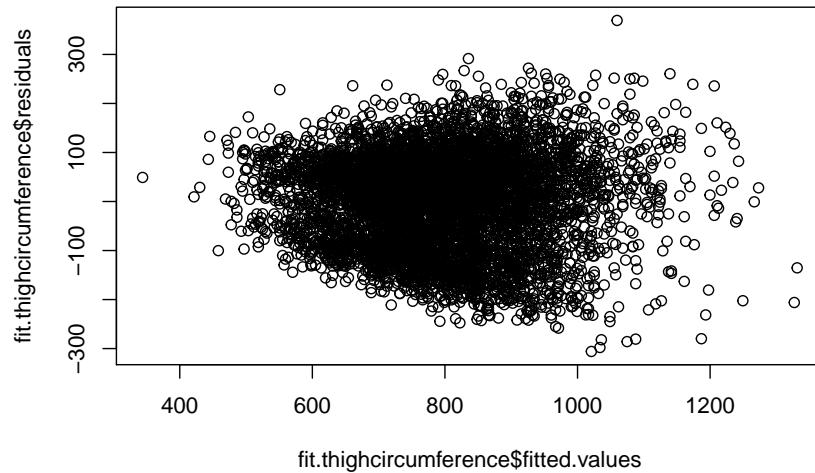


Figure 25: Reziduali modela kilaže i opsega bedra u ovisnosti o procjenama modela

```
##
## Lilliefors (Kolmogorov-Smirnov) normality test
##
## data: rstandard(fit.thighcircumference)
## D = 0.054098, p-value < 2.2e-16
```

Rezultati Kolmogorov-Smirnovljevog testa daju, osim u slučaju modela kilaže i opsega struka, male vrijednosti. No, budući da reziduali ne pokazuju preveliko odstupanje od normalnosti i da je t-test je robustan na normalnost, u analizi podataka se i dalje mogu donositi statistički zaključci iz regresijskih modela.

3.1.2 Ocjena kvalitete linearog modela i statističko zaključivanje o procijenjenom modelu

```
##
## Call:
## lm(formula = weightkg ~ stature, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -355.63  -81.63   -5.94   72.39  464.82
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.171e+03  2.879e+01 -40.69  <2e-16 ***
## stature      1.148e+00  1.677e-02   68.47  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 117.6 on 6066 degrees of freedom
## Multiple R-squared:  0.4359, Adjusted R-squared:  0.4359
## F-statistic: 4688 on 1 and 6066 DF,  p-value: < 2.2e-16
```

```

## 
## Call:
## lm(formula = weightkg ~ waistcircumference, data = data)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max
## -322.89  -50.10   -0.17   51.79  257.08
## 
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)            -2.961e+02  7.913e+00 -37.42   <2e-16 ***
## waistcircumference   1.195e+00  8.586e-03 139.22   <2e-16 ***
## ---                
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 76.43 on 6066 degrees of freedom
## Multiple R-squared:  0.7616, Adjusted R-squared:  0.7616 
## F-statistic: 1.938e+04 on 1 and 6066 DF,  p-value: < 2.2e-16

```

```

## 
## Call:
## lm(formula = weightkg ~ neckcircumference, data = data)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max
## -268.96  -59.48   -5.19   54.37  426.70
## 
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)           -421.89783  10.78352 -39.12   <2e-16 ***
## neckcircumference    3.24682   0.02856 113.67   <2e-16 ***
## ---                
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 88.48 on 6066 degrees of freedom
## Multiple R-squared:  0.6805, Adjusted R-squared:  0.6805 
## F-statistic: 1.292e+04 on 1 and 6066 DF,  p-value: < 2.2e-16

```

```

## 
## Call:
## lm(formula = weightkg ~ waistfrontlengthsitting, data = data)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max
## -372.65  -78.90  -10.37   67.95  513.55
## 
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)           -467.31440  17.68884 -26.42   <2e-16 ***
## waistfrontlengthsitting 3.35383   0.04676  71.73   <2e-16 ***
## ---                

```

```

## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 115.1 on 6066 degrees of freedom
## Multiple R-squared:  0.4589, Adjusted R-squared:  0.4589
## F-statistic:  5145 on 1 and 6066 DF,  p-value: < 2.2e-16

```

```

##
## Call:
## lm(formula = weightkg ~ thighcircumference, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -305.88  -72.16   12.08   67.47  369.32
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)           -544.28737  13.17970  -41.3  <2e-16 ***
## thighcircumference    2.15587   0.02109   102.2  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 94.88 on 6066 degrees of freedom
## Multiple R-squared:  0.6327, Adjusted R-squared:  0.6326
## F-statistic: 1.045e+04 on 1 and 6066 DF,  p-value: < 2.2e-16

```

Kao što je bilo vidljivo iz grafičkih prikaza, opseg struka ima izražen utjecaj na kilažu vojnika i objašnjava najveći postotak varijance (najveća vrijednost R^2). Također, u svim slučajevima koeficijenti uz nezavisnu varijablu su značajni, te F-testovi upućuju na to i da su svi modeli značajni.

3.1.3 Korelacijski koeficijenti u linearnom modelu

```

## [,1]
## data.stature          0.66
## data.waistcircumference 0.87
## data.neckcircumference 0.82
## data.waistfrontlengthsitting 0.68
## data.thighcircumference 0.80

```

Korelacijski koeficijenti potvrđuju da je utjecaj opsega struka na kilažu najizraženiji, a jak utjecaj imaju i opseg vrata kao i opseg bedra. Visina ima nešto manji, ali ipak značajan utjecaj.

3.2 Višestruka regresija

Najprije je potrebno odabratи podskup varijabli koje nisu previše međusobno korelirane.

	stature	waist circumference	neck circumference	front waist length sitting	thigh circumference
stature	1				
waist circumference	0.37	1			
neck circumference	0.63	0.68	1		
front waist length sitting	0.61	0.62	0.6	1	
thigh circumference	0.25	0.78	0.45	0.39	1

```
##
## Call:
## lm(formula = weightkg ~ waistcircumference + stature + waistfrontlengthsitting +
##      neckcircumference + thighcircumference, data = data)
##
## Residuals:
##       Min     1Q   Median     3Q    Max 
## -131.950 -19.366 -1.003  18.176 189.520 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -1.450e+03 7.883e+00 -183.878 < 2e-16 ***
## waistcircumference 3.689e-01 7.111e-03  51.871 < 2e-16 ***
## stature        4.697e-01 6.000e-03   78.285 < 2e-16 ***
## waistfrontlengthsitting 1.194e-01 1.789e-02   6.672 2.75e-11 ***
## neckcircumference 1.129e+00 1.546e-02   73.013 < 2e-16 ***
## thighcircumference 1.021e+00 1.066e-02   95.753 < 2e-16 ***
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 28.93 on 6062 degrees of freedom
## Multiple R-squared:  0.9659, Adjusted R-squared:  0.9659 
## F-statistic: 3.432e+04 on 5 and 6062 DF, p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = weightkg ~ waistcircumference + stature + waistfrontlengthsitting +
##      neckcircumference, data = data)
##
## Residuals:
##       Min     1Q   Median     3Q    Max 
## -140.211 -31.036 -1.967  29.314 189.240 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -1.173e+03 1.163e+01 -100.91 < 2e-16 ***
## waistcircumference 8.602e-01 7.803e-03  110.25 < 2e-16 ***
## stature        5.343e-01 9.450e-03   56.54 < 2e-16 ***
## waistfrontlengthsitting -1.877e-01 2.790e-02   -6.73 1.86e-11 ***
## neckcircumference 9.019e-01 2.422e-02   37.24 < 2e-16 *** 
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 45.84 on 6063 degrees of freedom
## Multiple R-squared:  0.9143, Adjusted R-squared:  0.9142
## F-statistic: 1.617e+04 on 4 and 6063 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = weightkg ~ stature + waistfrontlengthsitting + neckcircumference +
##      thighcircumference, data = data)
##
## Residuals:
##    Min      1Q  Median      3Q     Max 
## -156.979 -22.889 -1.664  21.030 227.856 
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -1.510e+03  9.368e+00 -161.16   <2e-16 ***
## stature       3.858e-01  6.942e-03   55.57   <2e-16 ***
## waistfrontlengthsitting 5.139e-01  1.946e-02   26.41   <2e-16 ***
## neckcircumference  1.514e+00  1.630e-02   92.87   <2e-16 ***
## thighcircumference 1.420e+00  8.868e-03  160.12   <2e-16 ***
## ---      
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 34.75 on 6063 degrees of freedom
## Multiple R-squared:  0.9507, Adjusted R-squared:  0.9507
## F-statistic: 2.925e+04 on 4 and 6063 DF,  p-value: < 2.2e-16
```

Model višestruke regresije koji uključuje sve varijable objašnjava cca. 97% varijance u podatcima.

3.2.1 Provjera normalnosti reziduala

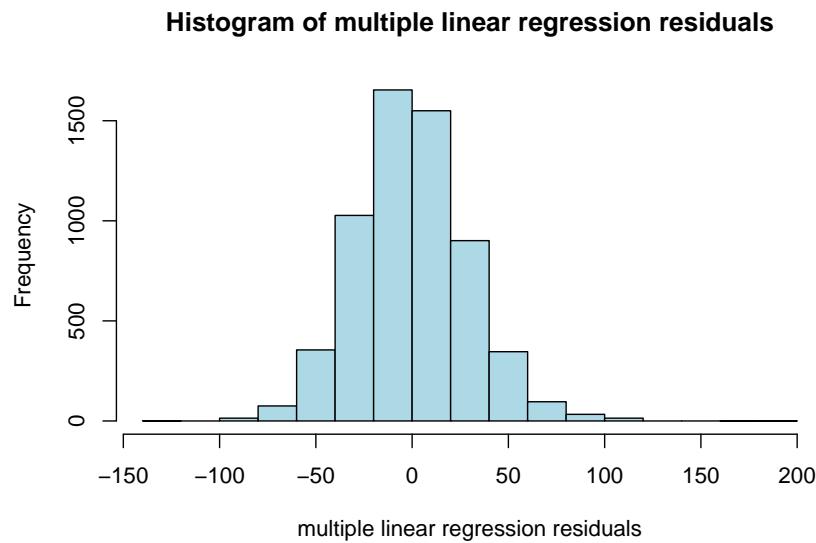


Figure 26: Histogram reziduala višestruke regresije

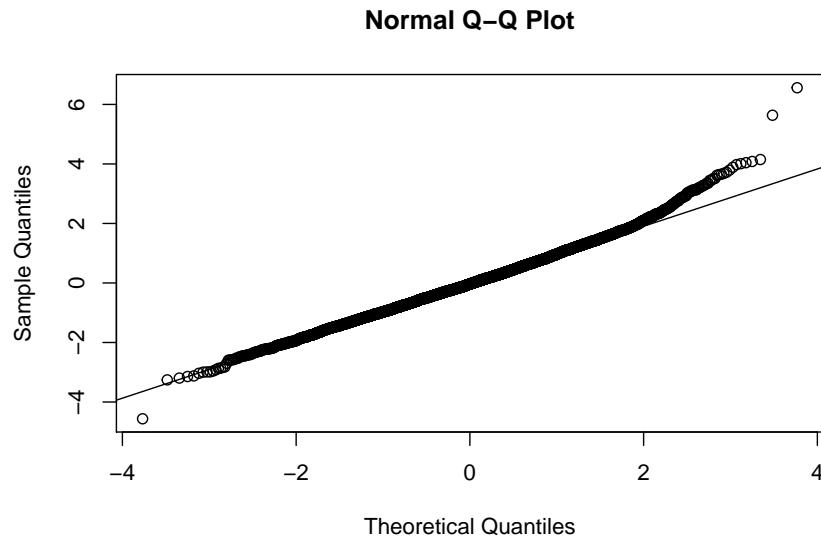


Figure 27: Prikaz g-g plota reziduala višestruke regresije

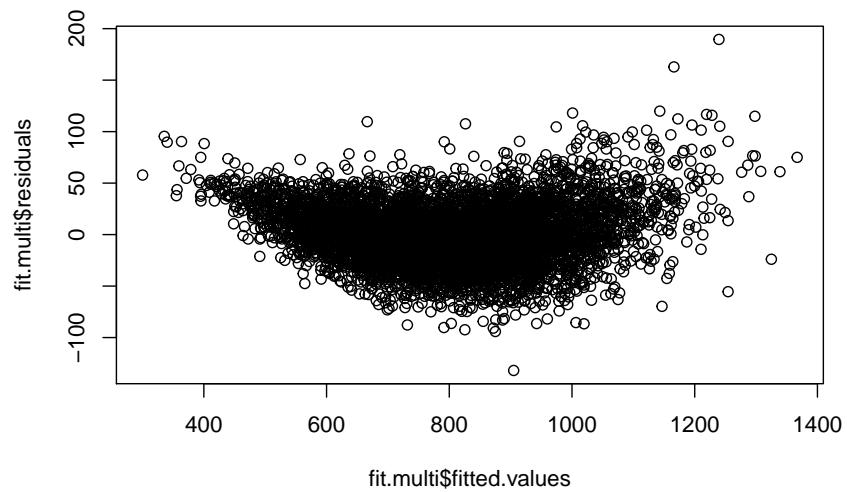


Figure 28: Reziduali višestruke regresije

```
##  
## Lilliefors (Kolmogorov-Smirnov) normality test  
##  
## data: rstandard(fit.multi)  
## D = 0.020985, p-value = 2.298e-06
```

Iz grafičkih prikaza vidi se da reziduali ne pokazuju preveliko odstupanje od normalnosti i u analizi podataka se i dalje mogu donositi statistički zaključci iz regresijskog modela.

3.2.2 Uključivanje kategorijskih varijabli

Kao regresor možemo uključiti i kategorijске varijable poput spola vojnika.

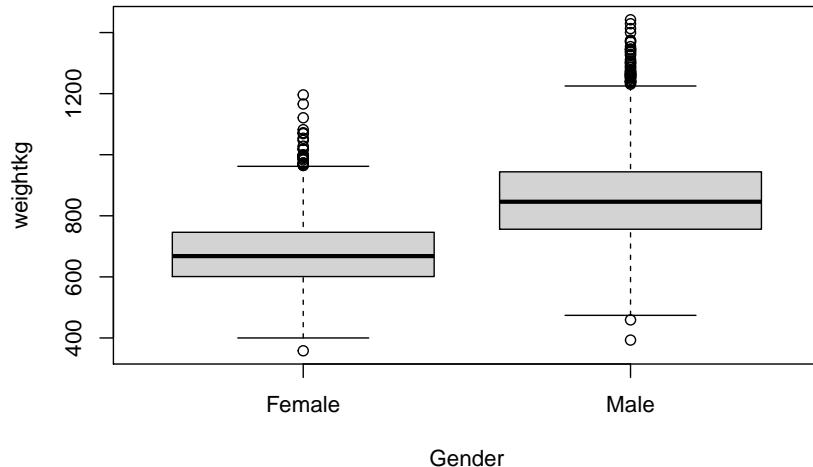


Figure 29: Kvadratni dijagram ovisnosti kilaže i spola

```
## Loading required package: fastDummies

## 
## Call:
## lm(formula = weightkg ~ waistcircumference + stature + waistfrontlengthsitting +
##     neckcircumference + thighcircumference + Gender_Female, data = data.d)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -131.482  -19.166   -0.945   18.264  188.000 
## 
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)             -1.478e+03  1.090e+01 < 2e-16 ***  
## waistcircumference      3.654e-01  7.161e-03  51.030 < 2e-16 ***  
## stature                  4.777e-01  6.348e-03  75.258 < 2e-16 ***  
## waistfrontlengthsitting 1.206e-01  1.787e-02   6.745 1.67e-11 ***  
## neckcircumference       1.195e+00  2.327e-02   51.372 < 2e-16 ***  
## thighcircumference      1.006e+00  1.134e-02   88.666 < 2e-16 ***  
## Gender_Female            6.551e+00  1.715e+00    3.820 0.000135 *** 
## ---                        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 28.89 on 6061 degrees of freedom
## Multiple R-squared:  0.966,  Adjusted R-squared:  0.9659 
## F-statistic: 2.867e+04 on 6 and 6061 DF,  p-value: < 2.2e-16
```

```

## 
## Call:
## lm(formula = weightkg ~ waistcircumference + stature + waistfrontlengthsitting +
##      neckcircumference + thighcircumference, data = data)
## 
## Residuals:
##       Min     1Q   Median     3Q    Max 
## -131.950 -19.366 -1.003  18.176 189.520 
## 
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)    
## (Intercept)             -1.450e+03 7.883e+00 -183.878 < 2e-16 ***
## waistcircumference      3.689e-01 7.111e-03   51.871 < 2e-16 ***
## stature                  4.697e-01 6.000e-03   78.285 < 2e-16 ***
## waistfrontlengthsitting 1.194e-01 1.789e-02    6.672 2.75e-11 ***
## neckcircumference        1.129e+00 1.546e-02   73.013 < 2e-16 ***
## thighcircumference       1.021e+00 1.066e-02   95.753 < 2e-16 ***  
## ---                        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
## 
## Residual standard error: 28.93 on 6062 degrees of freedom
## Multiple R-squared:  0.9659, Adjusted R-squared:  0.9659 
## F-statistic: 3.432e+04 on 5 and 6062 DF,  p-value: < 2.2e-16

```

3.2.3 Zaključak

Model koji uz svih pet navednih numeričkih varijabli uključuje i spol vojnika, daje nešto veći R^2 . Međutim, ako pogledamo prilagođeni koeficijent determinacije R_{adj}^2 koji penalizira dodatne parametre u modelu, vidimo da su isti za oba modela, što znači da varijabla spola nije toliko korisna i radi jednostavnosti ju možemo izbaciti.

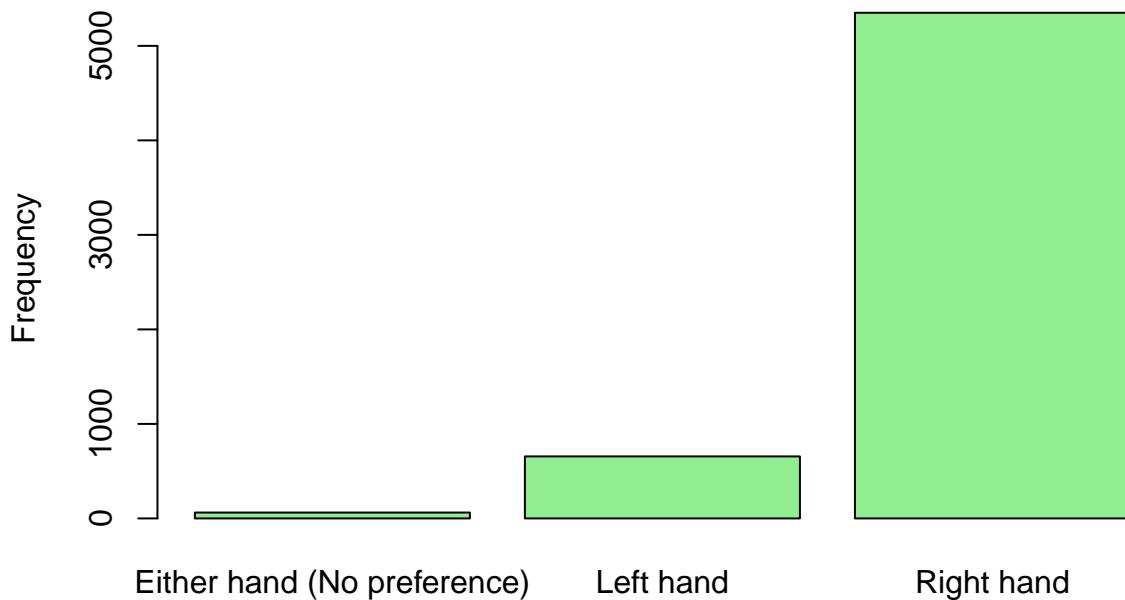
4 Opseg bicepsa

4.1 Prikaz raspodjele vojnika ovisno o preferenciji ruke i spolu

```
##  bicepscircumferenceflexed WritingPreference
## 1          315      Right hand
## 2          272      Right hand
## 3          300      Right hand
## 4          364      Right hand
## 5          320      Right hand
## 6          342      Left hand

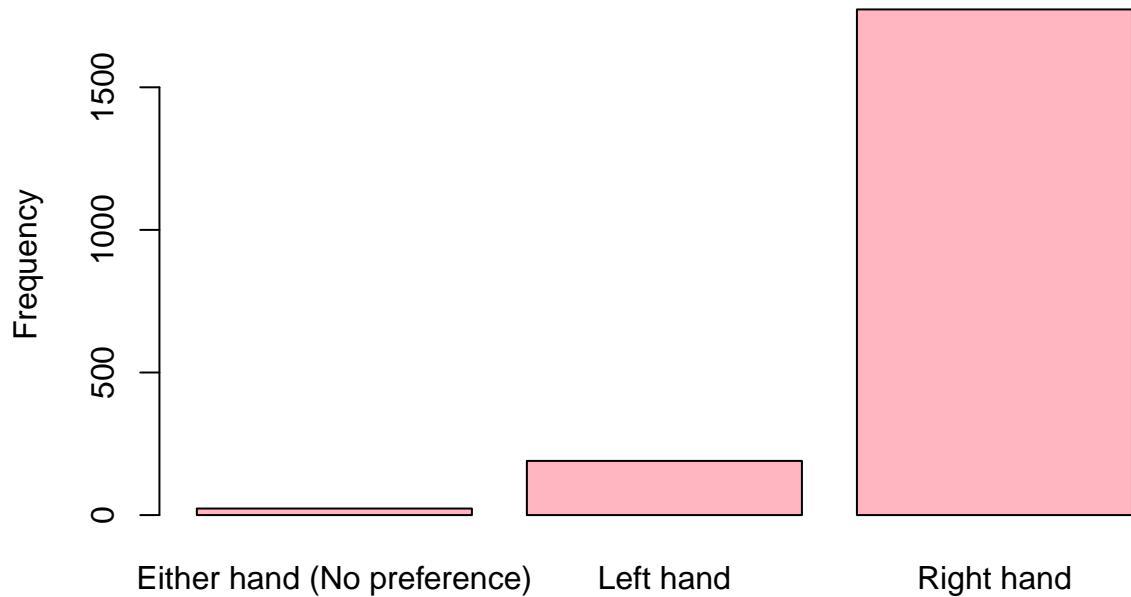
##
## Either hand (No preference)           Left hand
##                               62             656
## Right hand                         5350
```

Barplot of Soldiers Writing Preference Distribution



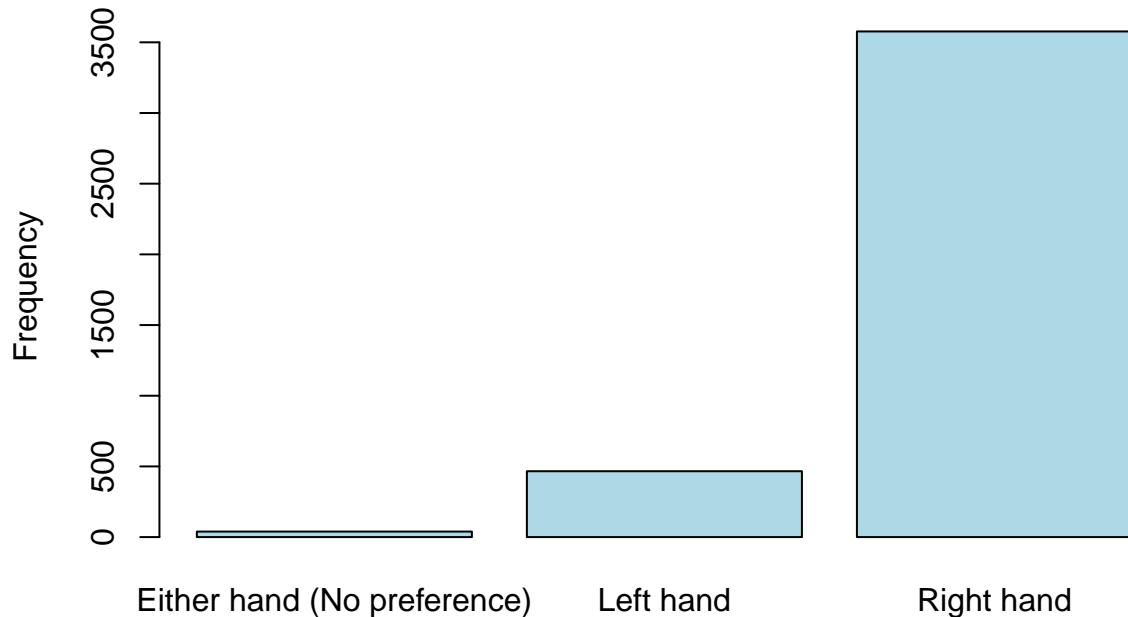
```
##
## Either hand (No preference)           Left hand
##                               23            190
## Right hand                         1773
```

Barplot of Female Soldiers Writing Preference Distribution



```
##  
## Either hand (No preference)          Left hand  
##                               39           466  
## Right hand                         3577
```

Barplot of Male Soldiers Writing Preference Distribution



4.2 Test o dvije proporcije: dva uzorka

4.2.1 Udjeli ljevaka ovisno o spolu

Želimo testirati hipotezu je li udio ljevakinja u odnosu na vojnikinje jednak udjelu ljevaka u odnosu na vojнике.

```
##  
## 2-sample test for equality of proportions with continuity correction  
##  
## data: c(190, 466) out of c(1986, 3682)  
## X-squared = 11.73, df = 1, p-value = 0.0003075  
## alternative hypothesis: less  
## 95 percent confidence interval:  
## -1.00000000 -0.01639446  
## sample estimates:  
##      prop 1      prop 2  
## 0.09566969 0.12656165
```

Zbog jako male p-vrijednosti od 0.0003075 možemo odbaciti H_0 hipotezu o jednakosti udjela u korist H_1 , odnosno možemo reći da je udio ljevakinja u ženskoj populaciji manji nego udio ljevaka u muškoj.

U 1940-ima znanstvenici su primjetili tu zanimljivu pojavu, a ona glasi da su žene rjeđe ljevakinje od muškaraca. Nisu sve studije zabilježile ovakve rezultate, ali daljnijim istraživanjem (od 2008. pa nadalje) ispostavilo se stvarno da je udio ljevakinja u ženskoj populaciji manji od udjela ljevaka u muškoj. Razlozi zašto je to tako još moraju biti istraženi, ali svakako je zanimljivo da se i u ovom datasetu vidi ta pojava.

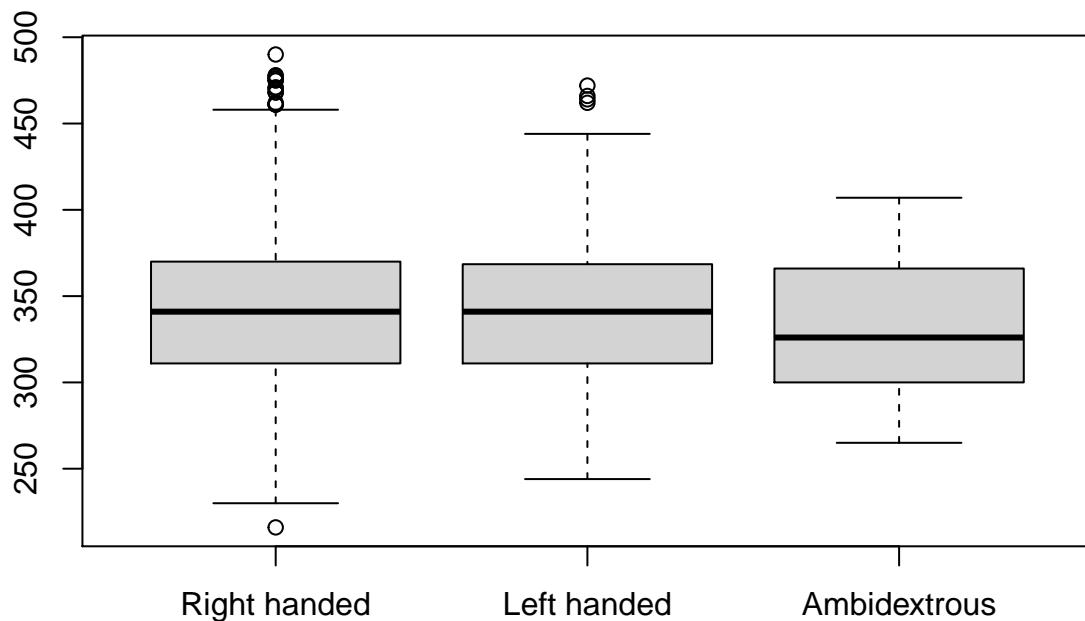
4.3 Opseg bicepsa ovisno o preferenciji ruke

```
##      Min. 1st Qu. Median     Mean 3rd Qu.    Max.
##    216.0   311.0  341.0   340.9  370.0   490.0
```

Prikaz boxplota i prosječnih opsega desnog bicepsa za svaku promatrano skupinu(dešnjaka, ljevaka i ambidekstrenih vojnika).

```
## Prosjecni opseg desnog bicepsa dešnjaka 341.0978
## Prosjecna opseg desnog bicepsa ljevaka 340.4162
## Prosjecna opseg desnog bicepsa ambidekstrenih vojnika 332.3065
```

Boxplot of Writing preference soldiers biceps circumferences



Želimo provjeriti imaju li vojnici dešnjaci veći opseg desnog bicepsa od vojnika ljevaka. Isto tako kasnije ćemo provjeriti imaju li vojnici dešnjaci veći opseg od ambidekstrenih vojnika.

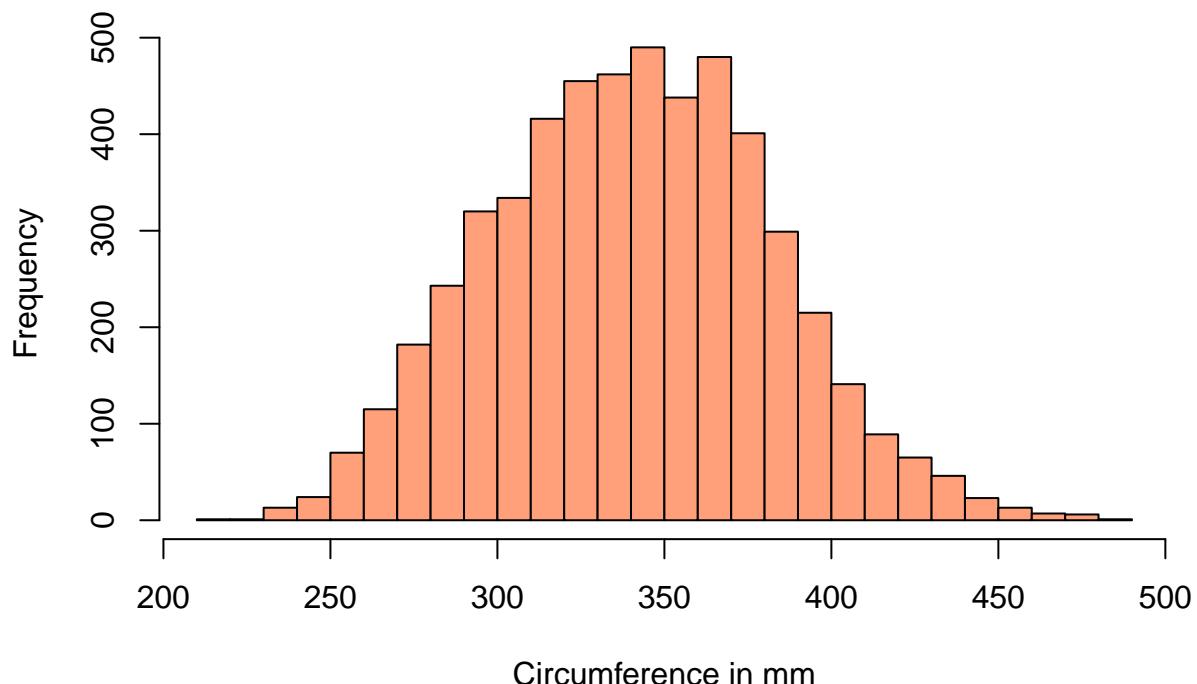
Koristit ćemo neupareni t-test. Kako bi mogli provesti test, moramo najprije provjeriti prepostavke normalnosti i nezavisnosti uzorka.

S obzirom kako je odabrana cijela populacija jedne baze iz koje promatramo ljevake i dešnjake i jer baza broji puno više od 30 članova smatramo kako su promatrani uzorci nezavisni. Iz istih razloga su nezavisni za dešnjake i ambidekstrene vojнике.

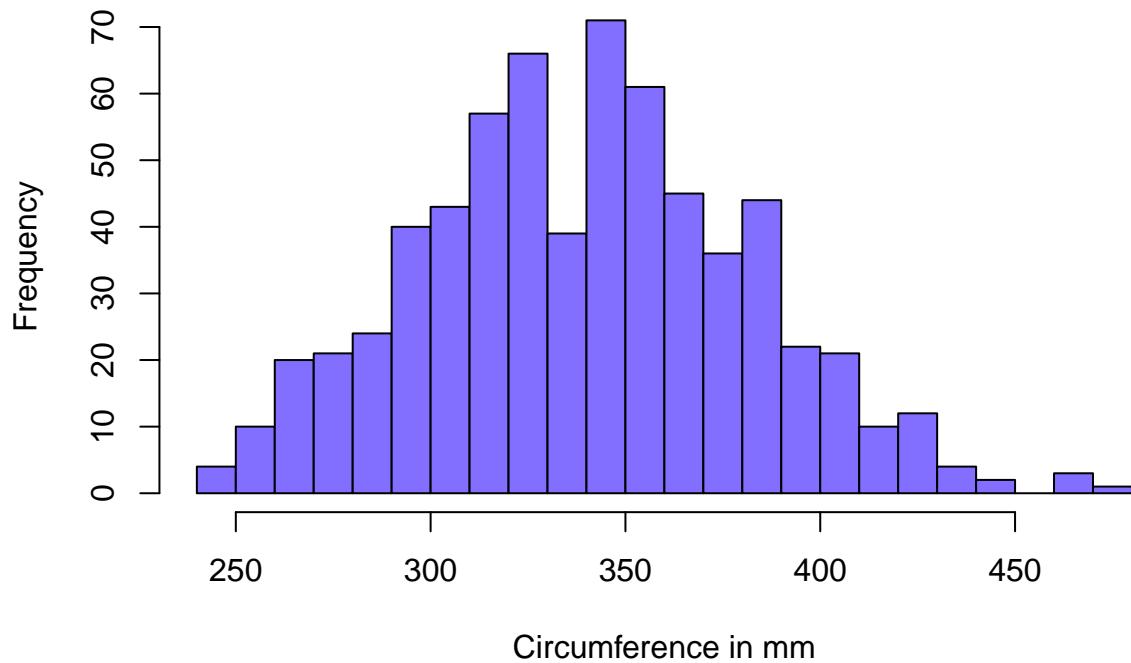
4.3.1 Opseg desnog bicepsa dešnjaka i ljevaka

4.3.1.1 Provjera normalnosti podataka Sljedeći korak je provjeriti normalnost podataka koju najčešće provjeravamo: histogramom, qq-plotom te KS-testom (kojim provjeravamo pripadnost podataka distribuciji).

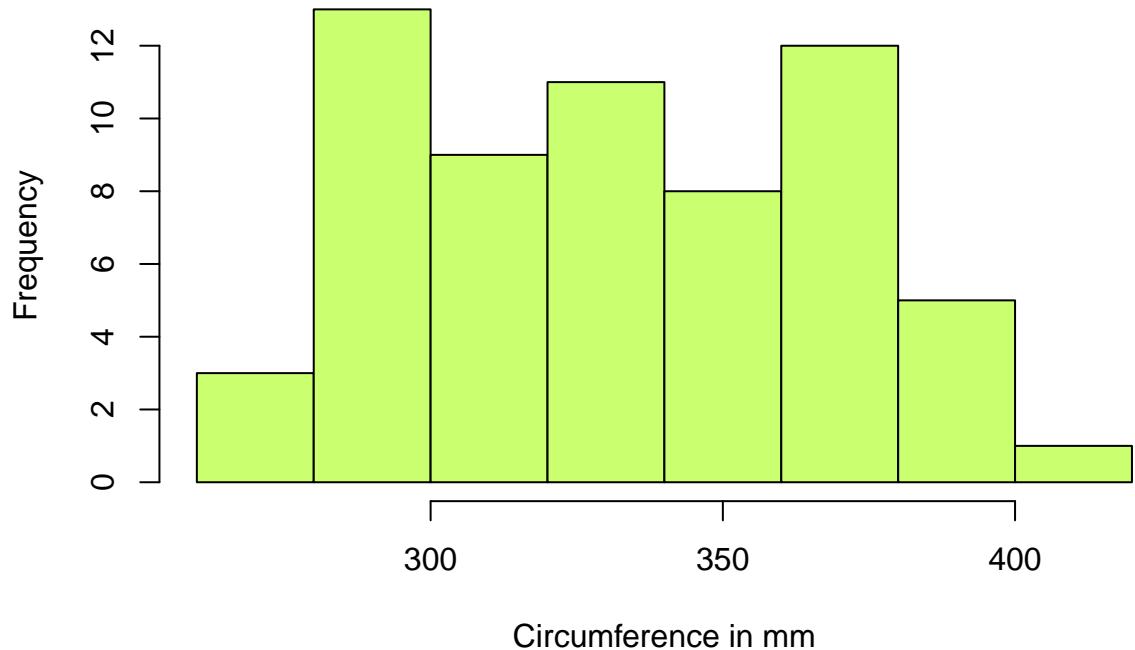
Histogram of biceps circumference of right handed soldiers



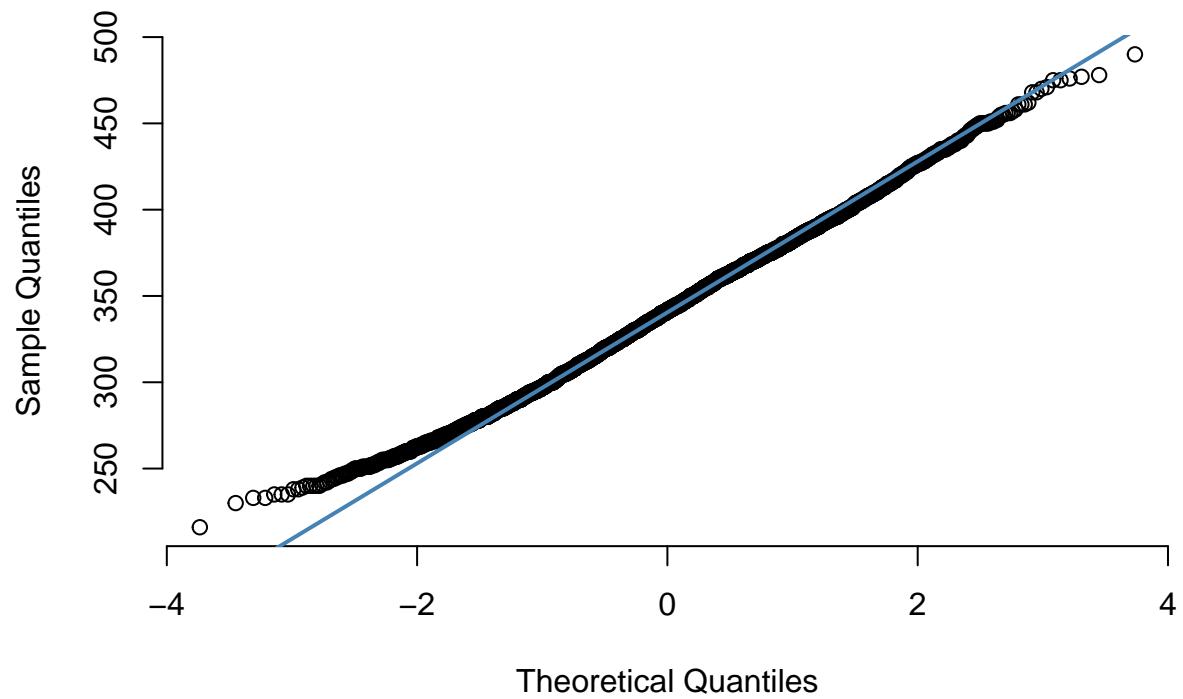
Histogram of biceps circumference of left handed soldiers



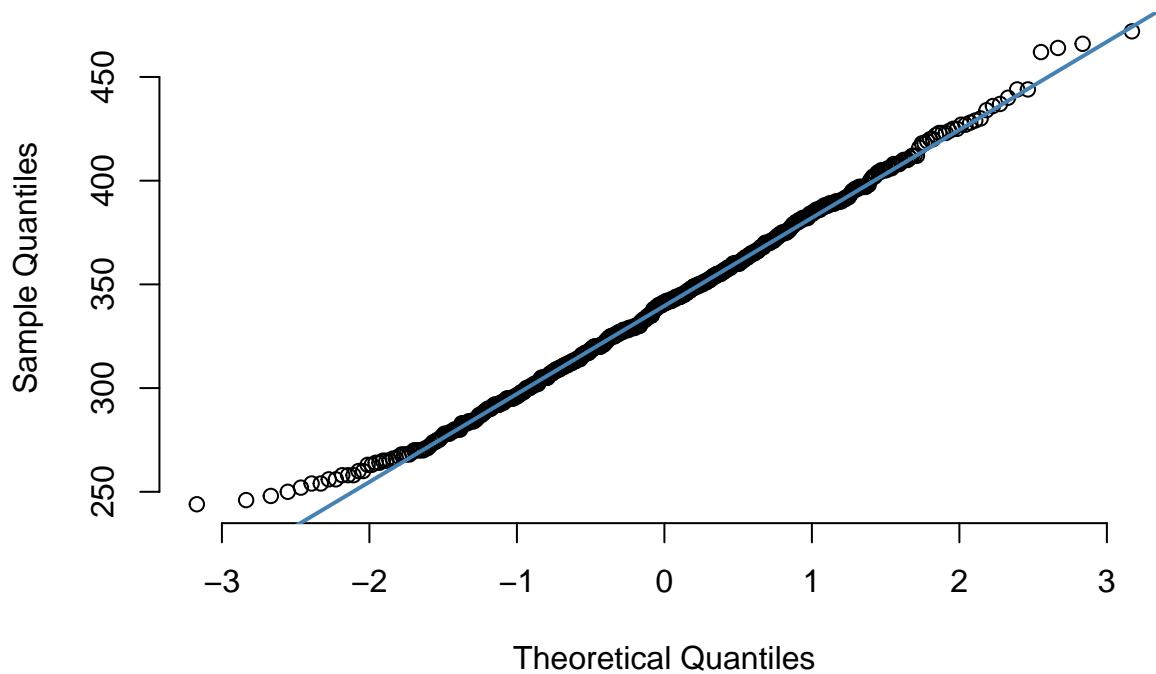
Histogram of biceps circumference of ambidextrous soldiers



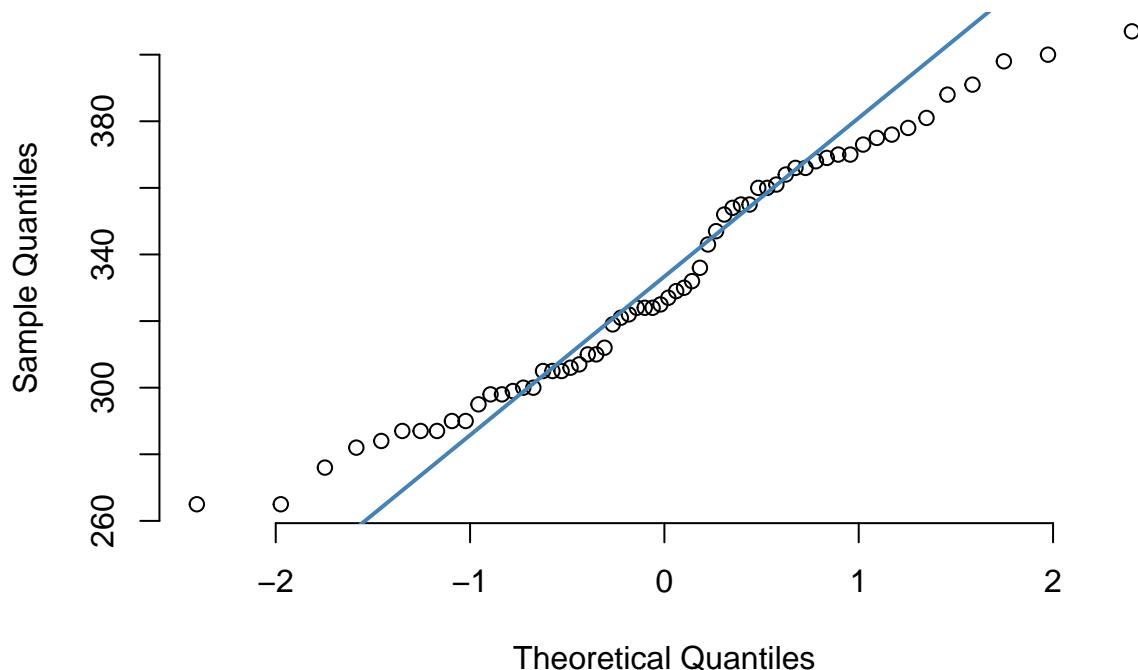
Right handed soldiers



Left handed soldiers



Ambidextrous soldiers



```
##  
## Lilliefors (Kolmogorov-Smirnov) normality test  
##  
## data: scale(right$bicepscircumferenceflexed)  
## D = 0.020079, p-value = 4.198e-05  
##  
## Lilliefors (Kolmogorov-Smirnov) normality test  
##  
## data: scale(left$bicepscircumferenceflexed, )  
## D = 0.032375, p-value = 0.09935  
##  
## Lilliefors (Kolmogorov-Smirnov) normality test  
##  
## data: scale(ambi$bicepscircumferenceflexed)  
## D = 0.097485, p-value = 0.1517
```

Histogrami upućuju na normalnost podataka kao i qq-plotovi.

Pod uvjetom da podatci zadovoljavaju sve pretpostavke možemo nastaviti sa t-testom kako bi ispitali imaju li dešnjaci veći opseg desnog bicepsa od ljevaka.

Kakve su varijance danih uzoraka?

```
## [1] 1722.682  
## [1] 1764.384
```

Jesu li varijance značajno različite? To ćemo ispitati testom o jednakosti varijanci.

4.3.1.2 Test o jednakosti varijanci Ako imamo dva nezavisna slučajna uzorka koji dolaze iz normalnih distribucija , tada slučajna varijabla ima Fisherovu distribuciju.

```
## 
## F test to compare two variances
##
## data: right$bicepscircumferenceflexed and left$bicepscircumferenceflexed
## F = 0.97636, num df = 5349, denom df = 655, p-value = 0.6708
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.868182 1.092309
## sample estimates:
## ratio of variances
## 0.9763647
```

p-vrijednost od 0.6708 nam govori da ne možemo odbaciti hipotezu H_0 da su varijance naša dva uzorka jednake.

4.3.1.3 t-test uz pretpostavku jednakosti varijanci H_0 prosječne vrijednosti opsega desnog bicepsa vojnika dešnjaka i vojnika ljevaka su jednake

H_1 prosječne vrijednosti opsega desnog bicepsa dešnjaka vojnika i ljevaka vojnika su različite (tj. dešnjaci vojnici imaju u prosjeku veći desni biceps od vojnika ljevaka).

```
## 
## Two Sample t-test
##
## data: right$bicepscircumferenceflexed and left$bicepscircumferenceflexed
## t = 0.39645, df = 6004, p-value = 0.3459
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## -2.146753      Inf
## sample estimates:
## mean of x mean of y
## 341.0978 340.4162
```

Zbog p-vrijednost 0.3459 ne možemo odbaciti H_0 hipotezu tj. ne možemo zaključiti da postoji značajna razlika između prosječnih vrijednosti opsega bicepsa vojnika dešnjaka i vojnika ljevaka.

4.3.2 Opseg desnog bicepsa dešnjaka i ambidekstrenih vojnika

Provjeravamo na isti način kao i u prethodnom primjeru jer su sve pretpostavke zadovoljene (normalnost i nezavisnost).

```
## [1] 1722.682
## [1] 1371.56
```

4.3.2.1 Test o jednakosti varijanci Testom o jednakosti varijanci ispitujemo jesu li varijance značajno različite.

```
## 
## F test to compare two variances
##
## data: right$bicepscircumferenceflexed and ambi$bicepscircumferenceflexed
## F = 1.256, num df = 5349, denom df = 61, p-value = 0.2517
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
```

```

##  0.8490384 1.7435240
## sample estimates:
## ratio of variances
##              1.256002

```

p-vrijednost od 0.2517 nam govori da nećemo odbaciti hipotezu H_0 da su varijance naša dva uzorka jednaka.

4.3.2.2 t-test uz pretpostavku jednakosti varijanci H_0 prosječne vrijednosti opsega desnog bicepsa vojnika dešnjaka i ambidekstrenih vojnika su jednake

H_1 prosječne vrijednosti opsega desnog bicepsa dešnjaka vojnika i ambidekstrenih vojnika su različite (tj. dešnjaci vojnici imaju u prosjeku veći desni biceps od ambidekstrenih vojnika).

```

##
## Two Sample t-test
##
## data: right$bicepscircumferenceflexed and ambi$bicepscircumferenceflexed
## t = 1.6601, df = 5410, p-value = 0.04847
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.07944869      Inf
## sample estimates:
## mean of x mean of y
## 341.0978 332.3065

```

Zbog jako male p-vrijednosti 0.04847 možemo odbaciti H_0 hipotezu o jednakosti prosječnih vrijednosti u korist H_1 , odnosno možemo reći da vojnici dešnjaci u prosjeku imaju značajno veći desni biceps od ambidekstrenih vojnika.

5 Mjere kaciga

```
##   headbreadth headcircumference headlength
## 1          141             548        191
## 2          138             535        180
## 3          146             588        207
## 4          153             593        206
## 5          140             522        181
## 6          146             570        194
```

5.1 Klase podataka

```
## [1] "integer"
## [1] "integer"
## [1] "integer"
```

5.2 Summary

```
##   headbreadth   headcircumference   headlength
## Min.    :131.0   Min.    :500       Min.    :168.0
## 1st Qu.:148.0   1st Qu.:557       1st Qu.:191.0
## Median  :152.0   Median  :570       Median  :197.0
## Mean    :152.2   Mean    :570       Mean    :196.3
## 3rd Qu.:156.0   3rd Qu.:582       3rd Qu.:202.0
## Max.    :180.0   Max.    :635       Max.    :225.0
```

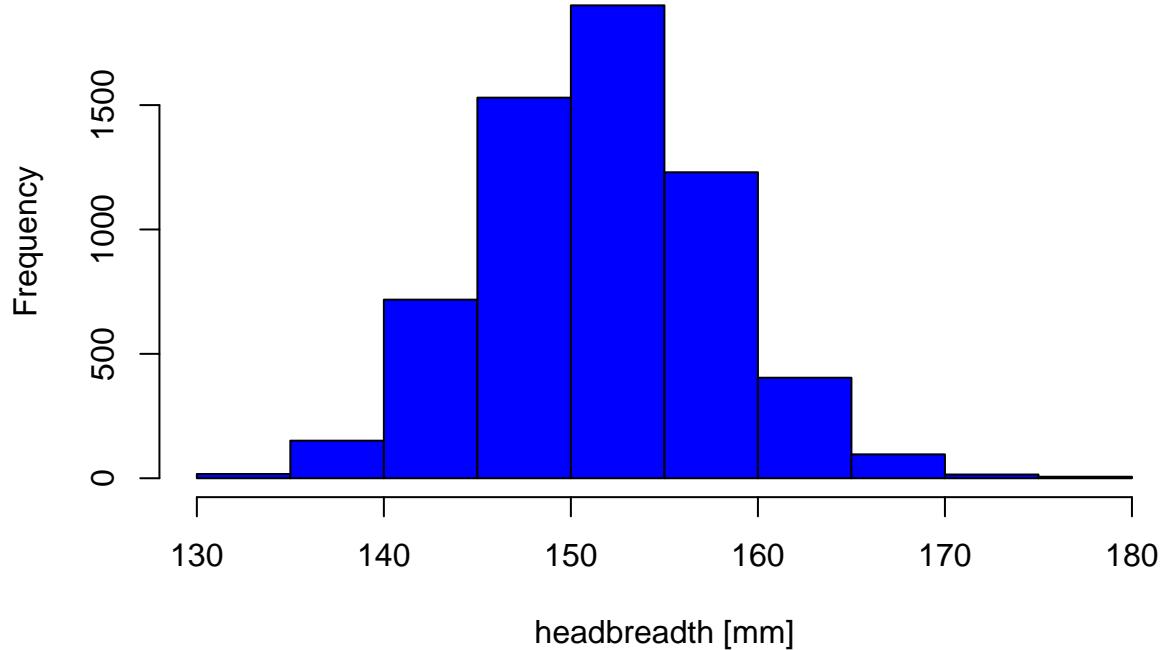
5.3 Mjere centralne tendencije i mjere rasipanja

	Vrijednost	headbreadth	headcircumference	headlength
## 1	Mod	153.000000	577.000000	200.000000
## 2	Podrezana srednja vrijednost	152.140857	569.96952	196.612301
## 3	Rang	49.000000	135.00000	57.000000
## 4	Interkvartilni rang	8.000000	25.00000	11.000000
## 5	Varijanca	38.883850	334.56576	72.001573
## 6	Standardna devijacija	6.235692	18.29114	8.485374

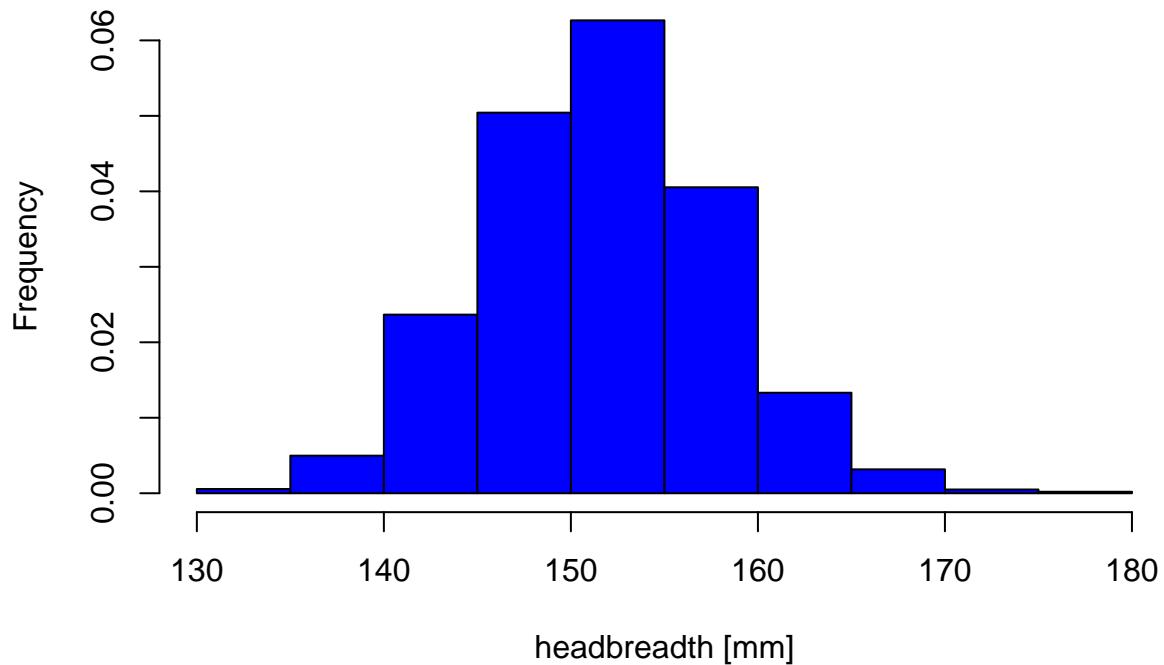
5.4 Vizualizacija podataka

- Histogram - pokazuje oblik distribucije i gustoću podataka, a zasnovan je na grupiranju varijabli u razrede
- Dijagram raspršenja (scatter plot) - jedan je od najvažnijih načina prikaza bivarijantnih podataka, te daje informaciju o povezanosti varijabli.

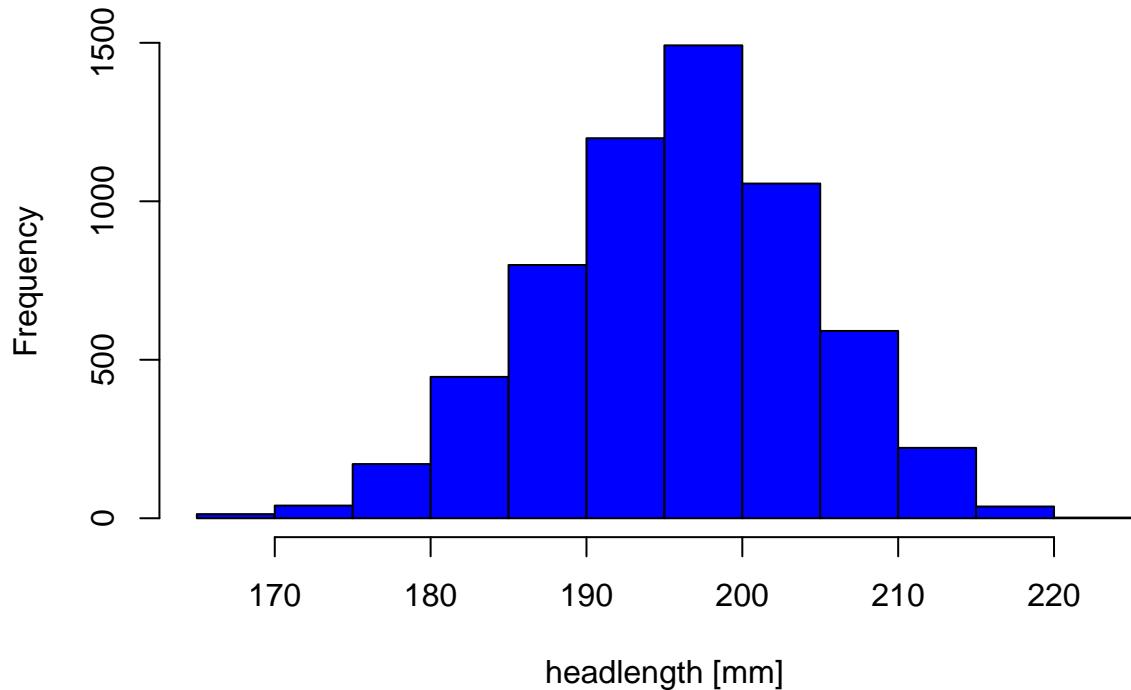
headbreadth histogram



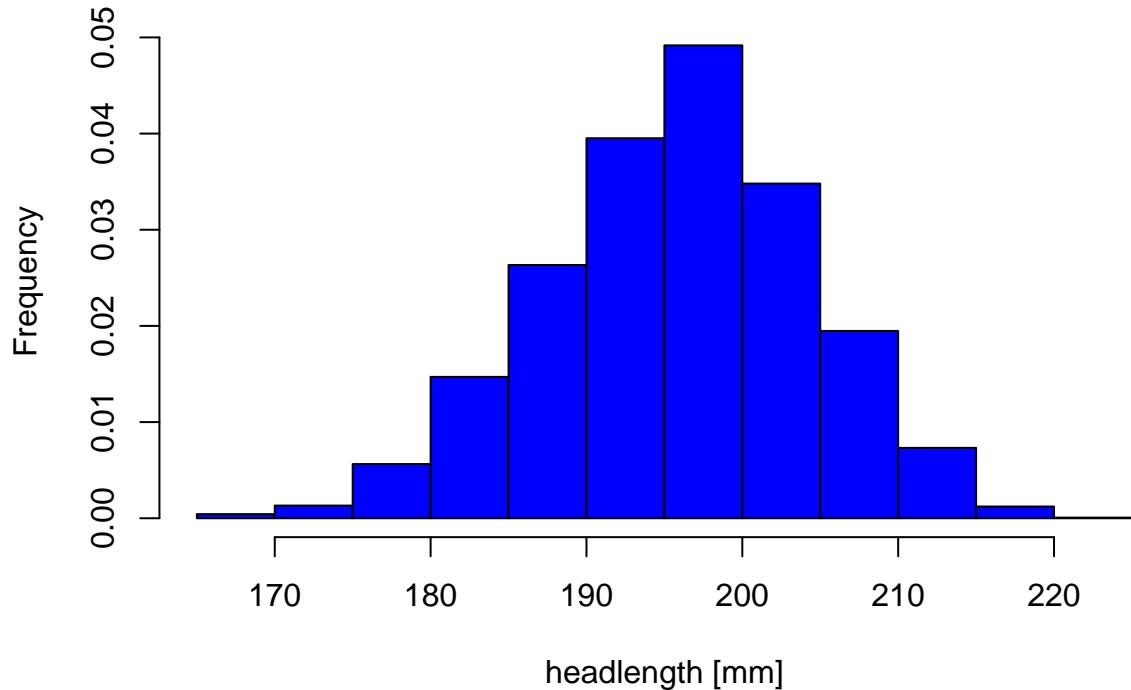
headbreadth histogram



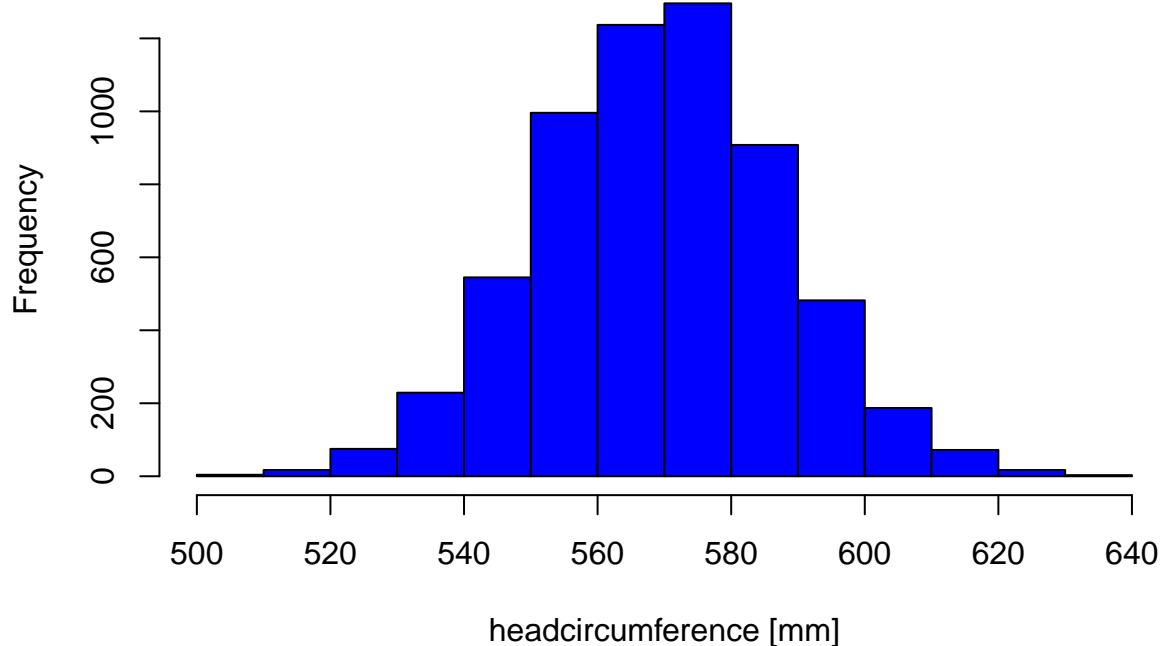
headlength histogram



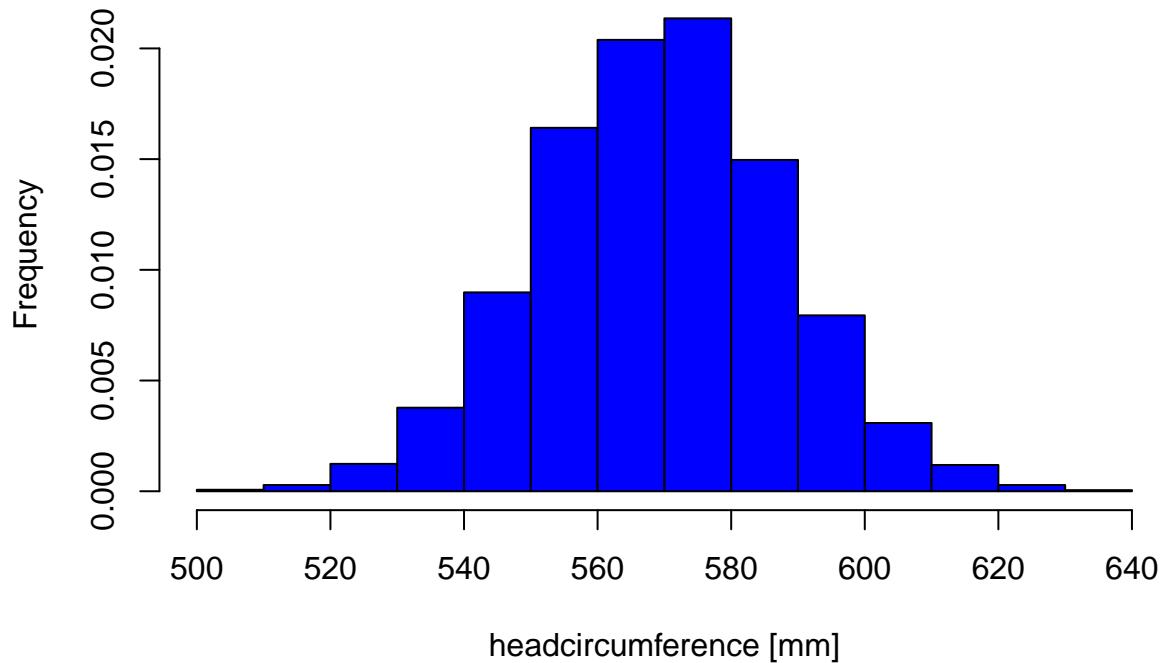
headlength histogram



headcircumference histogram



headcircumference histogram



5.5 Prikaz veličina kaciga

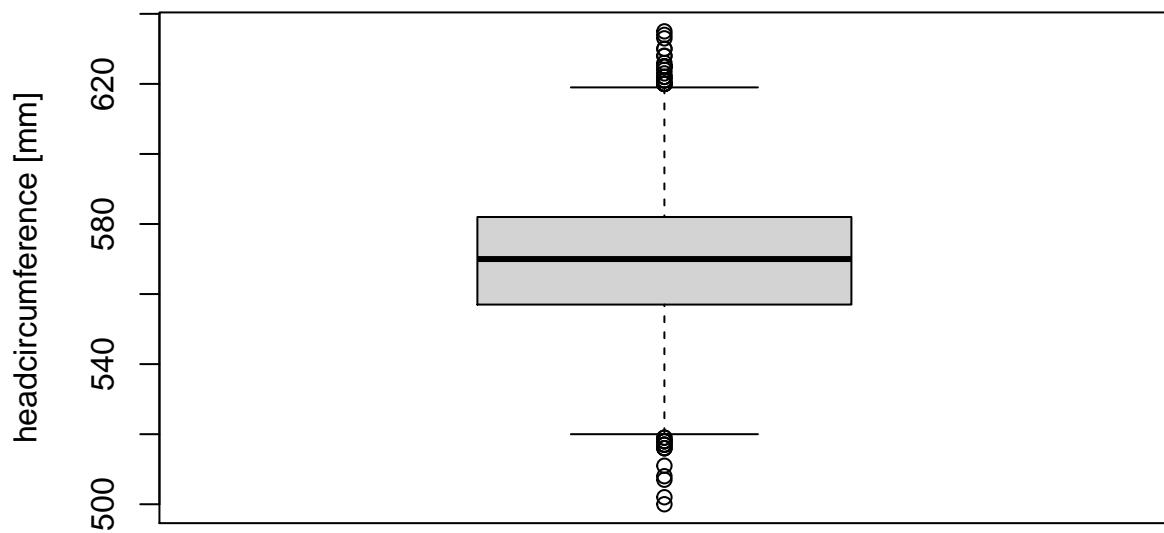
Daljna statistička analiza podataka će se referencirati na ove veličine kaciga.

Table 1. Head/Shell Sizing.

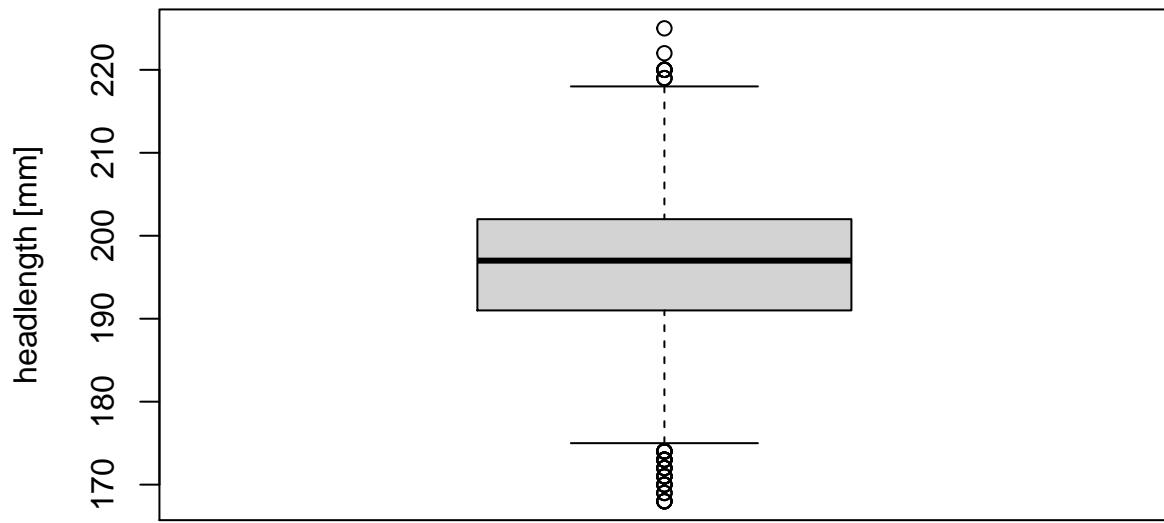
Helmet Shell	Head Length	Head Breadth (Width)	Head Circumference
Small Helmet Shell	Up to 7 $\frac{1}{4}$ inches (184 mm)	Up to 6 $\frac{1}{2}$ inches (162 mm)	Up to 21 $\frac{1}{4}$ inches (538 mm)
Medium Helmet Shell	From 7 $\frac{1}{4}$ inches (184 mm) up to 7 $\frac{3}{4}$ inches (198 mm)	Up to 6 $\frac{1}{2}$ inches (162 mm)	From 21 $\frac{1}{4}$ inches (538 mm) up to 22 $\frac{1}{2}$ inches (573 mm)
Large Helmet Shell	From 7 $\frac{3}{4}$ inches (198 mm) up to 8 $\frac{1}{4}$ inches (210 mm)	Up to 6 $\frac{1}{2}$ inches (162 mm)	From 22 $\frac{1}{2}$ inches (573 mm) up to 23 $\frac{1}{2}$ inches (597 mm)
Extra-Large Helmet Shell	8 $\frac{1}{4}$ inches (210 mm) and over	6 $\frac{1}{2}$ inches (162 mm) and over	23 $\frac{1}{2}$ inches (597 mm) and over

Figure 30: Helmet sizing

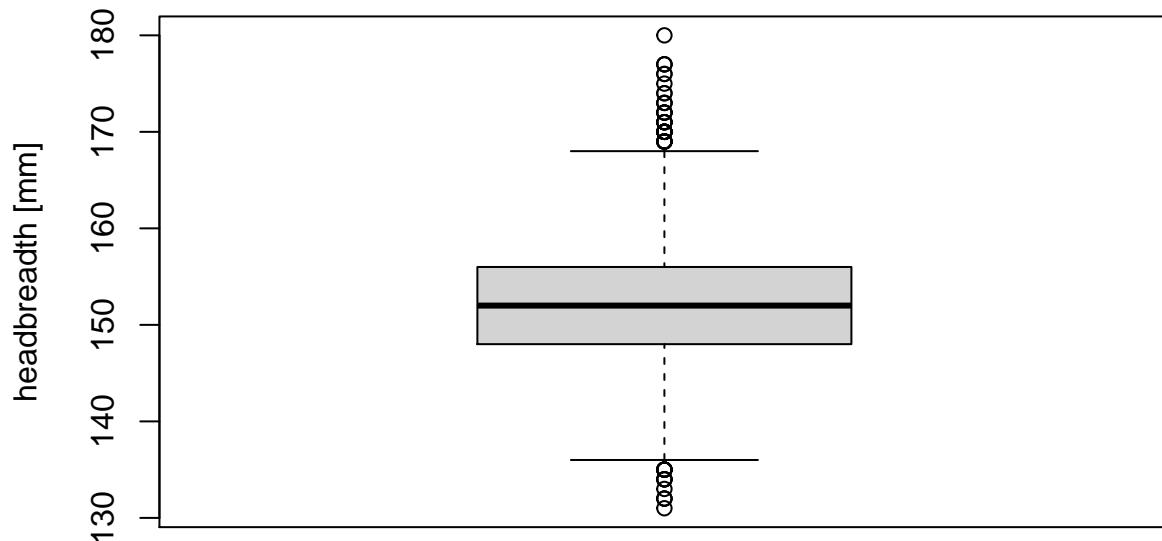
headcircumference box-plot



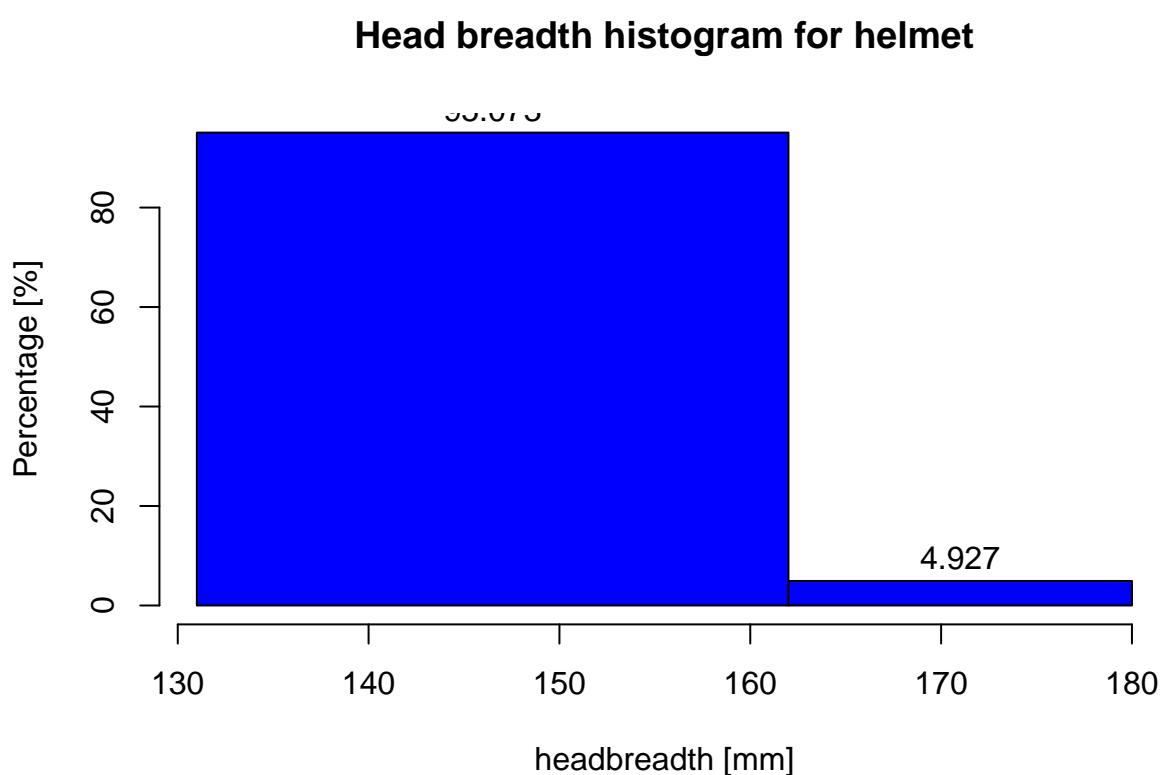
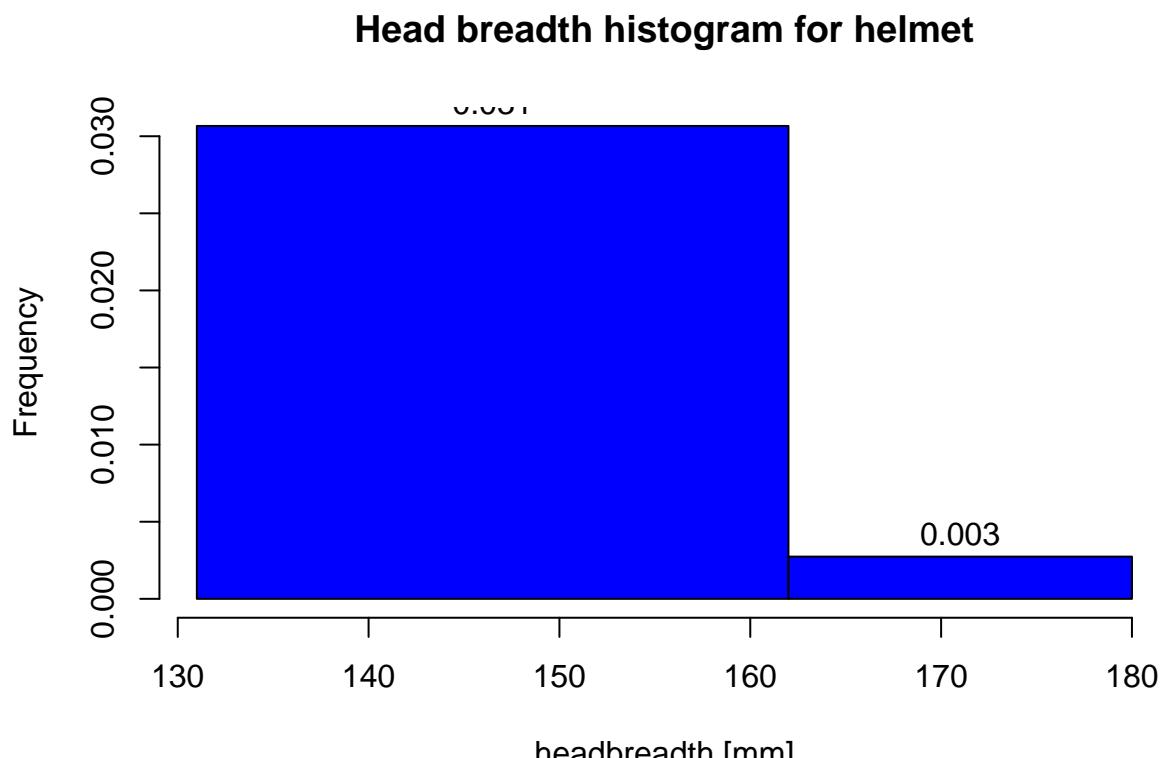
headlength box-plot



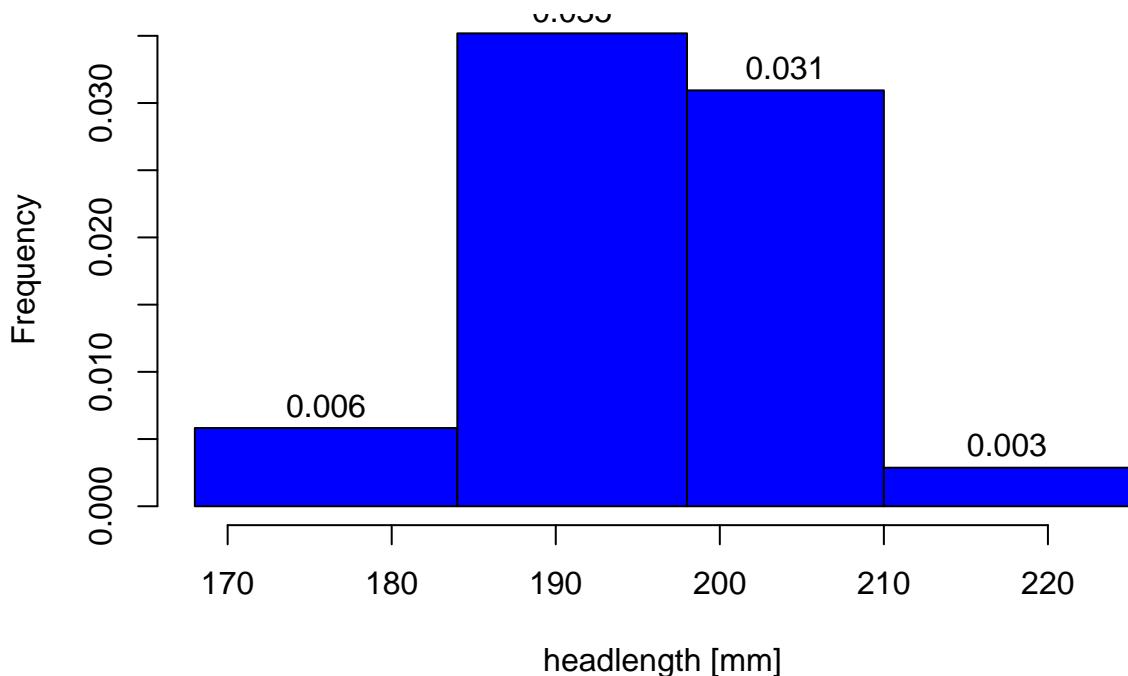
headbreadth box-plot



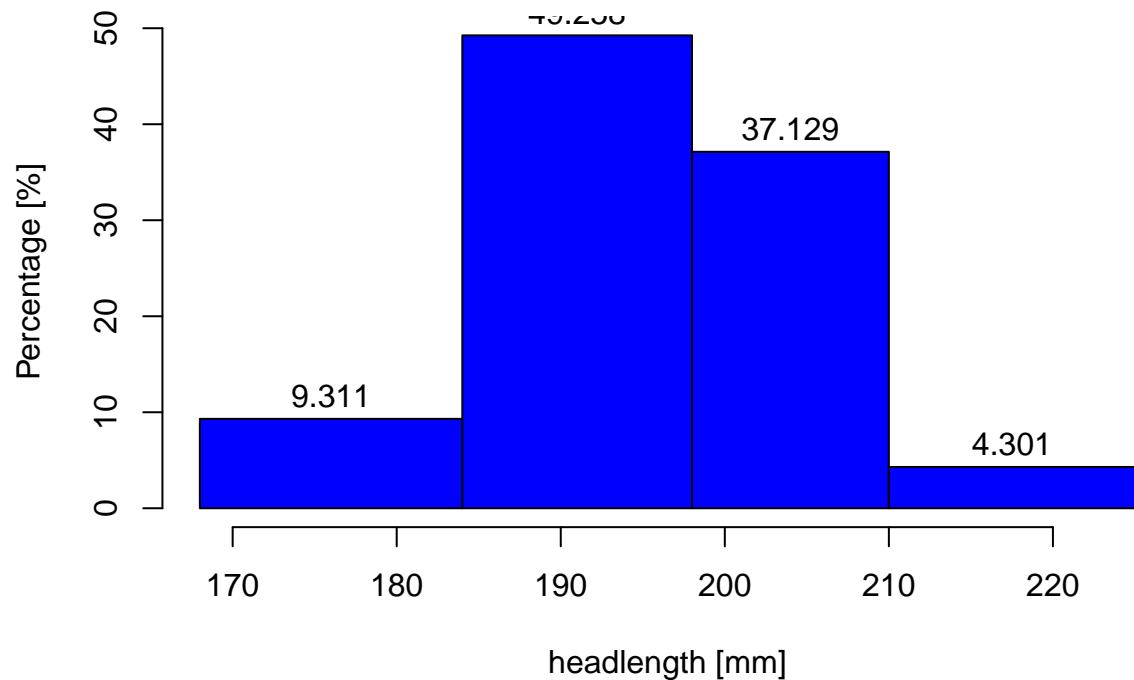
5.6 Histogrami s custom breakpointovima po veličinama kaciga



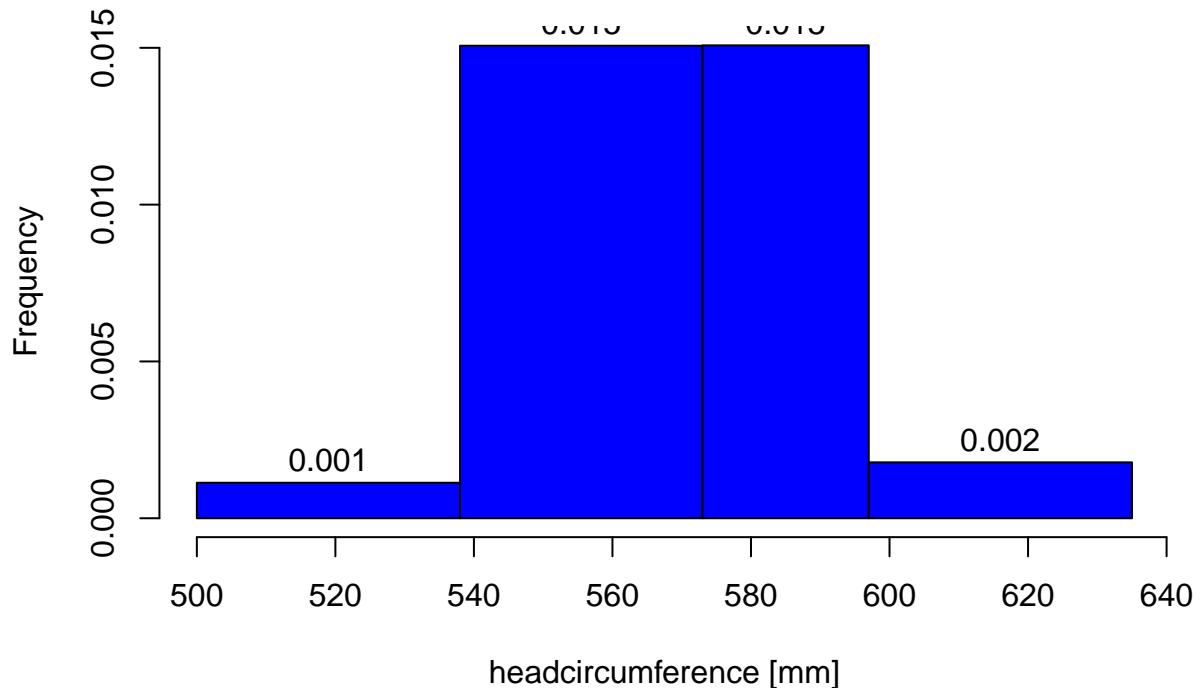
Head length histogram for helmet



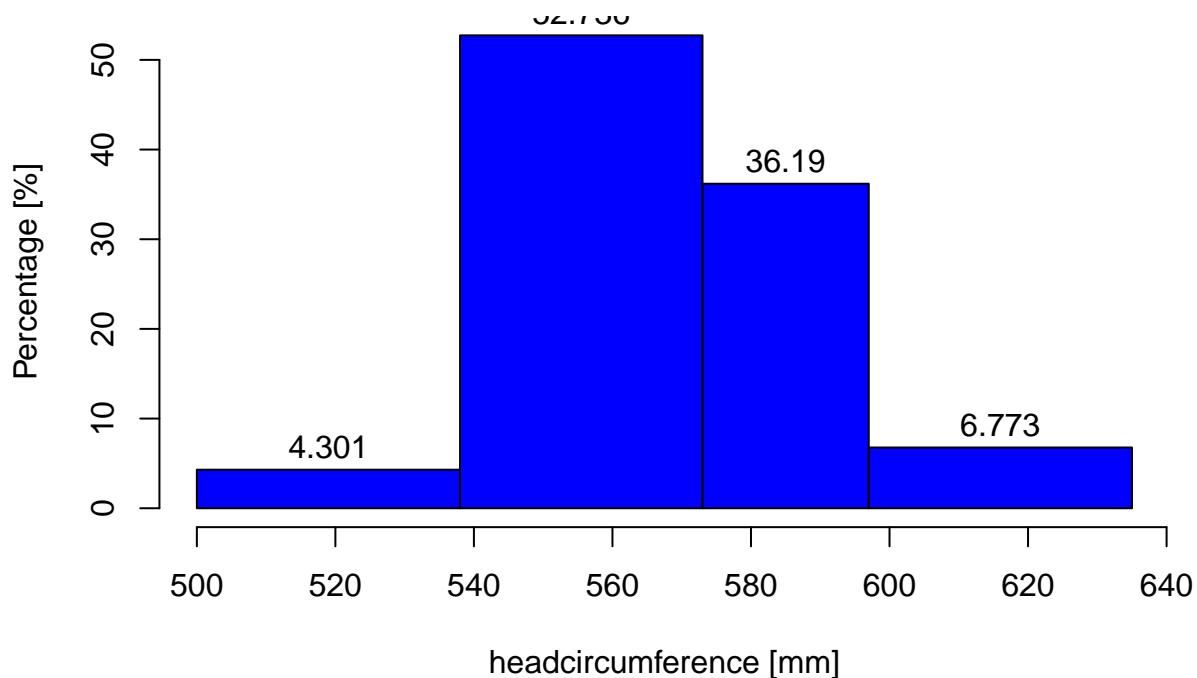
Head length histogram for helmet



Head circumference histogram for helmet



Head circumference histogram for helmet



5.7 Brojevi vojnika kojima treba odreena veličina kacige

```
## Broj vojnika kojima treba XL kaciga: 847  
##  
## Broj vojnika kojima treba L kaciga: 2543  
##  
## Broj vojnika kojima treba M kaciga: 2506  
##  
## Broj vojnika kojima treba S kaciga: 172
```

5.8 Postotci udjela veličina kaciga u vojski

Pie Chart velicina kaciga

