SAP - Projekt

Case study *Antropometrijski podaci američkih vojnih snaga*

Marija Dragošević, Petra Ilić, Petra Jović, Bartol Rod

17. January, 2021.

Sadržaj

1 Uvod **3**
1.1 Prikaz podataka . . . 3

2 Analiza vojnih kampova **6**
2.1 Priprema podataka . . . 6
2.2 Jednofaktorska ANOVA . . . 10
2.2.1 Normalna razdioba . . . 10
2.2.2 Homogenost varijanci među populacijama . . . 14
2.2.3 Provođenje ANOVA-e . . . 18
2.3 Dvofaktorska ANOVA . . . 19
2.4 Zaključak . . . 26

3 Predviđanje kilaže vojnika **27**
3.1 Jednostavna regresija . . . 27
3.1.1 Normalnost reziduala i homogenost varijance . . . 33
3.1.2 Ocjena kvalitete linearnog modela i statističko zaključivanje o procijenjenom modelu . . . 42
3.1.3 Korelacijski koeficijenti u linearnom modelu . . . 44
3.2 Višestruka regresija . . . 45
3.2.1 Provjera normalnosti reziduala . . . 47
3.2.2 Uključivanje kategorijskih varijabli . . . 49
3.2.3 Zaključak . . . 50

4 Opseg bicepsa **51**
4.1 Prikaz raspodjele vojnika ovisno o preferenciji ruke i spolu . . . 51
4.2 Test o dvije proporcije: dva uzorka . . . 53
4.2.1 Udio ljevaka ovisno o spolu . . . 53
4.3 Opseg bicepsa ovisno o preferenciji ruke . . . 54
4.3.1 Opseg desnog bicepsa dešnjaka i ljevaka . . . 54
4.3.1.1 Provjera normalnosti podataka . . . 54
4.3.1.2 Test o jednakosti varijanci . . . 61
4.3.1.3 t-test uz pretpostavku jednakosti varijanci . . . 61
4.3.2 Opseg desnog bicepsa dešnjaka i ambidekstrenih vojnika . . . 61
4.3.2.1 Test o jednakosti varijanci . . . 61
4.3.2.2 t-test uz pretpostavku jednakosti varijanci . . . 62

5 Mjere kaciga **63**
5.1 Klase podataka . . . 63
5.2 Summary . . . 63
5.3 Mjere centralne tendencije i mjere rasipanja . . . 63
5.4 Vizualizacija podataka . . . 63
5.5 Prikaz veličina kaciga . . . 69
5.6 Histogrami s custom breakpointovima po veličinama kaciga . . . 75
5.7 Brojevi vojnika kojima treba odreena veličina kacige . . . 79
5.8 Postotci udjela veličina kaciga u vojsci . . . 80

1 Uvod

U ovom projektu bavit ćemo se analizom antropometrsijskih podatke američkih vojnih snaga (ANSUR_II). Podatci se sastoje od 6068 vojnika i 107 značajki sakupljenih od 2010 d0 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
buttockpopliteallength	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
crotchlengthomphalion	557	549	643
crotchlengthposterioromphalion	310	329	374
earbreadth	35	32	36
earlength	65	60	65
earprotrusion	16	23	26
elbowrestheight	220	208	204
eyeheightsitting	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
interpupillarybreadth	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
radialestylionlength	235	259	258
shouldercircumference	1062	1014	1049
shoulderelbowlength	327	346	362
shoulderlength	148	142	164
sittingheight	803	835	904
sleevelengthspinewrist	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 Priprema podataka

Prvo čemo provjeriti broj opažanja i udio spolova u kampovima.

```
## # 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
```

```
## # A tibble: 12 x 3
##    Installation   pct.males pct.female
##    <chr>              <dbl>      <dbl>
##  1 Camp Atterbury      88.9       11.1
##  2 Camp Shelby         66.6       33.4
##  3 Fort Bliss          79.5       20.5
##  4 Fort Bragg          73.8       26.2
##  5 Fort Drum           81.1       18.9
##  6 Fort Gordon         38.7       61.3
##  7 Fort Hood           85.9       14.1
##  8 Fort Huachuca       80.5       19.5
##  9 Fort Lee             0        100
## 10 Fort McCoy          74.6       25.4
## 11 Fort Rucker        100          0
## 12 Fort Stewart        64.0       36.0
```

Prebrojavanjem vojnika u kampovima vidimo da se u kampu "Fort Rucker" nalazi samo jedna osoba, stoga taj kamp nećemo koristiti u našoj analizi.

Za analizu smo nasumično odabrali sljedeće atribute:

- duljina svoda stopala (balloffootlength)
- opseg prsa (chestcircumference)
- duljina ruke (handlength)
- širina kukova (hipbreadth)
- spol (Gender).

```
##  balloffootlength chestcircumference   handlength      hipbreadth
##  Min.   :151.0    Min.   : 695       Min.   :145.0   Min.   :264.0
##  1st Qu.:185.0    1st Qu.: 949       1st Qu.:181.0   1st Qu.:332.0
##  Median :195.0    Median :1021       Median :190.0   Median :347.0
##  Mean   :194.8    Mean   :1022       Mean   :189.3   Mean   :348.4
##  3rd Qu.:204.0    3rd Qu.:1092       3rd Qu.:197.0   3rd Qu.:364.0
##  Max.   :245.0    Max.   :1469       Max.   :239.0   Max.   :473.0
##
##     Gender             Installation
##  Male  :4081   Camp Shelby   :1160
##  Female:1986   Fort Bliss    : 963
##                Fort Gordon   : 669
##                Fort McCoy    : 452
##                Camp Atterbury: 441
##                Fort Hood     : 439
##                (Other)       :1943
```

Table 2: Srednja vrijednost pojedine mjere u zasebnom kampu

Installation	mean.balloffootlength	mean.chestcircumference	mean.handlength	mean.hipbreadth
Fort Hood	196.1367	1048.5877	192.0433	348.1276
Fort Bliss	195.8640	1042.4351	190.0530	348.5389
Camp Atterbury	196.9796	1063.5374	192.0068	350.3946
Fort Drum	195.7442	1047.2225	192.4015	350.9412
Fort McCoy	195.2235	1034.0354	190.8562	345.9358
Fort Lee	181.4342	927.5158	182.1684	347.5868
Fort Stewart	194.7493	1014.6077	188.3864	348.8525
Fort Bragg	197.4383	1019.4836	187.8866	342.5995
Fort Gordon	190.6637	970.6293	185.1749	347.4694
Fort Huachuca	198.1422	1034.2133	189.3601	347.3739
Camp Shelby	196.4767	1025.5379	190.3448	350.7155

Table 3: Medijan pojedine mjere u zasebnom kampu

Installation	med.balloffootlength	med.chestcircumference	med.handlength	med.hipbreadth
Fort Hood	196	1048.0	192	346
Fort Bliss	197	1046.0	190	348
Camp Atterbury	198	1065.0	192	349
Fort Drum	196	1036.0	193	351
Fort McCoy	196	1028.5	192	346
Fort Lee	181	923.0	182	346
Fort Stewart	195	1006.0	189	345
Fort Bragg	199	1021.0	187	341
Fort Gordon	189	966.0	184	345
Fort Huachuca	199	1038.5	190	347
Camp Shelby	198	1025.0	190	348

Table 4: Varijanca pojedine mjere u zasebnom kampu

Installation	var.balloffootlength	var.chestcircumference	var.handlength	var.hipbreadth
Fort Hood	149.62968	9292.841	113.1009	716.6960
Fort Bliss	157.75175	9077.791	121.8589	527.2280
Camp Atterbury	135.45186	8700.231	120.5613	573.6894
Fort Drum	156.66263	9849.738	133.0050	684.2452
Fort McCoy	168.40007	10801.183	140.4516	655.4526
Fort Lee	98.16716	5878.963	101.2486	626.4436
Fort Stewart	203.65588	9084.292	133.1254	784.5699
Fort Bragg	161.50439	7849.513	124.4240	565.4326
Fort Gordon	185.34630	7947.488	134.0906	702.8033
Fort Huachuca	173.21192	9366.462	118.5620	545.6691
Camp Shelby	199.01413	9755.866	128.5816	677.3271

Iz pregleda srednjih vrijednosti i medijana kampova, pošto su približno jednaki, možemo naslutiti na normalnu razdiobu vrijednosti. Također, ispisom varijance ne uočavamo veliko odstupanje u vrijednosti među kampovima, stoga, kako bi smo vidjeli imali li neke značajne razlike u antropometrijskim podatcima možemo provest jednofaktorski ANOVA model.

2.2 Jednofaktorska ANOVA

Pretpostavke ANOVE su:

- nezavisnost pojedinih podataka u uzorcima
- normalna razdioba podataka
- homogenost varijanci među populacijama.

Podatci su neovisni pošto se radi o pojedinim osobama.

2.2.1 Normalna razdioba

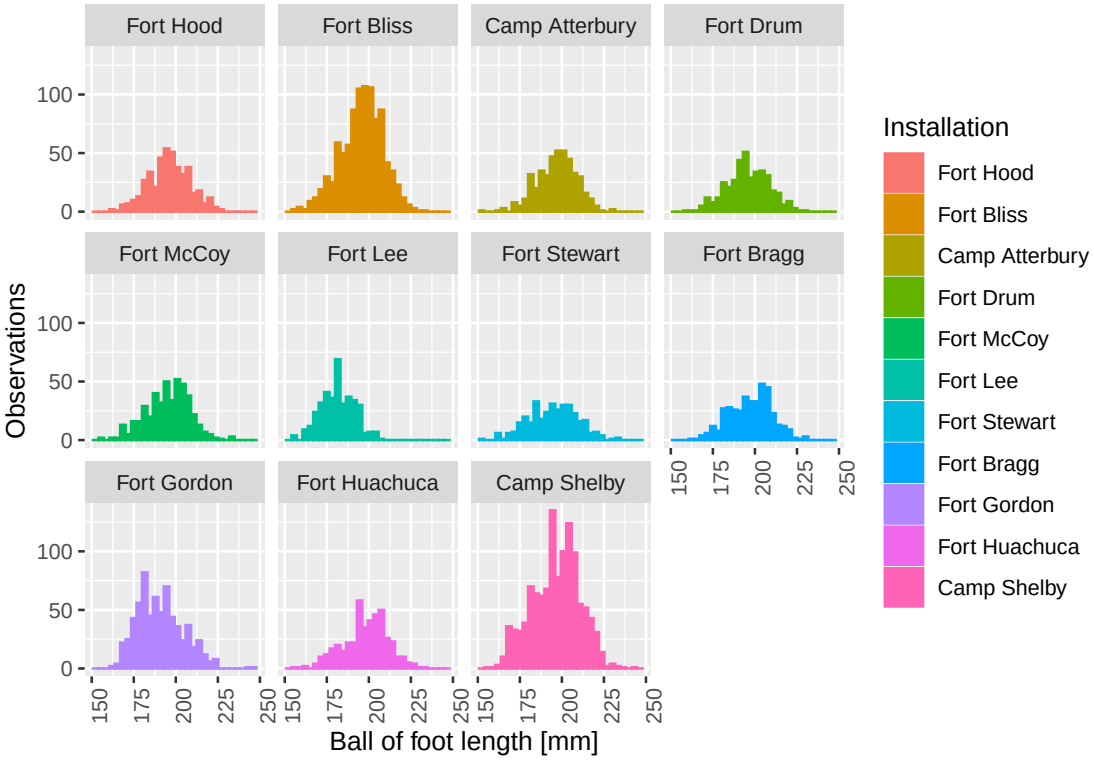


Figure 1: Histogram duljine svoda stopala

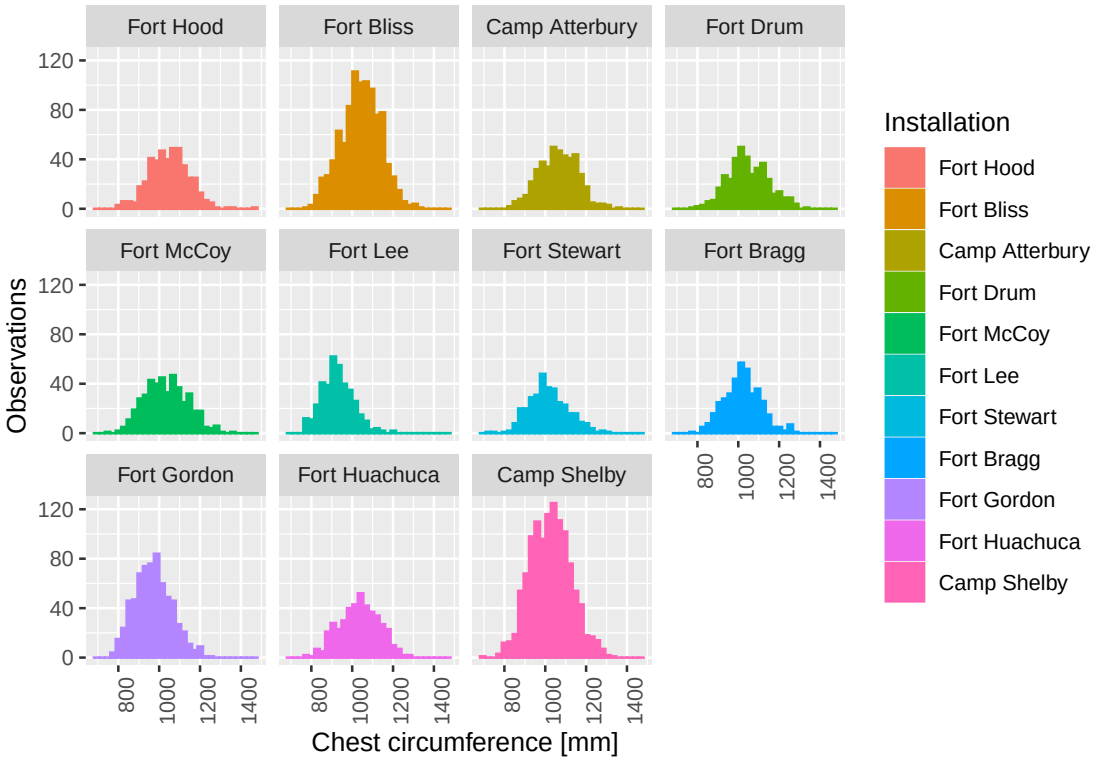


Figure 2: Histogram opsega prsa

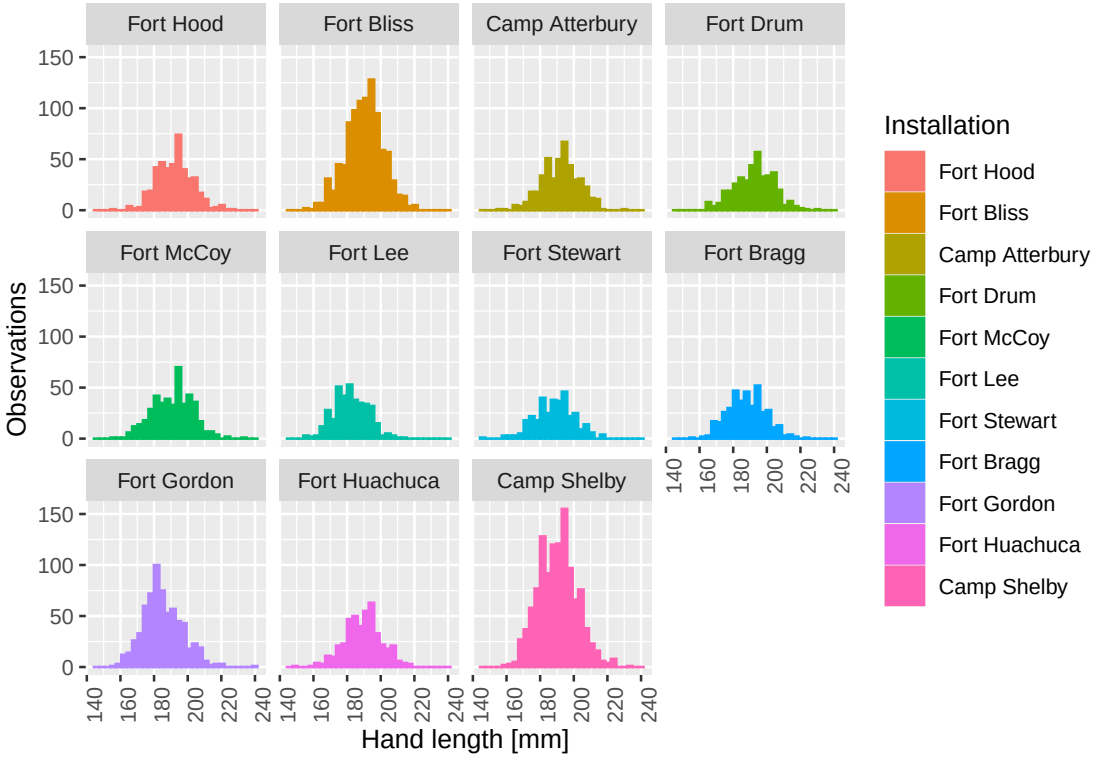


Figure 3: Histogram duljine ruke

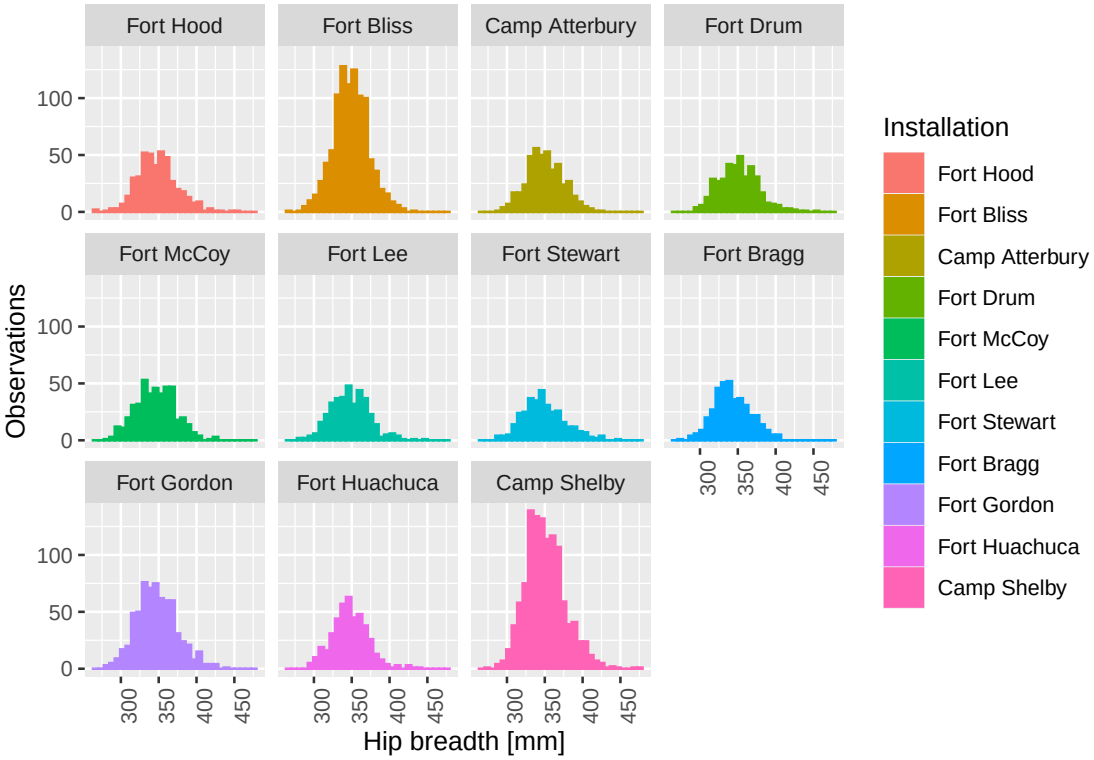


Figure 4: Histogram širine kukova

Table 5: P-vrijednosti Kolmogorov-Smirnovovog testa pojedinih vrijednosti u kampovima

Installation	balloffootlength.p	chestcircumference.p	handlength.p	hipbreadth
Ungrouped	0.0000000	0.0000466	0.0000000	0.0000000
Fort Hood	0.1088858	0.7832516	0.0334878	0.0000146
Fort Bliss	0.0000000	0.1792532	0.0000820	0.0257725
Camp Atterbury	0.0002280	0.6656600	0.0923188	0.0084249
Fort Drum	0.1923143	0.0047963	0.0237815	0.0390572
Fort McCoy	0.0000334	0.1274228	0.0000016	0.1074329
Fort Lee	0.2734808	0.0575193	0.0664283	0.5422562
Fort Stewart	0.0996757	0.0095202	0.0938587	0.0004391
Fort Bragg	0.0001751	0.7155254	0.1268504	0.0000558
Fort Gordon	0.0000009	0.0982468	0.0000000	0.0009352
Fort Huachuca	0.0003609	0.4262501	0.1010728	0.0508495
Camp Shelby	0.0000000	0.0406946	0.0019589	0.0000010

Testiranjem podataka na normalnost u većini slučajeva vidimo veoma visoke p-vrijednosti što uz određeni nivo značajnosti ukazuje na normalnost. P-vrijednosti negrupiranih pojedinih atributa bitno se razlikuju od grupiranih čime također možemo naslutiti na postojanje neke razlike među vojnim kampovima. Grafički prikaz podataka također nas navodi kako su podatci normalno distribuirani.

2.2.2 Homogenost varijanci među populacijama

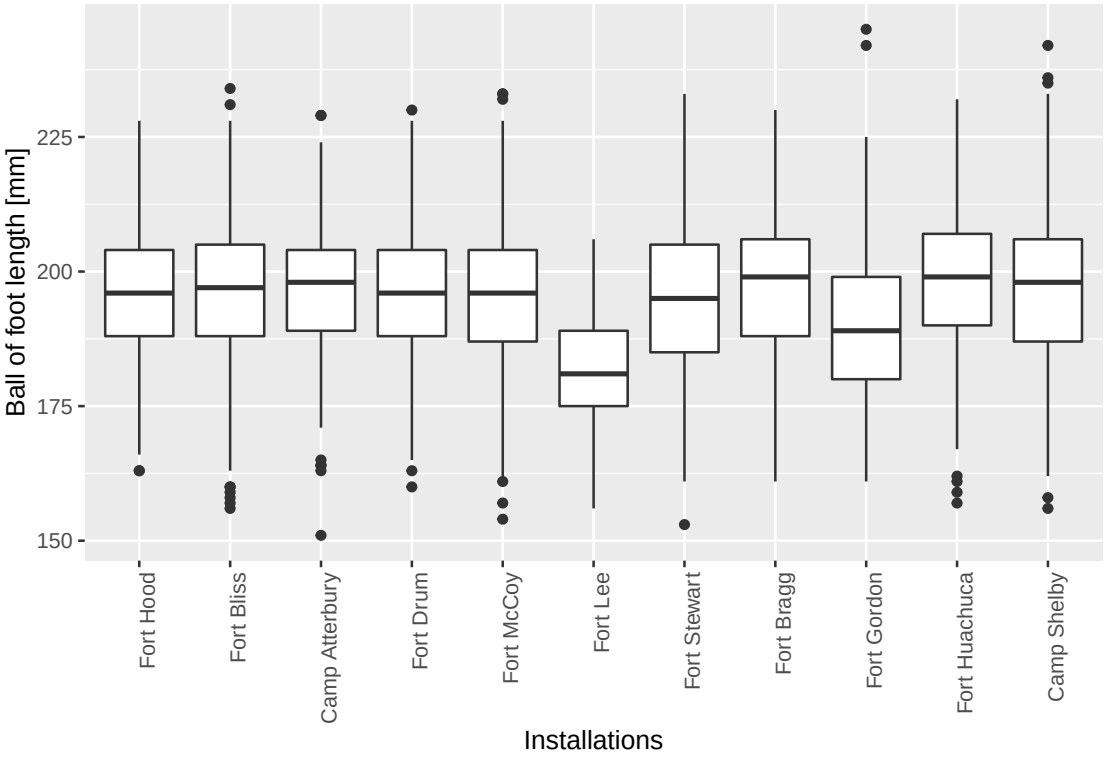


Figure 5: Boxplot-ovi duljine svoda stopala

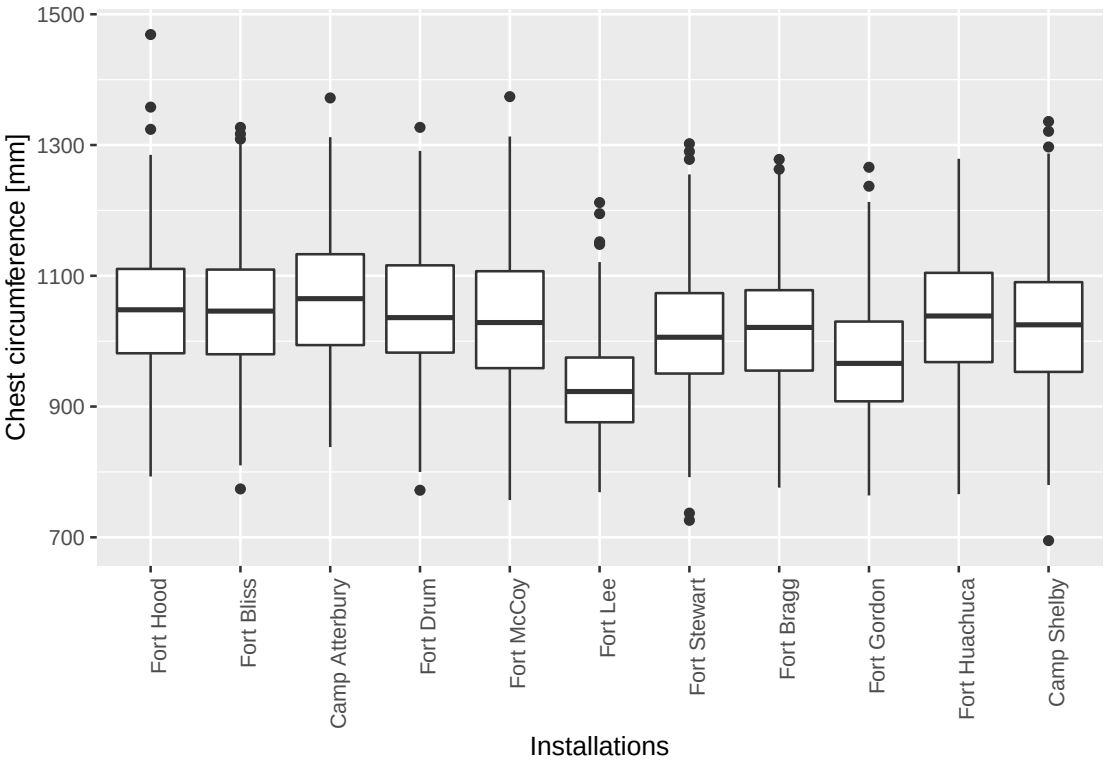


Figure 6: Boxplot-ovi opsega prsa

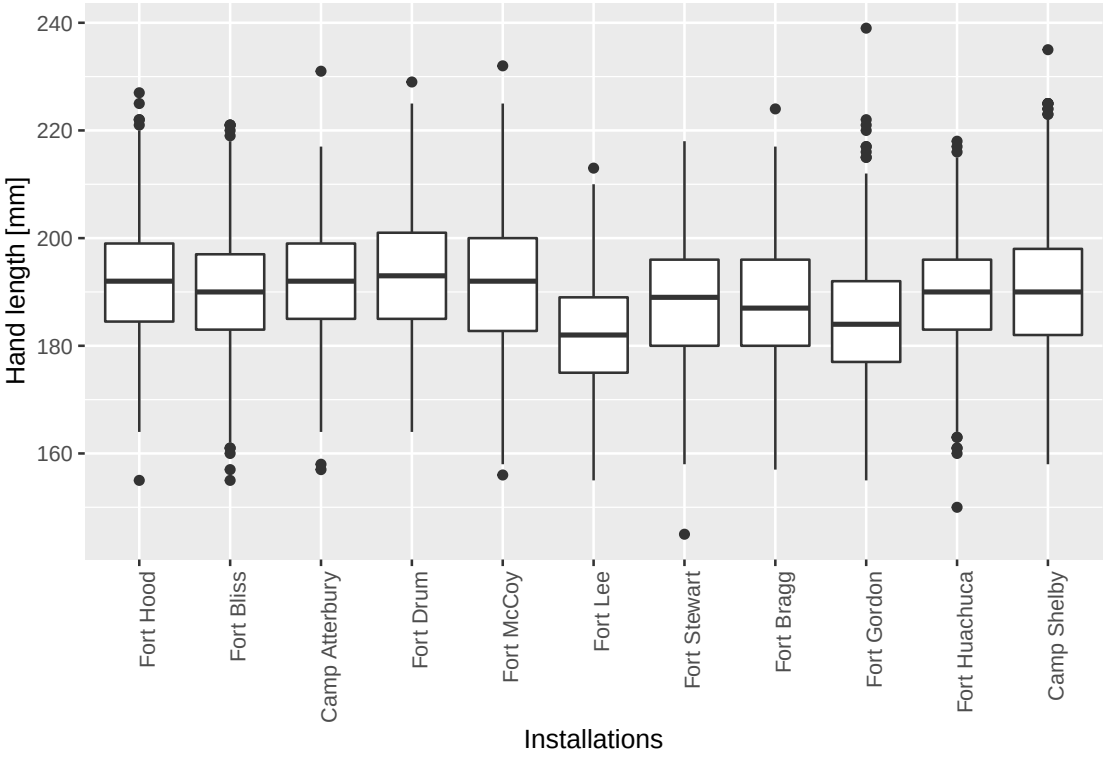


Figure 7: Boxplot-ovi duljine ruke

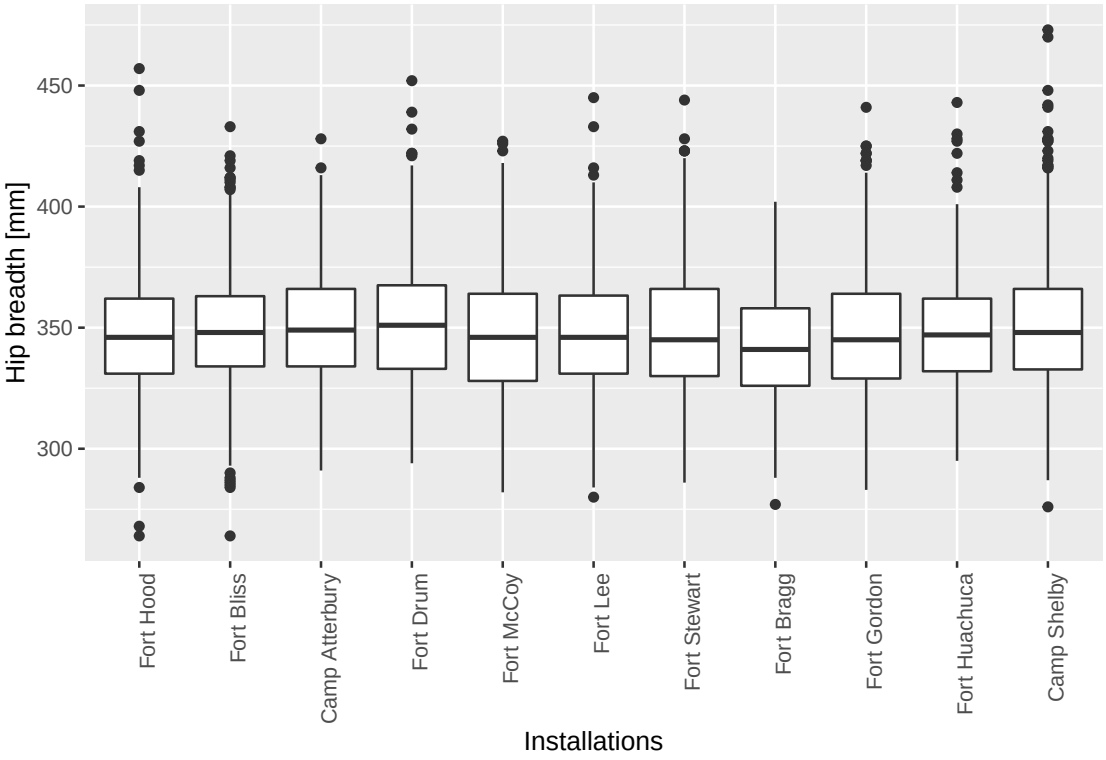


Figure 8: Boxplot-ovi širine kukova

```
##
##  Bartlett test of homogeneity of variances
##
## data:  df.camps$balloffootlength by df.camps$Installation
## Bartlett's K-squared = 89.255, df = 10, p-value = 7.524e-15
```

```
##
##  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
```

Iako Bartlettov test homogenosti varijanci ne daje povoljne p-vrijednosti, pošto se naše varijance ne razlikuju za veću dimenziju, provest ćemo ANOV-u.

2.2.3 Provođenje ANOVA-e

```
##                          Df  Sum Sq Mean Sq F value Pr(>F)
## df.camps$Installation   10   94613    9461   56.52 <2e-16 ***
## Residuals             6056 1013753     167
## ---
## 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  7045922  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
```

Provođenjem ANOVA-e uz beznačajno male p-vrijednosti zaključujemo kako postoji bitna razlika u varijancama antropometrijskih podataka pojedinog vojnog kampa.

2.3 Dvofaktorska ANOVA

Iz boxplot-a nekih antropometrijskih podataka možemo vidjeti znatno odskakanje srednje vrijednosti kampa "Fort Lee". Nadalje, iz tablice udjela spolova u kampovima vidljivo je da u kampu "Fort Lee" nema opažanja muških vojnika. Stoga provodimo dvofaktorsku ANOVA-u s faktorima spola i kampa.

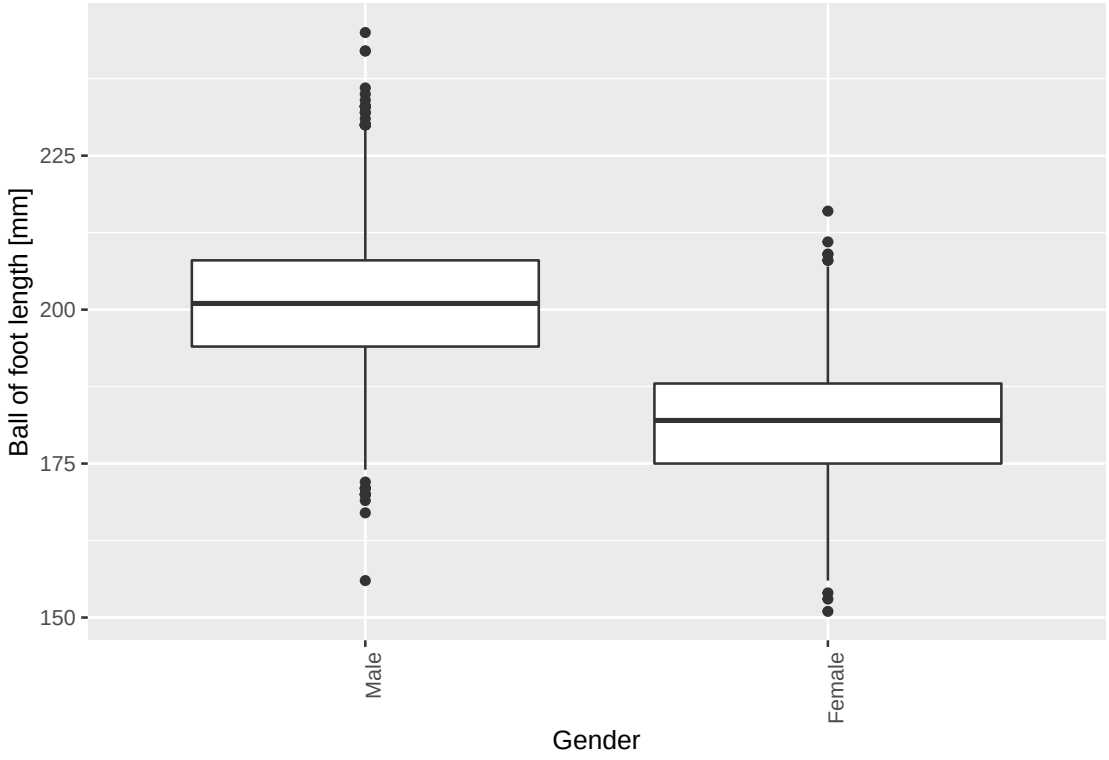


Figure 9: Boxplot-ovi duljine svoda stopala po spolovima

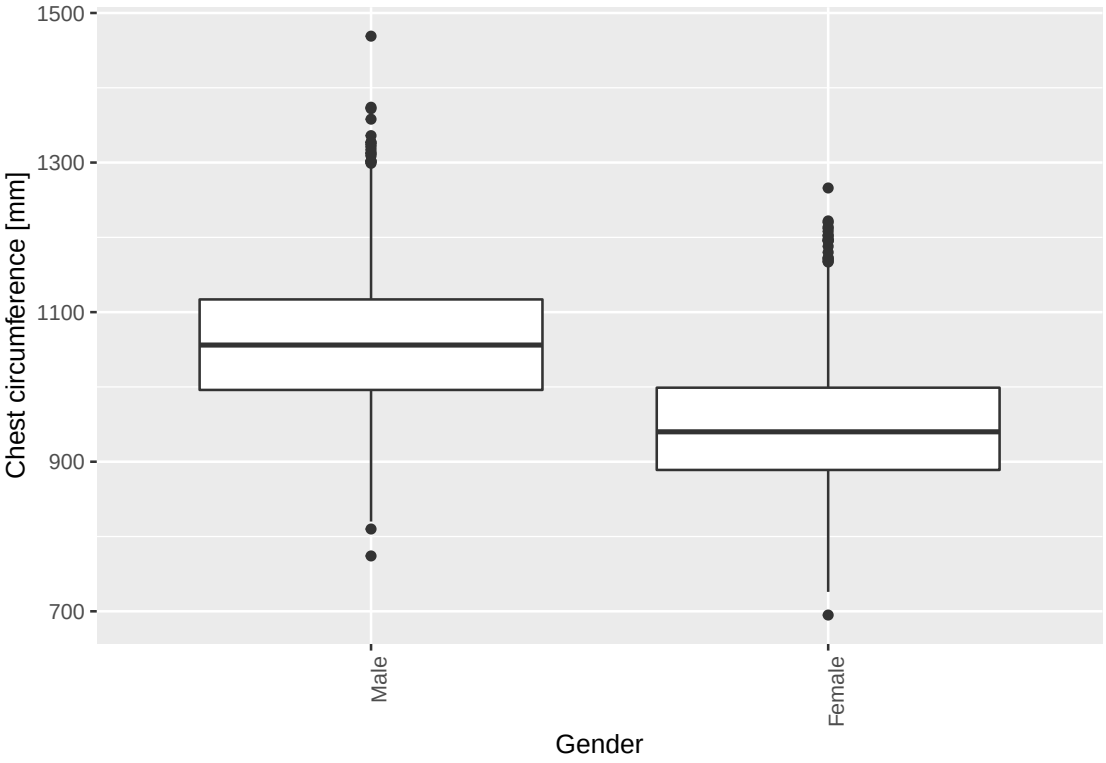


Figure 10: Boxplot-ovi opsega prsa po spolovima

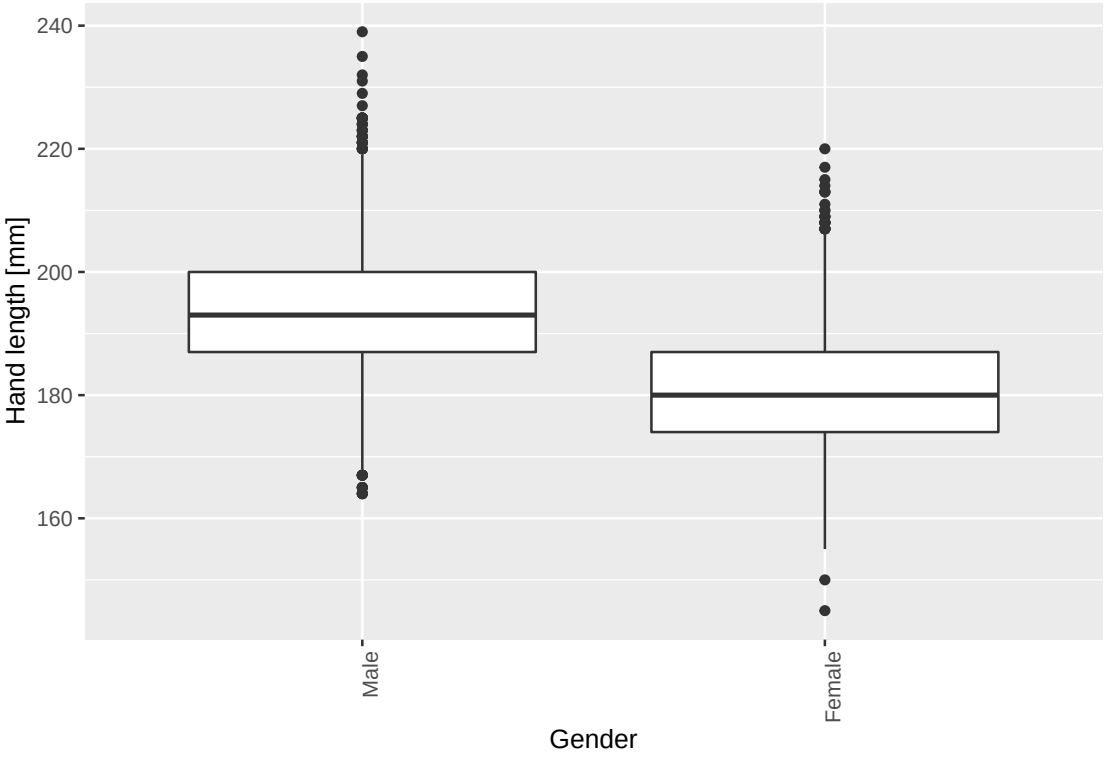


Figure 11: Boxplot-ovi duljine ruke po spolovima

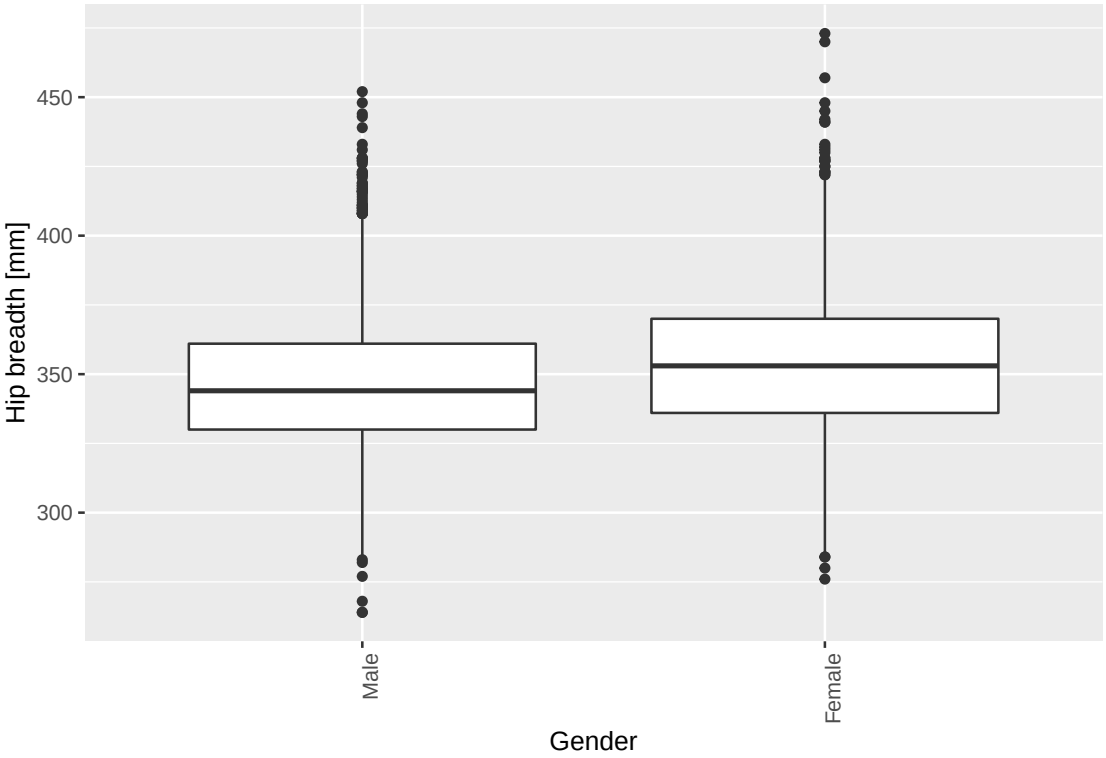


Figure 12: Boxplot-ovi širine kukova po spolovima

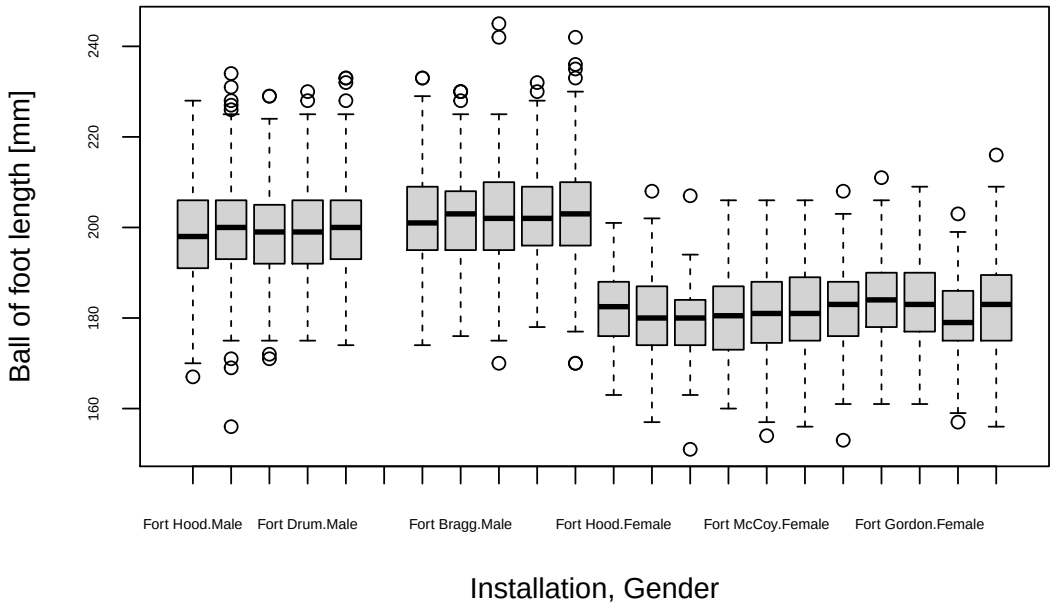


Figure 13: Boxplot-ovi duljine svoda stopala [Installation, Gender]

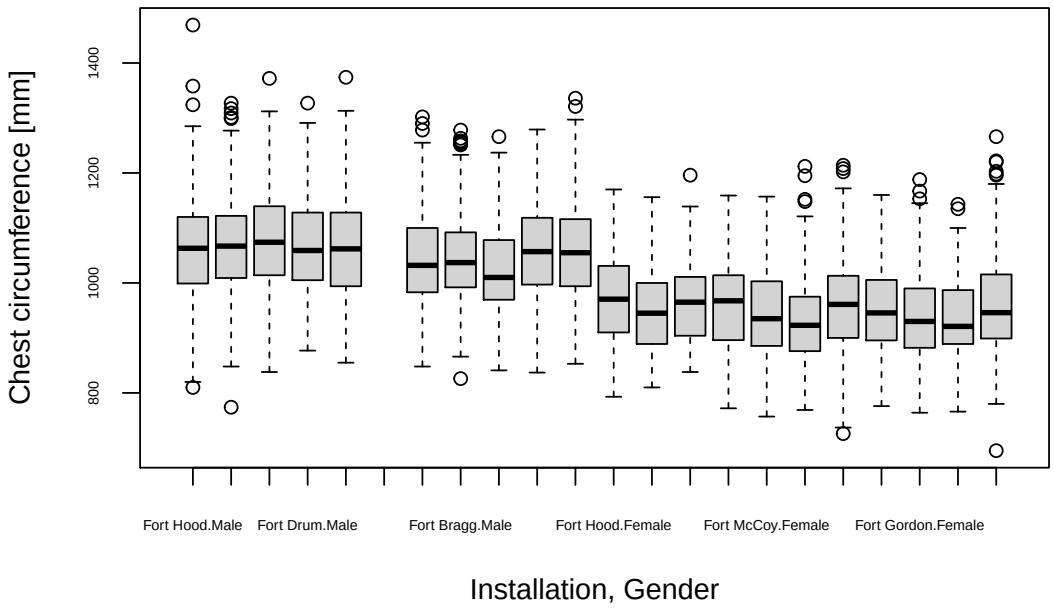


Figure 14: Boxplot-ovi opsega prsa po spolovima [Installation, Gender]

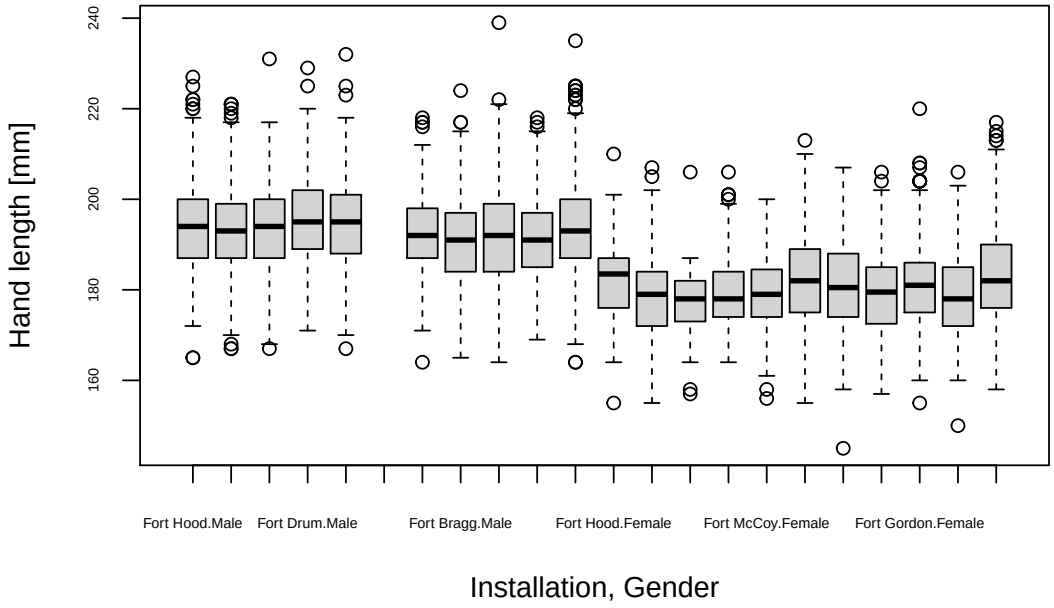


Figure 15: Boxplot-ovi duljine ruke po spolovima [Installation, Gender]

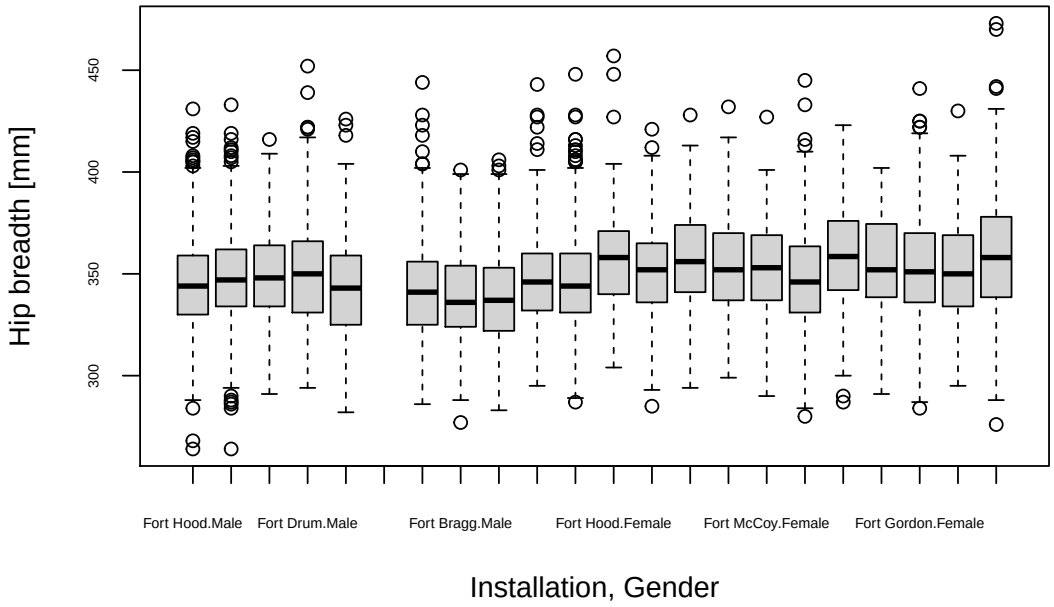


Figure 16: Boxplot-ovi širine kukova po spolovima [Installation, Gender]

Iz boxplotova vidljiva je znatna razlika u srednjoj vrijednosti po spolovima.

```
##
##  Bartlett test of homogeneity of variances
##
## data:  df.camps$balloffootlength by inter
## Bartlett's K-squared = 43.801, df = 20, p-value = 0.0016
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  df.camps$chestcircumference by inter
## Bartlett's K-squared = 42.709, df = 20, p-value = 0.002233
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  df.camps$handlength by inter
## Bartlett's K-squared = 31.026, df = 20, p-value = 0.05485
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  df.camps$hipbreadth by inter
## Bartlett's K-squared = 83.103, df = 20, p-value = 1.16e-09
```

```
##                       Df Sum Sq Mean Sq  F value Pr(>F)
## Installation          10  94613    9461   92.759 <2e-16 ***
## Gender                 1 395063  395063 3873.219 <2e-16 ***
## Installation:Gender    9   2006     223    2.185 0.0202 *
## Residuals           6046 616684     102
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##                       Df   Sum Sq  Mean Sq  F value Pr(>F)
## Installation          10  7045922   704592   97.294 <2e-16 ***
## Gender                 1 10500287 10500287 1449.943 <2e-16 ***
## Installation:Gender    9   156749    17417    2.405 0.0102 *
## Residuals           6046 43784303     7242
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##                       Df Sum Sq Mean Sq F value   Pr(>F)
## Installation          10  44991    4499   45.94  < 2e-16 ***
## Gender                 1 160815  160815 1642.12  < 2e-16 ***
## Installation:Gender    9   4936     548    5.60 9.79e-08 ***
## Residuals           6046 592093      98
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##                       Df  Sum Sq Mean Sq F value   Pr(>F)
## Installation          10   28010    2801   4.566 1.75e-06 ***
## Gender                 1  120939  120939 197.145  < 2e-16 ***
## Installation:Gender    9   24504    2723   4.438 8.13e-06 ***
## Residuals           6046 3708930     613
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2.4 Zaključak

Provođenjem jednofaktorske ANOVA-e zaključili smo kako postoji značajna razlika u antropometrijskim podacima populacija vojnih kampova. Daljnjom analizom grafički smo prikazali kako faktor spola utječe na neke antropometrijske mjere kao što su duljina ruke ili opseg prsa, ali iako je faktor spola podjelio podatke u njihovim srednjim vrijednostima, temeljem dvofaktorske ANOVA-e zaključili smo da još uvijek postoji značajna razlika u varijanci promatranih antropometrijskih podataka između istospolnih populacija kampova.

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

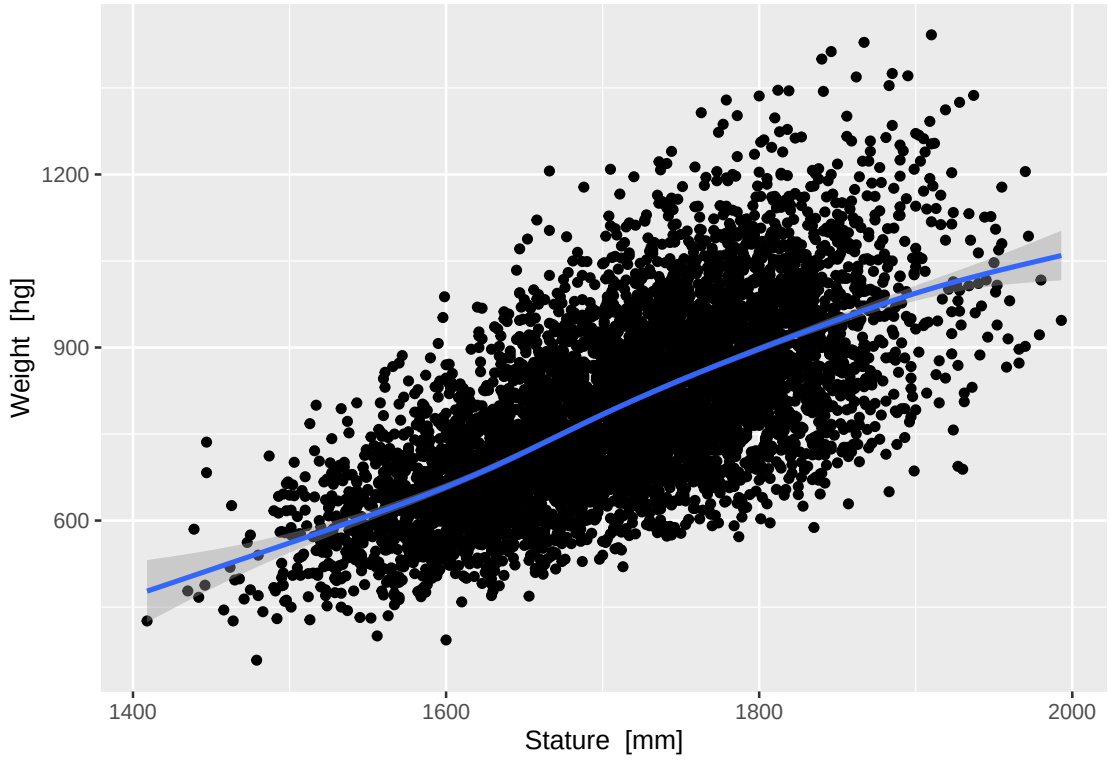


Figure 17: Utjecaj visine na kilažu

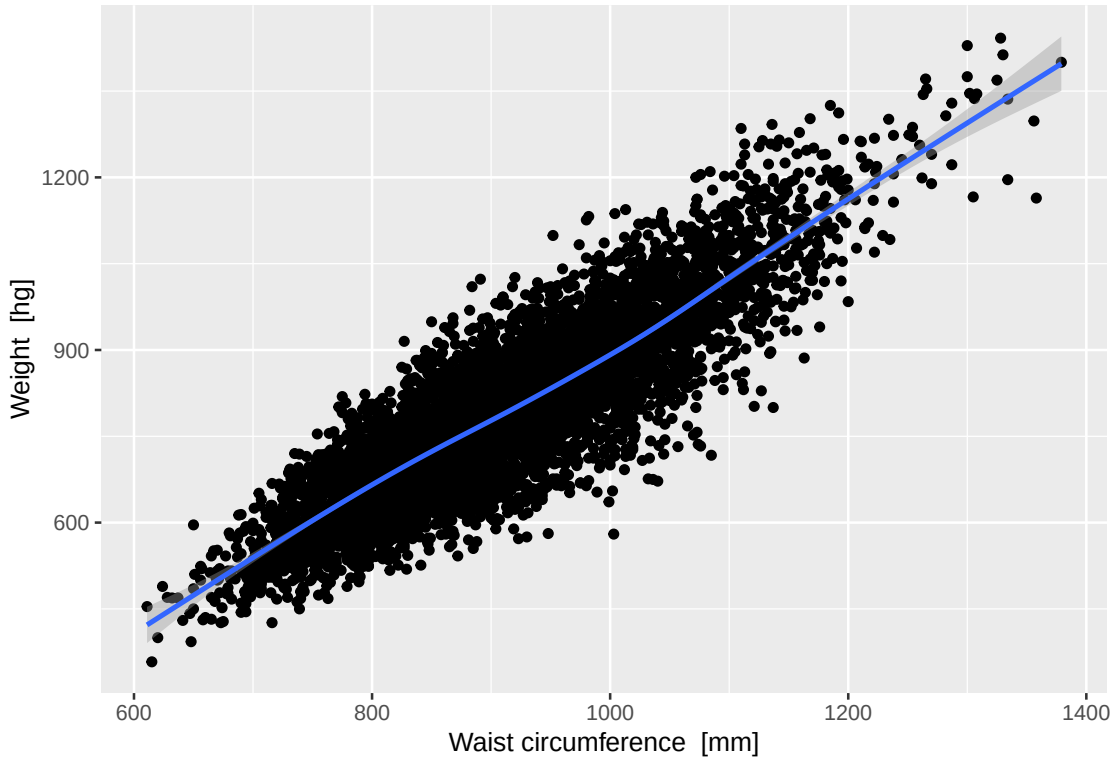


Figure 18: Utjecaj opsega struka na kilažu

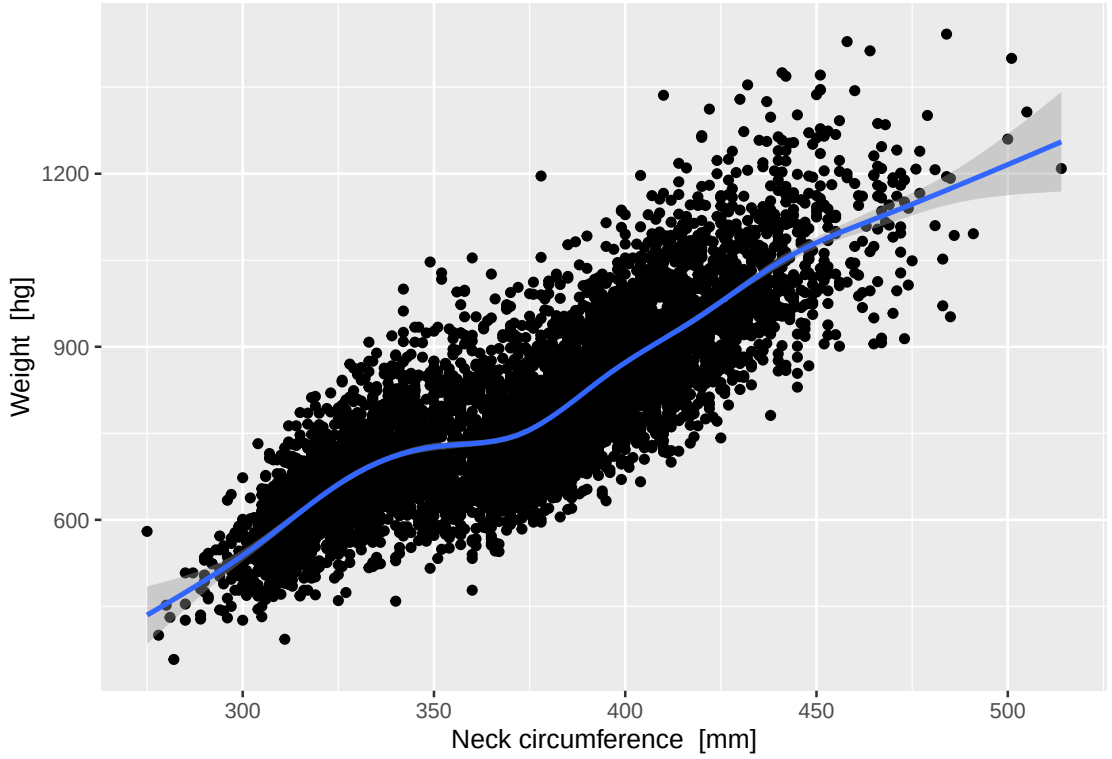


Figure 19: Utjecaj opsega vrata na kilažu

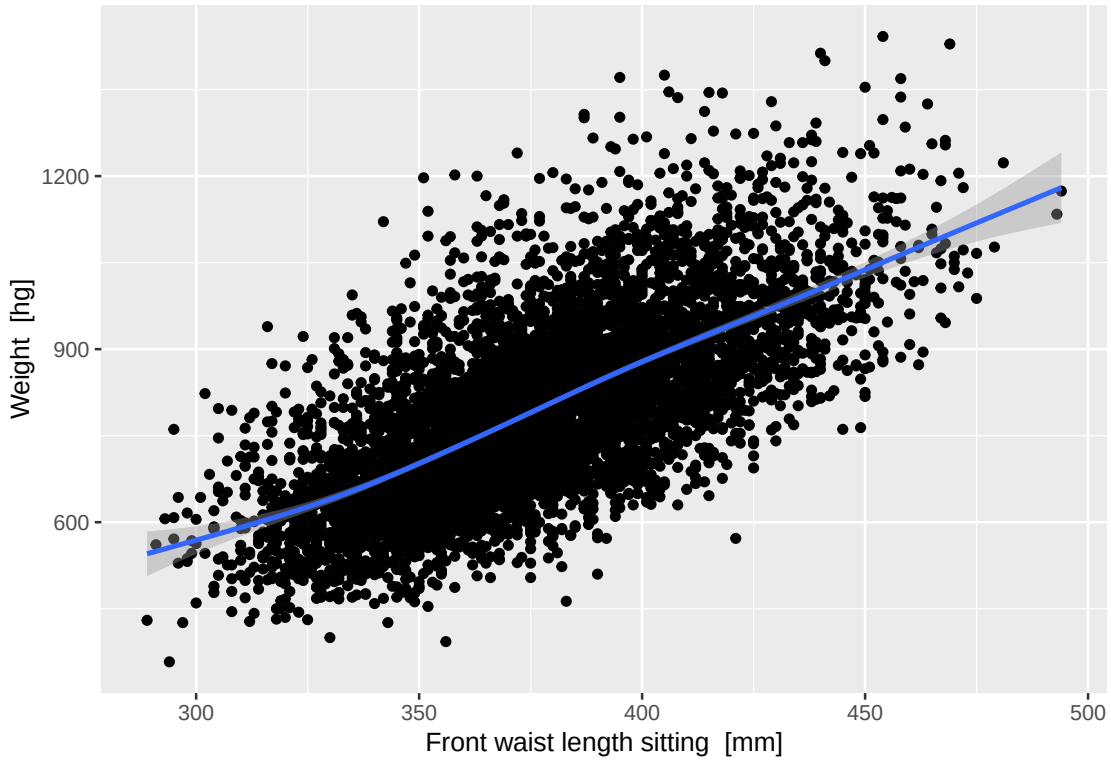


Figure 20: Utjecaj duljine prednjeg struka u sjedecem položaju na kilažu

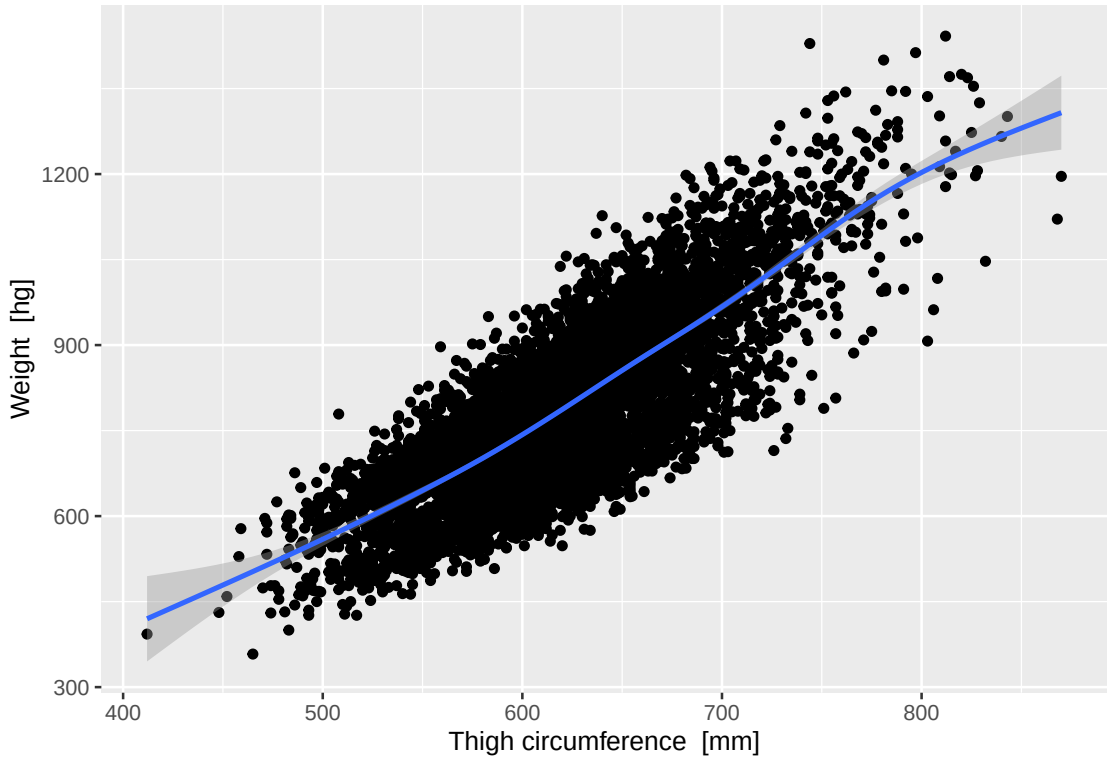


Figure 21: Utjecaj opsega bedra na kilažu

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.

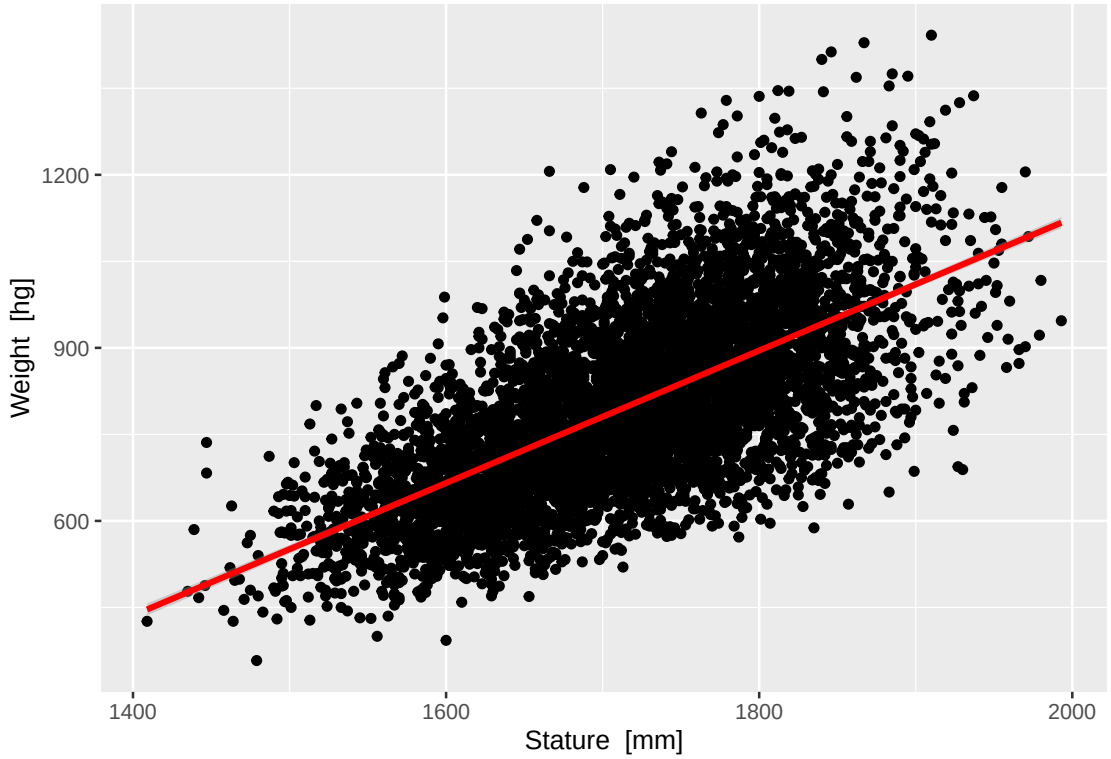


Figure 22: Linearni model kilaže i visine

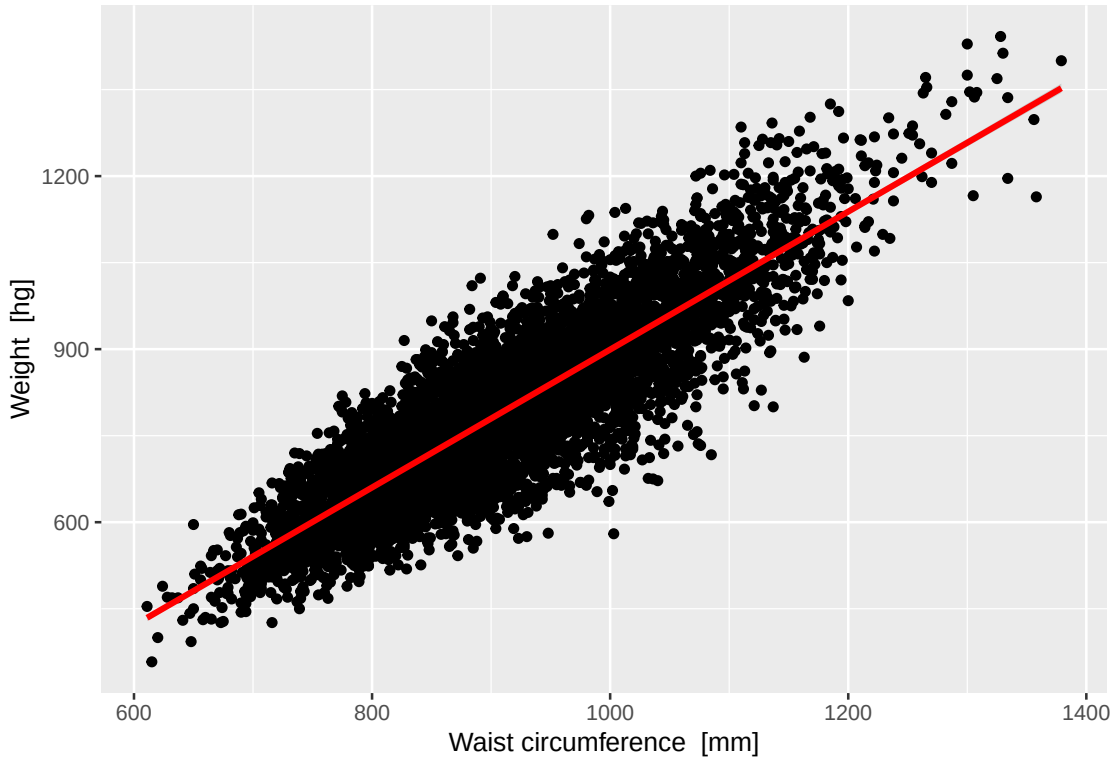


Figure 23: Linearni model kilaže i opsega struka

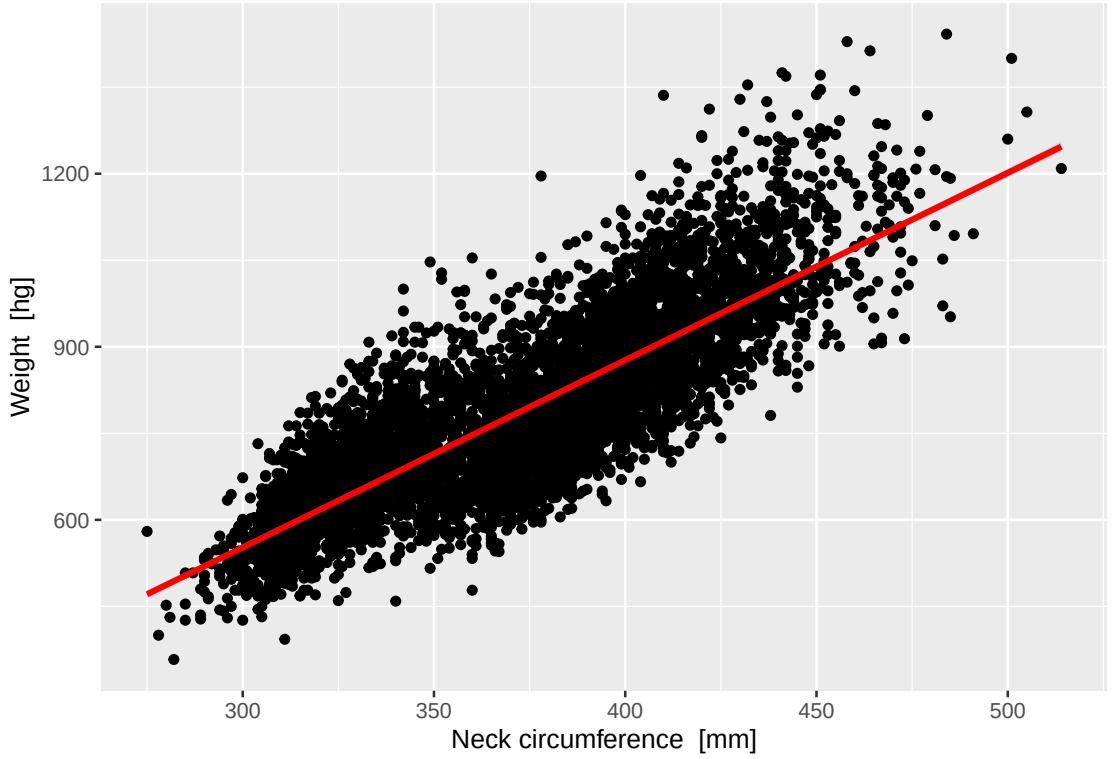


Figure 24: Linearni model kilaže i opsega vrata

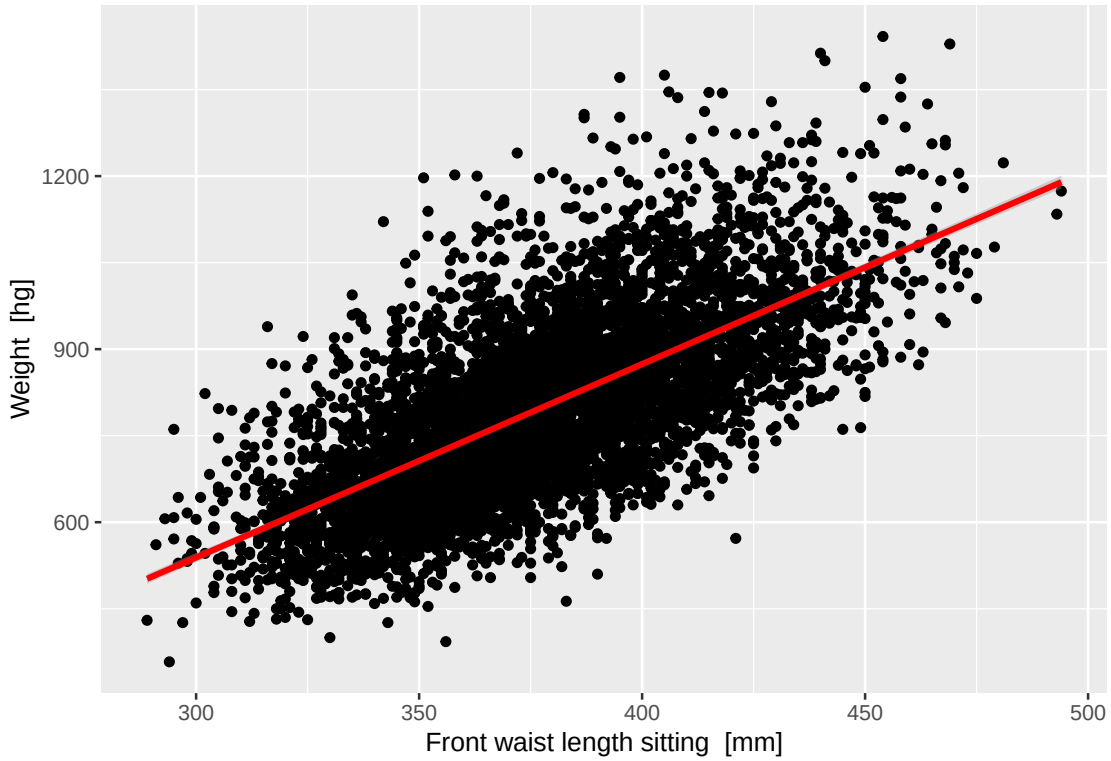


Figure 25: Linearni model kilaže i duljine prednjeg struka u sjedecem položaju

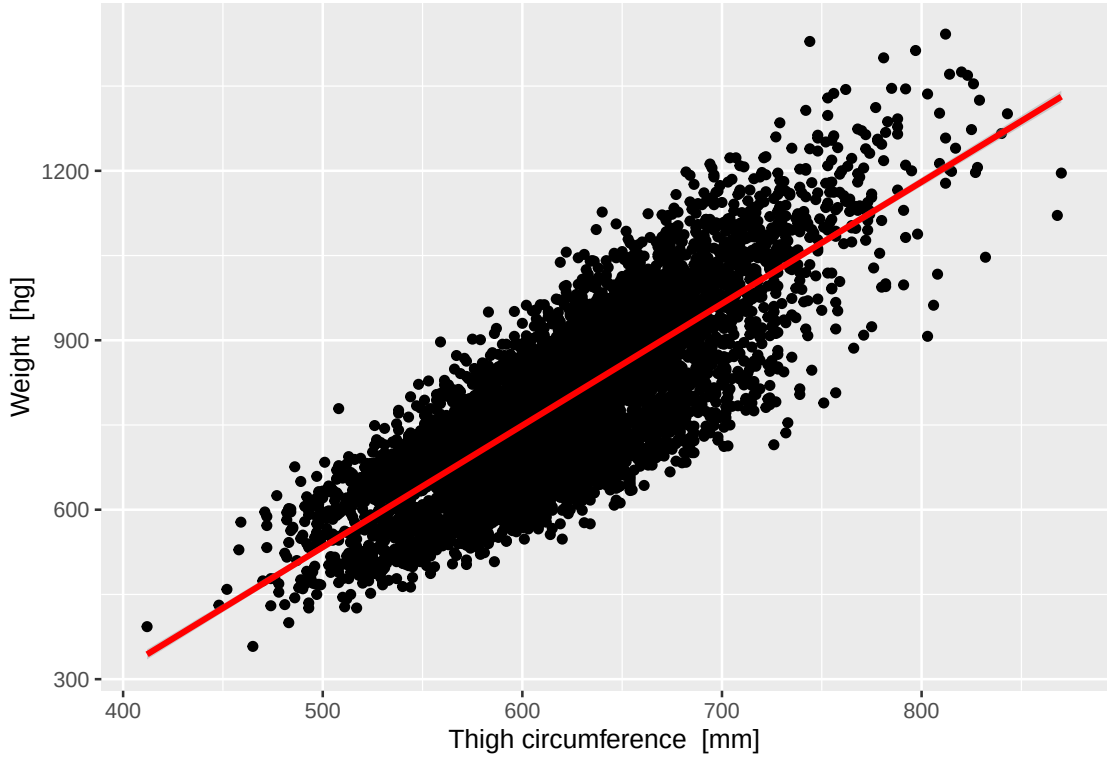


Figure 26: Linearni model kilaže i opsega bedra

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:

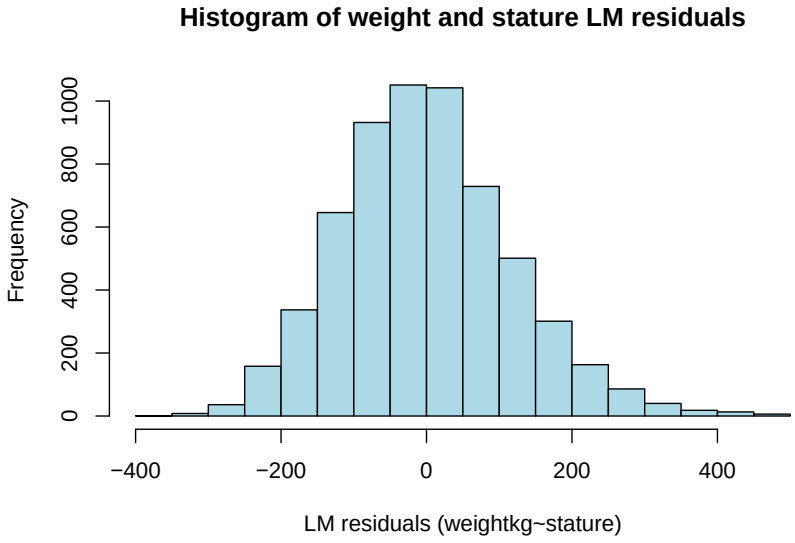


Figure 27: Histogram reziduala modela kilaže i visine

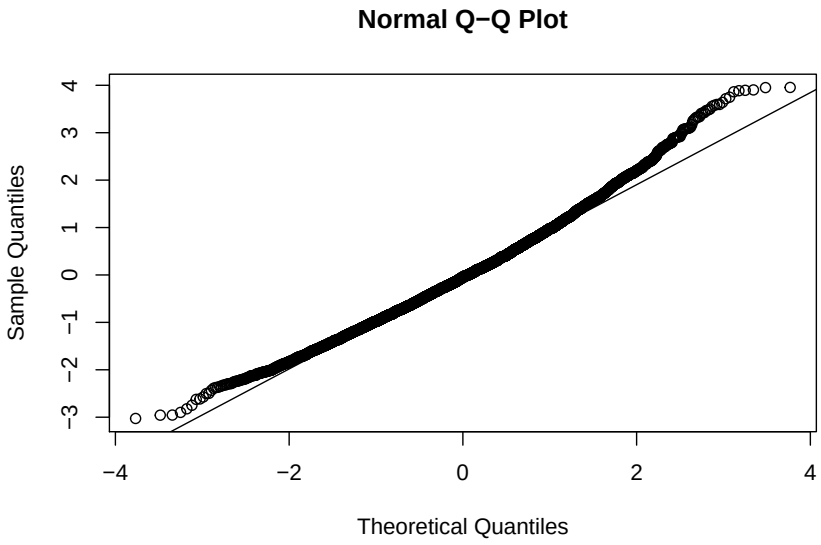


Figure 28: Prikaz q-q plota reziduala modela kilaže i visine s linijom normalne distribucije

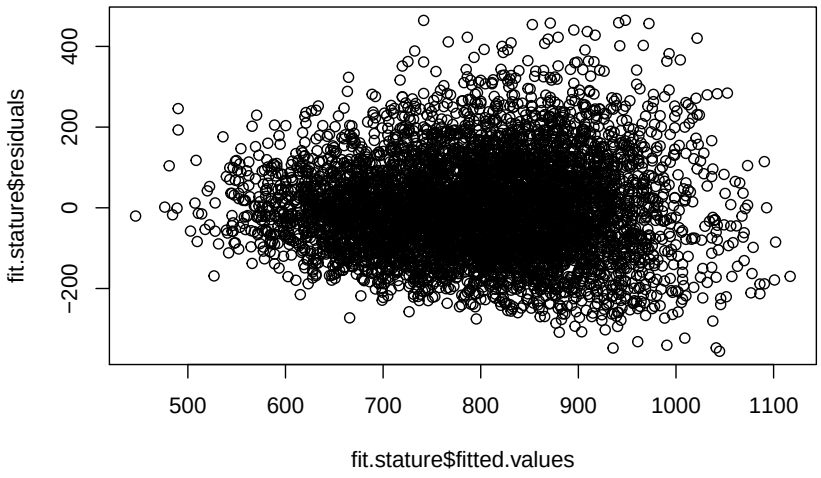


Figure 29: Reziduali modela kilaže i visine u ovisnosti o procjenama modela

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

Opseg struka:

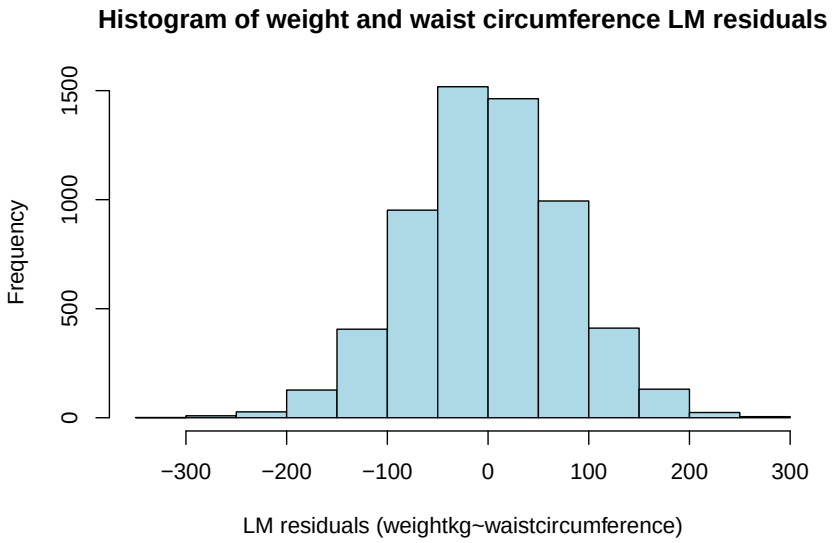


Figure 30: Histogram reziduala modela kilaže i opsega struka

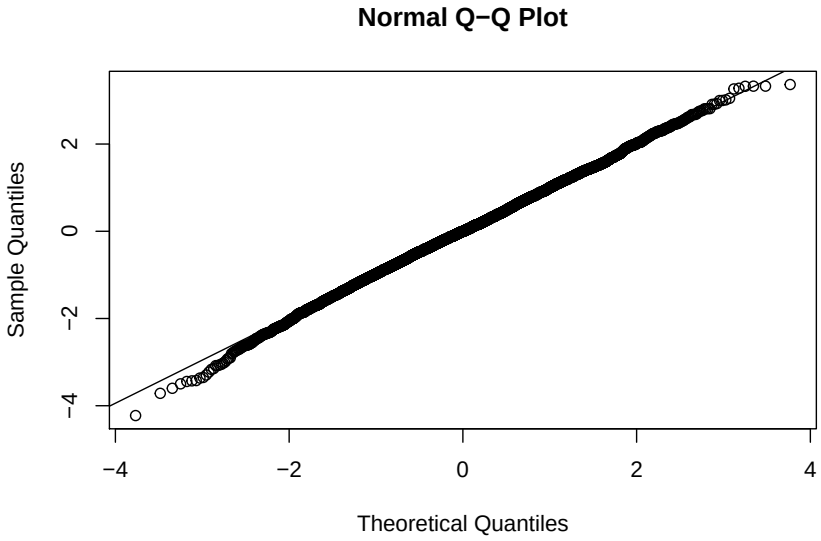


Figure 31: Prikaz q-q plota reziduala modela kilaže i opsega struka s linijom normalne distribucije

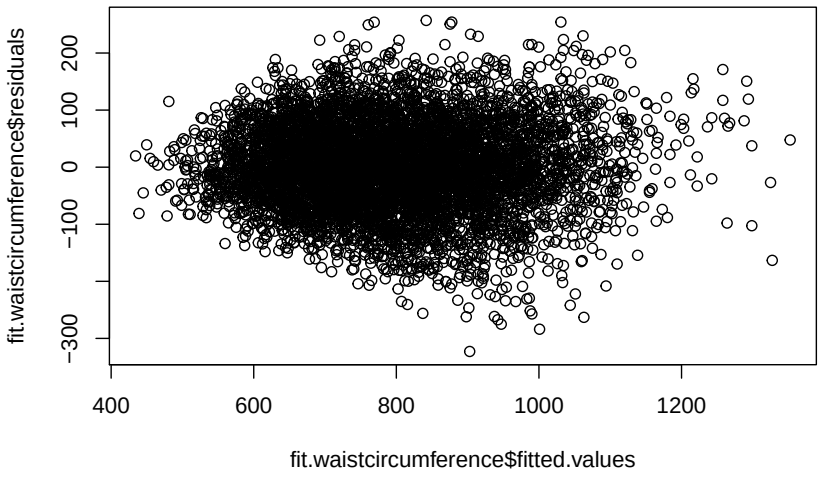


Figure 32: Reziduali modela kilaže i opsega struka u ovisnosti o procjenama modela

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

Opseg vrata:

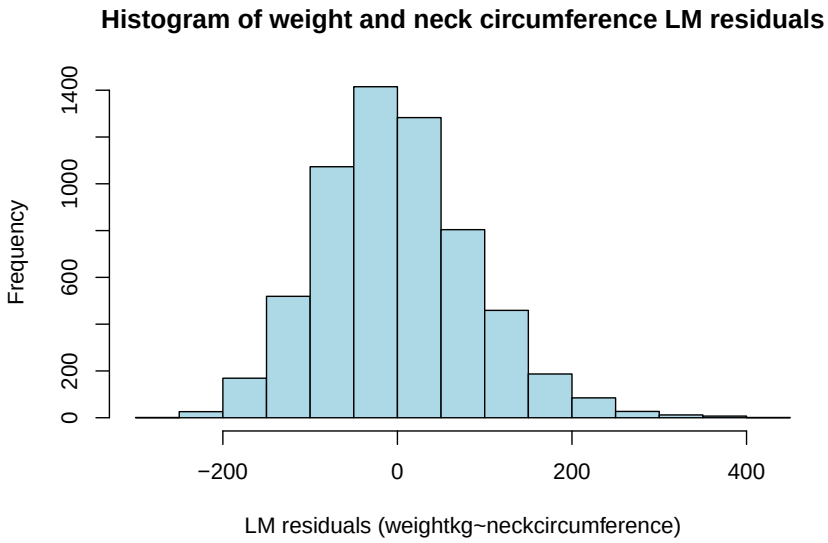


Figure 33: Histogram reziduala modela kilaže i opsega vrata

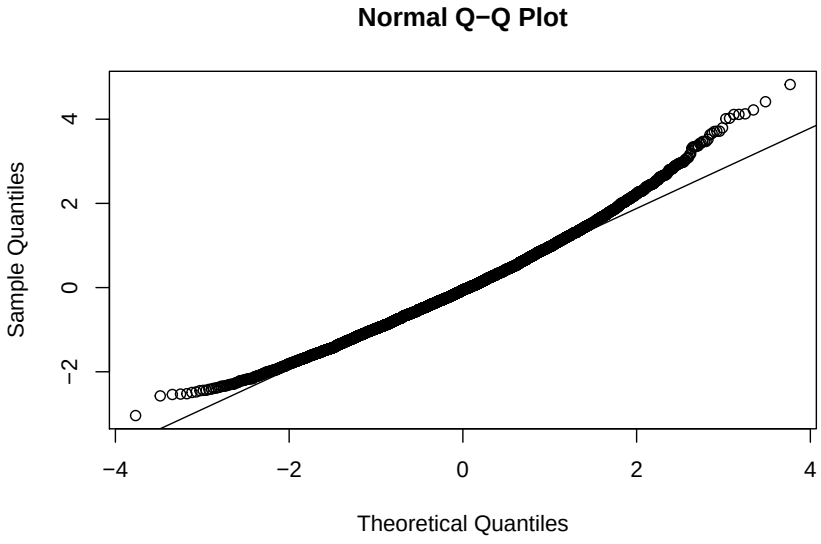


Figure 34: Prikaz q-q plota reziduala modela kilaže i opsega vrata s linijom normalne distribucije

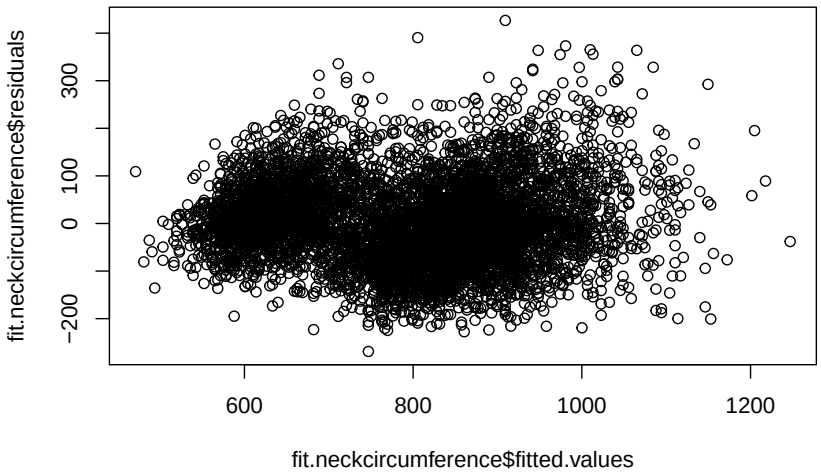


Figure 35: Reziduali modela kilaže i opsega vrata u ovisnosti o procjenama modela

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

Duljina prednjeg struka u sjedećem položaju:

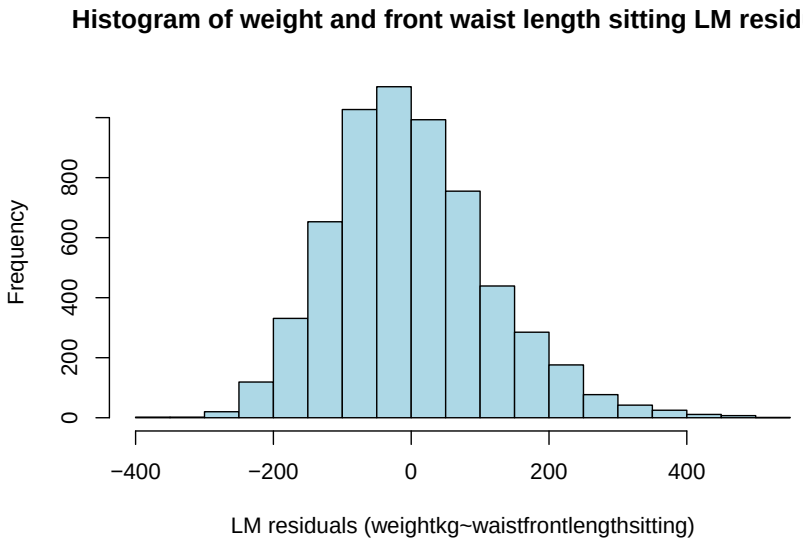


Figure 36: Histogram reziduala modela kilaže i duljine prednjeg struka u sjedecem položaju

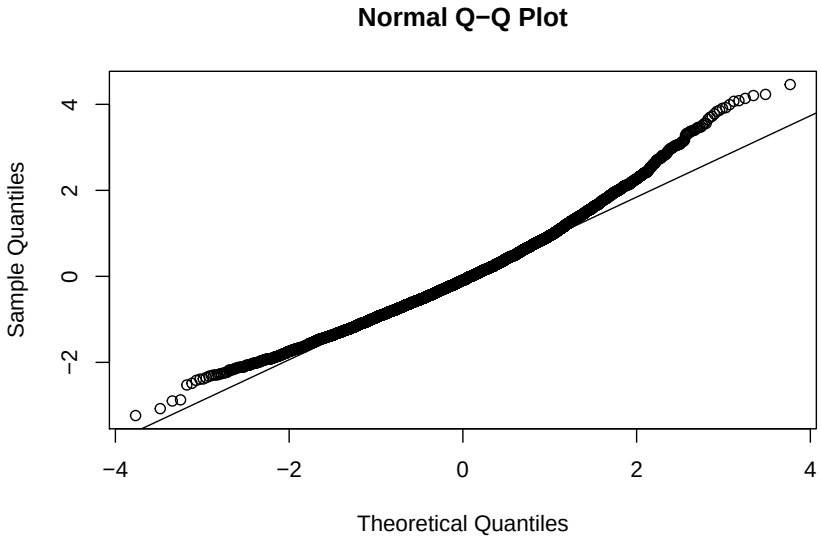


Figure 37: Prikaz q-q plota reziduala modela kilaže i duljine prednjeg struka u sjedecem položaju s linijom normalne distribucije

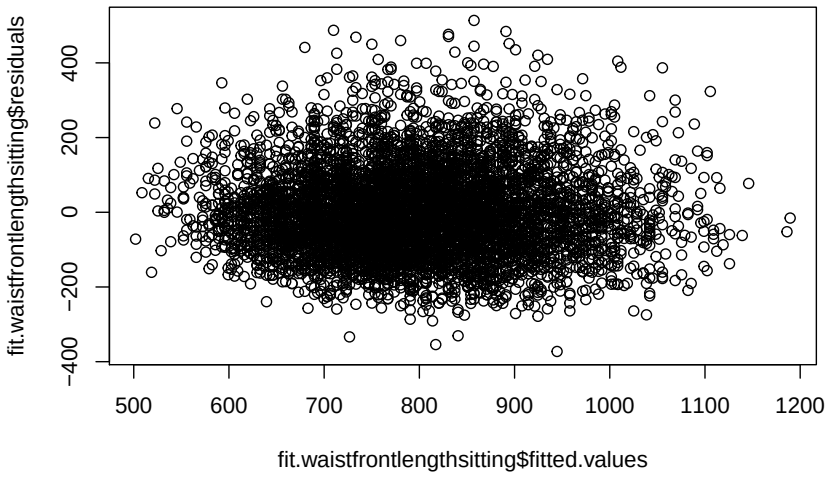


Figure 38: Reziduali modela kilaže i duljine prednjeg struka u sjedecem položaju u ovisnosti o procjenama modela

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

Opseg bedra:

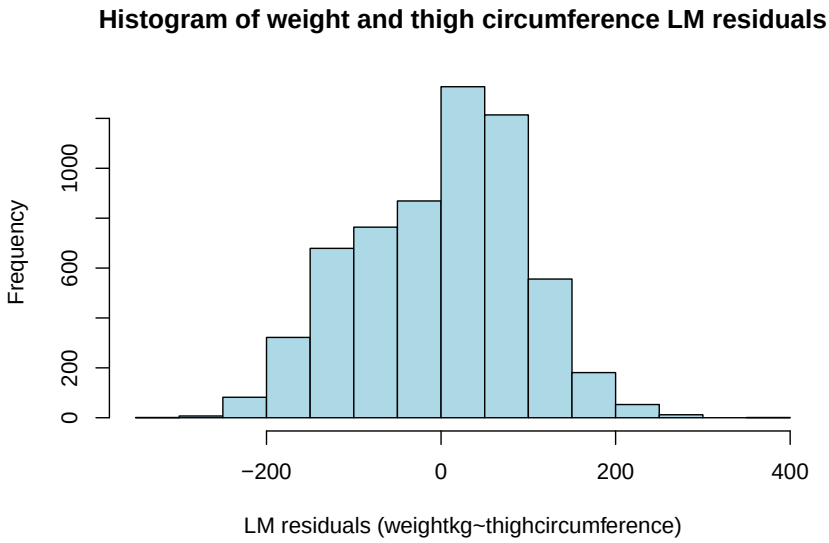


Figure 39: Histogram reziduala modela kilaže i opsega bedra

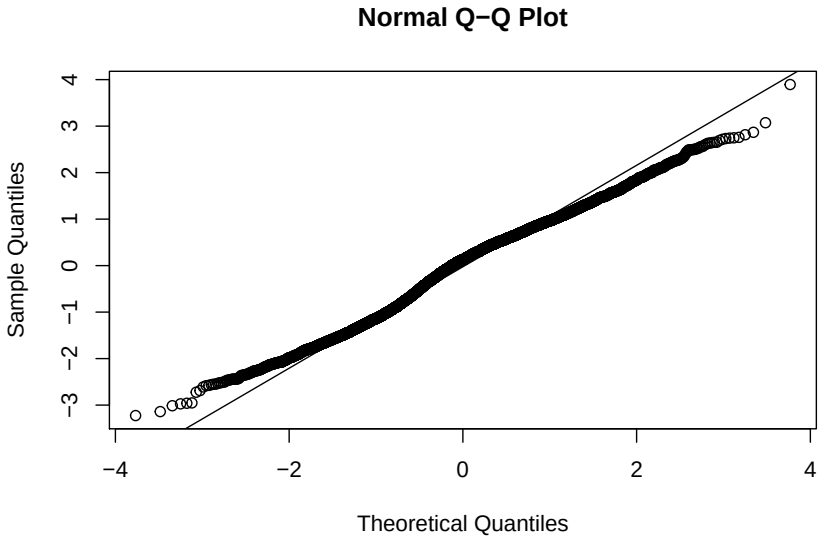


Figure 40: Prikaz q-q plota reziduala modela kilaže i opsega bedra s linijom normalne distribucije

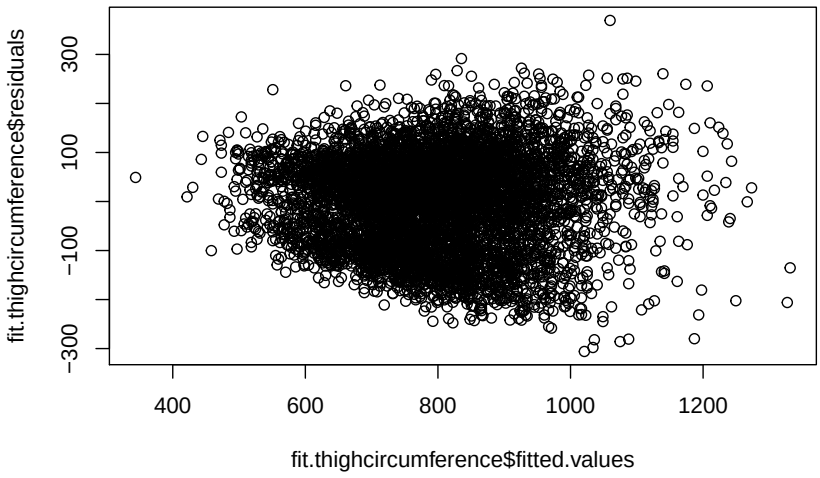


Figure 41: Reziduali modela kilaže i opsega bedra u ovisnosti o procjenama modela

```
##
##  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 p-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 linearnog 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 odabrati 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

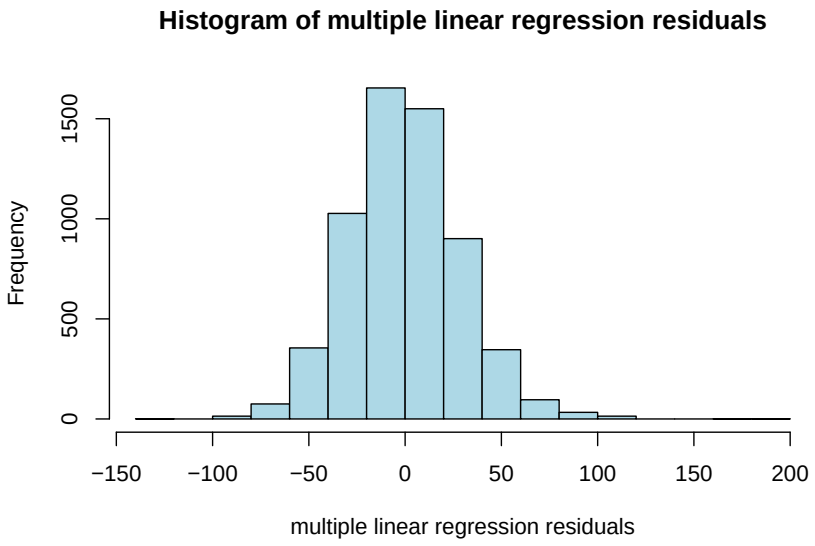


Figure 42: Histogram reziduala modela višestruke regresije

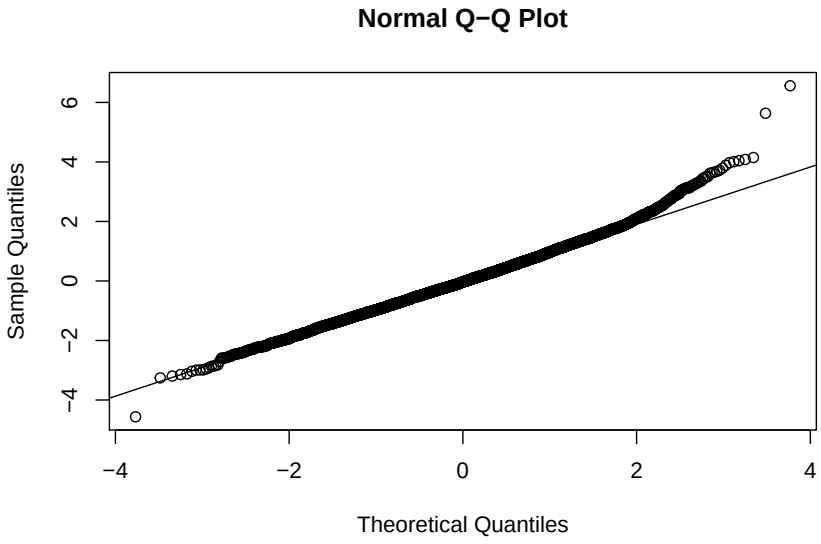


Figure 43: Prikaz q-q plota reziduala model višestruke regresije

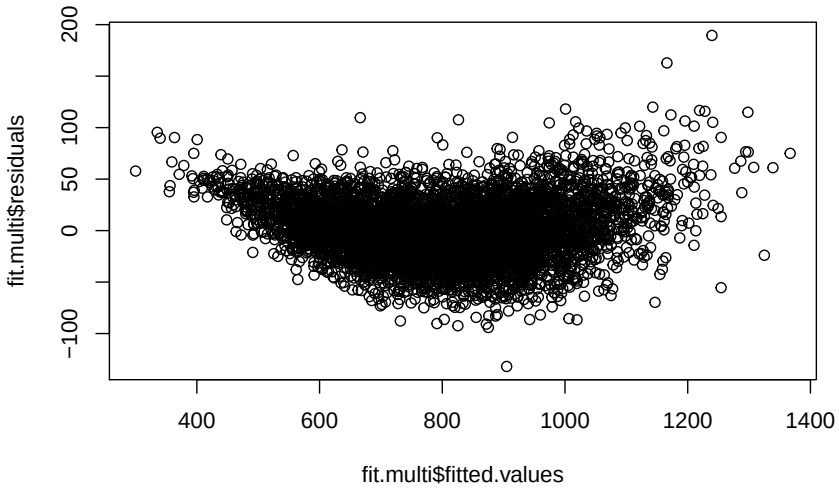


Figure 44: Reziduali modela višestruke regresije

```
##
##  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 kategorijske varijable poput spola vojnika.

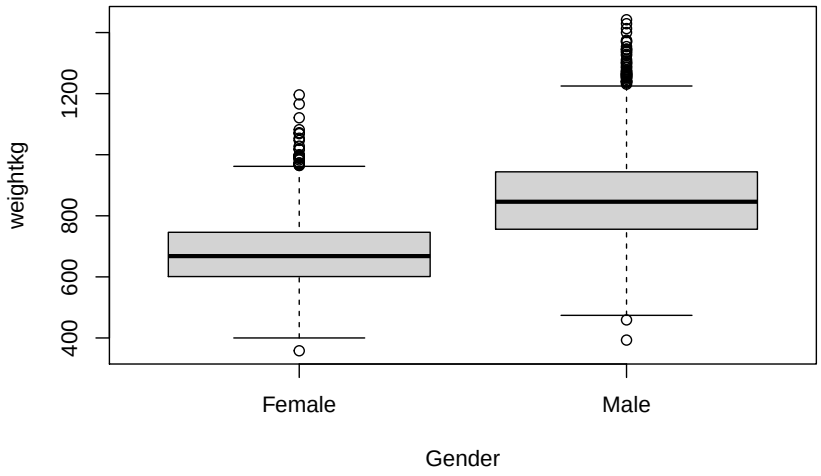


Figure 45: Kvadratni dijagram ovisnosti kilaže i spola

```
## Loading required package: fastDummies
## Warning: package 'fastDummies' was built under R version 4.0.3
##
## 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 -135.624  < 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

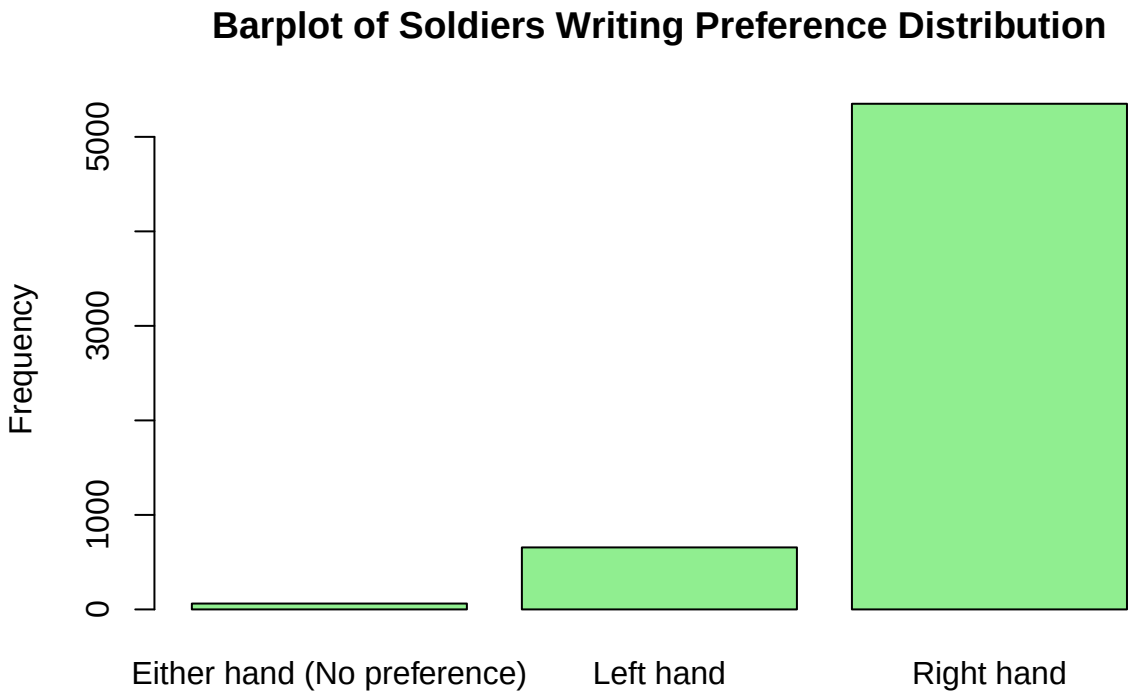
Model koji uz svih pet navednih numeričkih varijabli uključuje i spol vojnika, daje nešto veći R^2. Međutim, ako pogledamo prilagođeni koeficijent determinacije R^2_{adj} koji penalizira dodatne parametre u modelu, vidimo da su isti za oba modela, što znači da varijabla spola nije toliko korisna i radi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

4 Opseg bicepsa

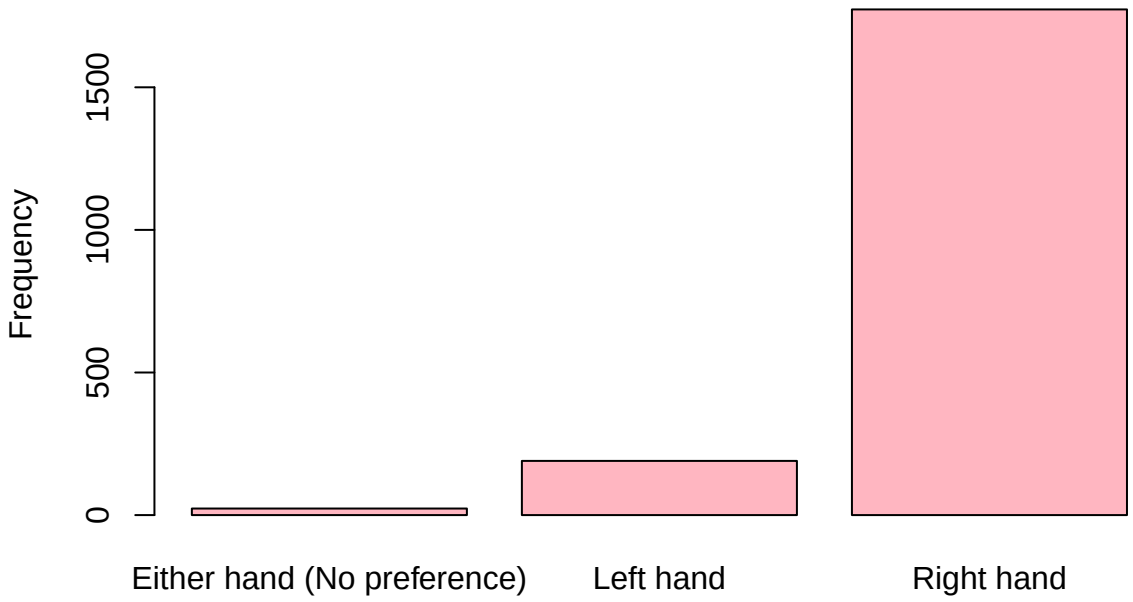
4.1 Prikaz raspodjele vojnika ovisno o preferenciji ruke i spolu

```
##   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
```

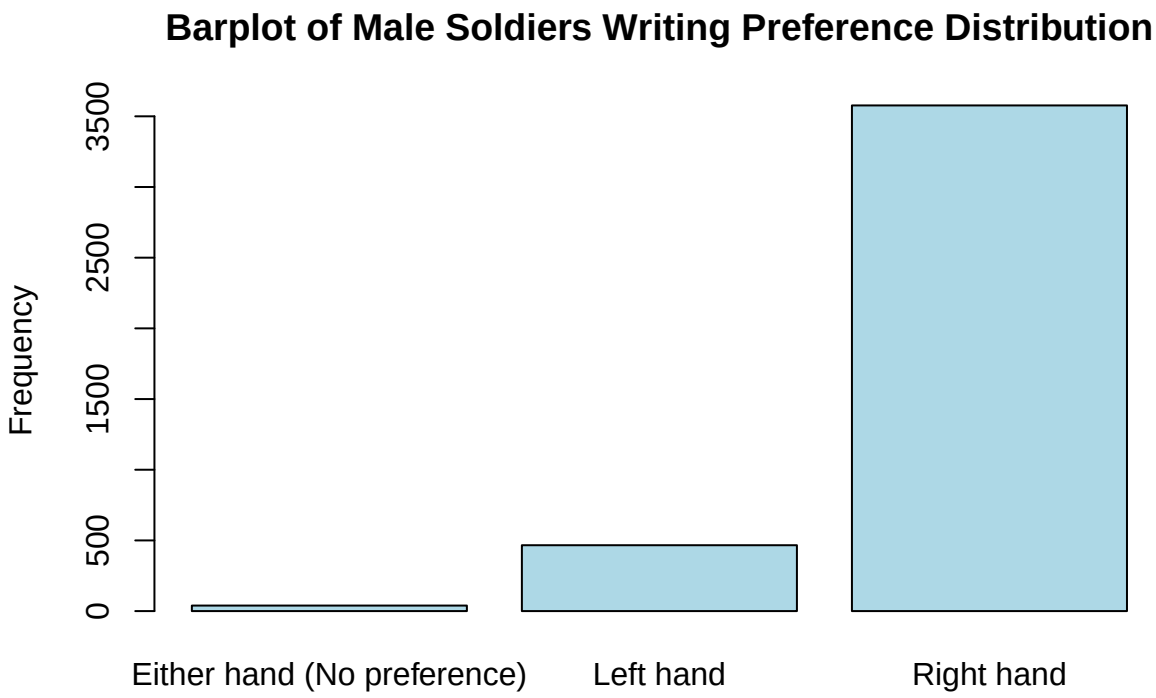
```
##
## Either hand (No preference)                   Left hand
##                          62                         656
##                  Right hand
##                        5350
```




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




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



4.2 Test o dvije proporcije: dva uzorka

4.2.1 Udio ljevaka ovisno o spolu

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

```
##
##  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 daljnjim 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

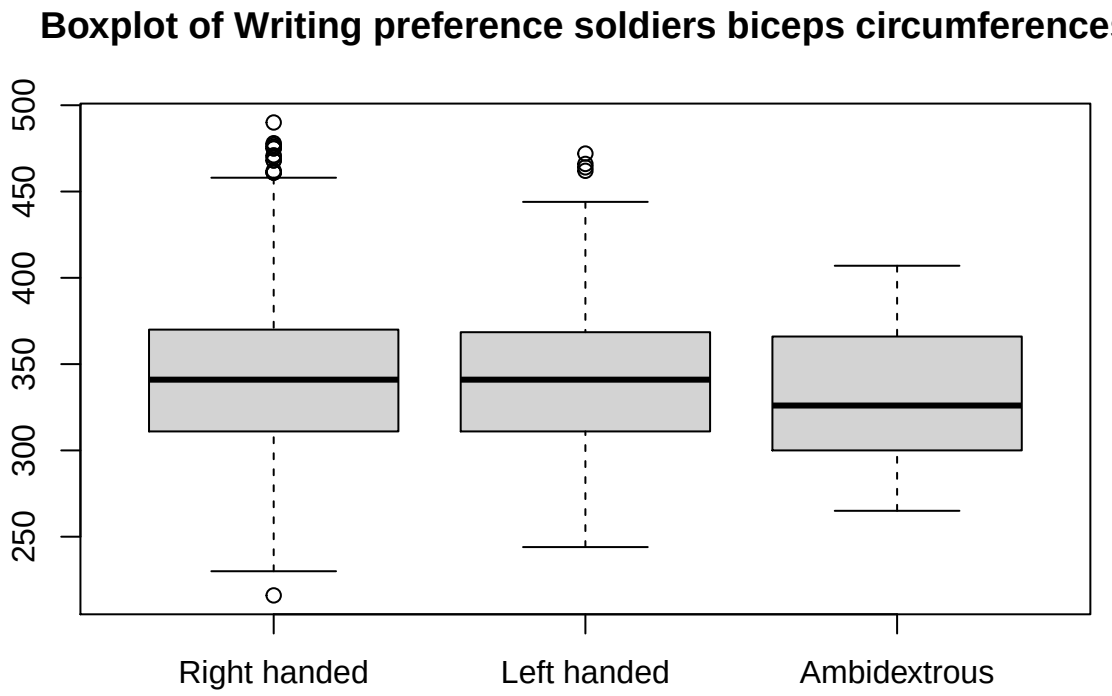
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   216.0   311.0   341.0   340.9   370.0   490.0
```

Prikaz boxplota i prosječnih opsega desnog bicepsa za svaku promatranu 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
```



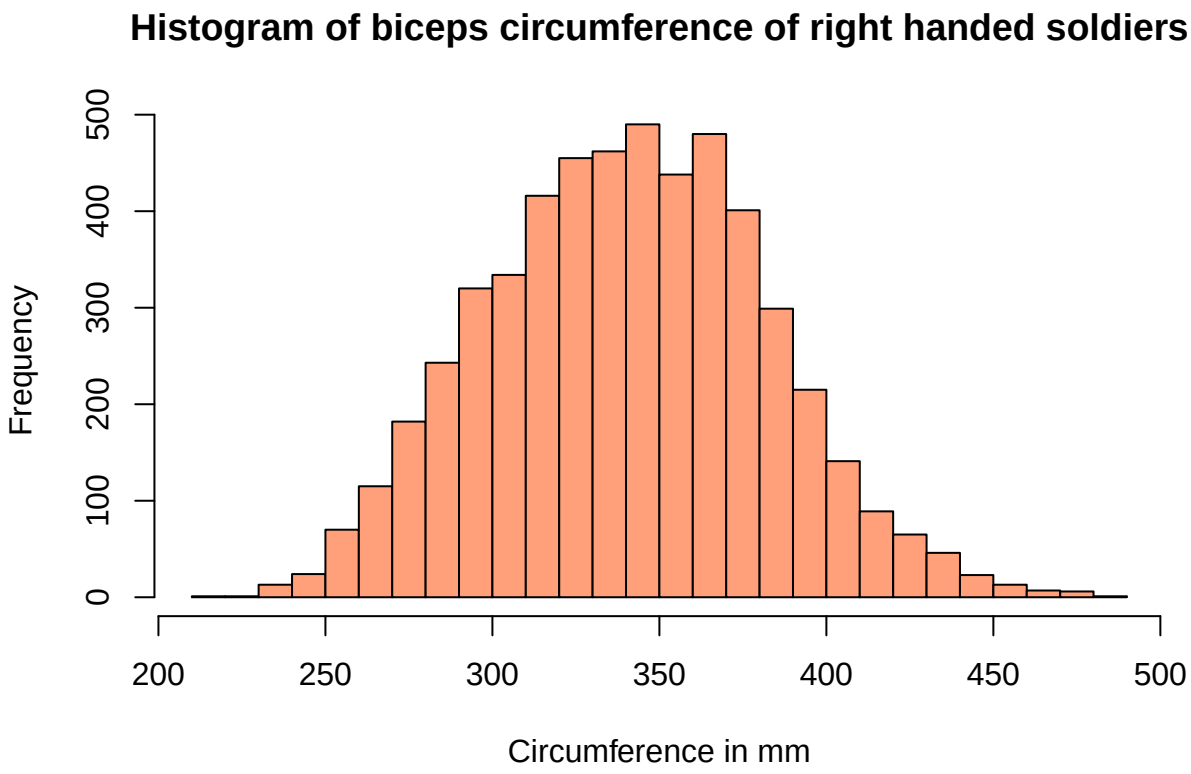
Želimo provjeriti imaju li vojnici dešnjaci veći opseg desnog bicepsa od vojnika ljevaka.Isto tako kasnije ćemo provjeriti imaju li vojnici dešnjaci veći opseg od ambidekstrenih vojnika.

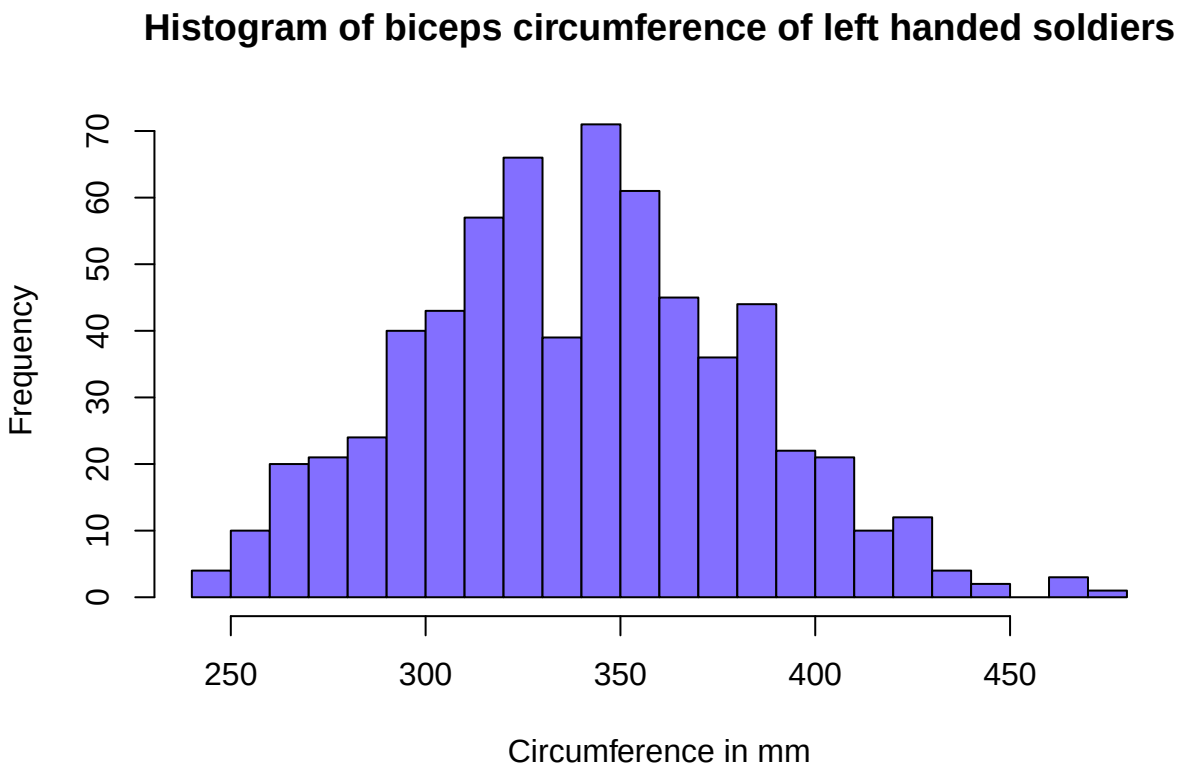
Koristit ćemo neupareni t-test. Kako bi mogli provesti test, moramo najprije provjeriti pretpostavke normalnosti i nezavisnosti uzorka.

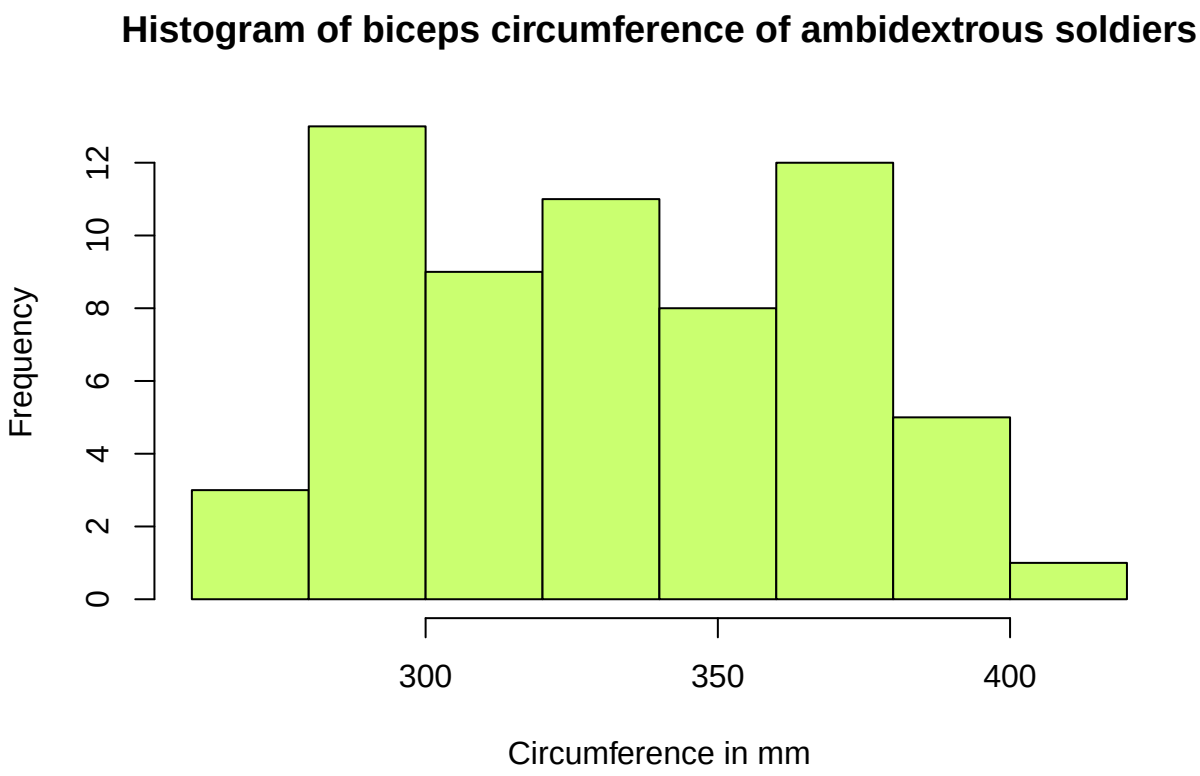
S obzirom kako je odabrana cijela populacija jedne baze iz koje promatramo ljevake i dešnjake i jer baza broji puno vise od 30 članova smatramo kako su promatrani uzorci nezavisni. Iz istih razloga su nezavisni za dešnjake i ambidekstrene vojnike.

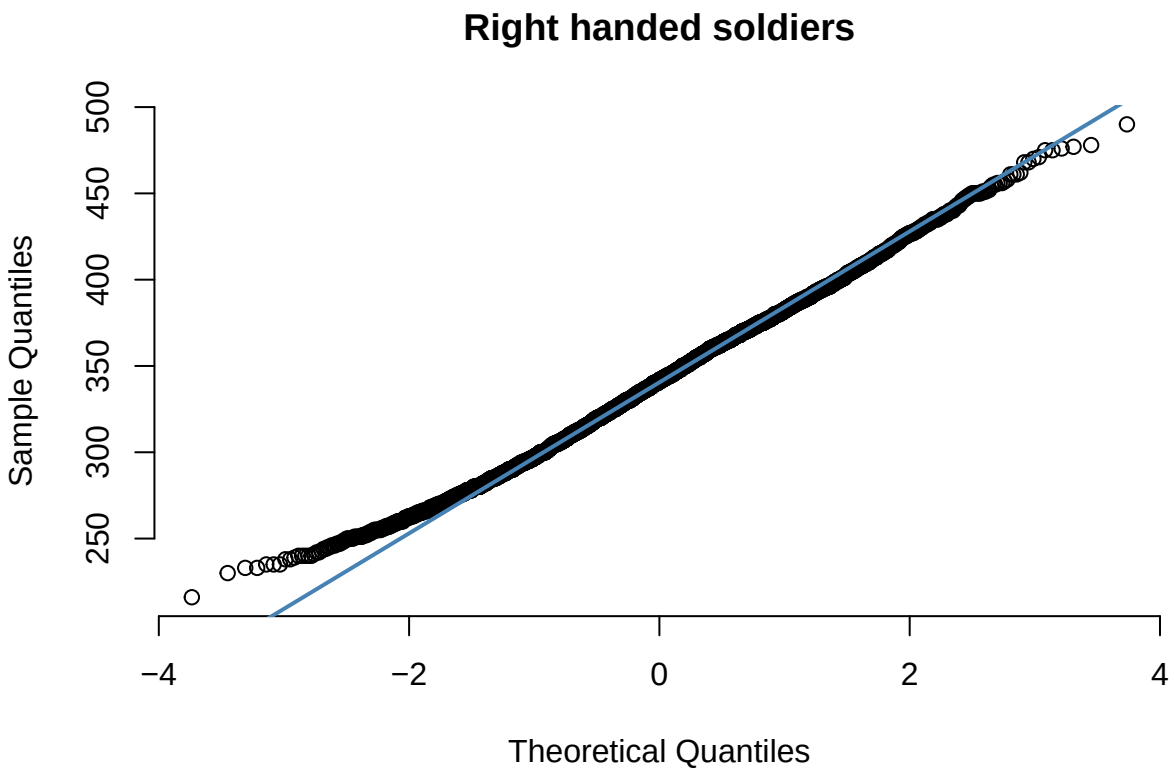
4.3.1 Opseg desnog bicepsa dešnjaka i ljevaka

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

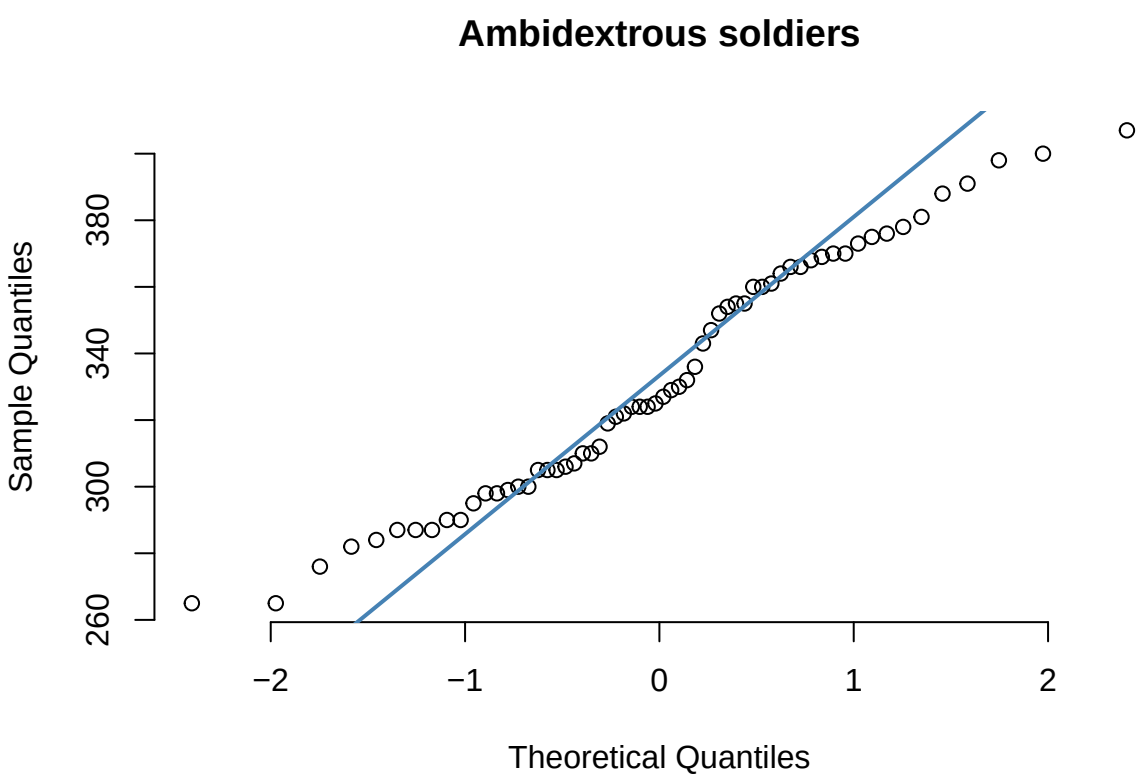

Histogram of biceps circumference of right handed soldiers
Frequency
0
100
200
300
400
500
200
250
300
350
400
450
500
Circumference in mm


Histogram of biceps circumference of left handed soldiers
Frequency
0
10
20
30
40
50
60
70
250
300
350
400
450
Circumference in mm


Histogram of biceps circumference of ambidextrous soldiers
Frequency
0
2
4
6
8
10
12
300
350
400
Circumference in mm


Right handed soldiers
Sample Quantiles
250
300
350
400
450
500
−4
−2
0
2
4
Theoretical Quantiles


Left handed soldiers
Sample Quantiles
250
300
350
400
450
−3
−2
−1
0
1
2
3
Theoretical Quantiles




```
##
##  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

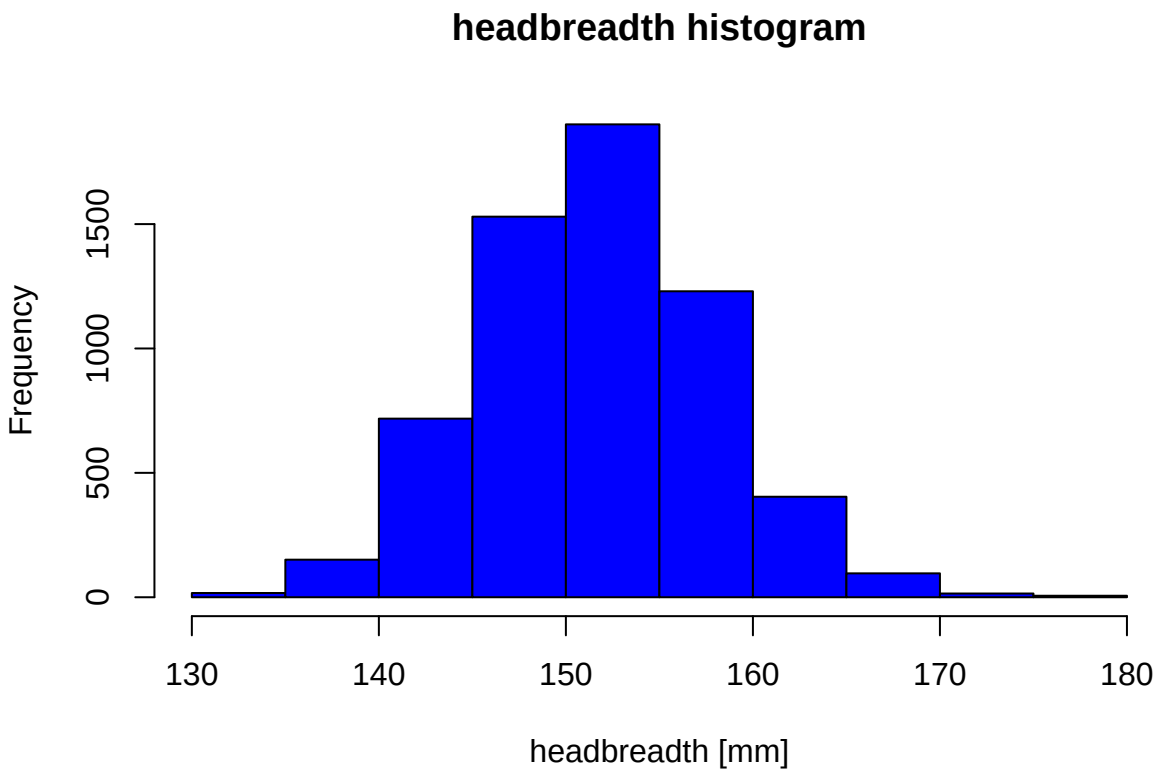
```
##   headbreadth    headcircumference   headlength
##  Min.   :131.0   Min.   :500       Min.   :168.0
##  1st Qu.:148.0   1st Qu.:557       1st Qu.:191.0
##  Median :152.0   Median :570       Median :197.0
##  Mean   :152.2   Mean   :570       Mean   :196.3
##  3rd Qu.:156.0   3rd Qu.:582      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

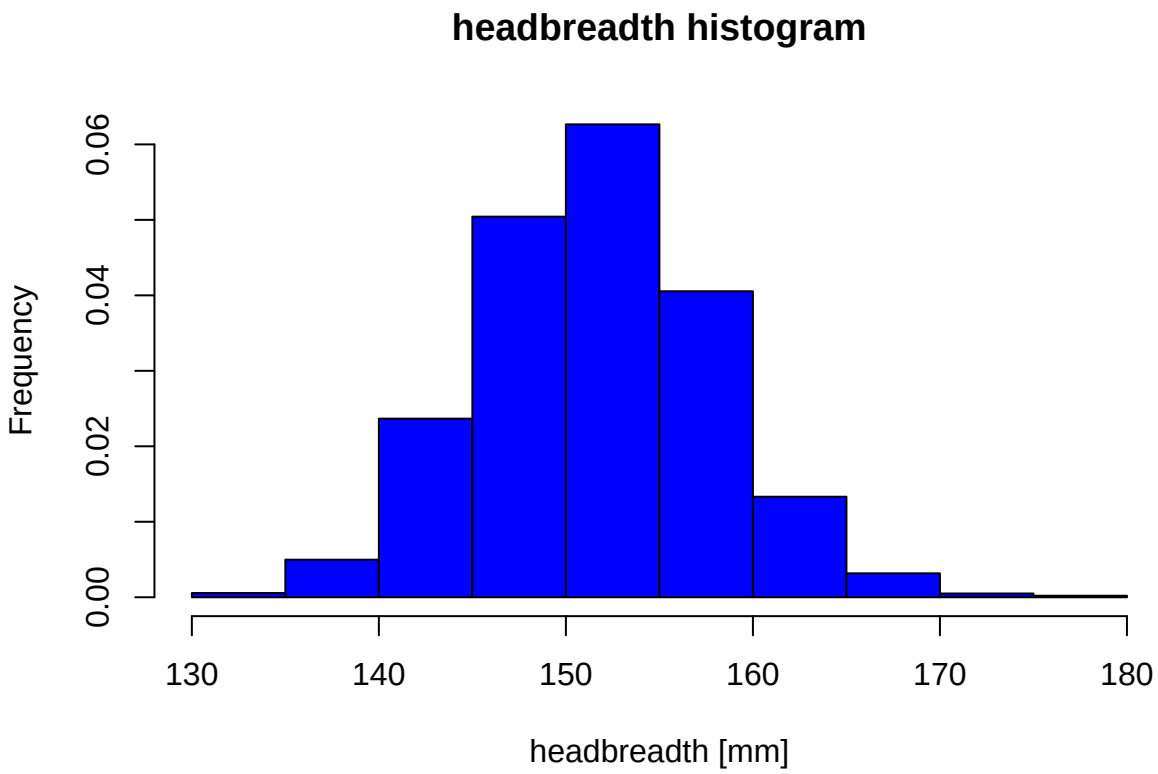
5.3 Mjere centralne tendencije i mjere rasipanja

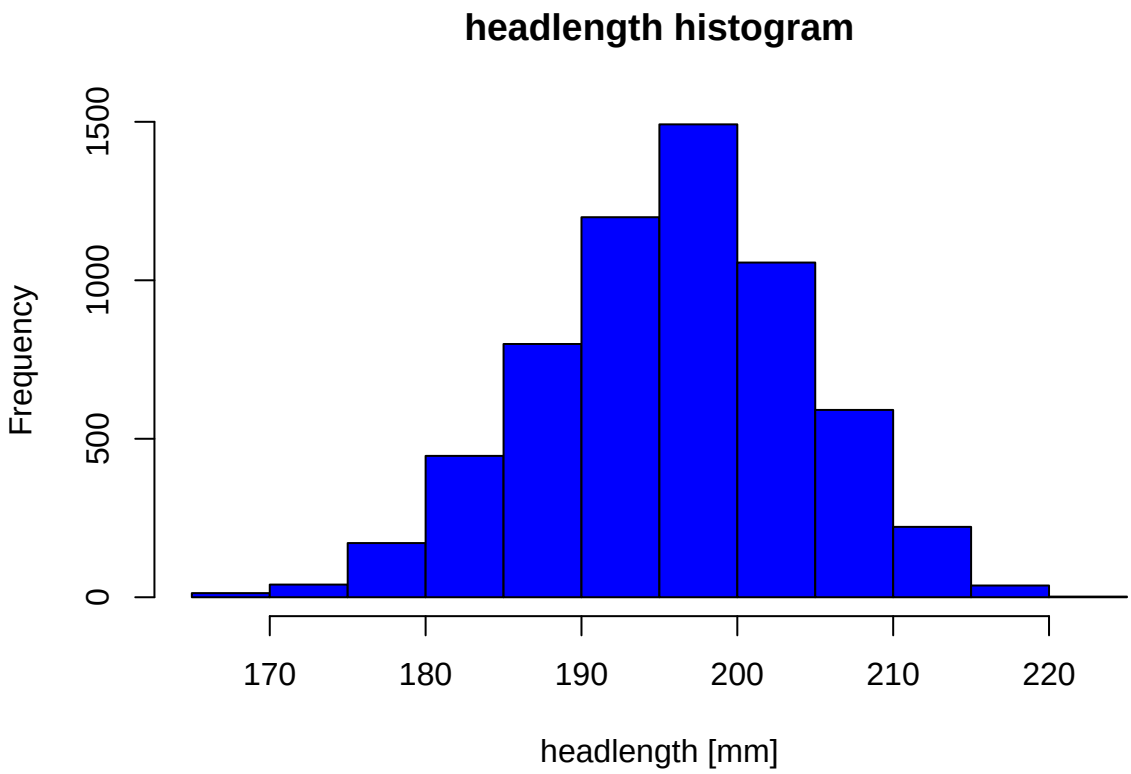
```
##                       Vrijednost headbreadth headcircumference headlength
## 1                            Mod  153.000000         577.00000 200.000000
## 2  Podrezana srednja vrijednost  152.140857         569.96952 196.612301
## 3                           Rang   49.000000         135.00000  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
```

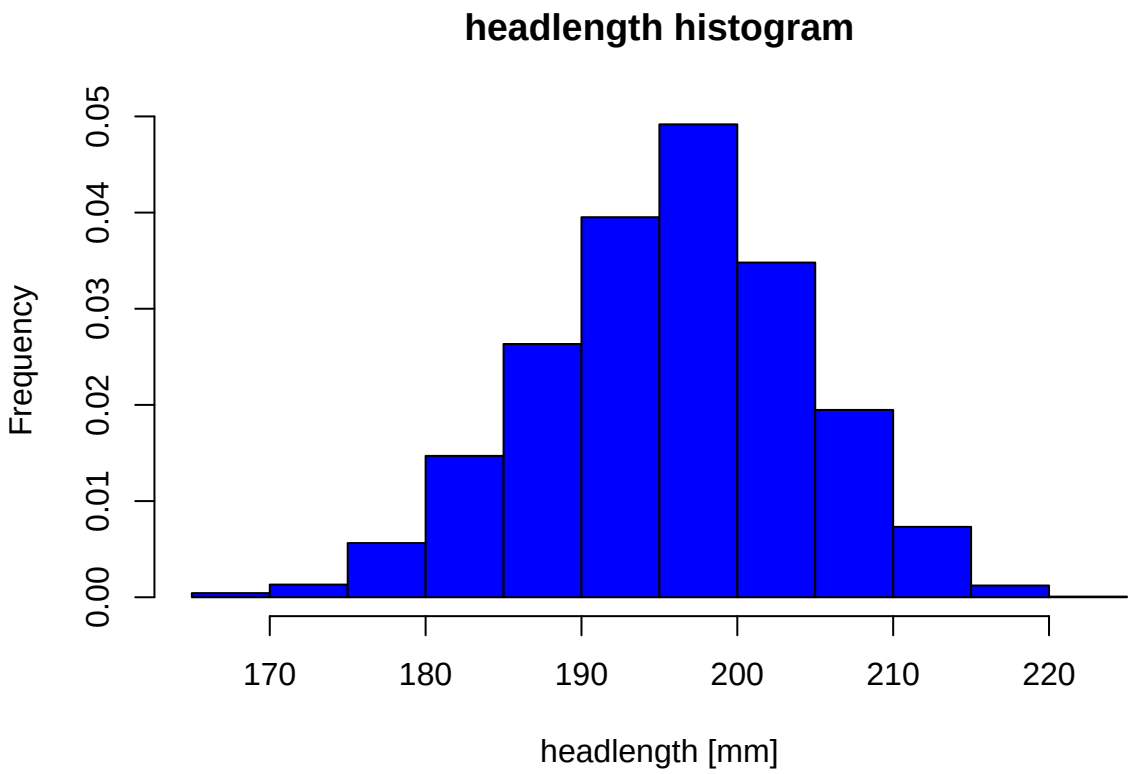
5.4 Vizualizacija podataka

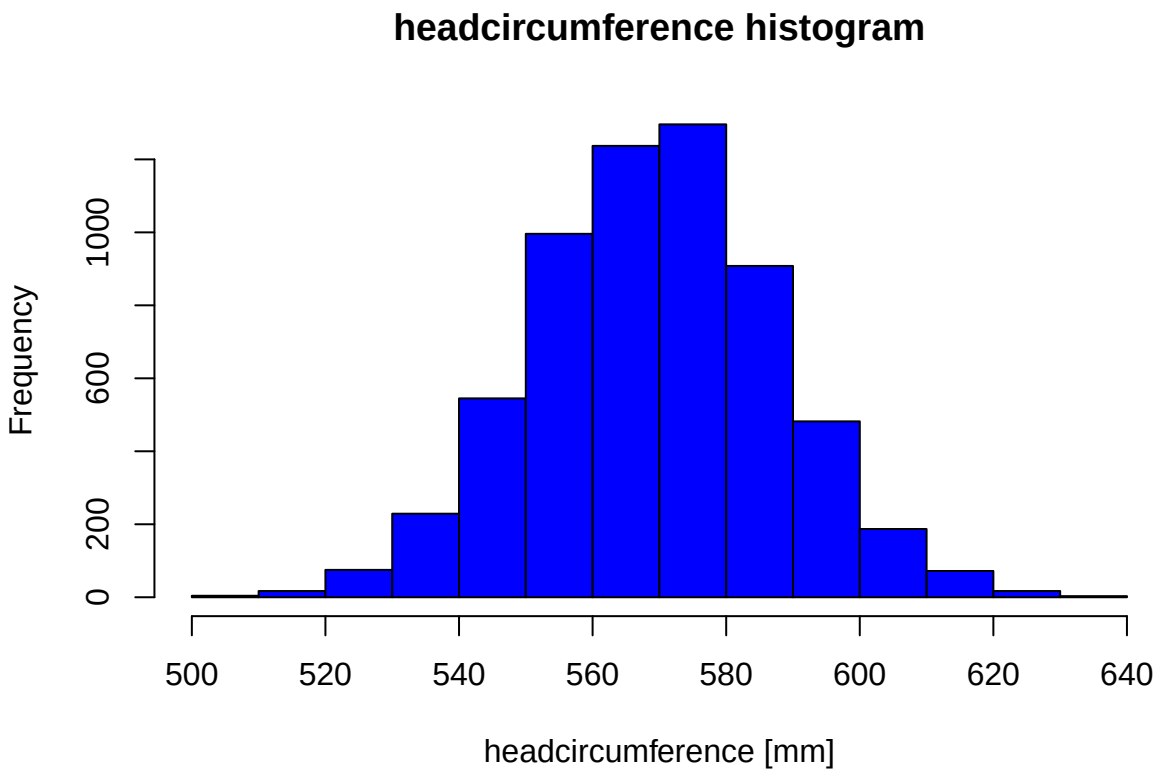
- Histogram - pokazuje oblik distribucije i gustoću podataka, a zasnovan je na grupiranju varijabli u razrede
- Dijagram raspršenja (scatter plot) - jedan je od najvažnijih načina prikaza bivarijantnih podataka, te daje informaciju o povezanosti varijabli.

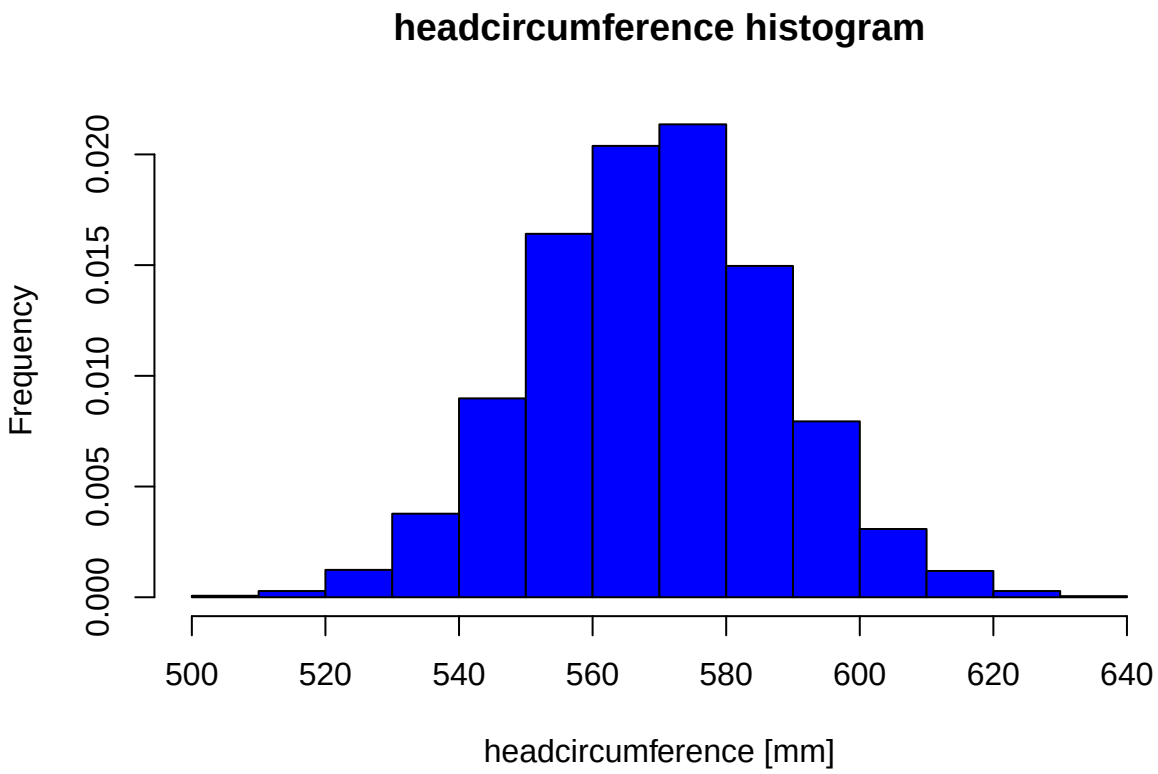

headbreadth histogram
Frequency
0
500
1000
1500
130
140
150
160
170
180
headbreadth [mm]


headbreadth histogram
Frequency
0.00
0.02
0.04
0.06
130
140
150
160
170
180
headbreadth [mm]


headlength histogram
Frequency
0
500
1000
1500
170
180
190
200
210
220
headlength [mm]


headlength histogram
Frequency
0.00
0.01
0.02
0.03
0.04
0.05
170
180
190
200
210
220
headlength [mm]


headcircumference histogram
Frequency
0
200
600
1000
500
520
540
560
580
600
620
640
headcircumference [mm]



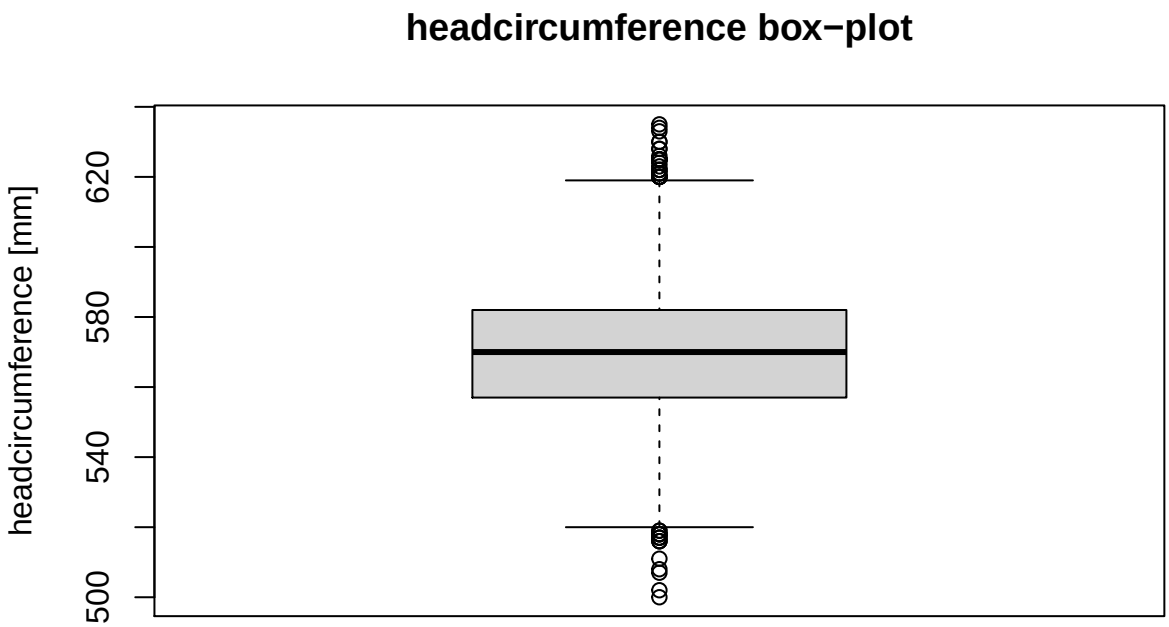
5.5 Prikaz veličina kaciga

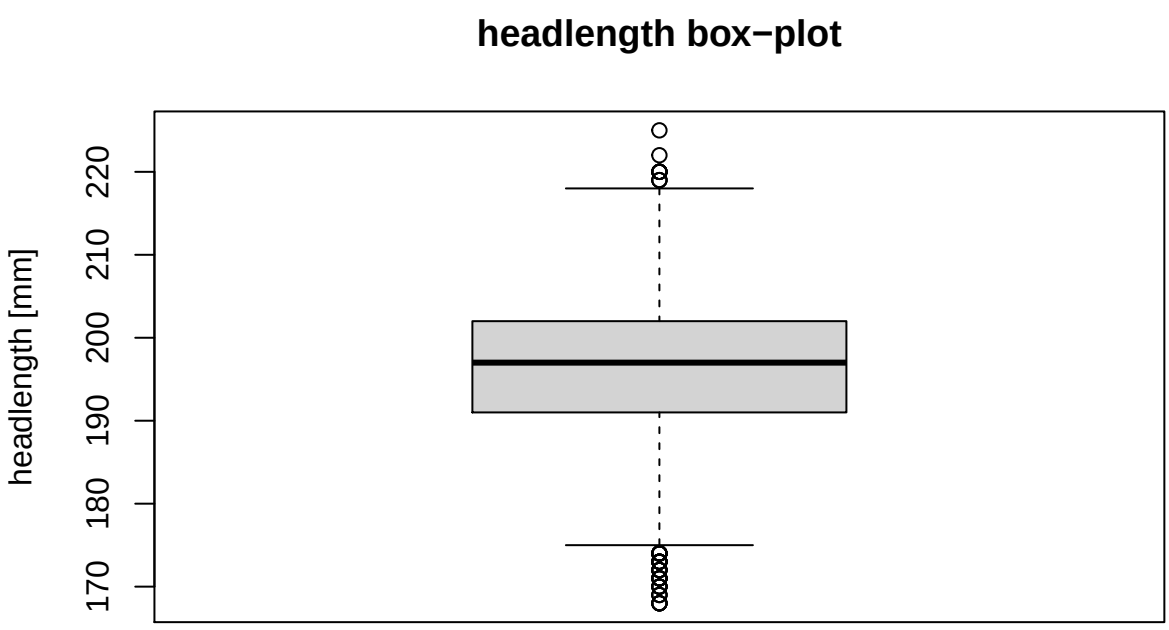
Daljna statistička analiza podataka će se referencirati na ove veličine kaciga.

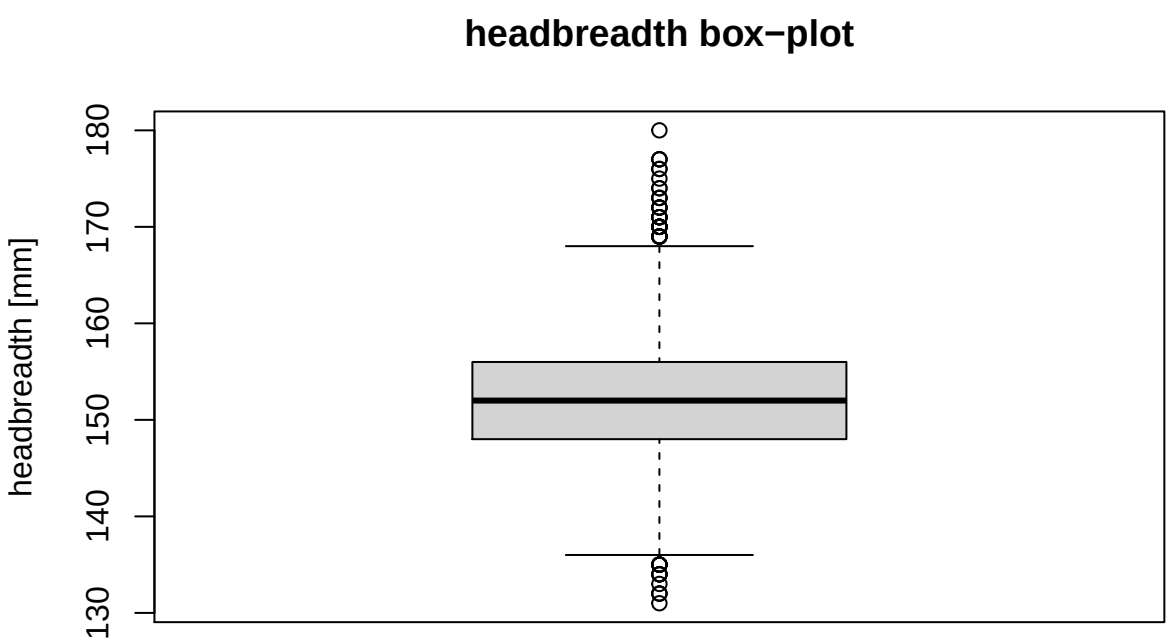
Table 1. Head/Shell Sizing.

Helmet Shell	Head Length	Head Breadth (Width)	Head Circumference
Small Helmet Shell	Up to 7 ¼ inches (184 mm)	Up to 6 ½ inches (162 mm)	Up to 21 ¼ inches (538 mm)
Medium Helmet Shell	From 7 ¼ inches (184 mm) up to 7 ¾ inches (198 mm)	Up to 6 ½ inches (162 mm)	From 21 ¼ inches (538 mm) up to 22 ½ inches (573 mm)
Large Helmet Shell	From 7 ¾ inches (198 mm) up to 8 ¼ inches (210 mm)	Up to 6 ½ inches (162 mm)	From 22 ½ inches (573 mm) up to 23 ½ inches (597 mm)
Extra-Large Helmet Shell	8 ¼ inches (210 mm) and over	6 ½ inches (162 mm) and over	23 ½ inches (597 mm) and over

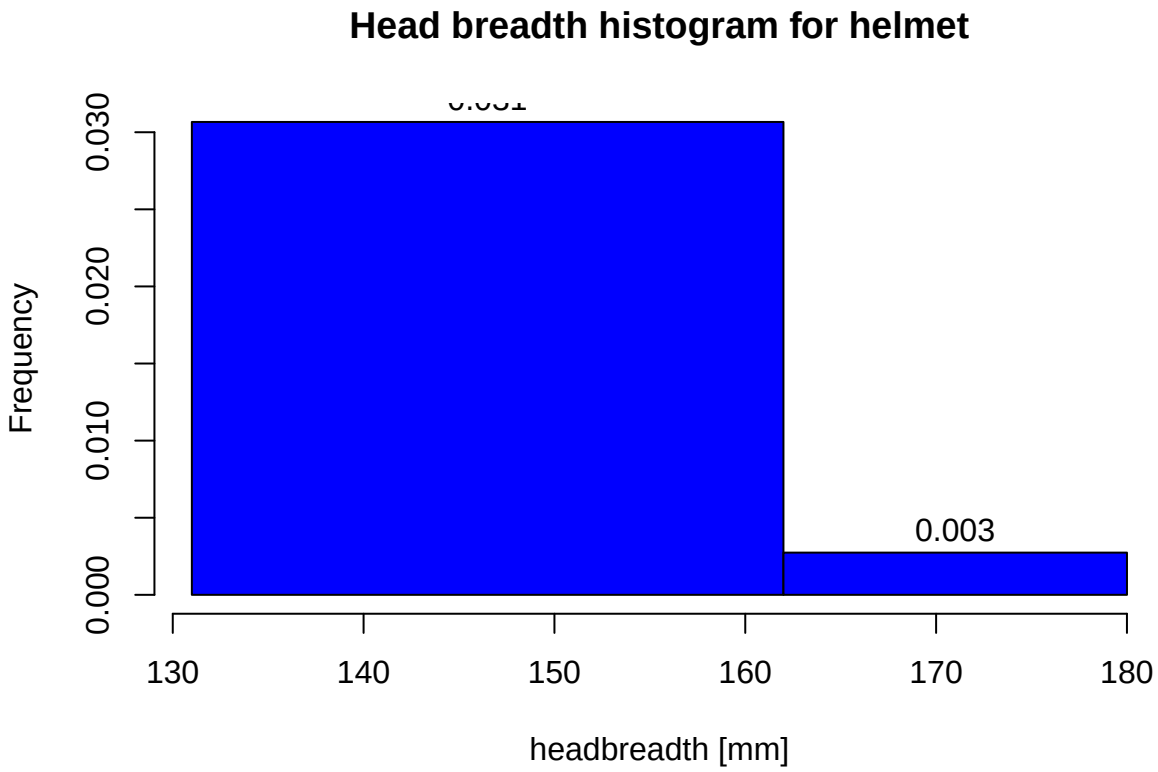
Figure 46: Helmet sizing

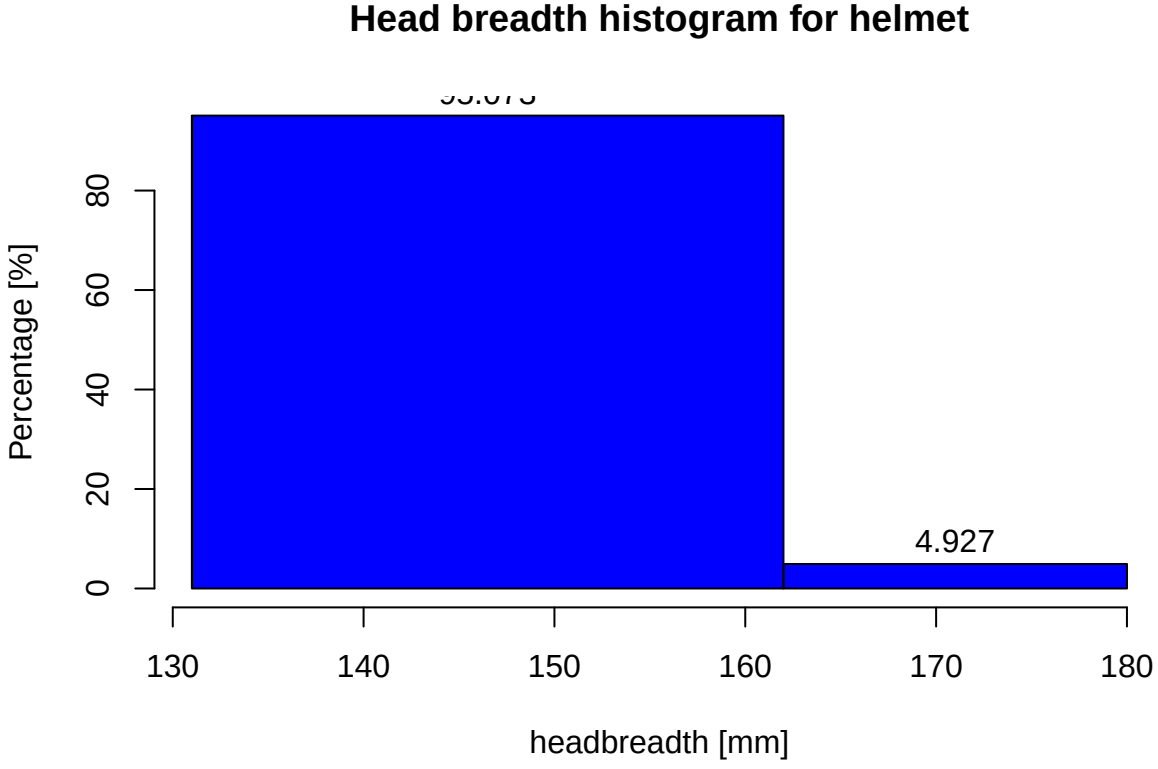

headcircumference box−plot
headcircumference [mm]
500
540
580
620

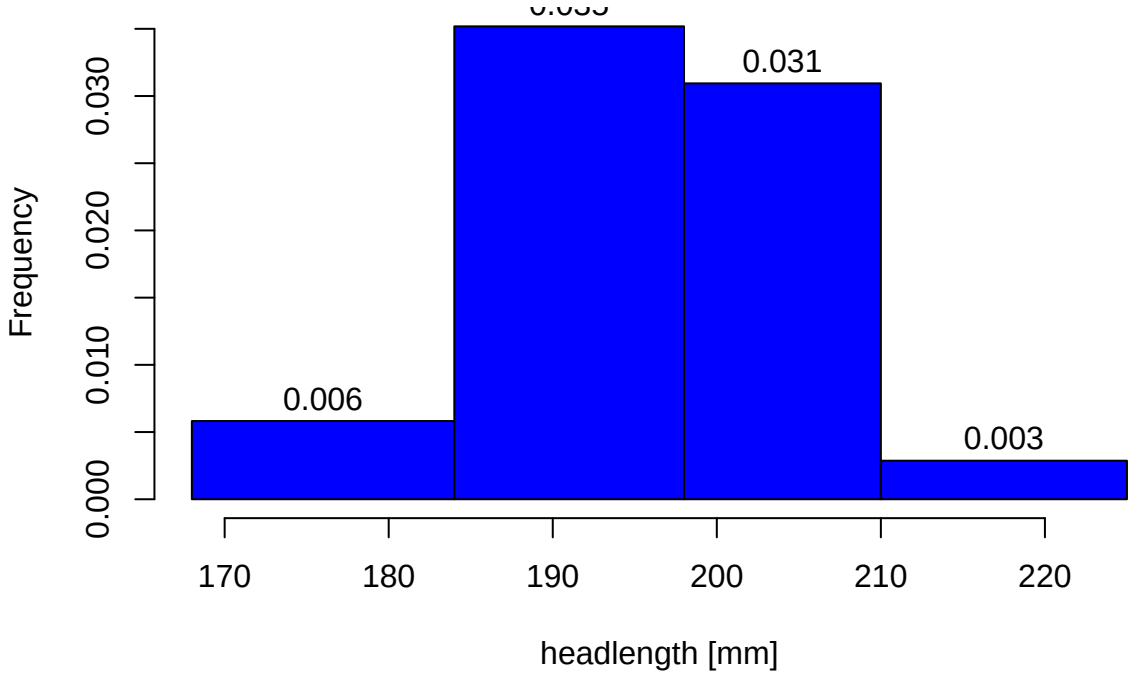

headlength box−plot
headlength [mm]
170
180
190
200
210
220

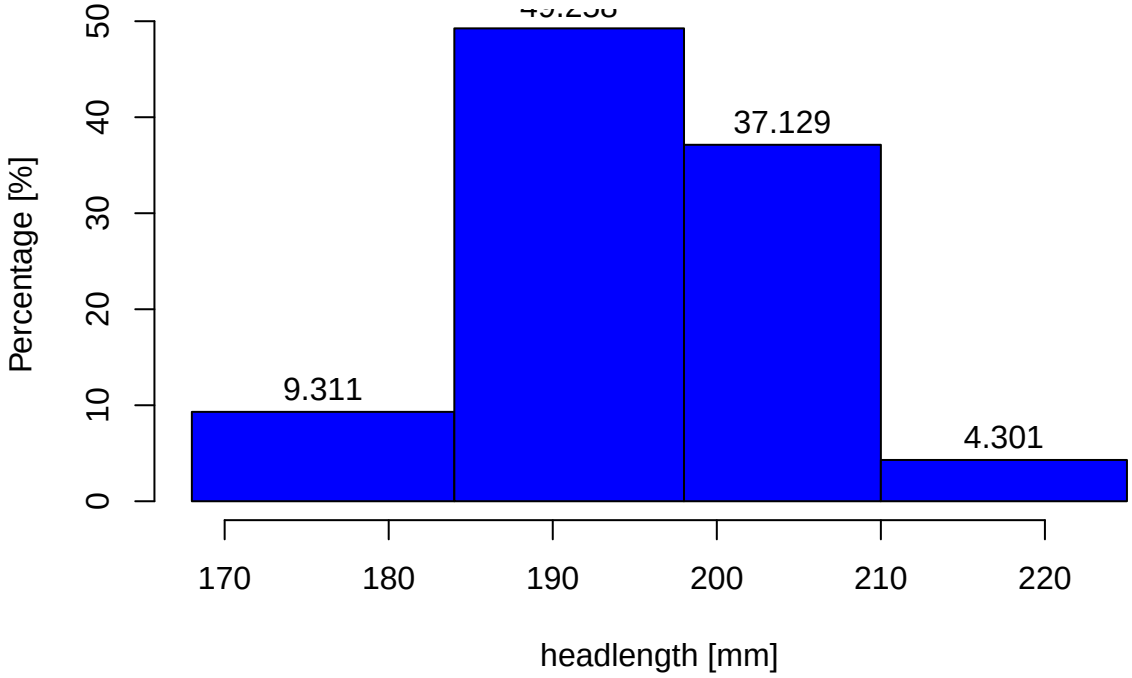

headbreadth box−plot
headbreadth [mm]
130
140
150
160
170
180

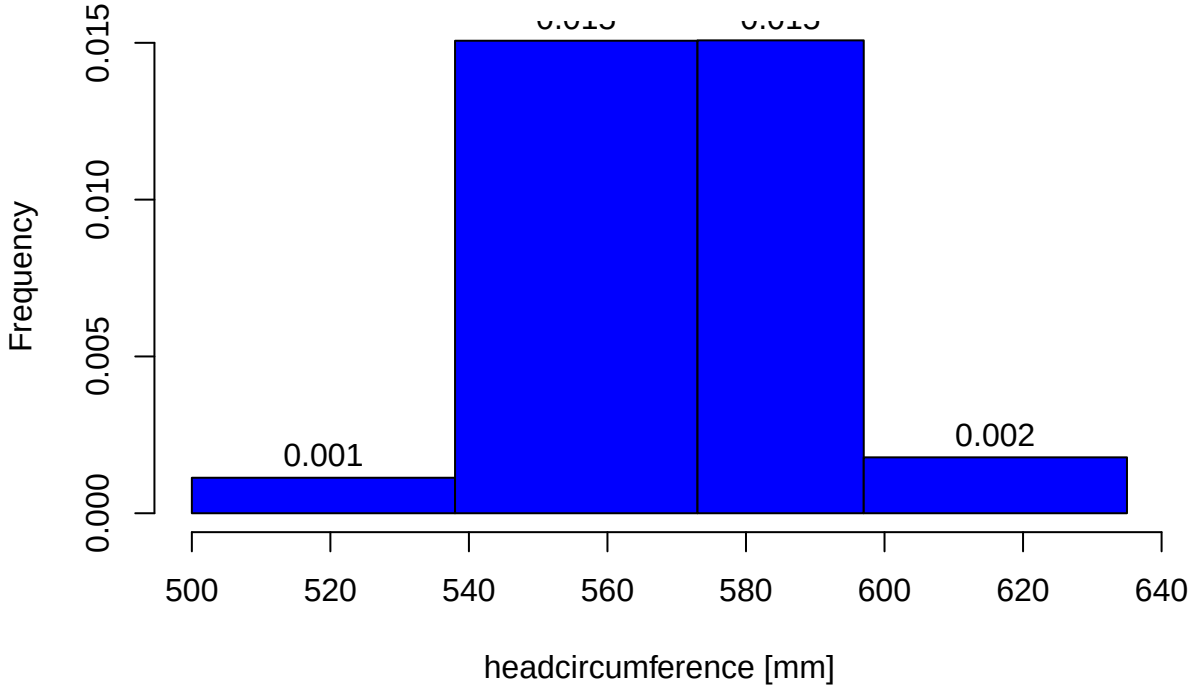
5.6 Histogrami s custom breakpointovima po veličinama kaciga

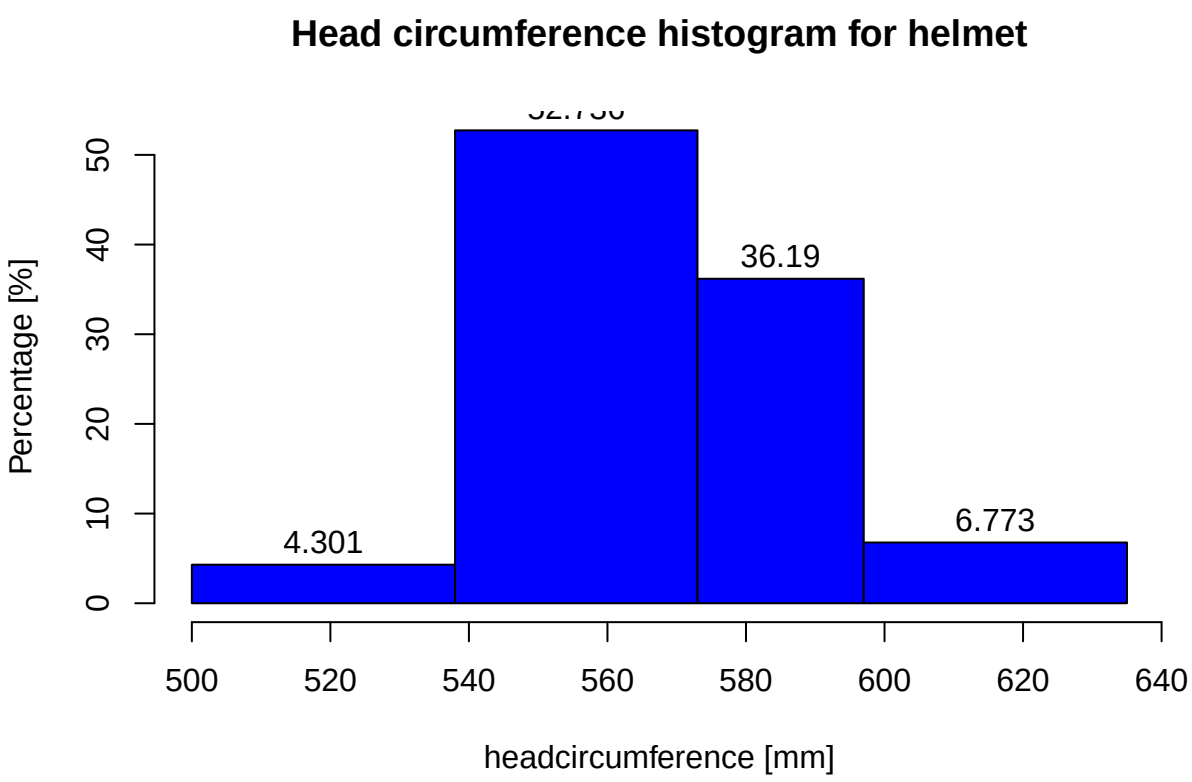





Head length histogram for helmet
Frequency
0.000
0.010
0.020
0.030
0.006
0.031
0.003
170
180
190
200
210
220
headlength [mm]


Head length histogram for helmet
Percentage [%]
0
10
20
30
40
50
9.311
37.129
4.301
170
180
190
200
210
220
headlength [mm]


Head circumference histogram for helmet
Frequency
0.000
0.005
0.010
0.015
0.001
0.002
500
520
540
560
580
600
620
640
headcircumference [mm]



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 vojsci