

Opisna statistika v programskem okolju R

KAZALO

1	UVOD	1
2	KVANTILI.....	2
2.1	Računanje kvantilov	2
2.2	Okvir z ročaji.....	3
3	SREDNJE VREDNOSTI.....	4
4	MERE VARIABILNOSTI	5

1 UVOD

V programskem okolju R lahko statistične izračune izvedemo s pomočjo računskih operacij (za opis glejte dokument *Uvod v R*), lahko pa uporabimo **statistične funkcije**, ki so v R že vgrajene.

R torej že v osnovni različici (brez dodatnih paketov) vsebuje funkcije za praktično vse osnovne statistike. Kadar pa željene funkcije ne najdemo v osnovni različici, je potrebno v R namestiti ustrezen paket (angl. *Package*).

Funkcijo kličemo tako, da navedemo njeno ime in potem v oklepajih () njene argumente, ločene z vejico. Nujno moramo navesti vse argumente, ki nimajo privzete vrednosti, ostale pa le, če nam privzeta vrednost ne ustreza.

Pri navajanju argumentov imamo dve možnosti:

- Argumente navajamo v takem vrstnem redu, kot so navedeni v funkciji. Ta način lahko uporabimo samo pri argumentih, ki so na začetku oziroma pri katerih smo navedli tudi vse argumente pred njimi.
- Argumente navajamo v poljubnem vrstnem redu skupaj z njihovimi imeni kot dvojice *ime=vrednost*.

V vsakem primeru argumente ločimo z vejicami. Drugi način je precej varnejši. Prvi se navadno uporablja le za prvi argument ali pri zelo pogosto uporabljenih funkcijah, čeprav tudi takrat ni priporočljiv.

V nadaljevanju si bomo ogledali uporabo statističnih funkcij. V ta namen bomo analizirali spremenljivko *x*, ki vsebuje vrednosti za višine 20 študentov (*v cm*):

180, 175, 178, 162, 170, 190, 171, 175, 181, 169, 177, 163, 197, 183, 172, 183, 173, 178, 199, 170

Spremenljivko najprej vnesemo v R:

```
x<-  
c(180,175,178,162,170,190,171,175,181,169,177,163,197,183,172,183,173,178,199,170)
```

2 KVANTILI

2.1 RAČUNANJE KVANTILOV

Za izračun kvantilov uporabimo funkcijo `quantile(x, probs)`, pri čemer je:

<code>x</code>	spremenljivka, za katero računamo kvantile
<code>probs</code>	argument, ki določa relativne range kvantilov, ki jih želimo izračunati

Na primeru spremenljivke `x` izračunajmo najprej vse kvartile:

```
quantile(x, prob=c(0.25, 0.5, 0.75))
```

Rezultat:

```
> quantile(x, prob=c(0.25, 0.5, 0.75))
      25%      50%      75%
170.75 176.00 181.50
```

Na primeru spremenljivke `x` izračunajmo najprej vse decile:

```
quantile(x, prob=seq(0.1, 0.9, by=0.1))
```

Rezultat:

```
> quantile(x, prob=seq(0.1, 0.9, by=0.1))
      10%      20%      30%      40%      50%      60%      70%      80%      90%
168.4 170.0 171.7 174.2 176.0 178.0 180.3 183.0 190.7
```

Na enak način lahko izračunamo tudi centile, percentile itd.

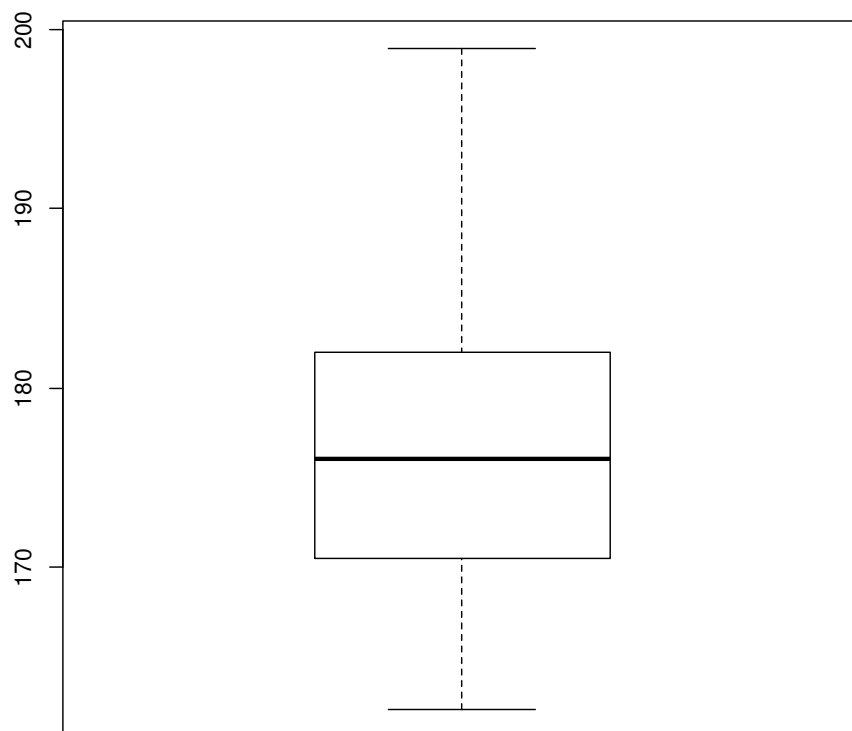
2.2 OKVIR Z ROČAJI

Za izris diagrama kvartilov uporabimo funkcijo `boxplot(x)`, pri čemer argument `x` določa spremenljivko, za katero želimo izrisati diagram kvartilov (okvir z ročaji).

Narišimo okvir z ročaji za obravnavano spremenljivko `x`:

```
boxplot(x)
```

Rezultat:



3 SREDNJE VREDNOSTI

Pregled statističnih funkcij za izračun srednjih vrednosti prikazuje Tabela 1.

Tabela 1: Pregled statističnih funkcij za izračun srednjih vrednosti v R

Funkcija	Opis
<code>mean(x)</code>	vrne povprečje spremenljivke x
<code>median(x)</code>	vrne vrednost mediane spremenljivke x
<code>summary(x)</code>	izračuna povzetek, ki je primeren glede na tip objekta/spremenljivke

Izračunajmo aritmetično sredino in mediano za našo spremenljivko x :

`mean(x)`

Rezultat:

```
> mean(x)
[1] 177.3
```

`median(x)`

Rezultat:

```
> median(x)
[1] 176
```

Za statistike je zelo koristna funkcija `summary`, ki izračuna povzetek, ki je primeren glede na tip objekta/spremenljivke. Funkcija se lahko uporabi na različnih podatkovnih strukturah (tudi na primer na celotnem podatkovnem okvirju).

Za spremenljivko x :

`summary(x)`

Rezultat:

```
> summary(x)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
162.0   170.8   176.0   177.3   181.5   199.0
```

Vse zgoraj našteje funkcije kot argument sprejmejo vektor (spremenljivko). V privzeti obliki ne dovoljujejo manjkajočih vrednosti (NA), saj v tem primeru tudi vrnejo NA. Vendar pa je mogoče nastaviti, da manjkajoče vrednosti ignorirajo s parametrom `na.rm=TRUE`.

4 MERE VARIABILNOSTI

V R lahko izračunamo vse obravnavane absolutne in relativne mere variabilnosti (s pomočjo funkcij ali s pomočjo izračunov). Funkcije najpogosteje uporabljenih mer variabilnosti prikazuje Tabela 2.

Tabela 2: Pregled statističnih funkcij za izračun mer variabilnosti v R

Funkcija	Opis
<code>min(x)</code>	vrne najmanjšo vrednost spremenljivke x
<code>max(x)</code>	vrne največjo vrednost spremenljivke x
<code>range(x)</code>	vrne razpon vrednosti spremenljivke x
<code>diff(range(x))</code>	vrne variacijski razmik spremenljivke x
<code>IQR(x)</code>	vrne kvartilni razmik spremenljivke x
<code>sd(x)</code>	vrne standardni odklon spremenljivke x
<code>var(x)</code>	vrne varianco spremenljivke x

Izračunajmo opisne statistike (variabilnost) spremenljivke x :

```
> min(x)
[1] 162
> max(x)
[1] 199
> range(x)
[1] 162 199
> diff(range(x))
[1] 37
> IQR(x)
[1] 10.75
> sd(x)
[1] 9.782369
> var(x)
[1] 95.69474
```

Funkcija *summary* sicer naredi zadosten povzetek, vendar pa pri številskih ne poda vseh želenih vrednosti (npr. standardnega odklona). V tem primeru lahko uporabimo funkcijo *describe*, ki jo najdemo v paketu *psych*. Ta funkcija poleg tega izračuna tudi mere asimetrije in sploščenosti.

Tudi funkcijo *describe* lahko uporabimo le na eni spremenljivki ali na celotnem podatkovnem okvirju. Je pa ta funkcija primerna le za intervalne in razmernostne spremenljivke. Izračun naredi sicer tudi za ostale, a ni smiselno in jih zato v rezultatih označi z "**".

V paketku *psych* najdemo tudi funkciji za koeficient asimetrije *skew* in sploščenosti *kurtosi* in še veliko drugih uporabnih funkcij (ki se uporabljajo predvsem v psihologiji in družboslovju).

V R namestimo paket *psych* na naslednji način:

```
install.packages('psych')
```

Pregled funkcij za osnovni opis spremenljivk v paketu *psych* prikazuje Tabela 3.

Tabela 3: Pregled osnovnih statističnih funkcij v paketu psych

Funkcija	Opis
<code>skew(x)</code>	vrne koeficient asimetrije spremenljivke <i>x</i>
<code>kurtosi(x)</code>	vrne koeficient sploščenosti spremenljivke <i>x</i>
<code>describe(x)</code>	izračuna opisne statistike spremenljivke <i>x</i> .

Izračunajmo še koeficienta asimetrije in sploščenosti ter pripravimo opisni izpis za spremenljivko *x*:

Pred uporabo funkcij, ki se nahajajo v paketu *psych*, moramo ta paket najprej naložiti v delovno okolje R:

```
library(psych)
```

Nato izračunamo željene mere:

```
skew(x)
```

```
kurtosi(x)
```

Rezultat:

```
> skew(x)
[1] 0.6406799
> kurtosi(x)
[1] -0.2125532
```

```
describe(x)
```

Rezultat:

```
> describe(x)
      vars   n  mean    sd median trimmed  mad min max range skew kurtosis   se
x1      1 20 177.3  9.78    176   176.56  8.15 162 199    37 0.64    -0.21 2.19
```