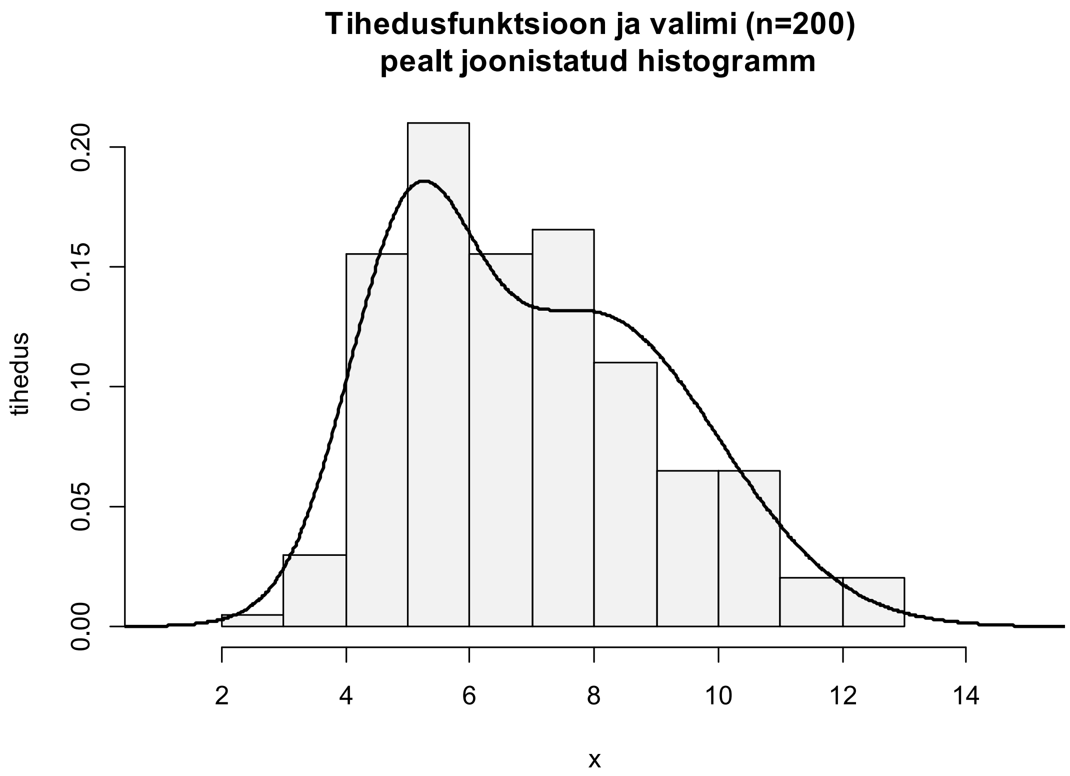
Biomeetria bioloogidele
3. praktikum

Näide: Tegelik jaotus, hinnatud jaotus ja jaotust kirjeldavad statistikud

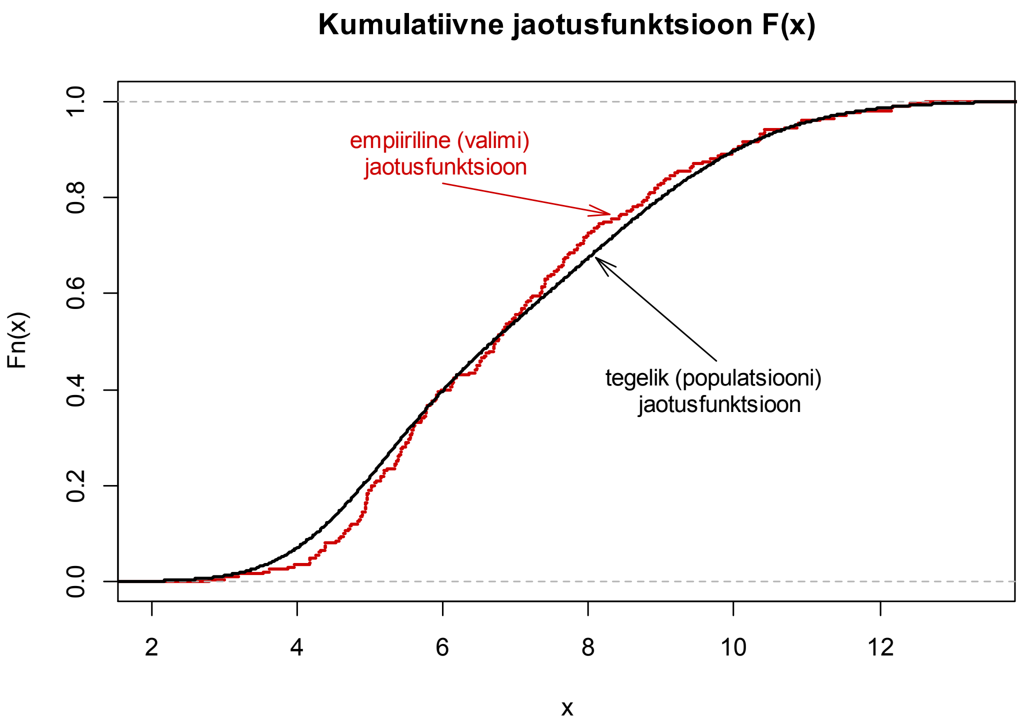


Tegelik (populatsiooni):

keskväärtus	EX = 6,95
mediaan	med = 6,68...
standardhälve	σ = 2,23...

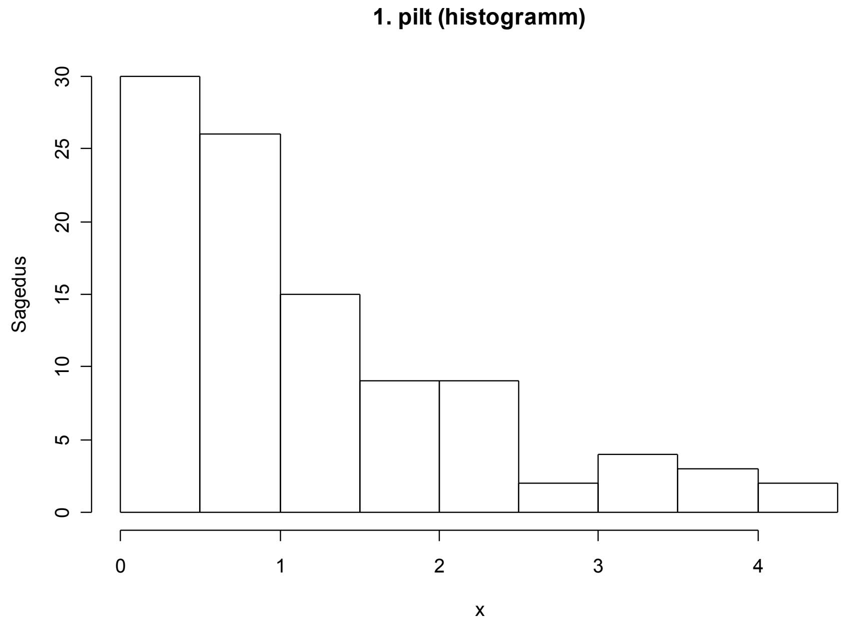
Valimi:

keskmine	$\bar{x}$ = 6,97...
mediaan	med = 6,74..
standardhälve	s = 2,07

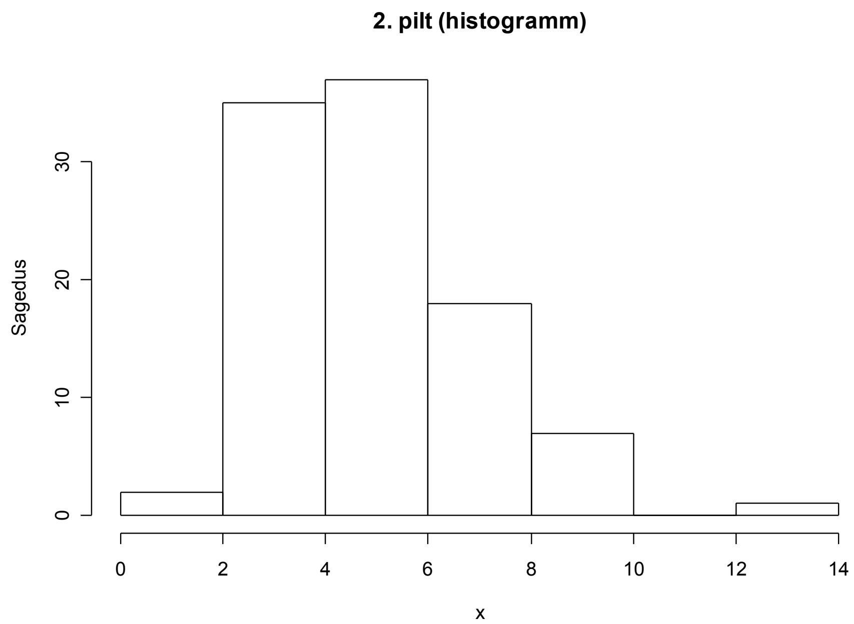


Ülesanne 1

Kasutades jooniseid ürita ära arvata statistikute väärtuseid

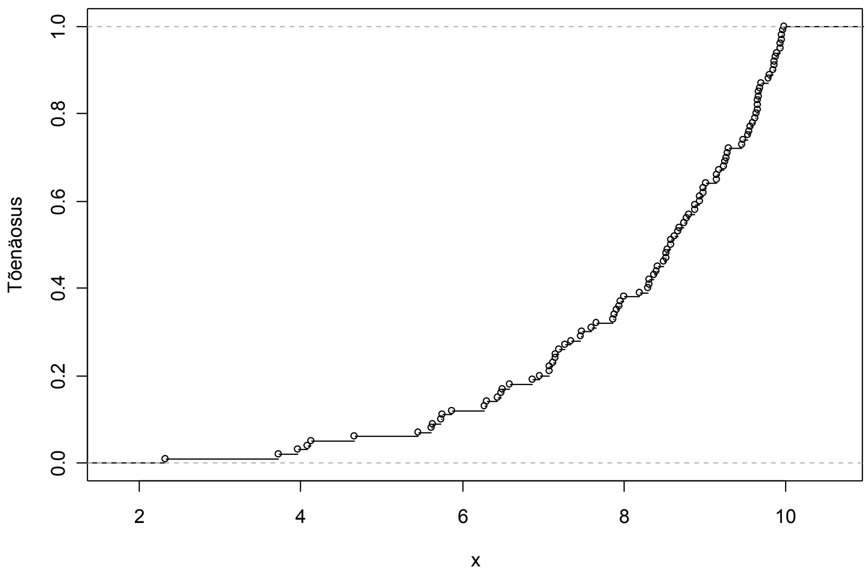


$\bar{x}$ = ...
med = ...
s = ...



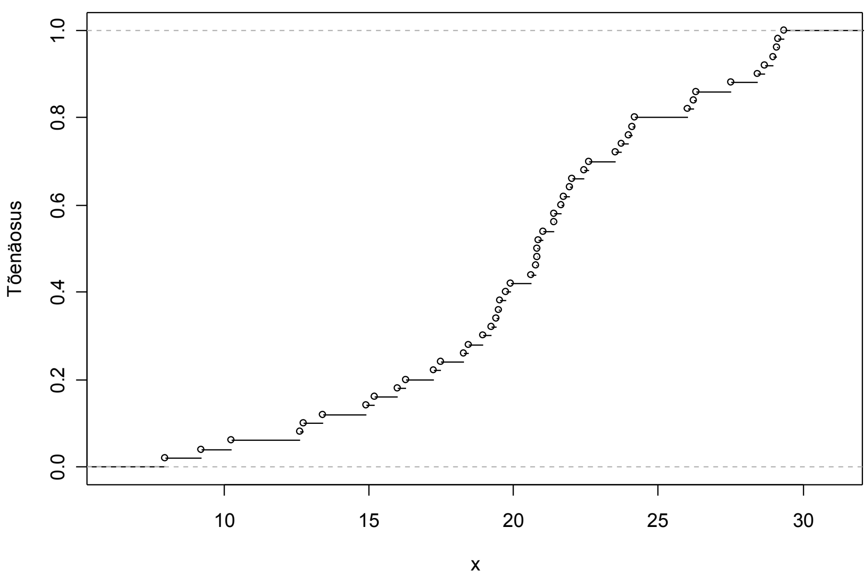
$\bar{x}$ = ...
med = ...
s = ...

3. pilt (kumulatiivne jaotusfunktsioon)



med = ...
s = ...
0,05-kvantiil = ...
0,95-kvantiil = ...

4. pilt (kumulatiivne jaotusfunktsioon)

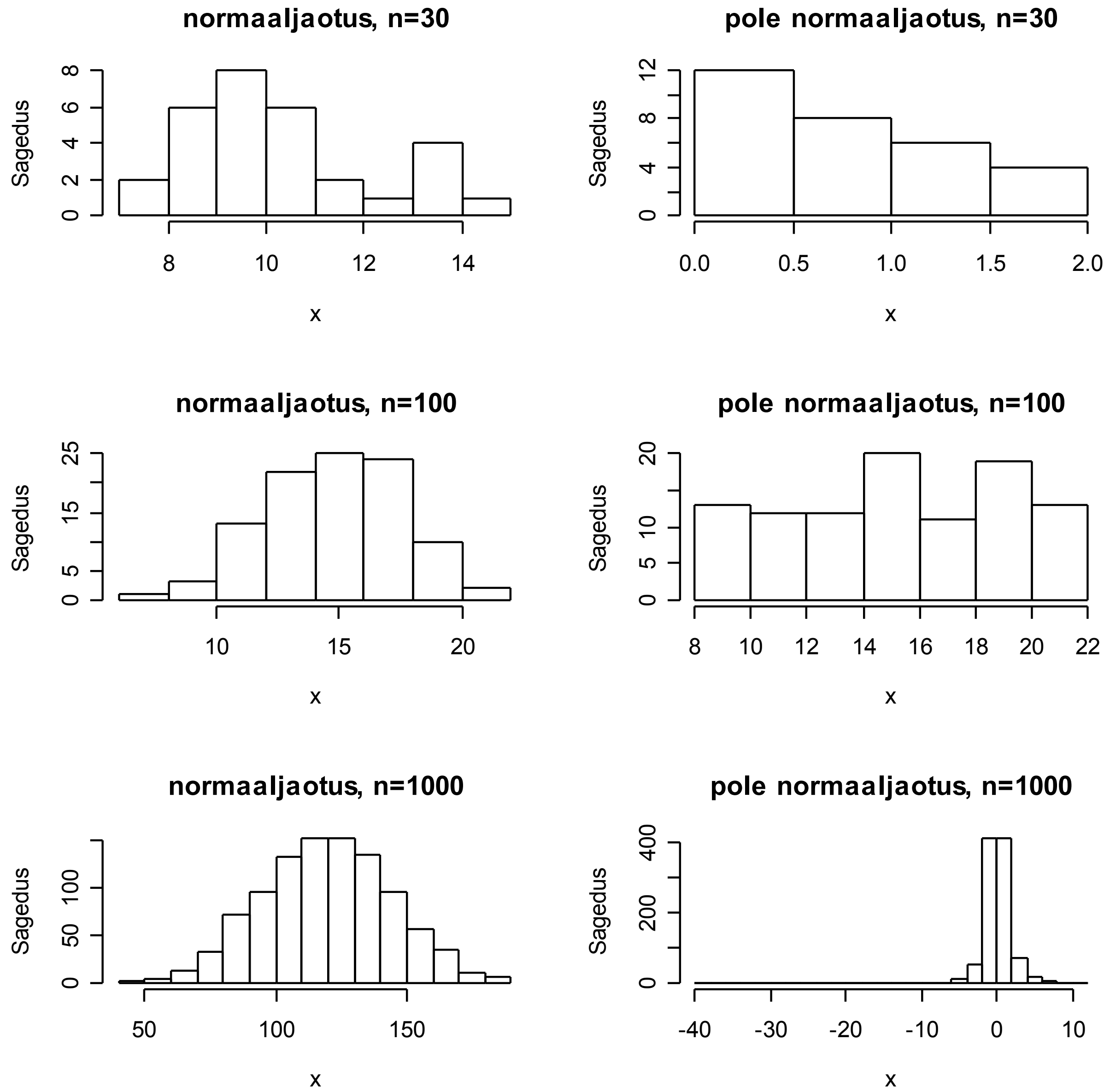


med = ...
s = ...
0,05-kvantiil = ...
0,95-kvantiil = ...

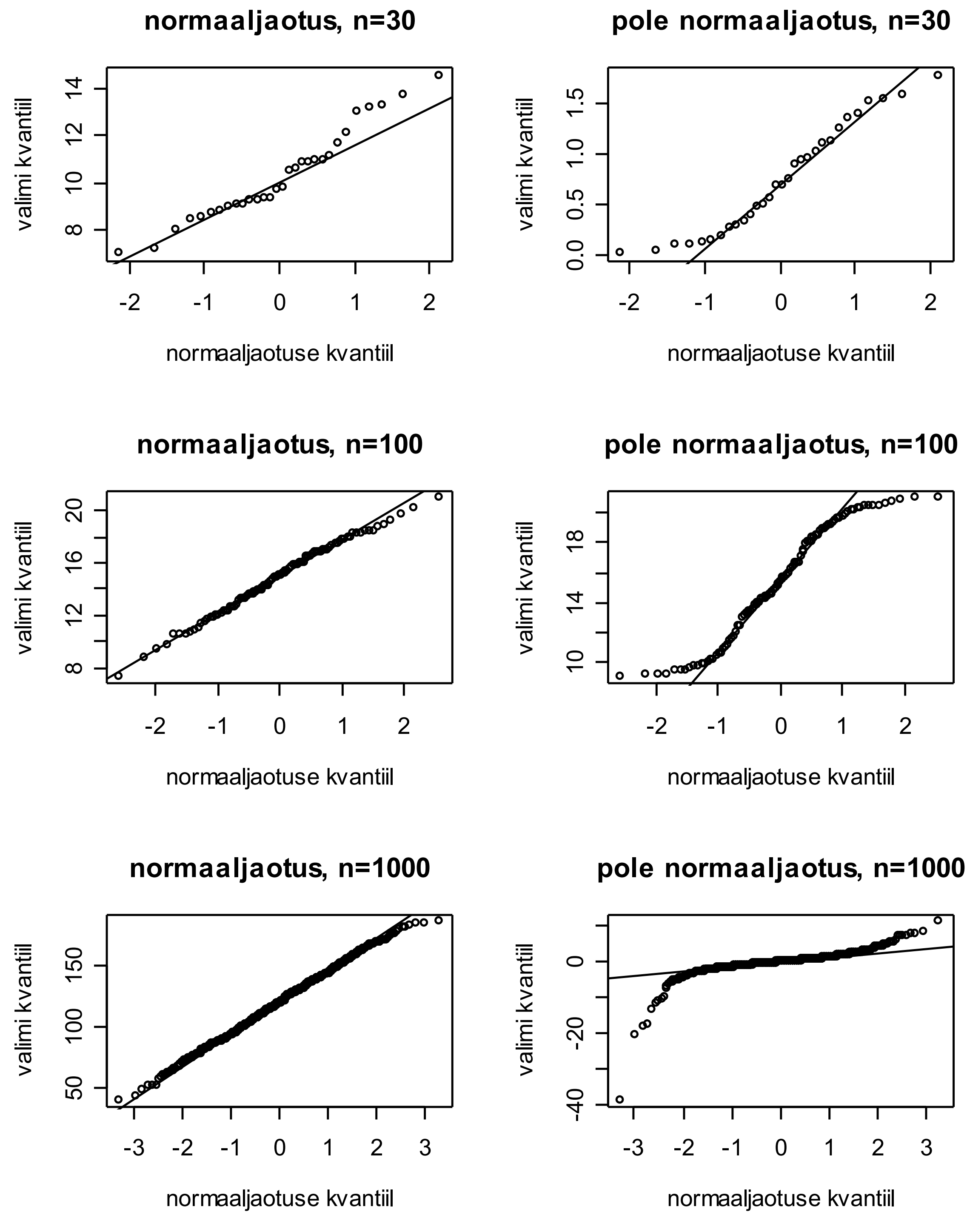
Kas uuritav tunnus on normaaljaotusega?

Vahel tuleb otsustada, kas uuritav tunnus võiks olla normaaljaotusega või mitte. Eriti väiksema valimi korral võib taolisele küsimusele vastamine keerukaks osutuda. Vaatame mõningaid võimalusi sellele küsimusele vastamiseks.

Üheks võimaluseks oleks teha otsus uuritava tunnuse histogrammi põhjal. Näiteid:



Histogramm ei pruugi olla kõige mugavam vahend normaaljaotuse tuvastamiseks. Kasutatakse ka tõenäosuspaberit (quantile-quantile plot, qq-plot) kus moodustatakse valimi kvantiilidest ja normaaljaotuse vastavatest kvantiilidest punktipaarid (näiteks: milline oli valimis 0,1-kvantiil ehk väärtus millest väiksemate väärtuste osakaal on 0,1 ja milline peaks olema normaaljaotusega juhusliku suuruse korral 0,1-kvantiil). Punktid kantakse graafikule ja juhul, kui uuritav tunnus on ka tegelikult normaaljaotusega, peaksid kõik punktid jääma enam-vähem ühele sirgele. Kui punktid kipuvad märgatavalt sirgest kõrvale kalduma, siis pole kardetavasti tegemist normaaljaotusega. Järgneval joonisel on antud tõenäosuspaberid kõigi 6 juhu tarvis, mille jaoks joonistasime ka histogrammid.



Kuidas joonistada tõenäosuspaberit tunnusele X?

`qqnorm(X)` - joonistab punktid
`qqline(X)` - lisab pildile silma abistamiseks võrdlusjoone

Järgnevatele küsimustele vastates kontrolli enne, milliseid graafikuid võid saada, kui uuritav tunnus on tegelikult normaaljaotusega. Seda tee nii:

1. vaata, kui suur on valim (mitut vaatlust oled oma graafiku joonistamiseks kasutanud). Oletame järgmises näiteprogrammis, et vaatluseid oli 42.
2. Tekita 42 vaatlusega valim garanteeritult normaaljaotusega uuritavast tunnusest. Normaaljaotusega populatsioonist võtab soovitud suurusega valimi R-i käsk `rnorm`. Joonista oma genereeritud andmete pealt graafik.
3. Korda sammu 2 mitu korda ja leia, kuivõrd kaugele „normist“ võib sellise suurusega valimi korral joonistatav graafik jääda. Vaata, kas tegelike andmete pealt joonistatud graafik tundub veidram kui genereeritud andmete pealt joonistatud graafikud.

Seda tööd võiks teha järgmine programm. Programmi võiksid esmalt kirja panna tekstiredaktoris (notepad) ja sealt kogu programmi korraga kopeerida R-i.

Histogrammide joonistamine:

Kordame sulgude { } vahel olevaid käske 4 korda

Võtame juhusliku valimi normaaljaotusega populatsioonist

Joonistame histogrammi

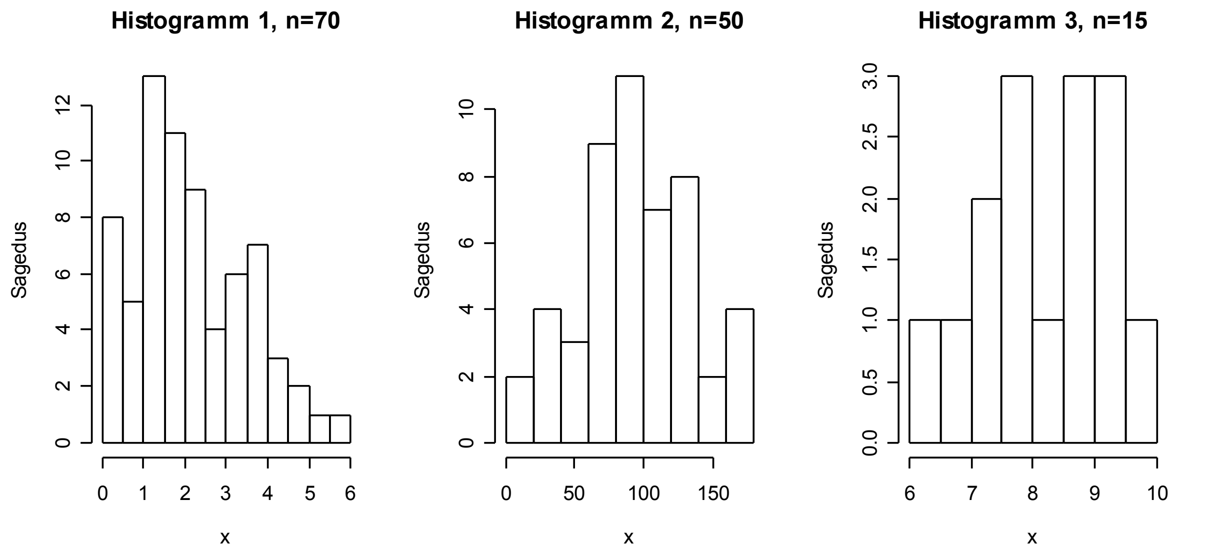
```
par(mfrow=c(2,2))
for (i in 1:4){
  valim=rnorm(42)
  hist(valim, main="normaaljaotus, n=42")
}
```

Tõenäosuspaberite joonistamine:

```
par(mfrow=c(2,2))
for (i in 1:4){
  valim=rnorm(42)
  qqnorm(valim, main="normaaljaotus, n=42")
  qqline(valim)
}
```

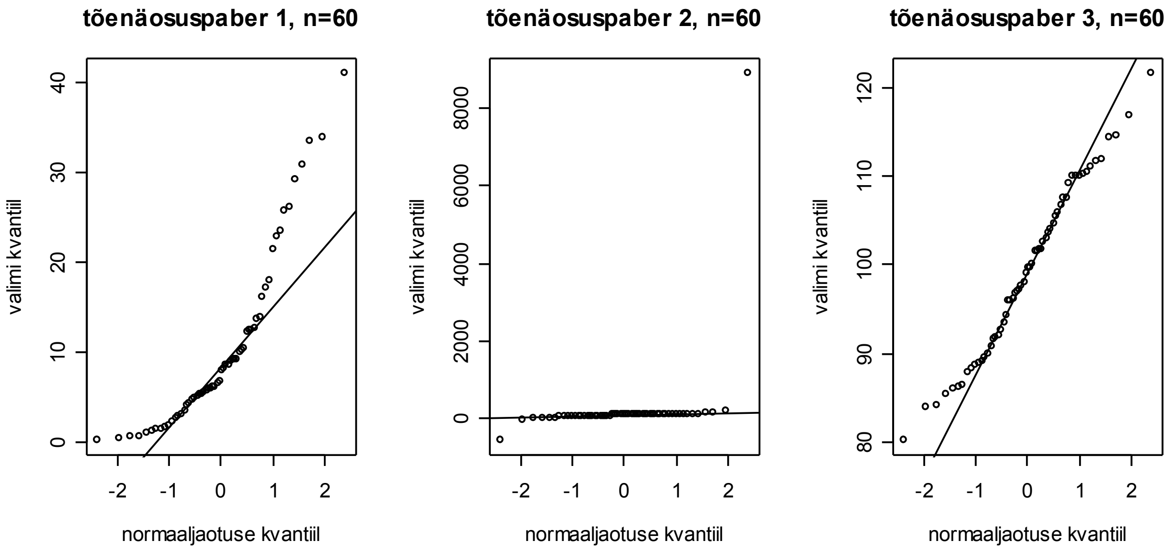
Ülesanne 2

Otsusta, kas järgmiste graafikute puhul võiks olla tegemist normaaljaotusega tunnustega:



Ülesanne 3

Vaata, millistel järgmistest graafikutest võiks olla uuritava tunnuse jaotuseks normaaljaotus?



Ülesanne 4

Loe sisse eelmises praktikumis kasutatud kalade andmestik. Seda sab teha käsuga:

```
load(url("http://www.ms.ut.ee/mart/biomeetria2012/kalamees.RData"))
```

Kontrolli, kas ahvenate (*Species*=7) kaal (*Weight*) on normaaljaotusega!

Valimi juhuslikkusest ja hinnangu täpsusest

Käsuga

```
rnorm(10, mean=150, sd=10)
```

saab võtta juhusliku valimi suurusega 10 (10 mõõtmist) tegelikult normaaljaotusega populatsioonist, mille keskväärtus (populatsiooni keskmine) on 150 ja milles uuritava tunnuse dispersioon on 10. Oletame, et kaks uurijat uurivad sama populatsiooni. Mõlemad võtavad samasuured valimid:

```
teadlane1_valim=rnorm(10, mean=150, sd=10)
teadlane2_valim=rnorm(10, mean=150, sd=10)
```

Mõlema teadlase valimid on erinevad (kuigi on võetud samast populatsioonist):

```
teadlane1_valim
teadlane2_valim
```

ja mõlemad saavad (veidi) erineva keskmise:

```
mean(teadlane1_valim)
```

```
mean(teadlane2_valim)
```

Oletame, et erinevaid teadlaseid, kes sama metoodika alusel – sama katseplaani kasutades, samasuuri valimeid võttes – uurivad sama populatsiooni. Näiteks vaatame 100 sama metoodikat kasutava teadlase hinnanguid keskväärtusele (populatsiooni keskmisele):

```
n=100
valimi_keskmine=rep(NA,n)
for (teadlase_nr in 1:n){
    valim=rnorm(10, mean=150, sd=10)
    valimi_keskmine[teadlase_nr]=mean(valim)
}
```

Nende 100 hüpoteetilise teadlase uuringutulemused on järgmised:

```
valimi_keskmine
```

Graafiliselt esitatult:

```
plot(valimi_keskmine, 1:n, ylab="teadlase nr",
          xlab="keskmine")
abline(v=150, lwd=2, col="red")
text(150, 90, "Tegelik keskväärtus", col="red", adj=c(0,1))
```

Kui erinevaid tulemusi eri teadlased saavad (milline on kasutatud metoodika täpsus)? Täpsuse kirjeldamiseks oli üheks võimaluseks (sobib nihketa hinnangute puhul) kirjeldada tehtavate vigade hajuvust – näiteks kasutades dispersiooni või standardhälvet:

```
sd(valimi_keskmine-10)
```

Tegelikult piisab valimikeskmiste standardhälbe leidmisest (Miks?):

```
sd(valimi_keskmine)
```

Seda suurust – metoodika täpsust – saame aga ka üheainsa valimi põhjal hinnata, ilma et laseksime sama populatsiooni uurida samal meetodil sadadel teadlastel. Selleks kasutame standardviga. Näiteks kahe esimese teadlase hinnangud kasutatud metoodika täpsusele oleksid järgmised:

```
sd(teadlane1_valim)/sqrt(10)
sd(teadlane2_valim)/sqrt(10)
```

Antud juhul – kuna teame uuritavat populatsiooni täpselt – saame leida ka kasutatud metoodika täpse „täpsuse“ (erinevate teadlaste tulemuste võimaliku kõikumise õigest väärtusest):

```
10/sqrt(10)
```

Kui täpselt suutsid teadlased enda poolt kasutatud metoodika täpsust kirjeldada?