Ülesanne 2

Jaagup Kippar

13 aprill 2016

Perioodide õpitulemuste erinevuse uuring

Ülesande kirjeldus

1. Loe jälle sisse ka 1. ülesandes kasutatud andmefail "student-mat.csv"
2. Kasuta funktsiooni str(), et näha andmefaili struktuuri. Kogu väljund peaks nüüd olema arusaadav...
3. Lisa andmefailile tulp "diff", mis on teise perioodi ja esimese perioodi hinde vahe.
4. Kas see on nii? Milline situatsioon andmetes valitseb? (Kasuta funktsioone summary() ja hist())
5. Milles on asi? Selle mõistmiseks võrdle muutujaid G1 ja G2. Seda on hea teha nt kõrvutades kaks histogrammi. Ühele joonisele kohakuti histogrammide saamiseks jooksuta kõigepealt kood par(mfrow = c(2, 1)) ja seejärel kummagi histogrammi kood. Et x-teljed oleksid võrreldavad, kasuta parameetrit xlim. Lisaks võib olla hea mõte muuta parameetrit breaks.
6. Muuda muutuja G2 0-väärused R-i puuduvaks väärtuseks (NA).
7. Kontrolli, et asi õnnestus: leia mitu puuduvat väärtust on nüüd antud tulbas. Vihje: hea viis seda saavutada on kasutada ära fakti, et sisemiselt on R-i jaoks FALSE 0 ja TRUE 1 (proovi konsoolis TRUE + 1 + FALSE). Seega saab vastuse lihtalt luues vektori, kus puuduv väärtus on TRUE ja mittepuuduv väärtus on FALSE ning leida nende väärtuste summa. (Loetavam alternatiiv on kasutada funktsiooni which(), et luua vektor ainult puuduvate väärtustega ja siis funktsiooni length(), et leida, kui pikk see vektor on ehk mitu neid on. Soovituslik harjutus on ka seda varianti proovida!) Ära unusta, et R-is on puuduvate väärtuste tuvastamiseks eraldi funktsioon!
8. Arvuta tulp "diff"" (teise ja esimese perioodi hinde vahe) uuesti ning tee uus histogram. (Selleks, et saada jälle kogu ala hõlmav joonis nagu tavaliselt, jooksuta kõigepealt käsk par(mfrow = c(1,1)) ; vastasel korral kehtib eelmine par() funktsiooni toime, mis jaotas "joonistusala" kaheks osaks.) Kas pole kenam?
9. Leia ka erinevuse keskmine.
10. Lisa andmetele tulp, kus erinevuse suuruse põhjal on loodud 4 gruppi. Kasuta selleks funktsiooni cut() ja gruppide loomise aluseks kvantiile. (Kasuta ainult funktsiooni cut() kaht esimest parameetrit.)
11. Mitu puuduvat väärust on loodud gruppide muutujas? Mõtle tagasi puuduvate väärtuste arvule muutujas G2 - kas tulemus on ootuspärane?
12. Leia muutuja diff väärtus, mille korral grupitunnus on puuduv väärtus. (Vali andmetest tulp diff ja rida, kus grupitunnus on puuduv ja diff ei ole puuduv)
13. Leia cut() funktsiooni help file-st parameeter, mis võiks selle olukorra põhjuseks olla, muuda selle väärtust ja arvuta grupitunnus uuesti; anna gruppidele ka nimed. Kontrolli, et grupitunnuse ja G2 (või diff) puuduvate väärtuste arv on võrdne.
14. Tee histogram madalama grupi tüdrukute puudumiste kohta
15. Leia kahe madalama grupi puudumiste keskmine.
16. Leia kahe madalama grupi puudumiste keskmine ainult poistel.
17. Leia mitme tüdruku puudumiste arv ületab poiste maksimum puudumiste arvu (Vihje: kasuta sama trikki nagu puuduvate väärtuste leidmisel või leia tingimusele vastav andmete alaosa ja siis selle suurus; võite proovida teha seda ühe reaga.)

Andmed sisse

```
andmed=read.table("student-mat.csv", header=TRUE, sep=";", dec=",")
```

str-käsu abil andmete struktuuri ülevaade

```
str(andmed)
```

```
## 'data.frame':    395 obs. of  33 variables:
##  $ school    : Factor w/ 2 levels "GP","MS": 1 1 1 1 1 1 1 1 1 1 ...
##  $ sex       : Factor w/ 2 levels "F","M": 1 1 1 1 1 2 2 1 2 2 ...
##  $ age       : int  18 17 15 15 16 16 16 17 15 15 ...
##  $ address   : Factor w/ 2 levels "R","U": 2 2 2 2 2 2 2 2 2 2 ...
##  $ famsize   : Factor w/ 2 levels "GT3","LE3": 1 1 2 1 1 2 2 1 2 1 ...
##  $ Pstatus   : Factor w/ 2 levels "A","T": 1 2 2 2 2 2 2 1 1 2 ...
##  $ Medu      : int  4 1 1 4 3 4 2 4 3 3 ...
##  $ Fedu      : int  4 1 1 2 3 3 2 4 2 4 ...
##  $ Mjob      : Factor w/ 5 levels "at_home","health",..: 1 1 1 2 3 4 3 3 4 3 ...
##  $ Fjob      : Factor w/ 5 levels "at_home","health",..: 5 3 3 4 3 3 3 5 3 3 ...
##  $ reason    : Factor w/ 4 levels "course","home",..: 1 1 3 2 2 4 2 2 2 2 ...
##  $ guardian  : Factor w/ 3 levels "father","mother",..: 2 1 2 2 1 2 2 2 2 2 ...
##  $ traveltime: int  2 1 1 1 1 1 1 2 1 1 ...
##  $ studytime : int  2 2 2 3 2 2 2 2 2 2 ...
##  $ failures  : int  0 0 3 0 0 0 0 0 0 0 ...
##  $ schoolsup : Factor w/ 2 levels "no","yes": 2 1 2 1 1 1 1 2 1 1 ...
##  $ famsup    : Factor w/ 2 levels "no","yes": 1 2 1 2 2 2 1 2 2 2 ...
##  $ paid      : Factor w/ 2 levels "no","yes": 1 1 2 2 2 2 1 1 2 2 ...
##  $ activities: Factor w/ 2 levels "no","yes": 1 1 1 2 1 2 1 1 1 2 ...
##  $ nursery   : Factor w/ 2 levels "no","yes": 2 1 2 2 2 2 2 2 2 2 ...
##  $ higher    : Factor w/ 2 levels "no","yes": 2 2 2 2 2 2 2 2 2 2 ...
##  $ internet  : Factor w/ 2 levels "no","yes": 1 2 2 2 1 2 2 1 2 2 ...
##  $ romantic  : Factor w/ 2 levels "no","yes": 1 1 1 2 1 1 1 1 1 1 ...
##  $ famrel    : int  4 5 4 3 4 5 4 4 4 5 ...
##  $ freetime  : int  3 3 3 2 3 4 4 1 2 5 ...
##  $ goout     : int  4 3 2 2 2 2 4 4 2 1 ...
##  $ Dalc      : int  1 1 2 1 1 1 1 1 1 1 ...
##  $ Walc      : int  1 1 3 1 2 2 1 1 1 1 ...
##  $ health    : int  3 3 3 5 5 5 3 1 1 5 ...
##  $ absences  : int  6 4 10 2 4 10 0 6 0 0 ...
##  $ G1        : int  5 5 7 15 6 15 12 6 16 14 ...
##  $ G2        : int  6 5 8 14 10 15 12 5 18 15 ...
##  $ G3        : int  6 6 10 15 10 15 11 6 19 15 ...
```

Tulemuste vahemikud kummalgi perioodil

```
summary(andmed$G1)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    3.00    8.00   11.00   10.91   13.00   19.00
```

```
summary(andmed$G2)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    9.00   11.00   10.71   13.00   19.00
```

Tulpade väärtuste vahe

```
andmed$diff=andmed$G2-andmed$G1
```

Vahe väljatrükk

```
andmed$diff
```

```
##   [1]   1   0   1  -1   4   0   0  -1   2   1  -2   2   0   0   2   0   1
##  [18]   2  -1   2   1   3   0   0  -1   3   0   1   0   2   2  -1  -1   2
##  [35]   2  -1   1   1   0  -1   3   0  -1   0   0   0   1   0   0   0   1
##  [52]   2   0   2   3   1   1   1   1   1   1  -2   2  -1   0  -1   0   0
##  [69]   1   0   2   0  -2   0   1   0   0   0   0   0   2  -1  -1   0   1
##  [86]   2  -1   1  -1  -2   0   1  -1  -1   2   3   4   1   3   2   0   1
## [103]   3  -1   2   1   1   2   3   1   1   3   3   1   0   0   2   1  -2
## [120]  -1  -1  -2   0  -3  -1   0   3   1  -3   0 -12  -8   3  -1  -9 -11
## [137] -10  -4  -2   0   2   0   2   0  -5   3   1   1  -1   1  -1   1   0
## [154]  -5   0  -3  -4  -1  -2   2  -1   4  -7   0   3  -1   0   1   1   0
## [171]  -1   2  -2  -1   1  -1   0  -1  -2   0  -1   1   1   0   1   0   1
## [188]   0  -1   1   1   0   1   1   1   1  -2   0   0   0   0   2   0  -1
## [205]   0  -1   0   1   0   0   0   0   1   1   2   1   0   0   0   1   0
## [222]  -1   0   1   0  -1  -1  -1  -2  -2   0   0  -2  -1  -2  -2  -1  -1
## [239]  -2   0   0   1  -6  -1  -7   0   0   2   2   2   2   3   3   1   4
## [256]   2  -2   0  -1  -1   1   0  -1  -1   1   0   0  -2  -1  -6   0  -1
## [273]   0  -1   0   0  -1   0  -1   1   0  -2   0   1  -1  -2   0  -1  -1
## [290]  -2  -1   0   0   0  -1  -2  -1  -2  -1  -1  -2   0  -3   0  -1  -2
## [307]   1   1  -3  -2   0  -2  -2  -3  -2  -2   0   1   0   0   0  -2   0
## [324]   2  -1   3   1  -1  -1   0  -1   2  -7   0  -1  -1  -1   1  -1   1
## [341]   1   0  -1  -1  -1   0  -1   0   2   2  -1   0  -1   0  -2  -1   1
## [358]   0   0  -2   0  -1   0  -1  -1   0   0  -1  -1  -2   0  -2  -2  -1
## [375]  -1   0  -1   1   0   0  -1  -1   0  -1  -1  -1  -1  -2   2  -1   0
## [392]   2  -2   1   1
```

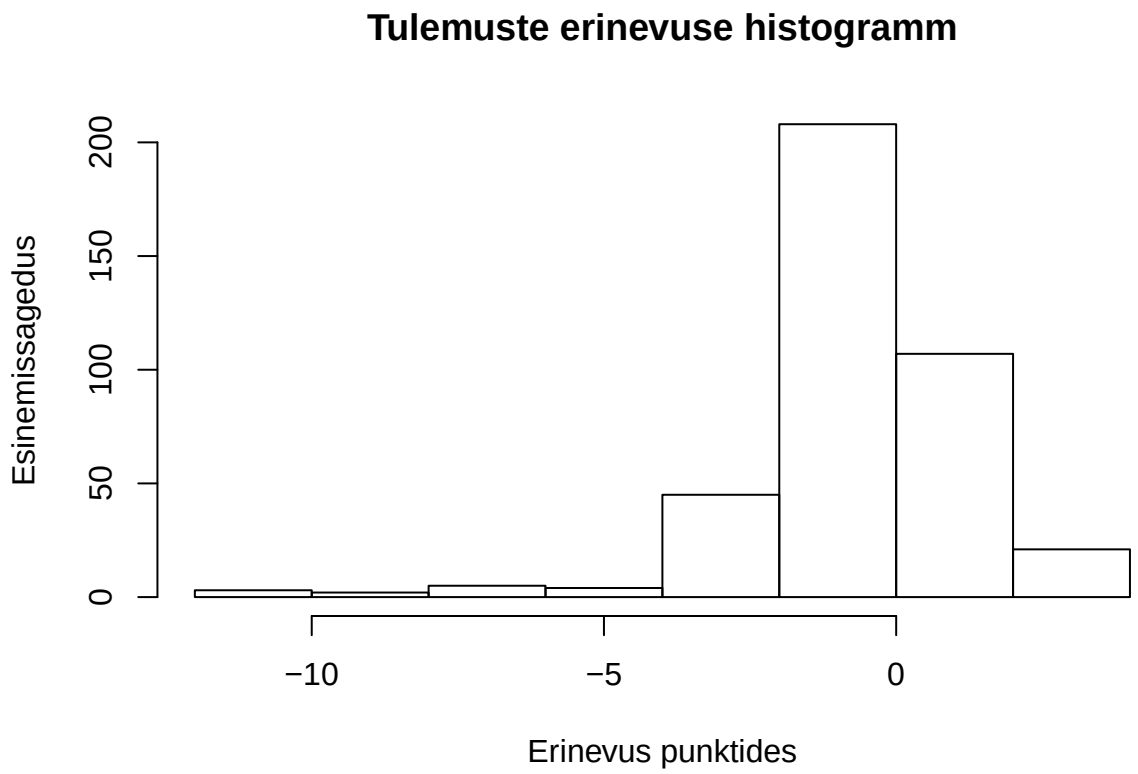
Jaotuse ülevaade

Arvudest paistab, et järgneval semestril läksid tulemused kuni 4 punkti tugevamaks või kuni 12 punkti nõrgemaks

```
summary(andmed$diff)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -12.0000  -1.0000   0.0000  -0.1949   1.0000   4.0000
```

```r
hist(andmed$diff, main="Tulemuste erinevuse histogramm", ylab="Esinemissagedus", xlab="Erinevus punktid
```



Tulemuste erinevuse väljatrükk

Ülemisel real on esinev väärtus, selle all väärtuse esinemissagedus Enim (117) jäi tulemusi samaks.

```r
table(andmed$diff)
```

```
##
## -12 -11 -10  -9  -8  -7  -6  -5  -4  -3  -2  -1   0   1   2   3   4
##   1   1   1   1   1   3   2   2   2   6  39  91 117  70  37  17   4
```

Keskmiselt tulemused halvenesid 0,19 punkti

```r
mean(andmed$diff)
```

```
## [1] -0.1949367
```

standardhälbega 1.97 punkti.

```r
sd(andmed$diff)
```

```
## [1] 1.971877
```

Mediaantulemus ei muutunud.

```
median(andmed$diff)
```

```
## [1] 0
```

Negatiivne assümmeetriakordaja (skewness) näitab, et jaotus on kallutatud vasakule, keskmisest tugevalt väiksemad väärtused nihutavad aritmeetilise keskmise mediaanist väiksemaks.

Funktsiooni *skewness* käivitamiseks tuleb laadida teek nimega e1091. Kui seda pole installeeritud, siis saab seda teha käsuga

install.packages("e1091")

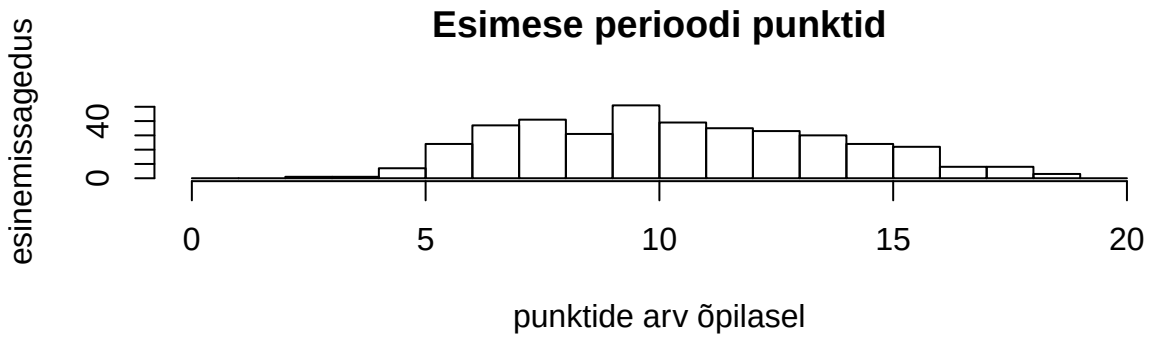
```
library(e1071)
skewness(andmed$diff)
```

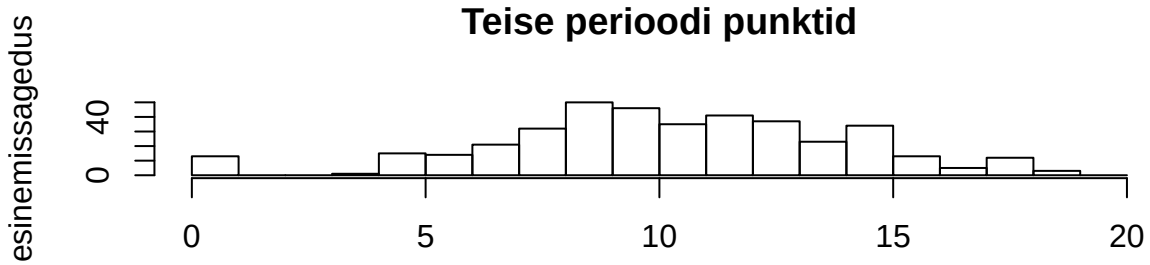
```
## [1] -1.890055
```

Tulemuste jaotus perioodiditi

Ühele joonistusalale üksteise alla kaks joonist (mfrow 2 rida, 1 veerg)

```
#knitr::opts_chunk$set(fig.asp = 2) #ei taha muutuda
#plot.window(c(0, 5), c(0, 10)) #ikka ei mõju
par(mfrow=c(2, 1))
hist(andmed$G1, xlim=c(0, 20), breaks=0:20, main="Esimese perioodi punktid", ylab="esinemissagedus", xla
hist(andmed$G2, xlim=c(0, 20), breaks=0:20, main="Teise perioodi punktid", ylab="esinemissagedus", xlab=
```





Joonistusrežiim tagasi ühele joonisele joonistusala kohta

```
par(mfrow=c(1,1))
```

Kahtlused lähteandmete tähenduses

Tulemuste esinemissagedused perioodiiti

```
  table(andmed$G1)
```

```
##
##  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19
##  1  1  7 24 37 41 31 51 39 35 33 30 24 22  8  8  3
```

```
  table(andmed$G2)
```

```
##
##  0  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19
## 13  1 15 14 21 32 50 46 35 41 37 23 34 13  5 12  3
```

Esimesel perioodil nullpunktilisi tulemusi polnud, teisel oli neid kogunisti kolmteist tükki. Võib kahtlustada, et nullid tähendavad hoopiski väärtuste puudumist.

Teisel perioodil puuduva tulemusega õpilaste esimese perioodi keskmine tulemus on kõikide õpilaste esimese perioodi keskmisest tulemusest rohkem kui kolme punkti jagu madalam.

```
mean( andmed$G1[andmed$G2==0])
```

```
## [1] 7.461538
```

```
mean(andmed$G1)
```

```
## [1] 10.90886
```

Võtame vastu otsuse asendada nullpunktilised tulemused puuduvate väärtustega (NA).

```
 andmed$G2[andmed$G2==0]=NA
 andmed$G2
```

```
##   [1]  6  5  8 14 10 15 12  5 18 15  8 12 14 10 16 14 14 10  5 10 14 15 15
##  [24] 13  9  9 12 16 11 12 11 16 16 10 14  7 16 16 12 13 10 12 18  8 10  8
##  [47] 12 19 15  7 13 13 11 10 13  9 15 15 10 16 11  8 10  9 10 15 13  7  9
##  [70] 16 15 10  6 12 12  9 11 11  8  5 12 10  6 15 10  9  7 14 10  6  7 17
##  [93]  6 10 13 10 15  9 14  9  7 17 13  6 18 11  8 18 13 15 19 10 13 19  9
## [116] 15 13 14  7 13 15 14 13 11  7 13 10  8  4 18 NA NA 13 11 NA NA NA NA
## [139] 12 16  9  9 11 14 NA 11  7 11  6  9  5 13 10 NA 11  8 12  8 15 12  6
## [162]  9 NA 10  8 11 10 15  7 14  5 15 11  7 11  9 13  5  8 10  8 13 17  9
## [185] 13 12 12 15  7  9 12  8  8  9 14 15 15  9 18  9 16 10  9  6 10  9  7
## [208] 12  9  7  8 12 13  7 10 15  6  6  7 10  6  5 16 13 13  8 15 11  8 10
## [231] 13 11  9 13  7  9 13 12 11  7 12 11 NA 12 NA 18 12  8  5 15  8 10  9
## [254]  9 12  9 12 11 14  9 18  8 12  9 10 17  9 10  9 NA  9 14 11 14 10 12
## [277]  9  9  8 11  8  9 12  9  9 10 18 12 14 13 11 15 12 18 13 12  9  8 13
## [300] 15 10 11 12 17 14 12 18  9 12 10  9 12 11 10 13 11  8 10 11 11 13  9
## [323] 11 14 15 12 15 10  9 14  8 14 NA  8  9 15 13  8 15 10 12 10 15  8 10
## [346] 13 15 10 15 13  7 13  7  8 11  9 13 12 10 16 13 12 11 15 11 10 13  6
## [369] 10 12  7 12 11  5 18  8 14  9 15 10 14  6 11  5  5  9  5  5  9  5  9
## [392] 16  8 12  9
```

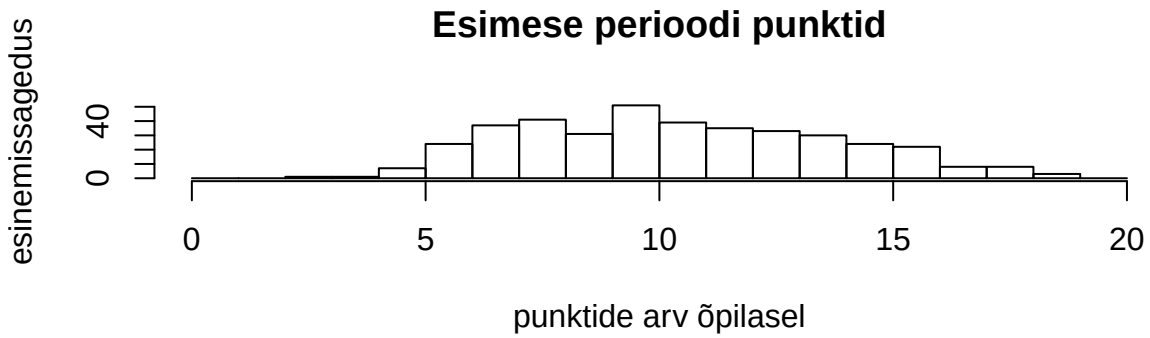
Tabelist võib näha, et neid on 13 tükki.

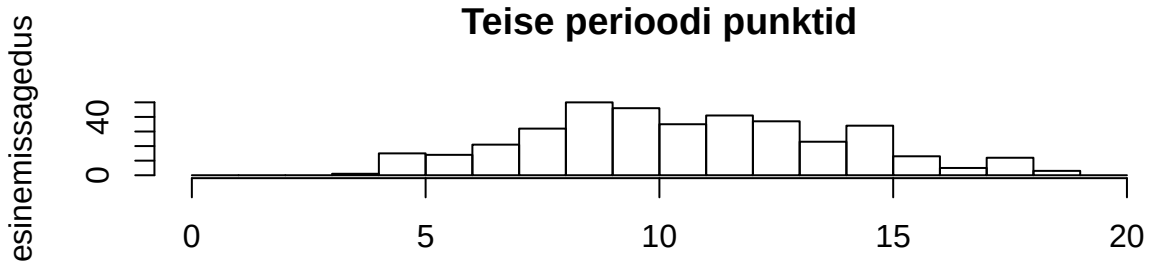
```
 table(andmed$G2, useNA="always")
```

```
##
##    4    5    6    7    8    9   10   11   12   13   14   15   16   17   18
##    1   15   14   21   32   50   46   35   41   37   23   34   13    5   12
##   19 <NA>
##    3   13
```

Nüüd histogrammi joonistades paistab, et pole enam teise perioodi nullilisi tulemusi.

```
par(mfrow=c(2, 1))
hist(andmed$G1, xlim=c(0, 20), breaks=0:20, main="Esimese perioodi punktid", ylab="esinemissagedus", xl
hist(andmed$G2, xlim=c(0, 20), breaks=0:20, main="Teise perioodi punktid", ylab="esinemissagedus", xlab=
```






```
par(mfrow=c(1,1))
```

Igaks juhuks puuduvate väärtuste arvu kontroll. Tulbas G2 oli endisi nulle ja nüüdseid NA-sid 13.

```
  sum(is.na(andmed$G2))
```

```
## [1] 13
```

```
  which(is.na(andmed$G2))
```

```
##  [1] 131 132 135 136 137 138 145 154 163 243 245 270 333
```

```
  length(which(is.na(andmed$G2)))
```

```
## [1] 13
```

Leiame õpilaste teise ja esimese perioodi punktide vahe uuesti. Kus üks väärtus puudub (NA), seal ka

```
andmed$diff=andmed$G2-andmed$G1
```

Kus üks väärtus puudub (NA), seal ka tulemuse juures NA

```
andmed$diff
```

```
##   [1]  1  0  1 -1  4  0  0 -1  2  1 -2  2  0  0  2  0  1  2 -1  2  1  3  0
##  [24]  0 -1  3  0  1  0  2  2 -1 -1  2  2 -1  1  1  0 -1  3  0 -1  0  0  0
##  [47]  1  0  0  0  1  2  0  2  3  1  1  1  1  1  1 -2  2 -1  0 -1  0  0  1
##  [70]  0  2  0 -2  0  1  0  0  0  0  0  2 -1 -1  0  1  2 -1  1 -1 -2  0  1
##  [93] -1 -1  2  3  4  1  3  2  0  1  3 -1  2  1  1  2  3  1  1  3  3  1  0
## [116]  0  2  1 -2 -1 -1 -2  0 -3 -1  0  3  1 -3  0 NA NA  3 -1 NA NA NA NA
## [139] -2  0  2  0  2  0 NA  3  1  1 -1  1 -1  1  0 NA  0 -3 -4 -1 -2  2 -1
## [162]  4 NA  0  3 -1  0  1  1  0 -1  2 -2 -1  1 -1  0 -1 -2  0 -1  1  1  0
## [185]  1  0  1  0 -1  1  1  0  1  1  1  1 -2  0  0  0  0  2  0 -1  0 -1  0
## [208]  1  0  0  0  0  1  1  2  1  0  0  0  1  0 -1  0  1  0 -1 -1 -1 -2 -2
## [231]  0  0 -2 -1 -2 -2 -1 -1 -2  0  0  1 NA -1 NA  0  0  2  2  2  2  3  3
## [254]  1  4  2 -2  0 -1 -1  1  0 -1 -1  1  0  0 -2 -1 NA  0 -1  0 -1  0  0
## [277] -1  0 -1  1  0 -2  0  1 -1 -2  0 -1 -1 -2 -1  0  0  0 -1 -2 -1 -2 -1
## [300] -1 -2  0 -3  0 -1 -2  1  1 -3 -2  0 -2 -2 -3 -2 -2  0  1  0  0  0 -2
## [323]  0  2 -1  3  1 -1 -1  0 -1  2 NA  0 -1 -1 -1  1 -1  1  1  0 -1 -1 -1
## [346]  0 -1  0  2  2 -1  0 -1  0 -2 -1  1  0  0 -2  0 -1  0 -1 -1  0  0 -1
## [369] -1 -2  0 -2 -2 -1 -1  0 -1  1  0  0 -1 -1  0 -1 -1 -1 -1 -2  2 -1  0
## [392]  2 -2  1  1
```

Väärtuste nüüdsed sagedused. Endiselt on enim õpilasi, kelle punktide arv jäi mõlemas perioodis samaks. Ühel õpilasel langes nelja võrra, neljal õpilasel tõusis nelja võrra, teistel muutus vähem.

```
table(andmed$diff)
```

```
##
##  -4  -3  -2  -1   0   1   2   3   4
##   1   6  39  91 117  70  37  17   4
```

Soovides näha ka puuduvaid väärtusi, tuleb see eraldi märkida

```
table(andmed$diff, useNA = "always")
```

```
##
##   -4   -3   -2   -1    0    1    2    3    4 <NA>
##    1    6   39   91  117   70   37   17    4   13
```

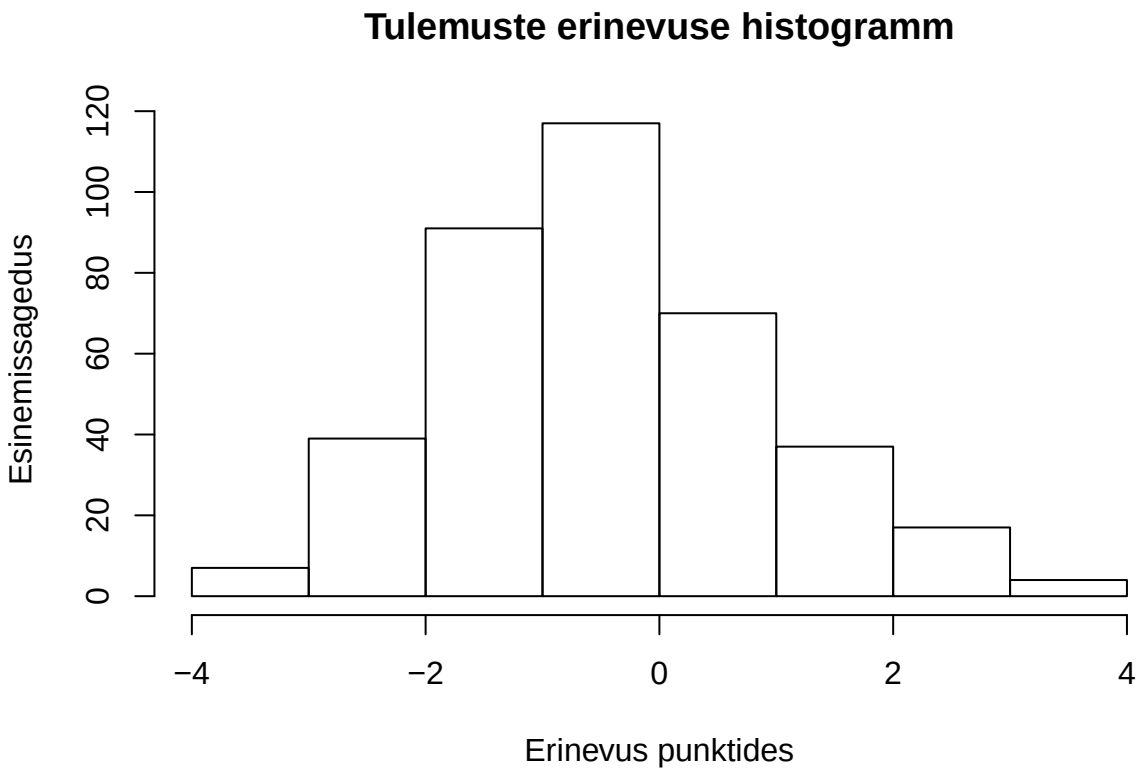
Erinevuste jaotus. Pikk miinustega "saba" on kadunud. Erinevuste nõrgalt positiivne aritmeetiline keskmine näitab, et punktide arv teisel perioodil on pigem veidi suurem.

```
summary(andmed$diff)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.     NA's
## -4.00000 -1.00000  0.00000  0.05236  1.00000  4.00000       13
```

ning histogramm

```
hist(andmed$diff, main="Tulemuste erinevuse histogramm", ylab="Esinemissagedus", xlab="Erinevus punktid
```



Keskmise

```
  mean(andmed$diff, na.rm=TRUE)
```

```
## [1] 0.05235602
```

standardhälbe

```
  sd(andmed$diff, na.rm=TRUE)
```

```
## [1] 1.403925
```

ja vahemiku leidmiseks peab eraldi ütlema, et puuduvad (NA) väärtused eemaldatakse arvutamisel.

```
  range(andmed$diff, na.rm=TRUE)
```

```
## [1] -4  4
```

Rühmitamine

Jagame vahed gruppidesse vastavalt kvantiilile

```
grupid=cut(andmed$diff, quantile(andmed$diff, na.rm=TRUE))
```

Igale algsele väärtusele vastab nüüd vahemik, millese ta kuulub.

```
grupid
```

```
##   [1] (0,1]   (-1,0]  (0,1]   (-4,-1] (1,4]   (-1,0]  (-1,0]  (-4,-1]
##   [9] (1,4]   (0,1]   (-4,-1] (1,4]   (-1,0]  (-1,0]  (1,4]   (-1,0]
##  [17] (0,1]   (1,4]   (-4,-1] (1,4]   (0,1]   (1,4]   (-1,0]  (-1,0]
##  [25] (-4,-1] (1,4]   (-1,0]  (0,1]   (-1,0]  (1,4]   (1,4]   (-4,-1]
##  [33] (-4,-1] (1,4]   (1,4]   (-4,-1] (0,1]   (0,1]   (-1,0]  (-4,-1]
##  [41] (1,4]   (-1,0]  (-4,-1] (-1,0]  (-1,0]  (-1,0]  (0,1]   (-1,0]
##  [49] (-1,0]  (-1,0]  (0,1]   (1,4]   (-1,0]  (1,4]   (1,4]   (0,1]
##  [57] (0,1]   (0,1]   (0,1]   (0,1]   (0,1]   (-4,-1] (1,4]   (-4,-1]
##  [65] (-1,0]  (-4,-1] (-1,0]  (-1,0]  (0,1]   (-1,0]  (1,4]   (-1,0]
##  [73] (-4,-1] (-1,0]  (0,1]   (-1,0]  (-1,0]  (-1,0]  (-1,0]  (-1,0]
##  [81] (1,4]   (-4,-1] (-4,-1] (-1,0]  (0,1]   (1,4]   (-4,-1] (0,1]
##  [89] (-4,-1] (-4,-1] (-1,0]  (0,1]   (-4,-1] (-4,-1] (1,4]   (1,4]
##  [97] (1,4]   (0,1]   (1,4]   (1,4]   (-1,0]  (0,1]   (1,4]   (-4,-1]
## [105] (1,4]   (0,1]   (0,1]   (1,4]   (1,4]   (0,1]   (0,1]   (1,4]
## [113] (1,4]   (0,1]   (-1,0]  (-1,0]  (1,4]   (0,1]   (-4,-1] (-4,-1]
## [121] (-4,-1] (-4,-1] (-1,0]  (-4,-1] (-4,-1] (-1,0]  (1,4]   (0,1]
## [129] (-4,-1] (-1,0]  <NA>    <NA>    (1,4]   (-4,-1] <NA>    <NA>
## [137] <NA>    <NA>    (-4,-1] (-1,0]  (1,4]   (-1,0]  (1,4]   (-1,0]
## [145] <NA>    (1,4]   (0,1]   (0,1]   (-4,-1] (0,1]   (-4,-1] (0,1]
## [153] (-1,0]  <NA>    (-1,0]  (-4,-1] <NA>    (-4,-1] (-4,-1] (1,4]
## [161] (-4,-1] (1,4]   <NA>    (-1,0]  (1,4]   (-4,-1] (-1,0]  (0,1]
## [169] (0,1]   (-1,0]  (-4,-1] (1,4]   (-4,-1] (-4,-1] (0,1]   (-4,-1]
## [177] (-1,0]  (-4,-1] (-4,-1] (-1,0]  (-4,-1] (0,1]   (0,1]   (-1,0]
## [185] (0,1]   (-1,0]  (0,1]   (-1,0]  (-4,-1] (0,1]   (0,1]   (-1,0]
## [193] (0,1]   (0,1]   (0,1]   (0,1]   (-4,-1] (-1,0]  (-1,0]  (-1,0]
## [201] (-1,0]  (1,4]   (-1,0]  (-4,-1] (-1,0]  (-4,-1] (-1,0]  (0,1]
## [209] (-1,0]  (-1,0]  (-1,0]  (-1,0]  (0,1]   (0,1]   (1,4]   (0,1]
## [217] (-1,0]  (-1,0]  (-1,0]  (0,1]   (-1,0]  (-4,-1] (-1,0]  (0,1]
## [225] (-1,0]  (-4,-1] (-4,-1] (-4,-1] (-4,-1] (-4,-1] (-1,0]  (-1,0]
## [233] (-4,-1] (-4,-1] (-4,-1] (-4,-1] (-4,-1] (-4,-1] (-4,-1] (-1,0]
## [241] (-1,0]  (0,1]   <NA>    (-4,-1] <NA>    (-1,0]  (-1,0]  (1,4]
## [249] (1,4]   (1,4]   (1,4]   (1,4]   (1,4]   (0,1]   (1,4]   (1,4]
## [257] (-4,-1] (-1,0]  (-4,-1] (-4,-1] (0,1]   (-1,0]  (-4,-1] (-4,-1]
## [265] (0,1]   (-1,0]  (-1,0]  (-4,-1] (-4,-1] <NA>    (-1,0]  (-4,-1]
## [273] (-1,0]  (-4,-1] (-1,0]  (-1,0]  (-4,-1] (-1,0]  (-4,-1] (0,1]
## [281] (-1,0]  (-4,-1] (-1,0]  (0,1]   (-4,-1] (-4,-1] (-1,0]  (-4,-1]
## [289] (-4,-1] (-4,-1] (-4,-1] (-1,0]  (-1,0]  (-1,0]  (-4,-1] (-4,-1]
## [297] (-4,-1] (-4,-1] (-4,-1] (-4,-1] (-4,-1] (-1,0]  (-4,-1] (-1,0]
## [305] (-4,-1] (-4,-1] (0,1]   (0,1]   (-4,-1] (-4,-1] (-1,0]  (-4,-1]
## [313] (-4,-1] (-4,-1] (-4,-1] (-4,-1] (-1,0]  (0,1]   (-1,0]  (-1,0]
## [321] (-1,0]  (-4,-1] (-1,0]  (1,4]   (-4,-1] (1,4]   (0,1]   (-4,-1]
## [329] (-4,-1] (-1,0]  (-4,-1] (1,4]   <NA>    (-1,0]  (-4,-1] (-4,-1]
## [337] (-4,-1] (0,1]   (-4,-1] (0,1]   (0,1]   (-1,0]  (-4,-1] (-4,-1]
## [345] (-4,-1] (-1,0]  (-4,-1] (-1,0]  (1,4]   (1,4]   (-4,-1] (-1,0]
## [353] (-4,-1] (-1,0]  (-4,-1] (-4,-1] (0,1]   (-1,0]  (-1,0]  (-4,-1]
## [361] (-1,0]  (-4,-1] (-1,0]  (-4,-1] (-4,-1] (-1,0]  (-1,0]  (-4,-1]
## [369] (-4,-1] (-4,-1] (-1,0]  (-4,-1] (-4,-1] (-4,-1] (-4,-1] (-1,0]
```

```
## [377] (-4,-1] (0,1]   (-1,0]  (-1,0]  (-4,-1] (-4,-1] (-1,0]  (-4,-1]
## [385] (-4,-1] (-4,-1] (-4,-1] (-4,-1] (1,4]   (-4,-1] (-1,0]  (1,4]
## [393] (-4,-1] (0,1]   (0,1]
## Levels: (-4,-1] (-1,0] (0,1] (1,4]
```

Gruppidesse jaotumise sagedus

```
table(grupid)
```

```
## grupid
## (-4,-1]  (-1,0]   (0,1]   (1,4]
##     136     117      70      58
```

Puuduvad väärtused tuleb eraldi välja küsida. Neid on aga üllatuslikult endisest ühe võrra rohkem.

```
table(grupid, useNA="always")
```

```
## grupid
## (-4,-1]  (-1,0]   (0,1]   (1,4]    <NA>
##     136     117      70      58      14
```

Kontrollime igaks juhuks üle. Käsk is.na annab tõeväärtusmassiivi, kas loetelus vastav väärtus puudub või mitte. Käsk which näitab tõeste väärtuste järjekorranumbrid ning length loeb lihtsalt kokku, et mitu neid on.

```
length(which(is.na(grupid)))
```

```
## [1] 14
```

Tasub uurima hakata, et milline väärtus siis veel lisaks kaduma läks peale algsete NA-de. Puuduva grupitunnusega elementide numbrid

```
which(is.na(grupid))
```

```
##  [1] 131 132 135 136 137 138 145 154 157 163 243 245 270 333
```

Puuduvad elemendid perioodi tulemuste vahedes.

```
which(is.na(andmed$diff))
```

```
##  [1] 131 132 135 136 137 138 145 154 163 243 245 270 333
```

Kuna nad on järjestatud, siis õnnestub ka näpuga järge ajades erinev koht üles leida. Et aga valmis olla suuremate andmete puhul sarnase erinevuse leidmiseks, siis selleks leiab hulkadevahelise erinevuse leidmise käskluse nimega setdiff

```
setdiff(which(is.na(grupid)), which(is.na(andmed$diff)))
```

```
## [1] 157
```

Uurime, et mis väärtus see erinevana leitud 157ndal kohal siis on

```
andmed$diff[157]
```

```
## [1] -4
```

Selgub, et alumises servas asuv väärtus. Hulgatunnustes pole ilmaasjata üks sulg kandiline ja teine ümar. Kandiline tähendab kaasavõetavat osa, ümarsulg välja jäetavat osa. Kõik vahemikud praegusel korral sarnaselt nõnda, et parempoolne väärtus võetakse sisse, vasakpoolne jäetakse välja. Nii jäi ka vähim väärtus gruppi määramata.

```
table(andmed$diff, useNA="always")
```

```
##
##   -4   -3   -2   -1    0    1    2    3    4 <NA>
##    1    6   39   91  117   70   37   17    4   13
```

Üksiku õnnetu väärtuse kaasa võtmiseks on cut-funktsioonil eraldi parameeter.

```
grupid=cut(andmed$diff, quantile(andmed$diff, na.rm=TRUE), include.lowest = TRUE)
```

Nüüd piirduvad puuduvad grupiväärtused vaid nendega, mis juba vahede tulbas puudu olid.

```
table(grupid, useNA = "always")
```

```
## grupid
## [-4,-1]  (-1,0]   (0,1]   (1,4]    <NA>
##     137     117      70      58      13
```

Väike kõrvalepõige

Omaette mõtisklemise koht neljaks jaotamise juures on, et suur ühesuguste nullide vägi ühendatakse nullist väiksemate grupiga, ehkki ei saa väita, et vastavate õpilaste tulemused kindla peale langenud oleksid. Mõned küll väidavad, et paigaltammumine tähendab tagasiminekut, kuid tõenäoliselt olid uue perioodi ülesanded ka eelmise omast raskemad.

Üheks võimaluseks ühesuguste väärtuste korraga ühte gruppi kukkumist tasandada on väärtustele lisada väike juhuslik komponent. Et juhuarvudega tehtud tulemused järgmisel näitamiskorral samasugused tuleksid, aitab käsklus set.seed miski numbriga.

```
set.seed(1500)
```

Nüüd paras luua väikese värina abil nihutatud tulemused

```
nihutatud=jitter(andmed$diff)
```

neid tulemusi endid vaadata

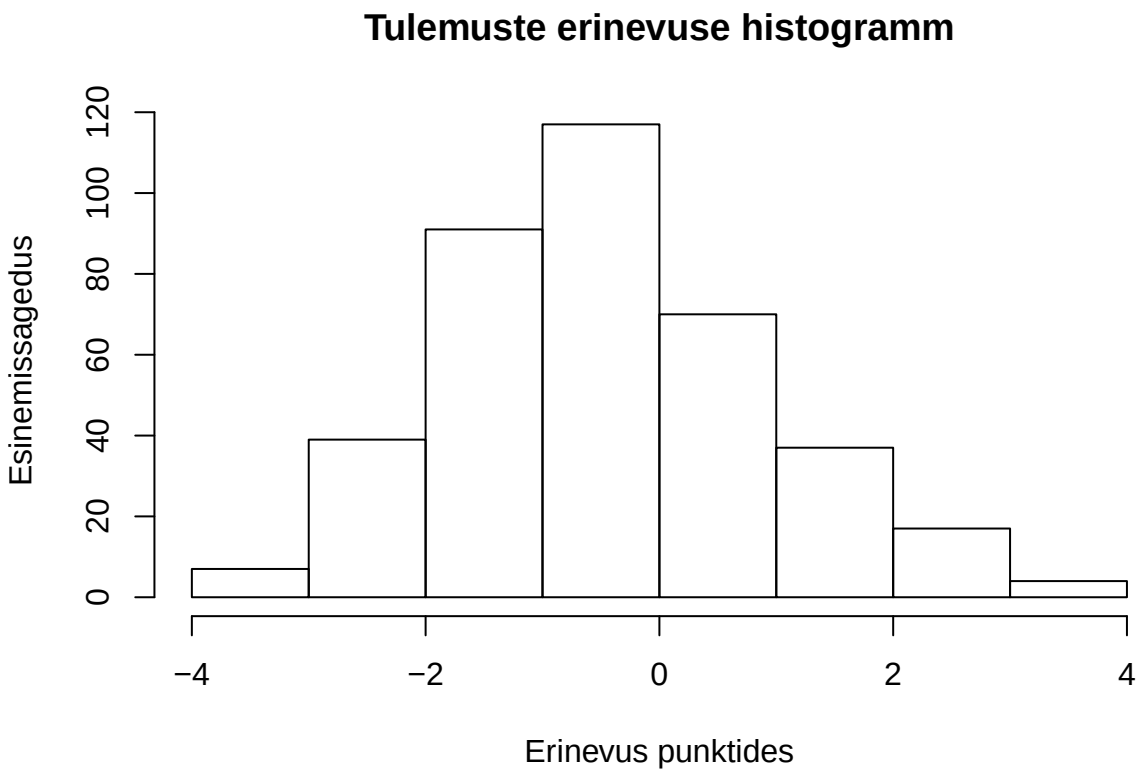
```
nihutatud
```

```
##   [1]  1.049581827  0.188887678  1.064013614 -1.063080093  3.924238487
##   [6] -0.096570377 -0.041450896 -0.814823377  2.179742473  1.125311131
##  [11] -2.117220134  2.156151024  0.004438476 -0.028330449  1.886066935
##  [16] -0.179910322  1.083940174  2.069104446 -0.908263080  2.005513802
##  [21]  0.996486250  2.902192330  0.188818343 -0.157646484 -1.040535865
##  [26]  3.092978026 -0.109472774  1.057225931  0.085056747  2.099271160
##  [31]  1.814486679 -1.041917764 -0.825031326  1.923438716  1.899510473
##  [36] -0.935799670  1.108903735  1.026485831 -0.132560918 -0.909421001
##  [41]  2.828549393 -0.198498877 -0.951647931  0.087302125 -0.088507898
##  [46] -0.174628084  1.193757961  0.193560861  0.056948496 -0.138764586
##  [51]  1.049723508  2.138367876  0.037937710  2.023505515  3.067908350
##  [56]  1.061396905  0.825406255  0.924207853  0.807013839  1.067472758
##  [61]  1.192093502 -2.002869319  2.154898640 -0.862637723 -0.132791445
##  [66] -1.158952599 -0.196025862  0.158851476  1.195388175  0.126270207
##  [71]  1.962813898  0.186114741 -1.884900367 -0.149366896  0.873756673
##  [76]  0.110183477 -0.105415265 -0.158542123  0.176203583 -0.169141130
##  [81]  1.893400438 -0.939360067 -0.937257849 -0.184952801  1.018607424
##  [86]  1.827666824 -0.988561171  0.941088121 -1.070696497 -2.173448403
##  [91]  0.179481953  1.094866905 -1.115336695 -0.842385381  1.946580031
##  [96]  3.054641535  3.881838909  1.020632992  3.104045319  1.933372540
## [101] -0.157285249  0.907362048  3.028761408 -1.092156368  2.152179573
## [106]  1.190796437  0.900028351  2.015536629  2.849004422  0.944960949
## [111]  0.909129729  2.927129033  2.883516716  1.031564905 -0.196947686
## [116] -0.066206524  2.133776593  1.048966224 -1.909644196 -1.141300569
## [121] -0.925687062 -1.986914271 -0.076645445 -2.879888311 -1.052957657
## [126] -0.087682587  3.194771710  1.116293303 -2.805280211  0.190804572
## [131]           NA           NA  3.064147940 -0.948641070           NA
## [136]           NA           NA           NA -2.018968777  0.046301419
## [141]  2.061026645  0.160205874  1.851239908 -0.041970396           NA
## [146]  2.810505467  1.075257839  0.970922963 -1.180722203  0.819362174
## [151] -0.967939083  1.165580549  0.167689631           NA -0.146068088
## [156] -2.970539807 -4.017320882 -1.027565933 -2.031482222  2.020901402
## [161] -1.028456900  4.100338343           NA  0.132564192  3.110923776
## [166] -1.133200169 -0.071114315  1.076753626  1.171970790 -0.049009511
## [171] -1.041152745  1.997411383 -1.810428361 -0.953512385  1.060370492
## [176] -0.844824756  0.174855424 -1.003930368 -2.044774393  0.023802605
## [181] -0.992008818  1.059947803  0.930816553  0.178842132  0.910565909
## [186] -0.096137356  1.086641320 -0.023042681 -0.898463427  1.157766152
## [191]  1.045520592  0.159553617  0.942235280  0.992694954  0.985038531
## [196]  0.909540773 -2.060635905 -0.008234704 -0.074390331 -0.087619125
## [201] -0.056047716  2.143712895 -0.107370792 -1.030442847  0.038966878
## [206] -0.828595263 -0.011125923  0.869242706  0.073927994  0.152096901
## [211] -0.108061080  0.111706065  1.136259206  1.075617559  2.036525962
## [216]  0.983750790 -0.058175294  0.054767810  0.120040402  1.019306634
## [221]  0.131610590 -1.023790877 -0.060661807  1.011982668  0.015115139
## [226] -0.956796957 -1.115785758 -0.998199879 -2.195668474 -2.047840808
## [231] -0.068814463 -0.125541728 -1.916616319 -1.104084857 -2.139201067
## [236] -1.829299191 -1.017242718 -0.937676231 -2.176680783 -0.109961166
## [241]  0.080512663  1.036147534           NA -1.112395199           NA
## [246] -0.192728626  0.161911982  1.948450134  1.971928771  2.173903164
## [251]  1.814104587  3.162521229  3.063506449  1.177350157  3.900799299
## [256]  2.092008380 -2.150352952 -0.010411702 -0.916696571 -1.019864017
## [261]  0.840004088 -0.002428806 -0.981703841 -1.184718540  0.908507863
## [266]  0.041902993 -0.178129492 -1.807195853 -0.988242771           NA
```

```
## [271] -0.164671437 -0.819687404 -0.198758093 -0.840157106 -0.159125021
## [276]  0.048909492 -1.032027326  0.167437002 -0.816816689  1.096353934
## [281]  0.068433780 -1.850352089 -0.157684339  0.843106793 -0.943119261
## [286] -2.000270512  0.078127599 -0.954454412 -0.884378560 -2.090916882
## [291] -1.147774499 -0.192382540  0.008273203 -0.173800771 -0.821897470
## [296] -1.850473359 -1.101933991 -1.957765079 -0.896329196 -0.977151691
## [301] -2.101599617 -0.068997154 -2.968787468 -0.178591401 -0.860148568
## [306] -2.064590754  1.051845327  0.984430888 -2.808695294 -1.969424480
## [311] -0.019008039 -2.119226486 -1.869637909 -3.017793390 -1.861523210
## [316] -1.818656204 -0.019525557  1.034116002  0.187374338 -0.002846245
## [321]  0.150080109 -1.977825812  0.084813505  2.182233269 -0.997278493
## [326]  3.095881122  1.100046629 -0.980539675 -1.005045310  0.097589922
## [331] -1.103768834  1.858268170           NA -0.045569695 -1.169269831
## [336] -0.971189181 -1.116109786  1.123213986 -0.864416280  1.169185386
## [341]  1.140851574 -0.069770072 -0.902827477 -0.968872402 -1.167788202
## [346]  0.125163250 -1.054364593  0.170456999  2.086060740  1.988819539
## [351] -1.164622905 -0.168708288 -1.081029833  0.042724032 -1.999957692
## [356] -0.923732447  1.068705360  0.081146581  0.055819002 -2.180104529
## [361] -0.035473796 -1.158446300 -0.012772622 -0.984497622 -0.831733395
## [366] -0.015324822  0.107720293 -0.944637359 -1.015927002 -1.922469319
## [371] -0.151553925 -1.910100480 -1.894280915 -0.968339850 -0.959710261
## [376] -0.105884782 -0.962877761  1.123243042  0.089720841 -0.104181824
## [381] -1.139252951 -1.045097766 -0.105270220 -0.805168929 -1.035716262
## [386] -0.829987923 -0.821897218 -2.106462735  2.099016678 -0.915180590
## [391] -0.138847488  2.130260871 -2.030152697  0.913385741  1.036624112
```

Ning jälgida, et jaotus iseenesest pole märgatavalt muutunud

```
hist(andmed$diff, main="Tulemuste erinevuse histogramm", ylab="Esinemissagedus", xlab="Erinevus punktid
```



Kvantiilide järgi gruppidesse paigutamine tuleb nüüd aga ühtlasem.

```
nihkegrupid=cut(nihutatud, quantile(nihutatud, na.rm=TRUE), include.lowest = TRUE)
```

Saab vaadata, kui palju kusagile gruppi läks ning suur nullide osa enam ei tekita anomaaliat ühes suunas.

```
table(nihkegrupid, useNA="always")
```

```
## nihkegrupid
##   [-4.02,-0.969] (-0.969,-0.0417]   (-0.0417,1.04]       (1.04,4.1]
##               96               95               95               96
##             <NA>
##               13
```

Tagasi algsete gruppide juurde

Gruppidele selgituseks nimed

```
grupid=cut(andmed$diff, quantile(andmed$diff, na.rm=TRUE), include.lowest = TRUE, labels=c("palju lange
```

Kontrolliks väärtuste väljatrükk

```
grupid
```

```
##   [1] tõusev            kuni veidi langev tõusev
##   [4] palju langev      palju tõusev      kuni veidi langev
##   [7] kuni veidi langev palju langev      palju tõusev
##  [10] tõusev            palju langev      palju tõusev
##  [13] kuni veidi langev kuni veidi langev palju tõusev
##  [16] kuni veidi langev tõusev            palju tõusev
##  [19] palju langev      palju tõusev      tõusev
##  [22] palju tõusev      kuni veidi langev kuni veidi langev
##  [25] palju langev      palju tõusev      kuni veidi langev
##  [28] tõusev            kuni veidi langev palju tõusev
##  [31] palju tõusev      palju langev      palju langev
##  [34] palju tõusev      palju tõusev      palju langev
##  [37] tõusev            tõusev            kuni veidi langev
##  [40] palju langev      palju tõusev      kuni veidi langev
##  [43] palju langev      kuni veidi langev kuni veidi langev
##  [46] kuni veidi langev tõusev            kuni veidi langev
##  [49] kuni veidi langev kuni veidi langev tõusev
##  [52] palju tõusev      kuni veidi langev palju tõusev
##  [55] palju tõusev      tõusev            tõusev
##  [58] tõusev            tõusev            tõusev
##  [61] tõusev            palju langev      palju tõusev
##  [64] palju langev      kuni veidi langev palju langev
##  [67] kuni veidi langev kuni veidi langev tõusev
##  [70] kuni veidi langev palju tõusev      kuni veidi langev
##  [73] palju langev      kuni veidi langev tõusev
##  [76] kuni veidi langev kuni veidi langev kuni veidi langev
##  [79] kuni veidi langev kuni veidi langev palju tõusev
##  [82] palju langev      palju langev      kuni veidi langev
##  [85] tõusev            palju tõusev      palju langev
##  [88] tõusev            palju langev      palju langev
##  [91] kuni veidi langev tõusev            palju langev
##  [94] palju langev      palju tõusev      palju tõusev
##  [97] palju tõusev      tõusev            palju tõusev
## [100] palju tõusev      kuni veidi langev tõusev
## [103] palju tõusev      palju langev      palju tõusev
## [106] tõusev            tõusev            palju tõusev
## [109] palju tõusev      tõusev            tõusev
## [112] palju tõusev      palju tõusev      tõusev
## [115] kuni veidi langev kuni veidi langev palju tõusev
## [118] tõusev            palju langev      palju langev
## [121] palju langev      palju langev      kuni veidi langev
## [124] palju langev      palju langev      kuni veidi langev
## [127] palju tõusev      tõusev            palju langev
## [130] kuni veidi langev <NA>              <NA>
## [133] palju tõusev      palju langev      <NA>
## [136] <NA>              <NA>              <NA>
## [139] palju langev      kuni veidi langev palju tõusev
## [142] kuni veidi langev palju tõusev      kuni veidi langev
## [145] <NA>              palju tõusev      tõusev
## [148] tõusev            palju langev      tõusev
## [151] palju langev      tõusev            kuni veidi langev
## [154] <NA>              kuni veidi langev palju langev
## [157] palju langev      palju langev      palju langev
## [160] palju tõusev      palju langev      palju tõusev
```

```
## [163] <NA>              kuni veidi langev palju tõusev
## [166] palju langev      kuni veidi langev tõusev
## [169] tõusev            kuni veidi langev palju langev
## [172] palju tõusev      palju langev      palju langev
## [175] tõusev            palju langev      kuni veidi langev
## [178] palju langev      palju langev      kuni veidi langev
## [181] palju langev      tõusev            tõusev
## [184] kuni veidi langev tõusev            kuni veidi langev
## [187] tõusev            kuni veidi langev palju langev
## [190] tõusev            tõusev            kuni veidi langev
## [193] tõusev            tõusev            tõusev
## [196] tõusev            palju langev      kuni veidi langev
## [199] kuni veidi langev kuni veidi langev kuni veidi langev
## [202] palju tõusev      kuni veidi langev palju langev
## [205] kuni veidi langev palju langev      kuni veidi langev
## [208] tõusev            kuni veidi langev kuni veidi langev
## [211] kuni veidi langev kuni veidi langev tõusev
## [214] tõusev            palju tõusev      tõusev
## [217] kuni veidi langev kuni veidi langev kuni veidi langev
## [220] tõusev            kuni veidi langev palju langev
## [223] kuni veidi langev tõusev            kuni veidi langev
## [226] palju langev      palju langev      palju langev
## [229] palju langev      palju langev      kuni veidi langev
## [232] kuni veidi langev palju langev      palju langev
## [235] palju langev      palju langev      palju langev
## [238] palju langev      palju langev      kuni veidi langev
## [241] kuni veidi langev tõusev            <NA>
## [244] palju langev      <NA>              kuni veidi langev
## [247] kuni veidi langev palju tõusev      palju tõusev
## [250] palju tõusev      palju tõusev      palju tõusev
## [253] palju tõusev      tõusev            palju tõusev
## [256] palju tõusev      palju langev      kuni veidi langev
## [259] palju langev      palju langev      tõusev
## [262] kuni veidi langev palju langev      palju langev
## [265] tõusev            kuni veidi langev kuni veidi langev
## [268] palju langev      palju langev      <NA>
## [271] kuni veidi langev palju langev      kuni veidi langev
## [274] palju langev      kuni veidi langev kuni veidi langev
## [277] palju langev      kuni veidi langev palju langev
## [280] tõusev            kuni veidi langev palju langev
## [283] kuni veidi langev tõusev            palju langev
## [286] palju langev      kuni veidi langev palju langev
## [289] palju langev      palju langev      palju langev
## [292] kuni veidi langev kuni veidi langev kuni veidi langev
## [295] palju langev      palju langev      palju langev
## [298] palju langev      palju langev      palju langev
## [301] palju langev      kuni veidi langev palju langev
## [304] kuni veidi langev palju langev      palju langev
## [307] tõusev            tõusev            palju langev
## [310] palju langev      kuni veidi langev palju langev
## [313] palju langev      palju langev      palju langev
## [316] palju langev      kuni veidi langev tõusev
## [319] kuni veidi langev kuni veidi langev kuni veidi langev
## [322] palju langev      kuni veidi langev palju tõusev
```

```
## [325] palju langev      palju tõusev      tõusev
## [328] palju langev      palju langev      kuni veidi langev
## [331] palju langev      palju tõusev      <NA>
## [334] kuni veidi langev palju langev      palju langev
## [337] palju langev      tõusev            palju langev
## [340] tõusev            tõusev            kuni veidi langev
## [343] palju langev      palju langev      palju langev
## [346] kuni veidi langev palju langev      kuni veidi langev
## [349] palju tõusev      palju tõusev      palju langev
## [352] kuni veidi langev palju langev      kuni veidi langev
## [355] palju langev      palju langev      tõusev
## [358] kuni veidi langev kuni veidi langev palju langev
## [361] kuni veidi langev palju langev      kuni veidi langev
## [364] palju langev      palju langev      kuni veidi langev
## [367] kuni veidi langev palju langev      palju langev
## [370] palju langev      kuni veidi langev palju langev
## [373] palju langev      palju langev      palju langev
## [376] kuni veidi langev palju langev      tõusev
## [379] kuni veidi langev kuni veidi langev palju langev
## [382] palju langev      kuni veidi langev palju langev
## [385] palju langev      palju langev      palju langev
## [388] palju langev      palju tõusev      palju langev
## [391] kuni veidi langev palju tõusev      palju langev
## [394] tõusev            tõusev
## Levels: palju langev kuni veidi langev tõusev palju tõusev
```

Puuduvate arvu lugemine - salamisi pole enam juurde tulnud

```
length(which(is.na(grupid)))
```

```
## [1] 13
```

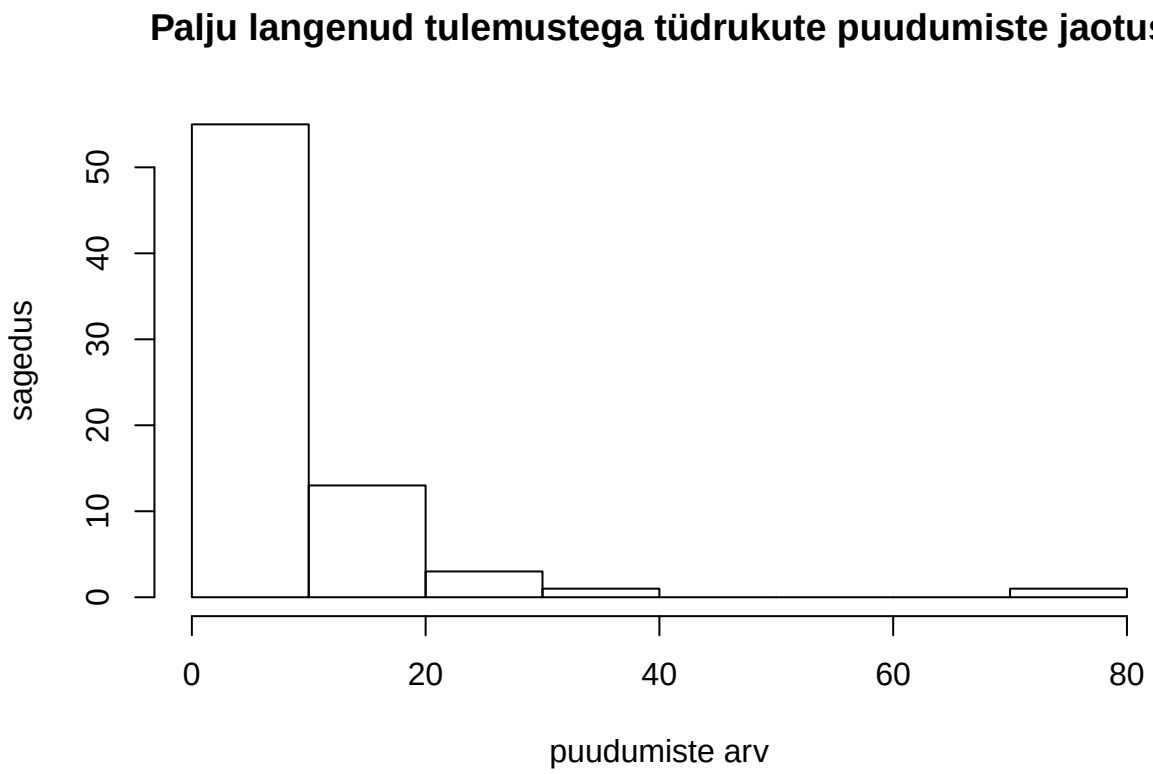
Edasise analüüsi huvides lisame grupimuutuja olemasoleva andmetabeli külge.

```
andmed$grupid=grupid
```

Statistika õpilaste kohta

Senine töö oli pelgalt ettevalmistus. Nüüd võimalik vaatama hakata, et kas ja mida võimalik neist andmetest välja lugeda.

```
hist(andmed[andmed$sex=="F" & andmed$grupid=="palju langev","absences"],  main="Palju langenud tulemust
```



Kahe madalama, langenud tulemustega grupi keskmine puudumiste arv

```
mean(andmed[andmed$grupid=="kuni veidi langev" | andmed$grupid=="palju langev","absences"], na.rm=TRUE)
```

```
## [1] 6.472441
```

Madalamate gruppide poiste tulemuste keskmine

```
mean(andmed[andmed$sex=="M" & (andmed$grupid=="kuni veidi langev" | andmed$grupid=="palju langev"),"abs
```

```
## [1] 6
```

Kaks alumist gruppi numbriloetelu abil välja võetuna

```
mean(andmed[andmed$sex=="M" & (as.numeric(andmed$grupid) %in% 1:2),"absences"], na.rm=TRUE)
```

```
## [1] 6
```

Poiste suurim puudumiste arv

```
max(subset(andmed, sex=="M")$absences)
```

```
## [1] 38
```

Tüdrukute puudumised, mis ületavad poiste puudumiste maksimumarvu

```
subset(andmed, sex=="F" & absences>max(subset(andmed, sex=="M")$absences))$absences
```

```
## [1] 54 56 75 40
```

Nende arv

```
length(subset(andmed, sex=="F" & absences>max(subset(andmed, sex=="M")$absences))$absences)
```

```
## [1] 4
```

Sama mitme lühema käskluse abil

```
pp=subset(andmed, sex=="M")$absences # poiste puudumised
tp=subset(andmed, sex=="F")$absences # tüdrukute puudumised
tp[tp>max(pp)] #tüdrukute puudumised, mis ületavad poiste suurimat puudumiste arvu
```

```
## [1] 54 56 75 40
```