

R fyrir byrjendur

Lýsistærðir

Anna Helga Jónsdóttir
Bjarki Þór Elvarsson
Sigrún Helga Lund

Tíðnitöflur: `table()`

- ▶ `table()` skipunin býr til einfaldar tíðnitöflur
- ▶ Mötum aðferðina með nöfnunum á einni eða fleiri flokkabreytum
- ▶ Gott skoða `help(table)`

Búum til nýja flokkabreytu sem lýsir líkamsræktarástundun:

```
puls <- mutate(puls, likamsraekt.flokkur=cut(likamsraekt,  
      c(0, 3, 5, max(likamsraekt, na.rm=T)+1), right=F),  
      likamsraekt.flokkur = fct_recode(likamsraekt.flokkur,  
      Lítil = "[0,3)", Meðal = "[3,5)", Mikil = "[5,26)"))
```

Tíðnitöflur: table()

```
table(puls$namskeid)

##
##  LAN203 STAE209
##    172    299

table(puls$namskeid,puls$likamsraekt.flokkur)

##
##           Lítil Meðal Mikil
##  LAN203      45    52    74
##  STAE209     95    84   116
```

Viljum heldur sýna hlutföll: `prop.table()`

prop.table()

```
prop.table(table(puls$namskeid,puls$likamsraekt.flokkur))
```

```
##  
##           Lítil      Meðal      Mikil  
## LAN203  0.09656652 0.11158798 0.15879828  
## STAE209 0.20386266 0.18025751 0.24892704
```

```
prop.table(table(puls$namskeid,puls$likamsraekt.flokkur),margin=1)
```

```
##  
##           Lítil      Meðal      Mikil  
## LAN203  0.2631579 0.3040936 0.4327485  
## STAE209 0.3220339 0.2847458 0.3932203
```

```
prop.table(table(puls$namskeid,puls$likamsraekt.flokkur),margin=2)
```

```
##  
##           Lítil      Meðal      Mikil  
## LAN203  0.3214286 0.3823529 0.3894737  
## STAE209 0.6785714 0.6176471 0.6105263
```

Helstu lýsistærðir

```
mean()      # meðaltal
median()    # miðgildi
quantile()  # fjórðungamörk
summary()   # fer eftir innihaldi...
sd()        # staðalfrávik
var()       # dreifni
range()     # spönn
min()       # lægsta gildi
max()       # hæsta gildi
```

Ef mælingar vantar þarf að gera ráðstafanir: `na.rm=T`

Lýsistærðir fyrir talnabreytur - miðja

```
mean(c(10,2,5,4,3))      # getum matað með vigri

## [1] 4.8

mean(puls$aldur)          # eða með nafni á breytu

## [1] 24.24416

mean(puls$haed)

## [1] NA

mean(puls$haed,na.rm=T)  # ef vantar gildi

## [1] 173.2532
```

Lýsistærðir fyrir talnabreytur - breytileiki

```
var(puls$haed, na.rm=T)
```

```
## [1] 95.53171
```

```
sd(puls$haed, na.rm=T)
```

```
## [1] 9.774032
```

```
quantile(puls$haed, na.rm=T)
```

```
##    0%   25%   50%   75%  100%  
##  150   166   172   181   198
```

tapply()

- ▶ Gott að nota ef við viljum reikna út gildin á lýsistærð talnabreytu fyrir hvern flokk flokkabreytu sem tilheyrir sömu gagnatöflu.
- ▶ Við mötum `tapply()` með nafni á talnabreytu, nafni á flokkabreytu og nafninu á aðferðinni sem við viljum beita.

```
tapply(puls$haed,puls$nameskeid,mean,na.rm=T)
```

```
##   LAN203   STAE209  
## 173.1657 173.3037
```

```
tapply(puls$thyngd,puls$likamsraekt.flokkur,sd,na.rm=T)
```

```
##   Lítil   Meðal   Mikil  
## 16.76336 13.29713 14.97487
```


summarise()

- ▶ Viljum við teikna út fleiri en eina lýsistærð og/eða ef við höfum fleiri en eina flokkabreytu má nota `summarise()` aðferðina.
- ▶ Tilheyrir `dplyr` pakkanum sem er hluti af `tidyverse`
- ▶ Nota má `group_by()` til að skipta gögnunum upp

summarise()

```
library(dplyr)
summarise(puls, meðalpúls = mean(fyrriPuls, na.rm=T),
          meðalþyngd = mean(thyngd, na.rm=T), fjöldi = n())

## # A tibble: 1 x 3
##   meðalpúls meðalþyngd fjöldi
##   <dbl>      <dbl>   <int>
## 1      72.0      71.5     471

summarise(filter(puls, !is.na(fyrriPuls), !is.na(thyngd)),
          meðalpúls = mean(fyrriPuls),
          meðalþyngd = mean(thyngd), fjöldi = n())

## # A tibble: 1 x 3
##   meðalpúls meðalþyngd fjöldi
##   <dbl>      <dbl>   <int>
## 1      72.3      71.3     444
```

summarise()

```
summarise(filter(puls, !is.na(fyrriPuls)), meðalpúls = mean(fyrriPuls),  
          staðalfrávik = sd(fyrriPuls), fjöldi = n())
```

```
## # A tibble: 1 x 3  
##   meðalpúls staðalfrávik fjöldi  
##   <dbl>      <dbl>   <int>  
## 1     72.0        11.9     454
```

summarise() með group_by()

```
puls.lik<-group_by(puls,likamsraekt.flokkur)
summarise(filter(puls.lik, !is.na(fyrriPuls)), meðaltal = mean(fyrriPuls),
           staðalfrávik = sd(fyrriPuls), fjöldi = n())
```

```
## # A tibble: 4 x 4
##   likamsraekt.flokkur meðaltal staðalfrávik fjöldi
##   <fct>                <dbl>         <dbl>  <int>
## 1 Lítil                72.7           12.4    136
## 2 Meðal                75.0           12.1    129
## 3 Mikil                69.1           10.7    184
## 4 <NA>                 80             8.69     5
```

summarise() með group_by() - pípuritháttur

```
puls %>%  
  filter(!is.na(fyrriPuls)) %>%  
  group_by(likamsraekt.flokkur) %>%  
  summarise(meðaltal = mean(fyrriPuls),  
            staðalfrávik = sd(fyrriPuls), fjöldi = n())  
  
## # A tibble: 4 x 4  
##   likamsraekt.flokkur meðaltal staðalfrávik fjöldi  
##   <fct>              <dbl>      <dbl>  <int>  
## 1 Lítil              72.7        12.4    136  
## 2 Meðal              75.0        12.1    129  
## 3 Mikil              69.1        10.7    184  
## 4 <NA>              80          8.69     5
```