

Aflevering 11

Lucas Bagge

Showhide: Opgave 1.6: Ændring i stormmønster

DMI vedligeholder en side med alle [storme i Danmark](#) fra 1891 og fremefter. Stormene klassificeres i fire styrke kategorier ud fra vindstyrken. I nedenstående tabel har jeg optalt antallet af storme i de forskellige kategorier for fire 30-års perioder.

Periode	Stormstyrke 1	Stormstyrke 2	Stormstyrke 3 og 4
1891-1920	39	16	4
1921-1950	21	8	8
1951-1980	14	12	5
1981-2010	18	12	10

- (a) Opstil den statistiske model, hvor antallet af storme for hver periode følger sin egen multinomialfordeling på de tre kategorier 1, 2 og 3+4. Angiv inden for den opstillede model hypotesen, at der er samme sandsynlighedsvektor for de tre styrke kategorier 1, 2 og 3+4 for de fire tidsperioder.

- (b) Undersøg, om data er i overensstemmelse med hypotesen om samme sandsynlighedsvektor for kategorierne 1, 2 og 3+4 for de fire tidsperioder.

For storme i Danmark finder I således ikke en ændring i fordeling på styrkekategori. Den næste tabel viser fordelingen af [hurricanes](#) fra verdenshavene på styrke kategorierne 1-3 og 4-5. Der er data fra to tidsperioder: 1975-1989 og 1990-2004. Data er fra artiklen [Changes in Tropical Cyclone Number, Duration, and Intensity in a Warming Environment](#).

Periode	Hurricanes 1-3	Hurricanes 4-5
1975-1989	617	171
1990-2004	508	269

- (c) Undersøg om fordelingen af hurricanes på de to styrke kategorier er den samme for de to tidsperioder.

For hurricanes ser det således ud til, at der er sket en ændring. I det sidste spørgsmål i denne opgave skal du se på, om ændringen er den samme i de forskellige verdenshave. Den næste tabel viser fordelingen af de største hurricanes (kategori 4-5) på fem verdenshave for de to tidsperioder.

Verdenshav	1975-1989	1990-2004
East Pacific Ocean	36	49
West Pacific Ocean	85	116
North Atlantic	16	25
Southwestern Pacific	10	22
Indian	24	57

- (d) Opstil model for data, og undersøg, om fordeling på de fem verdenshave er den samme i de to perioder.

a)

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.0 --

## v ggplot2 3.3.3      v purrr  0.3.4
## v tibble  3.1.0      v dplyr  1.0.5
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
data <- tribble(
  ~"Periode", ~"Stomr1", ~"Storm2", ~"Storm_3_4",
  "1891-1920", 39, 16, 4,
  "1921-1950", 21, 8, 8,
  "1951-1980", 14, 12, 5,
  "1981-2010", 18, 12, 10
)
data %>%
  summarise(across(where(is.numeric), ~ sum(.)))
```

```
## # A tibble: 1 x 3
##   Stomr1 Storm2 Storm_3_4
##   <dbl> <dbl>    <dbl>
## 1     92     48      27
```

$A_{1j} \sim \text{multinom}(92, \pi_{1j}) A_{2j} \sim \text{multinom}(48, \pi_{2j}) A_{3j} \sim \text{multinom}(27, \pi_{3j}) a_{ij} = (a_{1i}, \dots, a_{3i}) \pi_{ij} = (\pi_{i1}, \dots, \pi_{i3}) i = 1, \dots, 4 j =$

hypotese om samme sandynlgéh

$$\pi_{ij} = \pi_j$$

b)

Benytter af følgende kode:

```
hom_test <- function(obs){
  ex = outer(rowSums(obs), colSums(obs)) / sum(obs)
  obs1 <- ifelse(obs == 0, 1, obs)
  G <- 2 * sum(obs * log(obs1/ex))
  pval = 1- pchisq(G, (dim(obs)[1] - 1) * (dim(obs)[2] - 1))

  return(list(Forvendtede = ex, G = G, Pvaerdi = pval))
}

hom_test(rbind(c(39,
  16,
  4),
  c(21,
  8,
  8),
  c(14,
  12,
  5),
  c(18,
  12,
  10)))
```

```
## $Forvendtede
##           [,1]      [,2]      [,3]
```

```
## [1,] 32.50299 16.95808 9.538922
## [2,] 20.38323 10.63473 5.982036
## [3,] 17.07784 8.91018 5.011976
## [4,] 22.03593 11.49701 6.467066
##
## $G
## [1] 10.7671
##
## $Pvaerdi
## [1] 0.09584665
```

Her kan vi ikke forkaste H_0 . Dermed tyder det på at det er perioden der er afgørende.

c)

```
hom_test <- function(obs){
  ex = outer(rowSums(obs), colSums(obs)) / sum(obs)
  obs1 <- ifelse(obs == 0, 1, obs)
  G <- 2 * sum(obs * log(obs1/ex))
  pval = 1- pchisq(G, (dim(obs)[1] - 1) * (dim(obs)[2] - 1))

  return(list(Forvendtede = ex, G = G, Pvaerdi = pval))
}

hom_test(
  rbind
  (
    c(617,
      171),
    c(508,
      269)
  )
)
```

```
## $Forvendtede
##           [,1]      [,2]
## [1,] 566.4537 221.5463
## [2,] 558.5463 218.4537
##
## $G
## [1] 32.51158
##
## $Pvaerdi
## [1] 1.184847e-08
```

Hermed kan vi ikke forkaste H_0 og derfor må det være tidsperioden der er afgørende:

d)

Opstille modellen.

$A_{1j} \sim \text{multinom}(85, \pi_{1j}) A_{2j} \sim \text{multinom}(202, \pi_{2j}) A_{3j} \sim \text{multinom}(41, \pi_{3j}) A_{4j} \sim \text{multinom}(32, \pi_{3j}) A_{5j} \sim \text{multinom}(81, \pi_{5j})$

```
hom_test <- function(obs){
  ex = outer(rowSums(obs), colSums(obs)) / sum(obs)
  obs1 <- ifelse(obs == 0, 1, obs)
  G <- 2 * sum(obs * log(obs1/ex))
  pval = 1- pchisq(G, (dim(obs)[1] - 1) * (dim(obs)[2] - 1))

  return(list(Forvendtede = ex, G = G, Pvaerdi = pval))
}
```

```
hom_test(
  rbind
  (
    c(
      36,
      49
    ),
    c(
      85,
      116
    ),
    c(
      16,
      25
    ),
    c(
      10,
      22
    ),
    c(
      24,
      57
    )
  )
)
```

```
## $Forvendtede
##      [,1]      [,2]
## [1,] 33.03409 51.96591
## [2,] 78.11591 122.88409
## [3,] 15.93409 25.06591
## [4,] 12.43636 19.56364
## [5,] 31.47955 49.52045
##
## $G
## [1] 5.232063
##
## $Pvaerdi
## [1] 0.2643042
```

VI forkaster H_0 . dermed afhænger de ikke af verdenshavene men af tidsperioden.