

# Ideer til at introducere statistisk metode

Claus Thorn Ekstrøm

KU Biostatistik

[ekstrom@sund.ku.dk](mailto:ekstrom@sund.ku.dk)

Marts 18, 2019

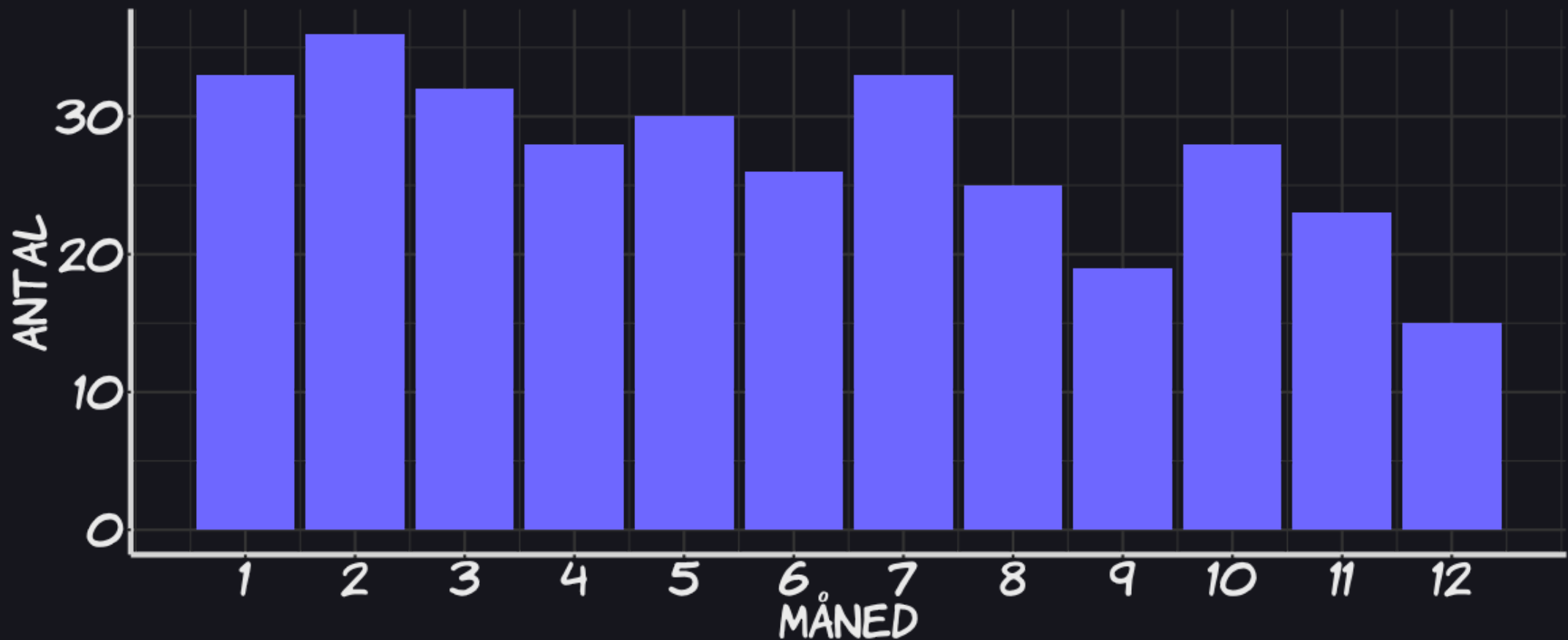
Slides @ [biostatistics.dk/talks/](https://biostatistics.dk/talks/)



# Fødselsdagspuklen



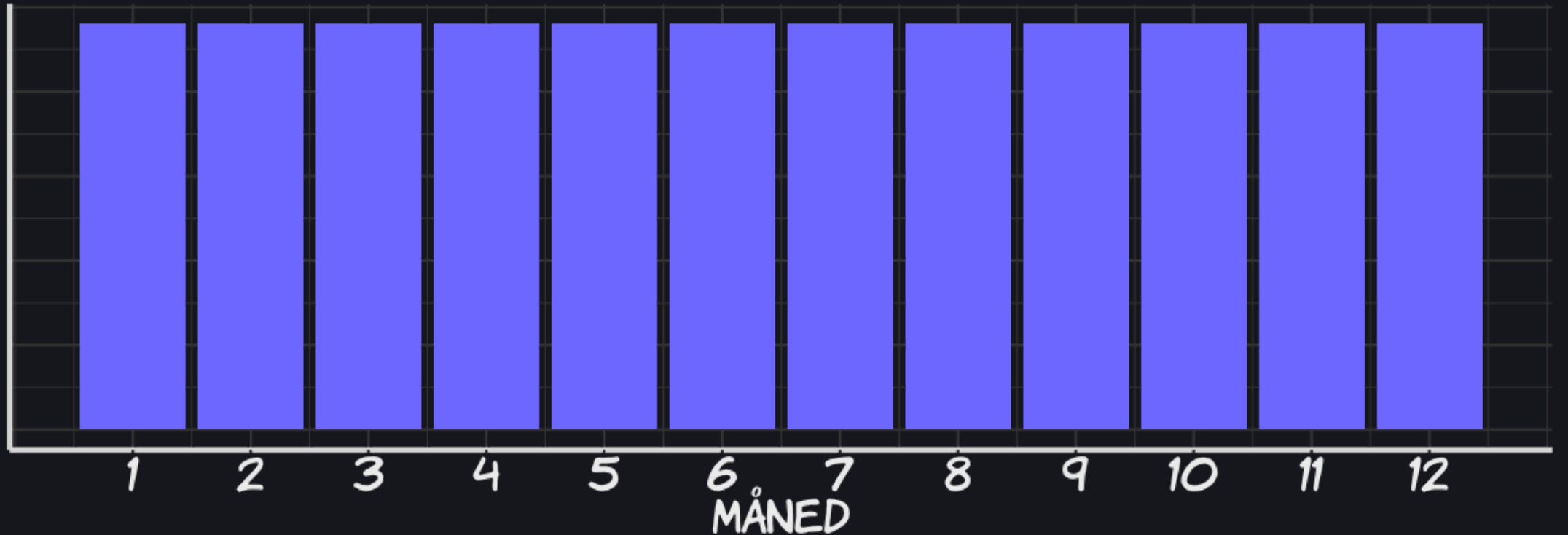
# Spillere fra 1. jan 1951 til 7. maj 2017 (N=334)



FRA DBU.DK

# Hvad er hypotesen?

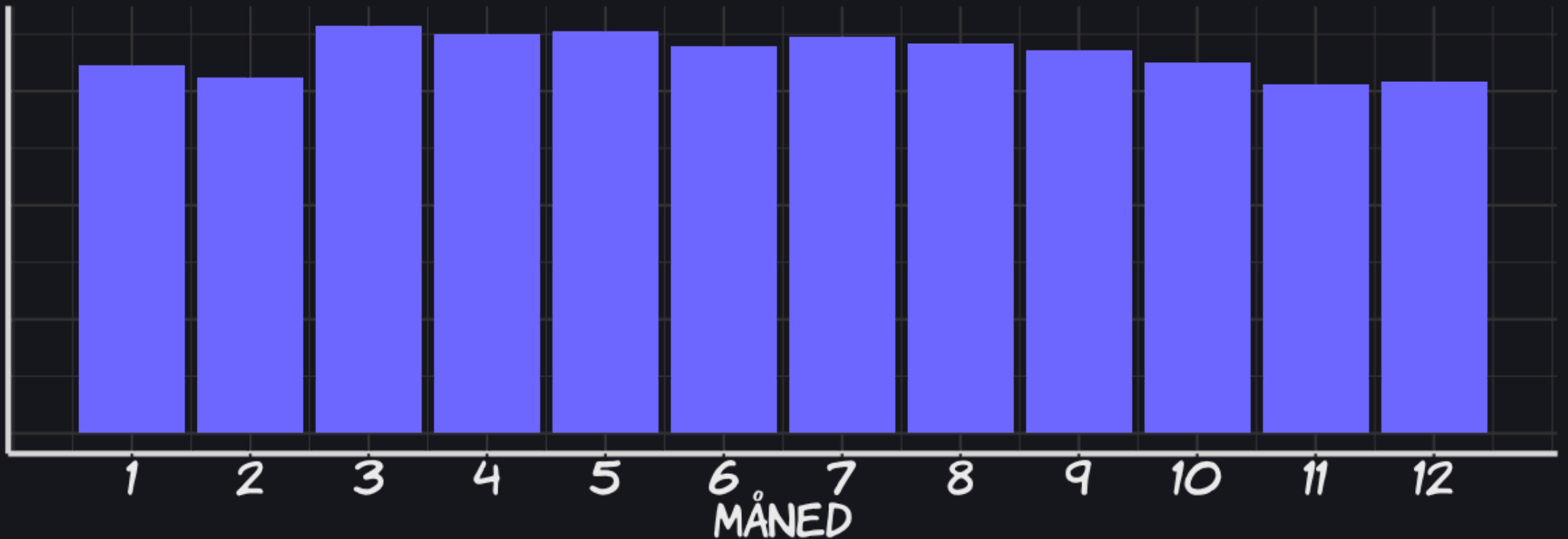
Hvordan burde fordelingen se ud? Og hvorfor?



FRA LOMMEFILOSOFI

# Hvad er hypotesen?

Hvordan burde fordelingen se ud? Og hvorfor?



FRA DST.DK

# Hvor godt passer data med hypotesen?

$$X^2 = \sum_i \frac{(\text{Obs}_i - \text{Forv}_i)^2}{\text{Forv}_i}$$

```
chisq.test(res$n, p=fdata$n)
```

```
##  
##      Chi-squared test for given probabilities  
##  
## data:  res$n  
## X-squared = 14.4, df = 11, p-value = 0.2117
```

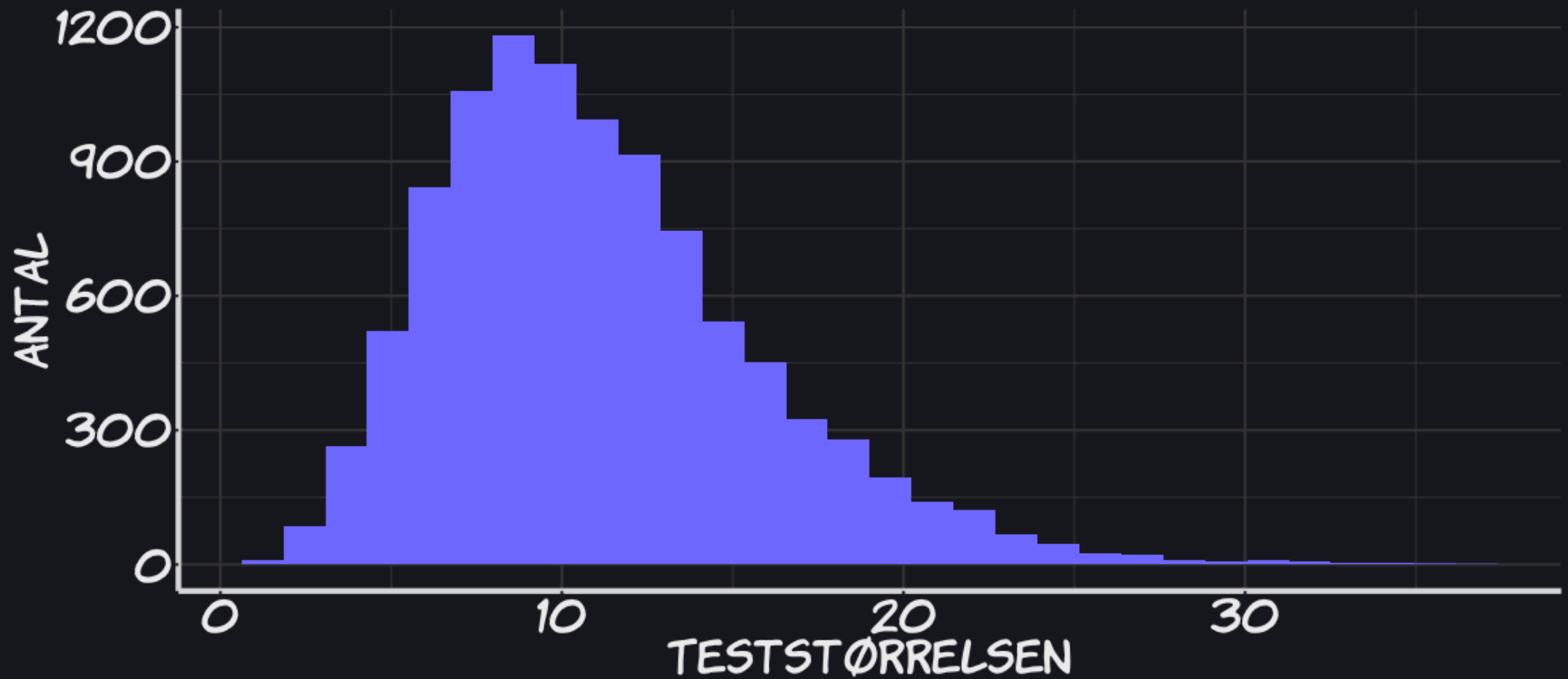
(Se *Aktuelt Naturvidenskab 3, 2017*)

# Simulationer

```
sim <- sapply(1:10000, function(i) {  
  chisq.test(table(sample(1:12,  
                        size=334,  
                        replace=TRUE,  
                        prob=fdata$n)),  
             p=fdata$n)$statistic } )
```



# Fordeling af teststørrelsen under $H_0$



Hvornår knækker kurven?

Det kan summeres sådan: "Selvom der fortsat er mange skader, er det glædeligt at antal i alt og antal alvorlige er faldet for tredje år i træk. Rakter har givet knap halvdelen af skaderne – også i år. Fyrværkeriskader sker for mænd, og de har i stort omfang ikke haft beskyttelsesbriller på. Kun to af 30 alvorlige skader skete for kvinder. Blandt de alvorlige skader er raketter og batterier ansvarlige for 2 af 3 skader, mens ulovligt fyrværkeri kun har givet få alvorlige skader. Vi finder det bekymrende, at knap halvdelen er under 18 år og en fjerdedel under 16 år, også blandt de alvorlige"

### Skader efter brug af fyrværkeri, som har medført sygehuskontakt i Danmark – samlet udvikling for de to døgn 31.12. og 01.01 i perioden 1995-1996 til og med 2018-2019

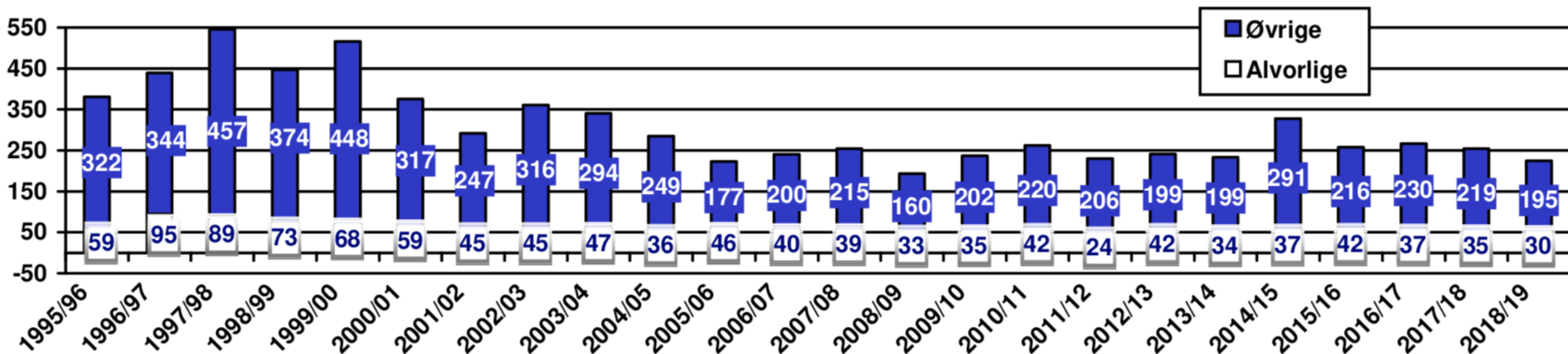




Foto: Martin Lehmann

Andelen af rygere i Danmark er steget signifikant inden for de seneste to år, viser den årlige undersøgelse af rygevaner.

SUNDHED 2. JAN. 2019 KL. 21.30



**HELENE NAVNE**  
Journalist



**ANDERS LEGARTH SCHMIDT**  
Journalist



**LARS IGUM RASMUSSEN**  
Sundhedsredaktør

## Første gang i 20 år: Flere danskere ryger

Prisen på cigaretter skal op. Sådan siger fagfolk på baggrund af årets undersøgelse af rygevaner, der viser, at andelen af rygere er steget markant på to år.

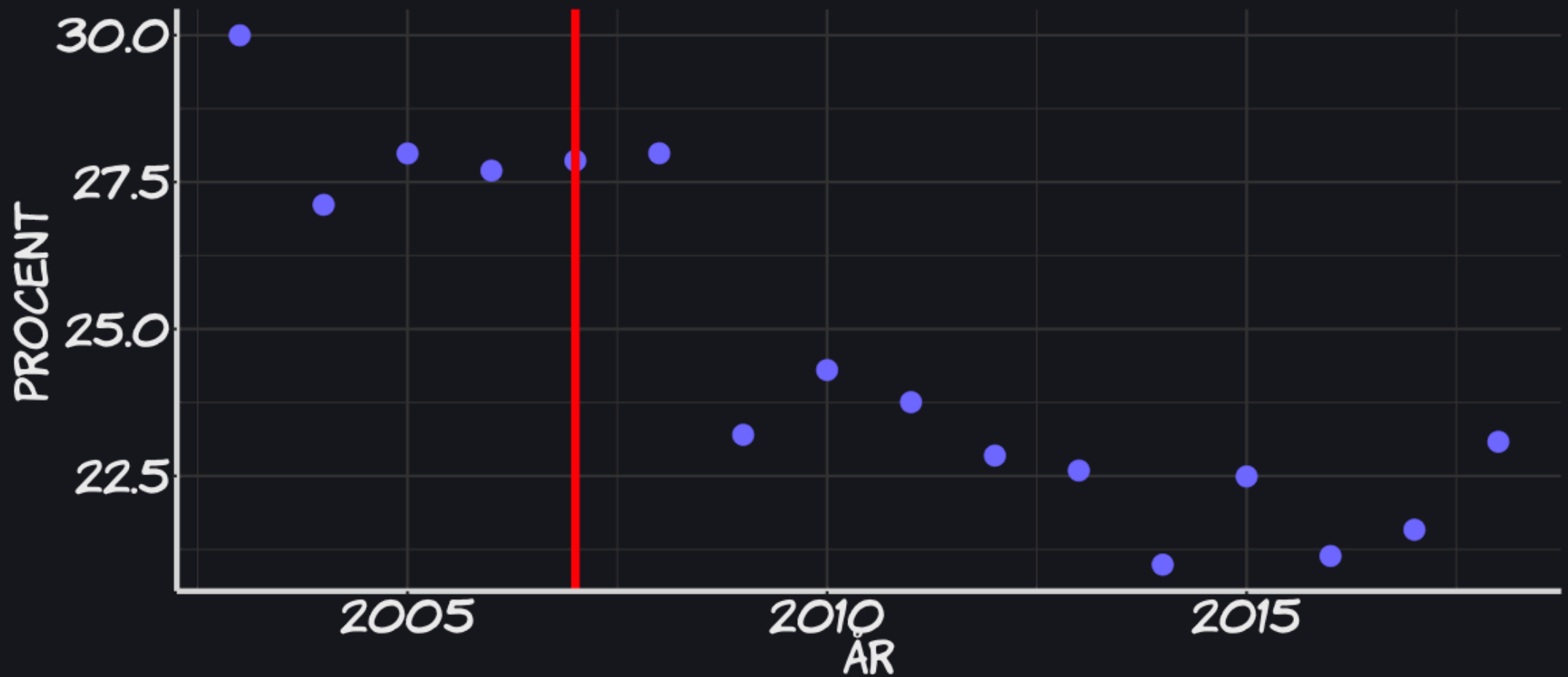
For første gang i to årtier er der en signifikant stigning i andelen af

### MEST LÆSTE

**DMI varsler vindstød af stormstyrke og voldsomt vejr i dele af landet**

**Eksklusivt uddrag fra ny bog om sygeplejerske-sagen: »Hun gør det fandeme. Jeg ved, hun gør det.«**

# Danskernes rygevaner



# Lineær regression

```
library("broom")  
lm(p ~ year, data=data) %>% tidy()
```

```
## # A tibble: 2 x 5  
##   term          estimate std.error statistic    p.value  
##   <chr>          <dbl>     <dbl>     <dbl>    <dbl>  
## 1 (Intercept)  1132.     149.      7.59  0.00000251  
## 2 year         -0.551    0.0742   -7.43  0.00000323
```

# Prædiktion (for 2018)

```
library("broom")  
lm(p ~ year, data=data, subset=(data$year<2018)) %>% tidy()
```

```
## # A tibble: 2 x 5  
##   term          estimate std.error statistic    p.value  
##   <chr>          <dbl>    <dbl>    <dbl>    <dbl>  
## 1 (Intercept)  1278.    140.     9.10  0.0000000533  
## 2 year         -0.624   0.0699  -8.92  0.0000000665
```

```
1278.07757 + 2018*-0.62354
```

```
## [1] 19.77385
```

# Prædiktion fortsat

En prædiktion fra lineære regression har spredning

$$\hat{y}_0 \pm 1.96 \cdot \hat{\sigma} \cdot \sqrt{1 + \frac{1}{N} + \frac{(x_0 - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}}$$

```
predict(res, newdata=data.frame(year=2018), interval="prediction")
```

```
##           fit          lwr          upr  
## 1 19.78148 16.90613 22.65684
```