

# FØDSELSDAGSBULEN

- Udarbejdet af Peter Arnborg Videsen og Henry Berthelsen, Viborg Gymnasium og HF for Aktuel naturvidenskab.

**Målgruppe:** Matematik på A og B-niveau. Kan bruges som SRP-emne i samarbejde med Idræt.

## TIL LÆREREN

Dette arbejdsark tager udgangspunkt i artiklen ["Fodboldlandsholdets fødselsdagsbule" i Aktuel Naturvidenskab 3-2017](#).

Arket kan indgå i forløb om  $X^2$ -test og Goodness of Fit.

Den sidste del af arbejdsarket, hvor Kolmogorov-Smirnov-fordelingen simuleres er ret teknisk, og vil nok kræve en del vejledning af eleverne.

## BEMÆRKNINGER:

- Artiklen skal læses først.
- Dataene i kolonnen "Befolkningen" i tabellen side 23 og kolonnen "Forventet" i tabellen side 24 summer ikke op til 100 %, men til 100,02 %. Det giver dog ikke de store forskelle, om man justerer dataene, så de summer op til 100 %. Det er der heller ikke gjort i regnearket, hvilket i praksis betyder, at december-andelen her bliver 0,02% mindre i simuleringen.
- Man kan ændre størrelsen af stikprøven ved at slette eller tilføje celler i kolonnerne "Simulering" i regnearket "Simulering".
- Funktionsknappen F9 genberegner regnearket.

## KOLMOGOROV-SMIRNOV versus $X^2$

I dette afsnit vil vi undersøge forskellen mellem de to teststørrelser og undersøge påstanden om, at Kolmogorov-Smirnov er mere følsom overfor rækkefølgen i fødselstallene.

I den grønne boks "Sammenligning af de to test" er angivet tabellen:

	Jan	Feb	Mar	Apr	Maj	Jun	Jul	Aug	Sep	Okt	Nov	Dec
Observeret	14	12	11	10	8	8	8	6	6	6	6	5
Forventet	7,96	7,74	8,91	8,86	8,89	8,62	8,82	8,62	8,43	8,04	7,55	7,58

- Beregn  $X^2$ -teststørrelsen for denne stikprøve.
- Beregn Kolmogorov-Smirnov-teststørrelsen for den samme stikprøve

$$KS = \max_{m \in \{1, \dots, 12\}} \left| \sum_{i=1}^m (Obs_i - Forvent_i) \right|$$

Nu bytter vi bare rundt på dataene (observeret og forventet) for Jan og Maj, Feb og Sep, Mar og Dec - observationerne og beregner teststørrelserne på ny.

	Jan	Feb	Mar	Apr	Maj	Jun	Jul	Aug	Sep	Okt	Nov	Dec
Observeret	8	6	5	10	14	8	8	6	12	6	6	11
Forventet	8,89	8,43	7,58	8,86	7,96	8,62	8,82	8,62	7,74	8,04	7,55	8,91

- Sammenlign teststørrelserne for de to stikprøver. Forklar hvorfor teststørrelserne ændrer sig som de gør?

## VURDERING AF KS-TESTSTØRRELSEN

Tilbage til den oprindelige tabel.

Når vi skal vurdere om KS-teststørrelsen er stor eller lille, er man ifølge artiklen nødt til at simulere data under nulhypotesen, at stikprøvens fordeling ligner befolkningens (populationens) fordeling.

Det kan man gøre ved hjælp af regnearket Fødselsdagsbullen.

Her er gennemført 1000 simuleringer af at tage en tilfældig stikprøve på 100 personer i befolkningen.

I arket "Beskrivelse af regnearket" kan du få hjælp til at forstå hvordan det virker.

Undersøg med både Goodness of Fit og KS-fordelingen, om stikprøven afviger fra befolkningens fordeling (signifikansniveau 5%).