

# Mini-Projekt $\chi^2$

## DEL 1: Simulering

### 1. Indledende strækøvelser.

- Forklar hvad der menes med et sandsynlighedsfelt.
- I roulette drejes en skive med 37 felter; 18 røde, 18 sorte og 1 grønt. Opstil en tabel hvor der til hvert udfald (rød, sort og grøn) er angivet *sandsynlighederne* for hvert udfald.
- Hvad er de forventede *hyppigheder* hvis man spiller 50 omgange roulette?

### 2. Chi-kvadratet $\chi^2$ . Efter 50 spil observerer vi hhv. 22, 24 og 4 røde, sorte og grønne. Hvad er $\chi^2$ -teststørrelsen for denne observation?

- Opskriv formelen for  $\chi^2$  og skriv med egne ord hvad det er et mål for.
- Regn det først chi-kvadratet,  $\chi^2$ , "i hånden" – husk mellemregninger.

Indlæs pakkerne:

**Statistics** og

**GymStat** (hentes i Lectio og lægges i program-folderen)

Dette gøres med **with(...)**

- Kontroller udregning i Maple vha. GymStat-pakkens:

**ChiKvadratet(<obs>, <forv>)** ,

her står <obs> for en liste af observerede hyppigheder (fx [12,23,2] ) og <forv> for en liste af forventede hyppigheder.

3. **Simulering.** Lav en simulering i Maple, vha. GymStat-pakken (hentes i lectio), hvor I:
- Definerer en tilfældighedsgenerator/fordeling, der giver tal mellem 1 og 37.
  - Laver et passende udfaldsrum hvor de røde, sorte og det grønne felt (vink: benyt intervaller fx [1..18, 19..36, 37..37])
  - Lav en simulering vha.  

$$\text{UdtrækChi2}(\langle \text{fordeling} \rangle, \langle \text{udfaldsrum} \rangle, \langle \# \text{ i spil} \rangle, \langle \# \text{ gentagelse af spil} \rangle)$$
 hvor  $\langle \text{fordelingen} \rangle$  er en fordeling og  $\langle \text{udfaldsrum} \rangle$  er en liste med udfaldsrummet,  $\langle \# \text{ i spil} \rangle$  er det antal gange man fx kaster terninger i et spil og  $\langle \# \text{ gent. af spil} \rangle$  er antallet af gange hvert spil skal gentages.
  - Lav en grafisk præsentation af de udtrukne  $\chi^2$ -værdier og se hvordan de fordeler sig (et histogram)
  - Man kan vise at den  $\chi^2$  fordeler<sup>†</sup> sig efter en matematisk formel:  

$$\text{ChiFordeling}(\langle t \rangle, \langle \# \text{ frihedsgrader} \rangle)$$
 Antallet af frihedsgrader er antallet af udfald minus en (fx med en terning er det 5).
4. **Refleksioner**
- Kommenter graferne.
  - Diskuter værdien af  $\chi^2$ -teststørrelsen, som I fandt i opgave 3, på baggrund i graferne. (Er det fundne  $\chi^2$  markant større/mindre end det vi simulere os til.)
5. **Signifikans.** Arealet under den teoretiske fordeling (af  $\chi^2$ -fordeling) fra den observerede  $\chi^2$  værdi  $t_0$  (spm 3) til uendelig ( $\infty$ ) angiver sandsynligheden for at få et resultat, som er lige så "godt" eller værre end det observerede  $\chi^2$ . Denne sandsynlighed kaldes for chi-kvadratets  $p$ -værdi:
- $$p = \int_{t_0}^{+\infty} f(t) dt$$
- Bestem  $p$ -værdien hørende til  $\chi^2$  fundet i spm 3.
6. Lav en **undersøgelse** (i bøger/ på nettet o.lign) af nedenstående begreber og relater dem til projektet:
- Hypoteser og test.
  - Signifikans og signifikans-niveau.
  - Acceptmængde og Kritisk mængde.
  - Fejl af type I og II.

---

<sup>†</sup> Med en matematisk fordeling af  $\chi^2$  menes en funktion, der fortæller med hvor stor sandsynlig  $\chi^2$  har værdier  $t$  i (værdien afhænger også af antallet .