

Pointer fra 3. uge

- Datatypen er afgørende for valg af analysemetode
- Sandsynligheder er frekvenser - bare for hele populationen, og derfor kender vi den (sædvanligvis) ikke
- Sandsynligheder for uafhængige hændelser skal ganges sammen
- Konfidensintervaller er vigtigere end P-værdier
- Der er ikke så meget information i en 0/1-variabel (f.eks. komplikationer ja/nej) som i en kvantitativ registrering (f.eks. FEV₁)

Fagligt har I hørt om:

- Summen af uafhængige 0/1-variable bliver Binomialfordelt
- Binomialfordelingen ligner Normalfordelingen, når N er stor, og p ikke er tæt på hverken 0 eller 1
- Testet for *uafhængighed* tester uafhængighed mellem gruppe (f.eks. behandling, skal være rækkerne i tabellen) og outcome (søjlerne i tabellen).
 - Fishers eksakte test kan bruges i tilfælde af små samle sizes (ellers kan det tage for lang tid)
 - χ^2 -testet kan bruges, når de forventede værdier er store nok (tom-melfingerregel: større end 5)
- Husk, at test kun er en (mindre) del af svaret, og at estimater (med konfidensintervaller) er vigtigere:
 - Differens mellem to sandsynligheder
 - Ratioen mellem to sandsynligheder (Relativ Risiko)
 - Odds Ratio for to sandsynligheder (svært forståeligt, men desværre nødvendigt)
- I større tabeller kan der blive tale om *test for trend*
- Parrede binære data

Næste gang skal I høre om:

- Lineær regression (find en linie til beskrivelse af en sammenhæng, f.eks. blodtryk som funktion af alder)
- Prediktionsintervaller
- Korrelation (som - næsten - er ubrugeligt)