## Konfidensintervaller 2025

**Problemstilling**

Vi vil på baggrund af interview af et lille antal personer (en stikprøve) fremkomme med et estimat (vurdering) over befolkningsandelen, der vil stemme på et bestemt parti. Estimat kommer far latin, hvor *aestimare* betyder *vurdere.*

**Opinionsundersøgelser og binomialfordelingen**

Opinionsundersøgelser baserer sig på omkring 1600 telefoninterviews. Undersøgelsen bygger på 4 antagelser

1. Der spørges kun om personerne stemmer på et parti (eller ikke)
2. Stikprøven opfattes som foretaget med tilbagelægning, selvom den selvfølgelig er uden.
3. De adspurgte svarer uafhængigt af hinanden (ingen ved hvad de andre svarer)
4. De adspurgte er tilfældigt udvalgt blandt hele befolkningen (repræsentativ)

De tre første betingelser gør det muligt at antage, at tilslutningen til et bestemt parti indenfor stikprøven er binomialfordelt. Hvis stikprøven i en opinionsundersøgelse ikke er repræsentativ, vil der opstå en systematisk fejl, der gør, at fx borgerlige partier overrepræsenteres.

**Konfidensinterval omkring befolkningens andel af til et bestemt parti.**

Vi ved, at der findes et interval omkring befolkningens *ukendte andel af vælgertilslutningen til et parti*, der med en vis sikkerhed indeholder stikprøvens andel. Kan ikke bruges til noget, da vi ikke kender denne andel. Når vi laver et konfidensinterval, finder vi et interval omkring *den observerede andel*, der med en vis sikkerhed indeholder befolkningens andel.

**Fortolkning af konfidensintervallet**

I 95 ud af 100 stikprøver vil konfidensintervallet omkring den observerede værdi indeholde den rigtige andel af vælgertilslutningen. Konfidensintervallet angiver *ikke* sandsynligheden for at befolkningens andel ligger i intervallet. Enten ligger andelen indenfor konfidensintervallet, eller også gør den ikke. Andelen er ikke en variabel, man kan flytte ind og ud af konfidensintervallet. Andelen er en ukendt konstant, og det er estimatet for andelen, der varierer. Derfor er konfidensintervallet selv en stokastisk variabel (variabel med en tilfældig variation)

**Eksempel 1.** *Konfidensinterval ved brug af binomialfordelingen*

I en stikprøve på 100 mennesker er der 50 der stemmer på *Rød Blok*. Procentdelen der stemmer på rød blok estimeres til 50 % og hvis vi anvender binomialfordelingen med antalsparameteren $n=100$ og sandsynligheden $p=0,5$, så bliver middelværdi og spredning

 $μ=n∙p=100∙0,5=50$ og $σ=\sqrt{n∙p∙\left(1-p\right)}=\sqrt{100∙0,5∙\left(1-0,5\right)}=5$

I en binomialfordeling hvor $n∙p\geq 5$ er 95 % af observationerne placeret i intervallet med 2 spredninger på hver side af middelværdien, dvs. som $50\pm 10$

$$95 \% konfidensinterval=\left\{40,41,…59, 60\right\}$$

Konfidensintervallet fortolkes som den estimerede andel ligger mellem 40 % og 60 %, hvilket svarer til at anvende formlen for konfidensintervallet, hvor $\hat{p}$ er den estimerede andel

$$\hat{p}\pm 2\sqrt{\frac{\hat{p}∙\left(1-\hat{p}\right)}{n}}=\frac{50}{100}\pm 2∙\sqrt{\frac{0,5∙\left(1-0,5\right)}{100}}=50 \%\pm 10 \%$$

**Formel til at bestemme et konfidensinterval ved brug af binomialfordelingen**

Hvi vi har en stikprøven med *n* personer, hvor der er *x* personer, der stemmer på et bestemt parti, så vil vi estimere procentdelen til $$\hat{p}= \frac{x}{n}$$

og hvis binomialfordelingen ikke er for skæv eller stikprøven er stor ($n∙p\geq 5$ ), så vil 95 % af stikprøverne give et resultat, der ligger indenfor 2 spredninger fra middeltallet (*95 % konfidensinterval)*

$$n∙\hat{p}\pm 2∙\sqrt{n∙\hat{p}∙\left(1-\hat{p}\right)} $$

Divideres med *n* fås den estimerede procentdel med usikkerhedsinterval

$$\hat{p}\pm \frac{2\sqrt{n∙\hat{p}∙(1-\hat{p})}}{n}=\hat{p}\pm \frac{2\sqrt{n∙\hat{p}∙(1-\hat{p})}}{\sqrt{n^{2}}}=\hat{p}\pm 2\sqrt{\frac{n∙\hat{p}∙\left(1-\hat{p}\right)}{n^{2}}}= \hat{p}\pm 2\sqrt{\frac{\hat{p}∙\left(1-\hat{p}\right)}{n}} $$

**Opgave 1.** *Hvorfor er konfidensintervallet størst, når stikprøveandelen er 50 %?*

En funktion er givet ved $f\left(x\right)=\sqrt{x∙\left(1-x\right)}$. Vis, at *f*(*x*) har maksimum, når *x* = 0,5, og derfor er usikkerheden størst når $\hat{p}=0,5$

**Opgave 2.** *Hvorfor bliver konfidensintervallet halvt så stort, når stikprøven bliver 4 gange større?*

**Opgave 3.** *Hvad kan man sige om bredden af**et 90 % konfidensinterval, et 95 % konfidensinterval og et 99 % konfidensinterval?*

$$90 \% konfidens \hat{p}\pm 1,645\sqrt{\frac{\hat{p}∙\left(1-\hat{p}\right)}{n}} og 99 \% konfidens \hat{p}\pm 2,58\sqrt{\frac{\hat{p}∙\left(1-\hat{p}\right)}{n}} $$

*Ovenfor er det væsentlige i spørgsmålet, nedenfor er vist hvad man kan medtage i dialogen*

**Eksempel 2.** *Hypotesetest og konfidensinterval*

Hvis *Rød Blok* ved sidste valg fik 50 % af stemmerne, bruger vi 50 % som nulhypotese. Den kritiske mængde *K* og acceptmængden *A*, beregnes via binomialfordelingen med *n* = 100 og *p* = 0,5, og et signifikansniveau på 5 %

$$K=\left\{0,2,3…39\right\}∪\left\{61, 62,…100\right\} og A=\left\{40,41,…59, 60\right\}$$

*Acceptmængden af en hypotesetest med observationen som nulhypotese svarer til konfidensintervallet.*

Hvis den observerede andel det stemmer rødt er på 50 %, så vil en acceptmængden til nulhypotese med *p* = 0,6 indeholde de observerede 50 % og det samme gælder for nulhypotesen *p* = 0,4

*Konfidensintervallet indeholder alle de nulhypoteser, der ikke forkastes af observationen.*

*Læg mærke til at konfidensintervallet er et symmetrisk interval omkring stikprøveandelen hvorimod acceptområdet ved en dobbeltsidet test er et symmetrisk interval omkring den hypotetiske andel.*

**Konfidensintervaller og hypotesetest**

1. Et 95 % konfidensinterval omkring den observerede andel indeholder alle de nulhypoteser, der *ikke* er forkastet på signifikansniveauet på 5 %.
2. Et 95 % konfidensinterval omkring den observerede andel svarer til acceptmængden i hypotesetest på signifikansniveau 5 %, hvor nulhypotesens påståede andel er lig den observerede andel.
3. Hvis nulhypotesens værdi ligger *udenfor* 95 % konfidensintervallet er det ensbetydende med at hulhypotesen *afvise*s på 5 % signifikansniveau

*Det sidste punkt ovenfor betyder, at man kan teste en hypotese ved brug af et konfidensinterval.*

Kilder

Andersen, E, *Teoretisk statistik for økonomer*, Akademisk forlag 2. udg. 1988

Ditlevsen&Sørensen., 2018, Introduktion til statistik 5. udg

chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://noter.math.ku.dk/ss18.pdf

Tan&Tan, The Correct Interpretation of Confidence Intervals, National Cancer center Singapore

<https://journals.sagepub.com/doi/pdf/10.1177/201010581001900316>