NORMALFORDELINGSAPPROKSIMATIONEN OG KONFIDENSINTERVAL

SÆTNING 1

Fra en population med succesandelen p (den sande andel for en population) udtages en tilfældig stikprøve, som udgør en ubetydelig del af populationen. Hvis stikprøven har størrelsen n og indeholder r succeser, kan det tilhørende 95% konfidensinterval for p beregnes som

$$\left[\ \hat{p} - 1{,}96 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 1{,}96 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\ \right] \tag{1}$$

Her kaldes $\hat{p} = \frac{r}{n}$ for **estimatet** for p og $1{,}96 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ for **usikkerheden**.

OPGAVER

Opgave/teori 1

En mønt giver krone 36 gange i 100 kast. Derfor er estimatet for p - altså $\hat{p}$ fra stikprøven givet ved $\hat{p} = \frac{36}{100} = 0{,}36$. Vores *estimat* fra eksperimentet - hvor $n = 100$ - er, at sandsynligheden for at få krone er 36%. Vi vil gerne angive den statistiske usikkerhed på estimatet. Det gør vi ved at beregne et konfidensinterval omkring 0,36. Det er teorien bag et dobbeltsidet binomialtest, der benyttes.

Vi vil nu vurdere, i hvilket interval vi kan forvente, at vi inkluderer populationens sande p.

a) Bestem et 95% konfidensinterval ved at bruge formlen (1) og ved 'automatisk' beregning med Nspire.

Opgave/teori 2

Der laves jævnligt holdningsundersøgelser om medlemskab af EU (se fx Europa-Parlamentets Eurobarometer). Vi ser på en tænkt undersøgelse. Borgerne i et EU-land har fået stillet spørgsmålet:

Synes du dit lands medlemskab af EU er en god ting?

Antag nu, at 900, 1800 og 3600 tilfældige mennesker i tre samtidige undersøgelser har fået spørgsmålet, og at 60,0% i alle tre undersøgelser svarede ja.
Kan vi bruge sætningen?

1. Stikprøverne er tilfældige - så det er fint.
2. Stikprøven er relativt lille - antallet adspurgte udgør kun en lille andel af alle borgere i EU.
3. $np \geq 5$ og $n(1-p) \geq 5$ er også klart opfyldt.

a) Bestem usikkerhed og et 95% konfidensinterval i de tre tilfælde ved brug af formel (191) i formelsamlingen.

Denne tabel skal altså udfyldes:

n	$\hat{p}$	Usikkerhed	95% konfidensinterval
900	0,600		
1800	0,600		
3600	0,600		

b) Hvad sker der med den statistiske usikkerhed, når stikprøvestørrelsen øges?
c) Kan du forklare det fra formlen?
d) Hvilken undersøgelse tænker du, at man vil foretrække?

Opgave 3

Hvilket parti vil du stemme på, hvis der var folketingsvalg i morgen?

Nedenfor er det spørgsmål besvaret for tre partier tilbage i 2018. Stikprøven omfattede 1485 personer.

Parti	Andel	Usikkerhed	95% konfidensinterval
A	26,2%		
I	3,9%		
Ø	9,3%		

a) Udfyld de tomme felter ved brug af formlen (1).
b) Udgør stikprøven en stor andel af den danske befolkning?
c) (Betragt igen formel (1) her på siden. Hvorfor bliver usikkerheden lille for små værdier af $\hat{p}$?)

Opgave 4

Nogle elever vil undersøge, hvor stor en andel af gymnasieungdommen i Danmark, der ryger. I en tilfældigt udtaget stikprøve på 150 elever viser det sig, at 35 er rygere.

a) Bestem et 95% konfidensinterval for andelen af rygere blandt gymnasieelever ud fra stikprøven. Benyt formlen (1) og beregn også 'automatisk' ved at bruge Nspire.
b) Eleverne er ikke tilfredse med usikkerheden i deres måling, som de gerne vil have ned på 4%. Hvor stor en stikprøve kræves for at opnå denne usikkerhed?
c) Eleverne overvejer, om de bare skal udtage en sådan stikprøve fra deres eget gymnasium. Overvej, om denne metode vil give et retvisende resultat.