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**.

Alternativt

$$\left[\ \hat{p} - 2 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 2 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\ \right] \text{og } 2 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

OPGAVE

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?