StatMet eksamensopgave, kladde

Morten Tulstrup

Opgave 1

Vi tager udgangspunkt i Theorem 2.1.5 (koncentrationsulighed baseret på Chebychevs ulighed):

$$P\left(|\hat{p} - p| > \delta\sqrt{\frac{p(1-p)}{n}}\right) \leq \frac{1}{\delta^2}$$

Indsætter $p = 0.2$ og $n = 80$:

$$P\left(|\hat{p} - 0.2| > \delta\sqrt{\frac{0.2(1-0.2)}{80}}\right) = P\left(|\hat{p} - 0.2| > \delta\sqrt{0.002}\right) \leq \frac{1}{\delta^2}$$

Vi vælger δ så

$$\delta\sqrt{0.002} = 0.1 \Rightarrow \delta = \frac{0.1}{\sqrt{0.002}} = 2.236$$

Dermed er:

$$P\left(|\hat{p} - 0.2| > 0.1\right) \leq \frac{1}{2.236^2} = 0.2 \Rightarrow$$
$$P(\hat{p} \in [0.1, 0.3]) \leq 0.2$$

Eftersom $0 \leq \hat{p} \leq 1$ og P er et sandsynlighedsmål er det ensbetydende med:

$$P(\hat{p} \in [0.1, 0.3]^C) = P(\hat{p} \in [0, 0.1) \cup (0.3, 1]) = P(\hat{p} \in [0.1)) + P(\hat{p} \in (0.3, 1])) > 0.8$$

Vi kan altså bruge koncentrationsuligheden til at sige, at sandsynligheden for at observere bivirkningen hos 10-30% af de vaccinerede er større end 0.8. Vi kan derimod ikke bruge koncentrationsuligheden til at sige, hvordan sandsynligheden fordeler sig mellem tilfældende hvor $\hat{p} < 0.1$ vs. $\hat{p} > 0.3$, så den egentlige sandsynlighed for at observere $\hat{p} > 0.1$ er altså større end 0.8.

Den faktiske sandsynlighed ud fra binomialfordelingen er 0.987:

```
1-pbinom(8, 80, 0.2)
```

```
## [1] 0.9869125
```

Som er et meget mere præcist tal, dvs.at koncentrationsuligheden er korrekt men meget upræcis.

Opgave 2:

Prædiktionsintervallerne bestemmes for $n = 40, 80, 120$. For kendt $p = 0.05$ er

$$\text{se} = \sqrt{\frac{0.05(1-0.05)}{n}} = \begin{cases} 0.06324555 & n = 40 \\ 0.04472136 & n = 80 \\ 0.03651484 & n = 120 \end{cases}$$

Baseret på den centrale grænseværdisætning er $\delta_{\text{gauss}} = 1.96$, jvf Tabel 2.1 i NRHAT (i nedenstående udregninger bruges dog det mere præcise tal opnået ved kommandoen qnorm(0.975) i R). Prædiktionsintervaller udregnet som $0.2 \pm \delta_{\text{gauss}} \times \text{se}$ bliver således:

$$\begin{array}{ll} [0.037, 0.362] & \text{for } n = 40 \\ [0.085, 0.315] & \text{for } n = 80 \\ [0.106, 0.294] & \text{for } n = 120 \end{array}$$

Det vil altså sige, at for en bivirkning med en sand sandsynlighed på 5% forventer vi, at det observerede $\hat{p}$ falder indenfor ovenstående intervaller med en sandsynlighed på 95%. Hvis vi f.eks. betragter tilfældet $n = 80$ bemærker vi, at prædiktionsintervallet er meget bredere end $(0.01, 0.1)$

Opgave 3:

Benytter NRHAT Definition 2.1.10 med:

$$\hat{p} = \frac{862}{1332719} = 0.000646798$$

Og

$$\hat{se} = \sqrt{\frac{\hat{p}(1-\hat{p})}{1332719}} = 2.20229e-05$$

Baseret på den Centrale Grænseværdisætning og $\alpha = 0.01$ er

$$\delta_{\text{gauss}} = 2.58$$

99-konfidensintervallet for $\hat{p}$ er da $[0.00059, 0.00070]$