

PUKTESTIMERING

Løsningsforslag fra Tor Slind og Hans Petter Hornæs.

6.2 Forventningsverdier og varianser:

$$E(\hat{\mu}_1) = E\left(\frac{2}{7}X_1 + \frac{5}{7}X_2\right) = \frac{2}{7}E(X_1) + \frac{5}{7}E(X_2) = \left(\frac{2}{7} + \frac{5}{7}\right)\mu = \underline{\mu}$$

$$E(\hat{\mu}_2) = E\left(\frac{1}{2}X_1 + \frac{1}{2}X_2\right) = \frac{1}{2}E(X_1) + \frac{1}{2}E(X_2) = \left(\frac{1}{2} + \frac{1}{2}\right)\mu = \underline{\mu}$$

$$E(\hat{\mu}_3) = E\left(175 + \frac{1}{100}X_1\right) = 175 + \frac{1}{100}E(X_1) = 175 + \frac{\mu}{100}$$

De 2 første estimatorene er forventningsrette, den 3. er bare forventningsrett hvis μ tilfeldigvis er 175.0/0.99. Vi sløyfer den i neste utregning.

$$Var(\hat{\mu}_1) = Var\left(\frac{2}{7}X_1 + \frac{5}{7}X_2\right) = \frac{4}{49}Var(X_1) + \frac{25}{49}Var(X_2) = \left(\frac{4}{49} + \frac{25}{49}\right)\sigma^2 = \underline{\frac{29}{49}\sigma^2}$$

$$Var(\hat{\mu}_2) = Var\left(\frac{1}{2}X_1 + \frac{1}{2}X_2\right) = \frac{1}{4}Var(X_1) + \frac{1}{4}Var(X_2) = \left(\frac{1}{4} + \frac{1}{4}\right)\sigma^2 = \underline{\frac{1}{2}\sigma^2}$$

Kommentar: Utregningen tar utgangspunkt i at høydene X_1 og X_2 har samme forventningsverdi μ og samme standardavvik σ .

Den beste estimatoren er den som gir den minste variansen. Siden $29/49 > 1/2$, velger vi estimator nr. 2. $\hat{\mu}_2$ er den beste estimatoren.

Kommentar: Det kan vises at estimatorer av typen $(aX_1 + (n-a)X_2)/n$ får minst varians når $a = n/2$:

$$\hat{\mu} = \frac{aX_1}{n} + \frac{(n-a)X_2}{n}, \quad E(\hat{\mu}) = \frac{a+n-a}{n}\mu = \underline{\mu}$$

$$Var(\hat{\mu}) = Var\left(\frac{a}{n}X_1 + \frac{n-a}{n}X_2\right) = \frac{a^2}{n^2}Var(X_1) + \frac{(n-a)^2}{n^2}Var(X_2) = \left(\frac{n^2 + 2a^2 - 2an}{n^2}\right)\sigma^2$$

Finner minimum for telleren ved derivasjon mhp. a :

$$\frac{d(n^2 + 2a^2 - 2an)}{da} = 4a - 2n = 0 \Rightarrow a = n/2$$

Den minste variansen blir dermed: $Var(\hat{\mu}) = \frac{n^2 + 2(n/2)^2 - n^2}{n^2}\sigma^2 = \underline{\frac{1}{2}\sigma^2} = Var(\hat{\mu}_2)$

Vi kan si at vi får det beste estimatet når vi legger like stor vekt på høydene til de to personene X_1 og X_2 .

6.5 Forventningsverdier og varianser

Binomisk fordeling: n enheter med defektandel p . X er # defekte enheter. $E(X)$ og $Var(X)$ gir forventningsverdi og varians til estimatoren $\hat{p}$ for den ukjente defektandelen p :

$$E(X) = np \quad Var(X) = np(1-p)$$

$$\hat{p} = \frac{X}{n} \quad E(\hat{p}) = \frac{1}{n}E(X) = p \quad Var(\hat{p}) = \frac{1}{n^2}Var(X) = \underline{\frac{p(1-p)}{n}}$$

Oppgitte estimatorer, setter $n_1 = 30$ og $n_2 = 70$:

$$E(\hat{p}_1) = E\left(\frac{X_1}{100} + \frac{X_2}{100}\right) = \frac{1}{100} E(X_1) + \frac{1}{100} E(X_2) = \frac{1}{100} n_1 p + \frac{1}{100} n_2 p = \frac{30 + 70}{100} p = p$$

$$E(\hat{p}_2) = E\left(\frac{3X_1}{580} + \frac{7X_2}{580}\right) = \frac{3}{580} n_1 p + \frac{7}{580} n_2 p = \frac{90 + 490}{580} p = p$$

Begge estimatorene er forventningsrette.

$$Var(\hat{p}_1) = Var\left(\frac{X_1 + X_2}{100}\right) = \frac{1}{10^4} (Var(X_1) + Var(X_2)) = \frac{1}{10^2} (n_1 p(1-p) + n_2 p(1-p)) =$$

$$\frac{30 + 70}{10^4} p(1-p) = \frac{1}{1000} p(1-p)$$

$$Var(\hat{p}_2) = Var\left(\frac{3X_1 + 7X_2}{580}\right) = \frac{9}{580^2} Var(X_1) + \frac{49}{580^2} Var(X_2) = \frac{270 + 3430}{580^2} (p(1-p)) =$$

$$\frac{3700}{580^2} p(1-p) = \frac{37}{3364} p(1-p)$$

Her har $\hat{p}_1$ den minste variansen. ($37/3364 > 1/1000$)

INTERVALLESTIMERING

6.12 Avstanden X er normalfordelt: $X \sim N(\mu, 2.04)$. Forventet avstand estimeres fra gjennomsnittet av $n = 16$ målinger. Gjennomsnittet er også normalfordelt:

$$\bar{X} = \frac{1}{16} \sum_{i=1}^{16} X_i, \quad E(\bar{X}) = \frac{1}{16} 16\mu = \mu, \quad Var(\bar{X}) = \frac{1}{16^2} 16\sigma^2 \Rightarrow \sigma_{\bar{X}} = \frac{\sigma}{4} = 0.51[m]$$

Et 95 % konf.intervall er gitt av 2.5%-kvantilen $z_{0.025} = 1.96$ (antall standardavvik):

$$[2316 - 1.96 \cdot 0.51, 2316 + 1.96 \cdot 0.51] = [2315, 2317][m]$$

Tilleggsoppgave:

Bredden L avhenger av # verdier n som brukes til å bestemme gjennomsnittsverdien:

$$[\bar{x} - z_{0.025} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{0.025} \frac{\sigma}{\sqrt{n}}] \Rightarrow L = 2z_{0.025} \frac{\sigma}{\sqrt{n}} \Rightarrow n = 4z_{0.025}^2 \frac{\sigma^2}{L^2}$$

$$n \geq 4 \cdot 1.96^2 \frac{2.04^2}{0.2^2} = 1598.7 \Rightarrow n \geq 1599$$

6.13 Bruker gjennomsnittet $\bar{x}$ av de 6 målingene som estimator for μ . Standardavviket σ estimeres med empirisk standardavvik s:

$$\bar{X} = \frac{1}{6} \sum_{i=1}^6 X_i, \quad E(\bar{X}) = \frac{1}{6} 6\mu = \mu, \quad Var(\bar{X}) = \frac{1}{6^2} 6\sigma^2 \Rightarrow \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{6}}$$

σ er ukjent, må bruke Student's t-fordeling med $6-1 = 5$ frihetsgrader. For et 95 % konfidensintervall er # standardavvik lik t-kvantilen

$$t_{0.025,5} = 2.571$$

Tallverdier fra målingene:

$$\bar{x} = 30.5[cm], \quad s = 1.0488[cm]$$

Et 95 % konfidensintervallet for μ blir:

$$\left[\bar{x} - t_{0.025,5} \frac{s}{\sqrt{6}}, \bar{x} + t_{0.025,5} \frac{s}{\sqrt{6}} \right] \Rightarrow [29.4, 31.6][cm]$$

6.14 Bredden L avhenger av # verdier n som brukes til å bestemme gjennomsnittsverdien:

$$\left[\bar{x} - t_{0.025,5} \frac{s}{\sqrt{n}}, \bar{x} + t_{0.025,5} \frac{s}{\sqrt{n}} \right] \Rightarrow L = 2t_{0.025,5} \frac{s}{\sqrt{n}} \Rightarrow n = 4t_{0.025,5}^2 \frac{s^2}{L^2}$$

$$n \geq 4 \cdot 2.571^2 \frac{1.0488^2}{0.2^2} = 727.09 \Rightarrow n \geq 728$$

Dette forutsetter at s er uendret. Std.avviket vil sannsynligvis ha en litt annen verdi i en ny undersøkelse. En bør måle flere enn 728 pizzaer.

6.16 Vi starter med en **binomisk fordeling**. X er # personer som foretrekker egen bil, utvalg på n personer, ukjent andel p foretrekker egen bil: $X \sim \text{bin}(n, p)$. Finner forventningsverdi og standardavvik for estimatoren $\hat{p} = X/n$ for den ukjente andelen p .

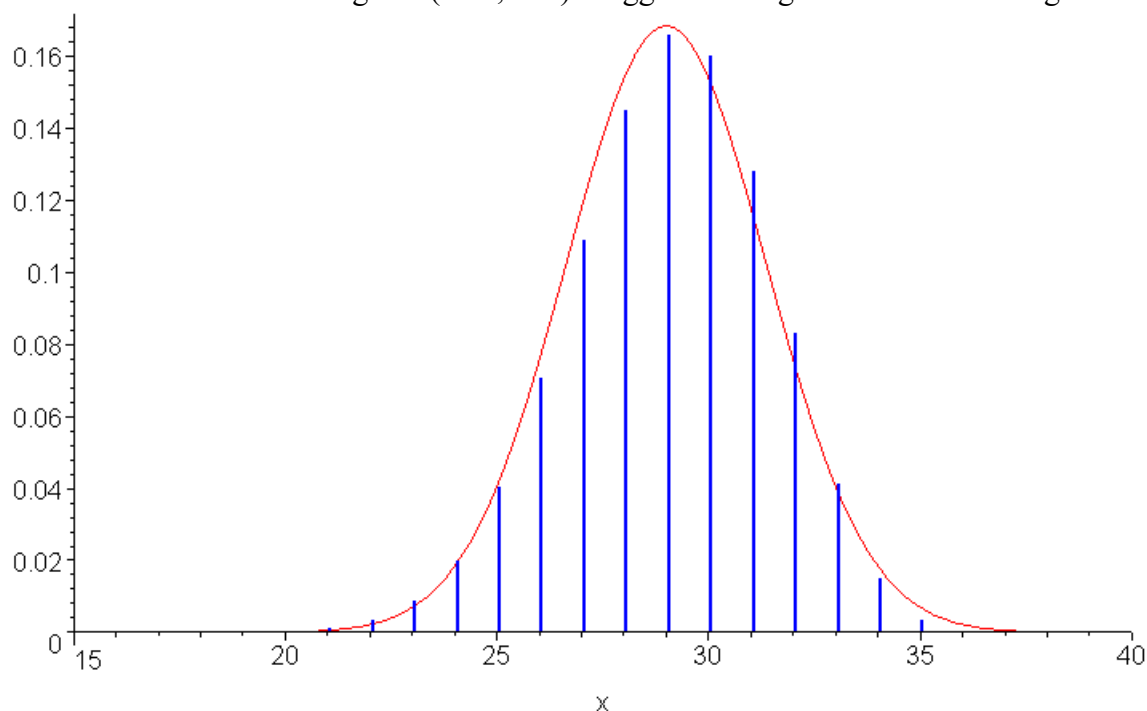
$$E(X) = np, \quad \text{Var}(X) = np(1-p)$$

$$\hat{p} = X/n, \quad E(\hat{p}) = \frac{1}{n} E(X) = p, \quad \text{Var}(\hat{p}) = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n}, \quad \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Vi bruker standardavviket til $\hat{p}$ for å regne ut konfidensintervallet, men σ avhenger av den ukjente verdien p som vi skal estimere. Erstatte p med $\hat{p}$ som vi finner fra forsøket:

$$\hat{p} = \frac{x}{n} = \frac{29}{36} = 0.806 \Rightarrow \sigma_{\hat{p}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.8055 \cdot 0.1945}{36}} = 0.066$$

I dette forsøket er $\text{Var}(X) \approx 36 \cdot 0.8055 \cdot 0.1945 = 5.6$, vi tror at den binomiske fordelingen kan erstattes med normalfordelingen $N(29.0, 2.37)$. Begge fordelingene er vist i neste figur:



Regner videre med $N(\hat{p}, \sigma_p) = N(0.806, 0.066)$ og finner et 90 % konfidensintervall for p :

$$[\hat{p} - z_{0.05}\sigma_{\hat{p}}, \hat{p} + z_{0.05}\sigma_{\hat{p}}] = [0.69, 0.92]$$

(Legg merke til at det standardavviket som brukes er standardavviket til den *estimatoren* som brukes.

Antallet som må spørres: se oppg. 12.