Oppgavegjennomgang BINOMISK FORDELING

5.51 Spør 20.000 studenter om de har stjålet. Andelen som stjeler: 0.2 = 20%
Trekker 10 tilfeldige fra disse, kan anta uavhengighet.

Klassisk binomisk "forsøk": $n = 10$ uavh. "forsøk", hvert med lik sannsynlighet for at personen stjeler eller ikke: P(stjele) = 0.2, P(ikke stjele) = 0.8
(2 mulige utfall, lik suksessannsynlighet)

Definerer en stokastisk variabel X = # som ikke stjeler

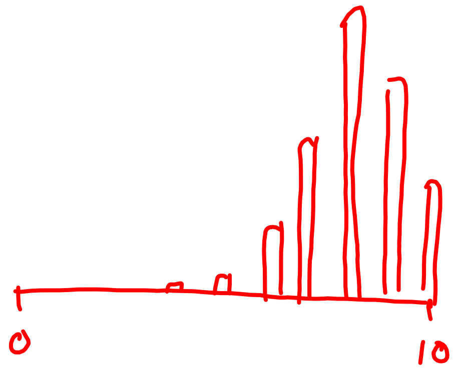
a) $X \sim \text{bin}(10, 0.8)$ $\quad P(X=x) = \binom{n}{x} 0.8^{x} \cdot 0.2^{(n-x)}$

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

$$\binom{10}{8} = \frac{10 \cdot 9 \cdot \cancel{8 \cdot 7 \cdots 1}}{\cancel{8 \cdot 7 \cdot 6 \cdots 1} \cdot 2 \cdot 1} = 45$$

x	0	1	2	3	4	5	6	7	8	9	10	sum
$P(X=x)$	0	0	0	0	0.01	0.03	0.09	0.20	0.30	0.27	0.11	1

$$\text{R> } \underbrace{\text{dbinom}}_{\text{"Regn ut sannsynligheten"}}(\underbrace{0:10}_{[0,1,2,3,4,5,6,7,8,9,10]},\ \underset{\text{"n=10"}}{10},\ \underset{\text{"p=0.8"}}{0.8})$$



$\binom{n}{x}$ $n=10$, $x=0$ $\binom{10}{0} = \frac{\cancel{10!}}{\cancel{10!}\cdot \underbrace{0!}_{=1}} = 1$

$n=10$, $x=1$ $\binom{10}{1} = \frac{10!}{1!\cdot 9!} = \frac{10\cdot \cancel{9\cdot 8\cdot 7\cdot 6 \cdots 1}}{1\cdot \cancel{9\cdot 8\cdot 7\cdots 1}} = 10$

b) P(minst 4 stjeler)

Y = # som stjeler

- n = 10 uavh forsøk
- To mulige utfall
- Lik sh for stjeling i hvert forsøk

$\Big\}\; Y \sim bin(10, 0.2)$

$P(\text{minst 4 stjeler}) = P(Y \geq 4) = 1 - P(Y < 4) = 1 - P(\underline{Y \leq 3})$

R> 1 − sum(dbinom(0:3, 10, 0.2)) ≈ <u>0.12</u>

6.20 Måler osteokalcin, $n = 31$ uavh. målinger

$\boxed{\text{KONFIDENSINTERVALL FOR } \mu}$

$\bar{X} = 33.4$ σ kjent, $\sigma = 19.6$

jfr situasjon 1 & 2 fra forelesningen på fredag.

Hvis enkeltobservasjonene, X_i-ene var $\sim N(\mu, \sigma)$ σ kjent:

Da vil også $\bar{X} \sim N$, men forventningen til $\bar{X} = \mu$ og standardavviket til $\bar{X} = \frac{\sigma}{\sqrt{n}}$, så

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \qquad \bar{X} \overset{\text{tiln}}{\sim} N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1) \qquad Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \overset{\text{tiln}}{\sim} N(0,1)$$

Hvis enkeltobservasjonene, X_i-ene var $\sim ?(\mu, \sigma)$ σ kjent, n stor

Da vil, i følge sentralgrenseteoremet,

95 % konfidensintervall for μ, "populasjonsgjennomsnittet" av osteokalcin
forventningsverdien

$$\boxed{\bar{X} \pm 1.96 \cdot \frac{19.6}{\sqrt{31}}} \longrightarrow \underline{[26.5, 40.3]}$$

But why?

Vi vet, vi tar utgangspunkt i $Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \overset{\text{(tiln)}}{\sim} N(0,1)$

$$P(-1.96 < Z < 1.96) = 95\%$$

$$P\left(-1.96 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < 1.96\right) = 0.95$$

vil ha denne i midten: "95% KI for μ"

$$P\left(-1.96 \cdot \frac{\sigma}{\sqrt{n}} < (\bar{X} - \mu) < 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

$$P\left(-\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} < -\mu < -\bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

$$P\left(\bar{X} + 1.96 \frac{\sigma}{\sqrt{n}} > \mu > \bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

95% KI for μ er gitt ved

$$\bar{X} \pm 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

$$E(\bar{X}) = E\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n}\sum E(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}\cdot n\cdot\mu = \mu$$

$$E\left(\frac{1}{n}\cdot X_1 + \frac{1}{n}X_2 + \frac{1}{n}\cdot X_3 + \dots + \frac{1}{n}\cdot X_n\right)$$

$$Var(\bar{X}) = Var\left(\frac{1}{n}\sum X_i\right) \overset{X_i\text{ uavh}}{=} \frac{1}{n^2}\cdot\sum Var(X_i) = \frac{1}{n^2}\cdot\sum\sigma^2 = \frac{1}{n^2}\cdot n\,\sigma^2 = \frac{\sigma^2}{n}$$

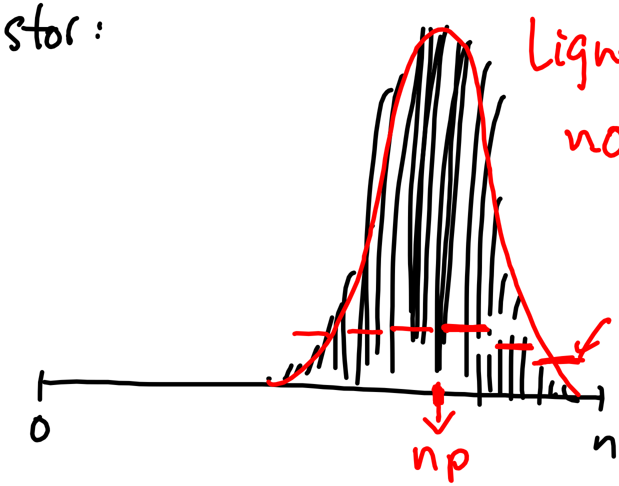
$$sd(\bar{X}) = \sqrt{Var(\bar{X})} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

KONFIDENSINTERVALL FOR p

Binomisk fordeling med n uavh forsøk, suksessannsynlighet p:

X = # suksesser på n forsøk:

$X \sim \text{Bin}(n, p)$ Hvis n stor:



Binomisk fordeling ligner på en normalfordeling

$\sigma = \sqrt{np(1-p)}$

Ok, så $X \overset{\text{tiln}}{\sim} N$, og hva er $E(X) = \mu$ og $sd(X) = \sigma$, her?

I en binomisk fordeling, så er $E(X) = n \cdot p$

$Var(X) = n \cdot p(1-p)$, $\sigma_X = \sqrt{n \cdot p \cdot (1-p)}$

Hvis

$X \sim \text{bin}(n, p)$ og n stor, og p ikke for nær 0 eller 1, så er

$$X \overset{\text{tiln}}{\sim} N\left(np, \sqrt{n \cdot p(1-p)}\right)$$

$$\frac{\sqrt{n \cdot p \cdot (1-p)}}{n} = \frac{\cancel{\sqrt{n}} \cdot \sqrt{p(1-p)}}{\cancel{\sqrt{n}} \cdot \sqrt{n}}$$

$$Z = \frac{X - np}{\sqrt{n \cdot p(1-p)}} \overset{\text{tiln}}{\sim} N(0,1)$$

$$\frac{\#\text{ suksesser}}{\#\text{ forsøk}} = \text{Andel i utvalg} = \hat{p}$$

$$Z = \frac{\frac{X}{n} - \frac{\cancel{n}p}{\cancel{n}}}{\frac{\sqrt{n \cdot p \cdot (1-p)}}{n}} \overset{\text{tiln}}{\sim} N(0,1)$$

$$Z = \frac{\hat{p} - p}{\frac{\sqrt{p(1-p)}}{\sqrt{n}}} \overset{\text{tiln}}{\sim} N(0,1)$$

jfr $Z = \dfrac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \overset{\text{(tiln)}}{\sim} N(0,1)$

95% Konfidensintervall for p :

$$P(-1.96 < Z < 1.96) \approx 0.95$$

$$P\left(-1.96 < \frac{\hat{p}-p}{\frac{\sqrt{p(1-p)}}{\sqrt{n}}} < 1.96\right) \approx 0.95$$

Trikser litt:

$$P\left(-1.96 < \frac{\hat{p}-p}{\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}} < 1.96\right) \approx 0.95$$

$$P\left(-1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} < (\hat{p}-p) < 1.96\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right) \approx 0.95$$

$$P\left(-\hat{p}-1.96\cdot\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} < -p < -\hat{p}+1.96\cdot\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right) \approx 0.95$$

$$P\left(\hat{p}+1.96\cdot SE(\hat{p}) > p > \hat{p}-1.96\cdot SE(\hat{p})\right) \approx 0.95$$

Altså : 95 % KI for p er gitt ved

$$\hat{p} \pm 1.96 \cdot \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$$

jfr 95 % KI for μ :

$$\bar{X} \pm 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

$$t \cdot \frac{sd}{\sqrt{n}}$$

$$\Big\downarrow n \to \infty$$

$$1.96 \cdot \frac{sd}{\sqrt{n}}$$

Oppgave 8.13 p = Andel av studenter som spiser frokost regelmessig

$n = 200$ $X = \#$ som spiser frokost regelmessig, av de $n = 200$

$$\frac{X}{n} = \hat{p} = \frac{84}{200}$$

Altså: $\hat{p} = 0.42$ er et estimat for p.

$$SE(\hat{p}) = \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} = \frac{\sqrt{0.42 \cdot 0.58}}{\sqrt{200}} = 0.035$$

95% KI for p: $\hat{p} \pm 1.96 \cdot 0.035 \rightarrow [0.35,\ 0.49]$

Dette betyr at basert på de $n = 200$ studentene som ble spurt, kan vi ikke utelukke at den samme andelen studenter som spiser frokost er så liten som 35%, eller så høy som 49%, men vårt beste estimat er 42%.