TMA4240 F20
28.10.2016

Parutvalg [9.9]

EKS: kroppsfettprosent før & etter pedikyr.

a) Uavhengige utvalg ⇒ Nei. Målingene er gjort på samme person før og etter pedikyr

X_{1i} = før
X_{2i} = etter
pedikyr, kroppsfettprosent
$i = 1, \dots, n$

b) $X_{1i} \sim N(\mu_1, \sigma_1^2)$, $X_{2i} \sim N(\mu_2, \sigma_2^2)$

Si noe om $\underbrace{\mu_2 - \mu_1}_{\delta}$ → estimator, KI.

$$\hat{\delta} = \hat{\mu}_2 - \hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} X_{1i} - \frac{1}{n}\sum_{i=1}^{n} X_{2i}$$

$$E(\hat{\delta}) = E(\hat{\mu}_2 - \hat{\mu}_1) = \mu_2 - \mu_1$$

$$Var(\hat{\delta}) = Var(\hat{\mu}_2) + Var(\hat{\mu}_1) + 2 \cdot 1 \cdot (-1)\, Cov(\hat{\mu}_2, \hat{\mu}_1)$$

$$Var(\hat{\delta}) = Var(\hat{\mu}_2) + Var(\hat{\mu}_1) - 2\underbrace{Cov(\hat{\mu}_2, \hat{\mu}_1)}_{>0}$$

$\Rightarrow$ Dvs at å bruke

$Var(\hat{\delta}) = Var(\hat{\mu}_2) + Var(\hat{\mu}_1)$ vil (oftest)

gi en alt for stor varians $\Rightarrow$ Ser heller

direkte på $D_i = X_{2i} - X_{1i}$

3) $D_1, D_2, .., D_n$ u.i.f $D_i \sim N(\mu_D, \sigma_D^2)$

$\Rightarrow$ gjør dette om til ett utvalg!

$$\hat{\mu}_d = \frac{1}{n}\sum_{i=1}^{n} D_i \quad , \quad S_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(D_i - \bar{D})^2$$

$(1-\alpha)100\%$ KI for μ_d

$$\left[\ \bar{d} \pm t_{\frac{\alpha}{2}, n-1} \, s_d/\sqrt{n}\ \right]$$

$$\frac{\bar{D} - \mu_d}{S_d/\sqrt{n}} \sim t_{n-1}$$

EKS: 95%

$n = 40 \rightarrow t_{0.025,39} = 2.030$

$\bar{d} = 46.1 - 45.8 = 0.3$

$s_d = 1.10$

$$\left[0.3 \pm 2.030 \cdot \frac{1.10}{\sqrt{40}} \right] = \underline{\underline{[-0.05, 0.65]}}$$

$\Rightarrow$ ingen bevist effekt av pedikyr på måling kroppsfett-prosent fordi 95% KI inneholder 0

Estimering av andeler [9.10, 9.11]

p

Ett utvalg

EKS:

$n = 331$ med sertifikat

$x = 170$ sier bedre enn gjisn til å kjøre bil

Situasjon: - n spurt, uavhengig
(antar ∞ befolkning) - "$\underbrace{\text{bedre enn}}_{\text{suksess}}$ gjisn eller ikke"

- $P(\text{suksess}) = p$

$X = \#\text{suksesser} \sim \text{bin}(n, p)$

kjent, ukjent

Estimator for p: $\hat{p} = \frac{X}{n}$ (SME)

EKS: $\hat{p} = \frac{170}{331} = 0.51$

Kan lage KI basert på binomisk fordeling (litt teknisk).

Når n er stor kan vi bruke sentralgrense-teoremet.

$$E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{1}{n} E(X) = \frac{1}{n} \cdot n \cdot p = p$$

$$Var(\hat{p}) = Var\left(\frac{X}{n}\right) = \left(\frac{1}{n}\right)^2 Var(X)$$

$$= \frac{n \cdot p(1-p)}{n^2} = \frac{p(1-p)}{n}$$

Når n er stor:

$$Z = \frac{\hat{p} - E(\hat{p})}{SD(\hat{p})} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

$$P\Bigg(\underbrace{-z_{\frac{\alpha}{2}} < \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}}_{1} \; \underbrace{< z_{\frac{\alpha}{2}}}_{2}\Bigg) = 1-\alpha$$

Kan løse de to ulikhetene, får 2gradlign i p (metode 2 i læreboka).

Vi bruker et tilnærmet resultat og bytter ut p med $\hat{p}$ i nevneren.

$$P\left(-z_{\frac{\alpha}{2}} < \frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} < z_{\frac{\alpha}{2}}\right) \approx 1-\alpha$$

(metode 1 i læreboka)

$\left[\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$ er et tilnærmet $(1-\alpha)\,100\%$ KI for p når n er stor (og p ikke er veldig nært 0 eller 1)

EKS: 99% KI

$\hat{p} = 0.51$, $n = 331$, $z_{0.005} = 2.576$

$$\left[0.51 \pm 2.576 \cdot \sqrt{\frac{0.51 \cdot (1-0.51)}{331}}\right]$$

$$= \underline{\underline{[0.48,\ 0.54]}}$$

To utvalg: differense mellom andeler

$X_1 \sim bin(n_1, p_1)$
$X_2 \sim bin(n_2, p_2)$ uavhengig

$(p_1 - p_2)$ = differanse mellom andeler

1) Estimator: $\hat{p}_1 - \hat{p}_2 = \frac{X_1}{n_1} - \frac{X_2}{n_2}$

2) KI ⇒ trenger størrelse med kjent fordeling der $p_1 - p_2$ inngår.

$$E(\hat{p}_1 - \hat{p}_2) = E(\hat{p}_1) - E(\hat{p}_2) = p_1 - p_2$$

$$Var(\hat{p}_1 - \hat{p}_2) = Var(\hat{p}_1) + Var(\hat{p}_2)$$

$$= \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$$

SGT

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \approx N(0,1)$$

$(1-\alpha)\cdot 100\%$ KI (setter $\hat{p}_1$ og $\hat{p}_2$ inn for p_1 og p_2 i nevner)

$$\left[\hat{p}_1-\hat{p}_2 \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}\right]$$

Dette er et tilnærmet intervall!

EKS:

1: menn $n_1 = 200$, $\hat{p}_1 = 0.605$

2: kvinner $n_2 = 131$, $\hat{p}_2 = 0.374$

Estimat for $p_1 - p_2 = 0.605 - 0.374 = 0.23$

99% tilnærmet KI:

$$\underbrace{(\hat{p}_1-\hat{p}_2)}_{0.23} \pm \underbrace{z_{0.005}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}_{0.05}$$

$$\underline{\underline{[0.18\ ,\ 0.28]}}$$

på forelesning skrev jeg 0.20, 0.26 og det var jo galt...

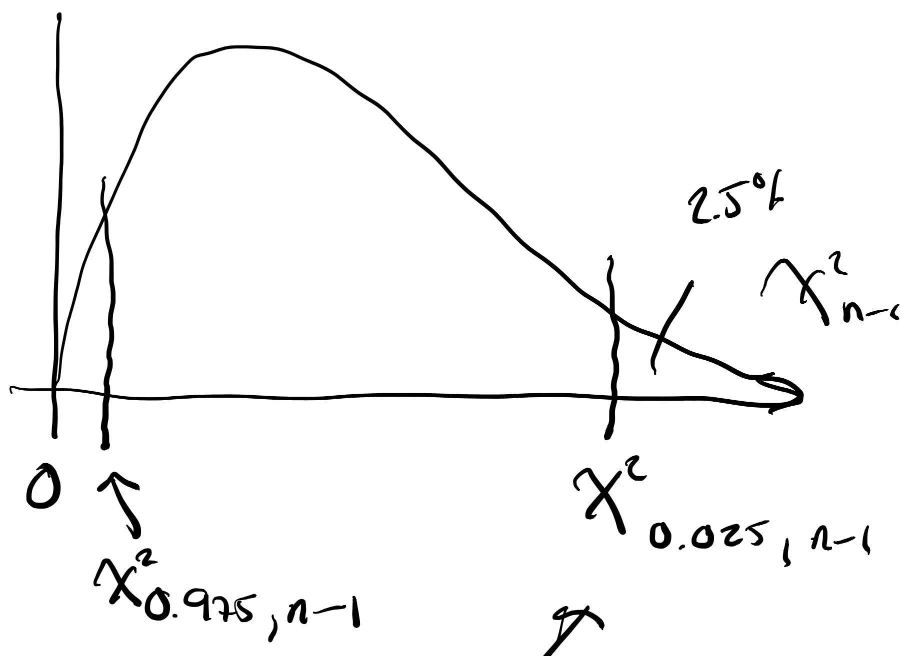
Konfidensintervall for varians [9.12]

$X_1, X_2, \ldots, X_n$ u.i.f $N(\mu, \sigma^2)$

Ønsker 95% KI for σ^2.

1) Estimator: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

2) Fordeling: $V = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$



begge må finnes i tabellen

$$P\left(\chi^2_{0.975,\,n-1} < V < \chi^2_{0.025}\right) = 0.95$$

$$V = \frac{(n-1)S^2}{\sigma^2}$$

$\Rightarrow$ løser ut

Se video! August 2015 # 3a