

- Kva er gjennomsnittshøgda for kvinnelege NTNU-studentar. **Estimering**
- Er eg høgare enn gjennomsnittleg kvinnelege NTNU-studentar. **Hypotesetest**
- Kva er forskjellen i kvalitet på laks for lagringsmetode 1 og lagringsmetode 2. **Estimering**
- Er kvaliteten på laks ved lagringsmetode 2 betre enn ved lagringsmetode 1. **Hypotesetest**

- Har data $x_1, x_2, \dots, x_n$
- Antar fordeling frå kjennskap til fenomenet; $x_i \sim f(x; \theta)$.
- *Estimerer/ berekner θ frå data v.h.a. ein estimator $\hat{\theta}$.*
 - Kva er ein god estimator (i dag)
 - Korleis finne ein estimator (i dag)
 - Korleis kvantifisere usikkerheita i estimat (torsdag)

Du har fått sommarjobb på vindusfabrikken. Dei tar i bruk ny produksjonsteknikk, og ønsker å finne defektsannsynet p .

- $m_1 = 40$ vindu produsert i første skift
- $m_2 = 60$ vindu i andre skift.
- $u = 5$ defekte i første skift.
- $v = 15$ defekte i andre skift.

Foreslå minst to måtar å *estimere* p på.

Kva sannsynsfordeling har U og V ?

- La $X_\nu = 0$ dersom vindu nr ν er ikkje-defekt, og $X_\nu = 1$ dersom defekt. $\nu = 1, 2, \dots, m_1 + m_2$
- $P(X_\nu = x_\nu) = p^{x_\nu}(1 - p)^{1-x_\nu}$; sannsynsfordeling
- $U = \sum_{\nu=1}^{m_1} X_\nu$; antall defekte i første skift.
- $V = \sum_{\nu=m_1+1}^{m_1+m_2} X_\nu$; antall defekte i andre skift.

Foreslå minst to *estimatorar* for p

Bernoulli prosess

- 1 n uavhengige forsøk
- 2 Kvant forsøk resulterer i suksess, $I_i = 1$ eller ikkje-suksess $I_i = 0$.
- 3 Suksess-sannsynet $p = P(I_i = 1)$ er konstant.

Binomisk fordeling

Ser på antall suksess i ein Bernoulli prosess, $X = \sum_{i=1}^n I_i$. X er då binomisk fordelt;

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

Dersom $X \sim \text{Bin}(n, p)$ så er

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

- $\hat{p}(X_1, X_2, \dots, X_{m_1+m_2}) = \frac{U + V}{m_1 + m_2}$
- $\hat{\hat{p}}(X_1, X_2, \dots, X_{m_1+m_2}) = 0.5\left(\frac{U}{m_1} + \frac{V}{m_2}\right)$

Forventningsrette:

- $E(\hat{p}) = p$
- $E(\hat{\hat{p}}) = p$

Viktige rekneregler

Forventning til lineærkombinasjon:

$$E(a_0 + a_1 X_1 + \dots + a_n X_n) = a_0 + a_1 E(X_1) + \dots + a_n E(X_n)$$

Varians til lineærfunksjon:

$$\text{Var}(a + bX_1) = b^2 \text{Var}(X_1)$$

Dersom $X_1, X_2, \dots, X_n$ er **uavhengige**:

$$\text{Var}(a_0 + a_1 X_1 + a_2 X_2 + \dots + a_n X_n) = a_1^2 \text{Var}(X_1) + a_2^2 \text{Var}(X_2) + \dots + a_n^2 \text{Var}(X_n)$$

Antar uavhengighet.

- $Var(\hat{p}) = Var(\frac{U + V}{m_1 + m_2}) = \frac{p(1-p)}{m_1 + m_2}$
- $Var(\hat{\hat{p}}) = Var(0.5(\frac{U}{m_1} + \frac{V}{m_2})) = 0.25p(1-p)(\frac{1}{m_1} + \frac{1}{m_2})$

Kan vise at

$$Var(\hat{\hat{p}}) - Var(\hat{p}) = p(1-p) \frac{(m_1 - m_2)^2}{4m_1 m_2 (m_1 + m_2)} \geq 0$$

Altså $Var(\hat{\hat{p}}) \geq Var(\hat{p}) \Rightarrow$ foretrekker $\hat{\hat{p}}$.

Finn den verdien for parameteren θ (studentjul $\theta = p$) som gjev høgast sannsyn for å observere dei dataene vi har observert.

OPPSKRIFT

- Finn likelihoodfunksjonen

$$L(\theta; x_1, x_2, \dots, x_n) = f(x_1, x_2, \dots, x_n; \theta) \stackrel{\text{uavh}}{=} \prod_{i=1}^n f(x_i; \theta)$$

- Finn toppunkt av likelihoodfunksjonen:

- Tar ln av L ; $l(\theta; x_1, x_2, \dots, x_n) = \ln(L(\theta; x_1, x_2, \dots, x_n))$.
Reknetriks som nesten alltid blir brukt. L og l har same toppunkt.

- Deriverer og set lik 0; $\frac{\partial}{\partial \theta} l(\theta; x_1, x_2, \dots, x_n) = 0$

- Løyer ut for θ . *Skal eigentleg sjekke endepunkt og at det er eit topp-punkt og.*

- Har data $x_1, x_2, \dots, x_n$
- Antar fordeling frå kjennskap til fenomenet; $x_i \sim f(x; \theta)$.
- *Estimerer/ berekner θ frå data v.h.a. ein estimator $\hat{\theta}$.*
 - Kva er ein god estimator (forventningsrett og minst mogeleg varians).
 - Korleis finne ein estimator (SME)
 - Kvantifisere usikkerheita i estimat (konfidensintervall)