

Valg av testobservator i en del hypotesetestingssituasjoner

TMA4240/TMA4245 Statistikk

Håkon Tjelmeland

Institutt for matematiske fag

Norges teknisk-naturvitenskapelige universitet

Hypotesetesting

- ★ Formulere to hypoteser, H_0 og H_1
 - ønsker å undersøke om det er grunnlag for å påstå at H_1 er sann
- ★ Velge en testobservator, T
 - skal være en observator: kan ikke inneholde ukjente parametre
 - må ha kjent fordeling når H_0 er sann
- ★ Bestem beslutningsregel
 - forkast H_0 hvis $T \geq k$
 - forkast H_0 hvis $T \leq k$
 - forkast H_0 hvis $T \leq k_\ell$ eller $T \geq k_u$
- ★ To metoder videre
 - forkastningsområdemetoden
 - p-verdi-metoden
- ★ Fokus i denne videoen: Valg av testobservator

Hvilke situasjoner skal vi se på?

- ★ $X_1, X_2, \dots, X_n$ tilfeldig utvalg fra $N(\mu, \sigma^2)$ -populasjon
 - hypotesetest om μ, σ^2 er kjent
 - hypotesetest om μ, σ^2 er ukjent
 - hypotesetest om σ^2, μ er kjent
 - hypotesetest om σ^2, μ er ukjent
- ★ X er binomisk fordelt med parametre n og p
 - hypotesetest om p
- ★ To-utvalg: $X_1, X_2, \dots, X_n$ tilfeldig utvalg fra $N(\mu_1, \sigma_1^2)$ -populasjonen og $Y_1, Y_2, \dots, Y_m$ tilfeldig utvalg fra $N(\mu_2, \sigma_2^2)$ -populasjonen, X_i -ene og Y_i -ene uavhengige
 - hypotesetest om $\mu_1 - \mu_2, \sigma_1^2$ og σ_2^2 kjente
- ★ Merk: Dette er de samme situasjonene som vi diskuterte da vi så på konfidensintervall
 - vi skal bruke egenskaper vi utledet da vi diskuterte konfidensintervall
 - merk sammenheng: valg av pivotal for konfidensintervall — valg av testobservator

Hypotesetesting i normalfordeling, en ukjent parameter

★ Situasjon

- la $X_1, X_2, \dots, X_n$ være tilfeldig utvalg fra $N(\mu, \sigma^2)$ -populasjonen
- verdien til μ er ukjent, hypotesetest med $H_0: \mu = \mu_0$, antar verdien til σ^2 kjent

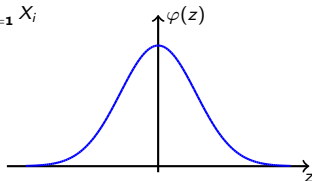
★ Tar utgangspunkt i estimator for μ : $\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

★ Vi vet at

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0, 1)$$

★ Testobservator

$$Z = \frac{\bar{X} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}, \quad Z \sim N(0, 1) \text{ når } H_0 \text{ sann}$$



★ Situasjon

- la $X_1, X_2, \dots, X_n$ være tilfeldig utvalg fra $N(\mu, \sigma^2)$ -populasjonen
- verdien til σ^2 er ukjent, nullhypotese med $H_0: \sigma^2 = \sigma_0^2$, antar verdien til μ kjent

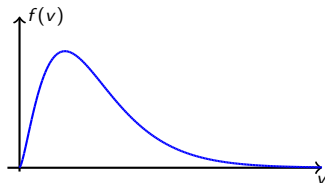
★ Tar utgangspunkt i SME for σ^2 : $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$

★ Vi vet at

$$\frac{n\hat{\sigma}^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi_n^2$$

★ Testobservator

$$V = \frac{n\hat{\sigma}^2}{\sigma_0^2} = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma_0} \right)^2, \quad V \sim \chi_n^2 \text{ når } H_0 \text{ sann}$$



Hypotesetesting om μ når også verdien til σ^2 er ukjent

★ Situasjon

- la $X_1, X_2, \dots, X_n$ være tilfeldig utvalg fra $N(\mu, \sigma^2)$ -populasjonen
- verdien til μ er ukjent, hypotesetest med $H_0: \mu = \mu_0$, antar verdien til σ^2 ukjent

★ Da verdien til σ^2 var kjent tok vi utgangspunkt i

$$Z = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0, 1)$$

★ Siden σ^2 er ukjent erstatter vi denne med estimator, og får

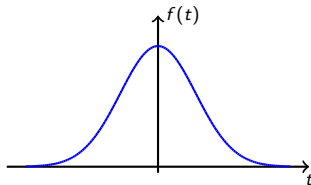
$$\frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}} \quad - \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

★ Vi vet da at

$$\frac{\bar{X} - \mu}{\sqrt{\frac{S^2}{n}}} \sim t_{n-1}$$

★ Testobservator

$$T = \frac{\bar{X} - \mu_0}{\sqrt{\frac{S^2}{n}}}, \quad T \sim t_{n-1} \text{ når } H_0 \text{ sann}$$



Hypotesetesting om σ^2 når også verdien til μ er ukjent

★ Situasjon

- la $X_1, X_2, \dots, X_n$ være tilfeldig utvalg fra $N(\mu, \sigma^2)$ -populasjonen
- verdien til σ^2 er ukjent, hypotesetest med $H_0 : \sigma^2 = \sigma_0^2$, antar verdien til μ ukjent

★ Da verdien til μ var kjent tok vi utgangspunkt i

$$\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi_n^2$$

★ Siden μ er ukjent erstatter vi denne med en estimator, og får

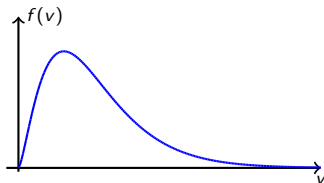
$$\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 \quad - \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

★ Vi vet da at

$$\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 \sim \chi_{n-1}^2$$

★ Testobservator

$$V = \frac{(n-1)S^2}{\sigma_0^2} = \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma_0} \right)^2, \quad V \sim \chi_{n-1}^2 \text{ når } H_0 \text{ sann}$$



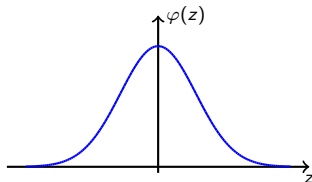
Hypotesetest om p i binomisk fordeling

★ Situasjon

- la X være binomisk fordelt med n forsøk og sannsynlighet for suksess lik p
- verdien til p er ukjent, hypotesetest med $H_0 : p = p_0$

★ Antar n stor slik at

$$X \approx N(np, np(1-p)) \Rightarrow \frac{X - np}{\sqrt{np(1-p)}} \approx N(0, 1)$$



★ Testobservator

$$Z = \frac{X - np_0}{\sqrt{np_0(1-p_0)}} = \frac{\frac{X}{n} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}, \quad Z \approx N(0, 1) \text{ når } H_0 \text{ sann}$$

To-utvalg: hypotesetest om $\mu_1 - \mu_2$ når verdiene til σ_1^2 og σ_2^2 er kjente

★ Situasjon

- la $X_1, X_2, \dots, X_n$ være tilfeldig utvalg fra $N(\mu_1, \sigma_1^2)$ -populasjonen
- la $Y_1, Y_2, \dots, Y_m$ være tilfeldig utvalg fra $N(\mu_2, \sigma_2^2)$ -populasjonen
- anta X_i -ene og Y_i -ene er uavhengig av hverandre
- verdiene til μ_1 og μ_2 er ukjente, hypotesetest med $H_0 : \mu_1 - \mu_2 = 0$
- antar verdiene til σ_1^2 og σ_2^2 kjente

★ Estimatorer for μ_1 og μ_2

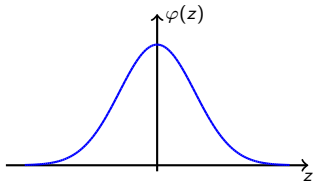
$$\hat{\mu}_1 = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{og} \quad \hat{\mu}_2 = \bar{Y} = \frac{1}{m} \sum_{i=1}^m Y_i$$

★ Vi vet at

$$\bar{X} \sim N\left(\mu_1, \frac{\sigma_1^2}{n}\right) \quad \text{og} \quad \bar{Y} \sim N\left(\mu_2, \frac{\sigma_2^2}{m}\right), \quad \bar{X} \text{ og } \bar{Y} \text{ er uavhengige}$$

★ Estimator for $\mu_1 - \mu_2$

$$\begin{aligned} \bar{X} - \bar{Y} &\sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right) \\ &\Downarrow \\ \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} &\sim N(0, 1) \end{aligned}$$



★ Testobservator

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}, \quad Z \sim N(0, 1) \text{ når } H_0 \text{ sann}$$

Oppsummering

- ★ Har sett på hvilke testobservatorer vi kan bruke i hypotesetester
- ★ $X_1, X_2, \dots, X_n$ tilfeldig utvalg fra $N(\mu, \sigma^2)$ -populasjon
 - hypotesetest om μ , σ^2 er kjent
 - hypotesetest om μ , σ^2 er ukjent
 - hypotesetest om σ^2 , μ er kjent
 - hypotesetest om σ^2 , μ er ukjent
- ★ X er binomisk fordelt med parametre n og p
 - hypotesetest om p
- ★ To-utvalg: $X_1, X_2, \dots, X_n$ tilfeldig utvalg fra $N(\mu_1, \sigma_1^2)$ -populasjonen og $Y_1, Y_2, \dots, Y_m$ tilfeldig utvalg fra $N(\mu_2, \sigma_2^2)$ -populasjonen, X_i -ene og Y_i -ene uavhengige
 - hypotesetest om $\mu_1 - \mu_2$, σ_1^2 og σ_2^2 kjente

