

STÆ209G formúlublað

Vinsamlegast ekki skrifa á formúlublaðið

Lýsistærðir

$$\text{Miðja spannar} = \frac{x_{\min} + x_{\max}}{2}$$

$$\text{Spönn} = x_{\max} - x_{\min}$$

$$\text{Sæti í röð miðgildi} = 0.5 \cdot (n + 1)$$

$$\text{Sæti í röð } Q_1 = 0.25 \cdot (n + 1)$$

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\text{Sæti í röð } Q_3 = 0.75 \cdot (n + 1)$$

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

$$\bar{x}_w = \frac{w_1 \cdot x_1 + w_2 \cdot x_2 + \dots + w_n \cdot x_n}{w_1 + w_2 + \dots + w_n} = \frac{\sum_{i=1}^n w_i \cdot x_i}{\sum_{i=1}^n w_i}$$

$$IQR = Q_3 - Q_1$$

$$r = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

Strjálur líkindadreifingar

$$P(X \leq k) = 1 - P(X > k)$$

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

$$P(X < k) = 1 - P(X \geq k)$$

$$\text{Tvíkostadreifing: } P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, E[X] = np, \text{ VAR}[X] = np(1-p)$$

$$P(X \geq k) = 1 - P(X < k)$$

$$P(X > k) = 1 - P(X \leq k)$$

$$\text{Poisson dreifing: } P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}, E[X] = \lambda, \text{ VAR}[X] = \lambda$$

Samfelldar líkindadreifingar

$$P(a < X < b) = P(X < b) - P(X < a) \quad \text{Normaldreifingin: } f(x) = \phi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$P(X > a) = 1 - P(X < a)$$

$$Z = \frac{X-\mu}{\sigma}, \text{ p.s. } X \sim N(\mu, \sigma^2)$$

$$x = \mu + z\sigma$$

Höfuðsetning tölfræðinnar

Ef X_1, \dots, X_n eru óháðar og einsdreifðar slembistærðir þá fylgir \bar{X} normaldreifingu með væntigildi μ og dreifni σ^2/n

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

óháð dreifingu X_1, \dots, X_n ef n er nógu stórt. Staðalskekkjan (staðalfrávik \bar{X}) er $\frac{\sigma}{\sqrt{n}}$.

Framkvæmd tilgátuprófa

- 1 Ákveða hvaða tilgátupróf er viðeigandi fyrir gögnin okkar.
- 2 Ákveða hæstu ásættanlegu villulíkur.
- 3 Setja fram núlltilgátu og ákveða um leið áttun prófsins (einhliða/tvíhliða).
- 4 Reikna prófstærðina sem svarar til tilgátuprófsins.
- 5a Kanna hvort prófstærðin falli á höfnunarsvæði tilgátuprófsins.
- 5b Kanna p-gildi tilgátuprófsins.
- 6 Draga ályktun.

Villa af gerð I og II

	H_0 er sönn	H_0 er röng
Hafna H_0	Villa af gerð I Líkur: α	Rétt ályktun Líkur: $1 - \beta$
Hafna ekki H_0	Rétt ályktun Líkur: $1 - \alpha$	Villa af gerð II Líkur: β

Ályktanir fyrir hlutföll

Öryggisbil og tilgátupróf fyrir hlutfall þýðis (p)

Séu skilyrðin fyrir að nota normalnálgunina uppfyllt má reikna neðra öryggismark fyrir p með:

$$\hat{p} - z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

og efra öryggismarkið með

$$\hat{p} + z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

þar sem $\hat{p} = \frac{x}{n}$.

Prófstærðin er:

$$Z = \frac{X - np_0}{\sqrt{np_0(1-p_0)}}$$

Höfnunarsvæðin eru:

Gagntilgáta	Hafna H_0 ef:
$H_1 : p < p_0$	$Z < -z_{1-\alpha}$
$H_1 : p > p_0$	$Z > z_{1-\alpha}$
$H_1 : p \neq p_0$	$Z < -z_{1-\alpha/2}$ eða $Z > z_{1-\alpha/2}$

Öryggisbil og tilgátupróf fyrir hlutföll tveggja þýða (p_1 og p_2)

Séu skilyrðin fyrir að nota normalnálgunina uppfyllt má reikna neðra öryggismark fyrir muninn á p_1 og p_2 með:

$$\hat{p}_1 - \hat{p}_2 - z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

og efra öryggismarkið með

$$\hat{p}_1 - \hat{p}_2 + z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

þar sem $\hat{p}_1 = \frac{x_1}{n_1}$, $\hat{p}_2 = \frac{x_2}{n_2}$.

Séu skilyrðin fyrir að nota normalnálgunina uppfyllt má nota eftirfarandi. Prófstærðin er:

$$Z = \frac{\frac{X_1}{n_1} - \frac{X_2}{n_2}}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \quad \text{þar sem } \hat{P} = \frac{X_1 + X_2}{n_1 + n_2}$$

Höfnunarsvæðin eru:

Gagntilgáta	Hafna H_0 ef:
$H_1 : p_1 < p_2$	$Z < -z_{1-\alpha}$
$H_1 : p_1 > p_2$	$Z > z_{1-\alpha}$
$H_1 : p_1 \neq p_2$	$Z < -z_{1-\alpha/2}$ eða $Z > z_{1-\alpha/2}$

Kí-kvaðrat próf

Prófstærðin er:

$$\chi^2 = \sum_{i=1}^l \sum_{j=1}^d \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

þar sem l er fjöldi lína, d er fjöldi dálka, o er mæld tíðni og e er væntanleg tíðni.

Sé núlltilgátan sönn fylgir prófstærðin χ^2 -dreifingu með $(l-1) \cdot (d-1)$ fjölda frígráða. Hafna skal núlltilgátunni sé $\chi^2 > \chi_{1-\alpha, ((l-1) \cdot (d-1))}^2$.

Ályktanir fyrir dreifni

Öryggisbil og tilgátupróf fyrir dreifni þýðis (σ)

Neðra mark $1 - \alpha$ öryggisbils er:

$$\frac{(n-1) \cdot s^2}{\chi_{1-\alpha/2, (n-1)}^2}$$

Efra mark $1 - \alpha$ öryggisbils er:

$$\frac{(n-1) \cdot s^2}{\chi_{\alpha/2, (n-1)}^2}$$

Prófstærðin er:

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

Gagntilgáturnar og höfnunarsvæðin eru:

Gagntilgáta	Hafna H_0 ef:
$H_1 : \sigma^2 < \sigma_0^2$	$\chi^2 < \chi_{\alpha, (n-1)}^2$
$H_1 : \sigma^2 > \sigma_0^2$	$\chi^2 > \chi_{1-\alpha, (n-1)}^2$
$H_1 : \sigma^2 \neq \sigma_0^2$	$\chi^2 < \chi_{\alpha/2, (n-1)}^2$ eða $\chi^2 > \chi_{1-\alpha/2, (n-1)}^2$

Tilgátupróf fyrir dreifni tveggja normaldreifðra þýða (σ_1^2 og σ_2^2)

Mögulegar gagntilgátur, prófstærðir og höfnunarsvæði eru:

Gagntilgáta	Prófstærð	Hafna H_0 ef:
$H_1 : \sigma_1^2 < \sigma_2^2$	$F = \frac{S_2^2}{S_1^2}$	$F > F_{1-\alpha, (n_2-1, n_1-1)}$
$H_1 : \sigma_1^2 > \sigma_2^2$	$F = \frac{S_1^2}{S_2^2}$	$F > F_{1-\alpha, (n_1-1, n_2-1)}$
$H_1 : \sigma_1^2 \neq \sigma_2^2$	$F = \frac{S_M^2}{S_m^2}$	$F > F_{1-\alpha/2, (n_M-1, n_m-1)}$

Í tvíhliða prófinu skal ávalt velja úrtakið með hærri dreifni sem úrtak M og úrtakið með lægri dreifni sem úrtak m .

Ályktanir fyrir meðaltöl

Öryggisbil og tilgátupróf fyrir meðaltal þýðis (μ): σ er þekkt eða n er stórt

Pegar σ er þekkt skal skipta s út fyrir σ hér að neðan.

Neðra mark $1 - \alpha$ öryggisbils:

$$\bar{x} - z_{1-\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

Efra mark $1 - \alpha$ öryggisbils:

$$\bar{x} + z_{1-\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

Prófstærðin er:

$$Z = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

Höfnunarsvæðin eru:

Gagntilgáta	Hafna H_0 ef:
$H_1 : \mu < \mu_0$	$Z < -z_{1-\alpha}$
$H_1 : \mu > \mu_0$	$Z > z_{1-\alpha}$
$H_1 : \mu \neq \mu_0$	$Z < -z_{1-\alpha/2}$ eða $Z > z_{1-\alpha/2}$

Öryggisbil og tilgátupróf fyrir meðaltal þýðis (μ): normaldreift þýði eða n er stórt

Neðra mark $1 - \alpha$ öryggisbils:

$$\bar{x} - t_{1-\alpha/2, (n-1)} \cdot \frac{s}{\sqrt{n}}.$$

Efra mark $1 - \alpha$ öryggisbils:

$$\bar{x} + t_{1-\alpha/2, (n-1)} \cdot \frac{s}{\sqrt{n}}.$$

Prófstærðin er:

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

Höfnunarsvæðin eru:

Gagntilgáta	Hafna H_0 ef:
$H_1 : \mu < \mu_0$	$T < -t_{1-\alpha, (n-1)}$
$H_1 : \mu > \mu_0$	$T > t_{1-\alpha, (n-1)}$
$H_1 : \mu \neq \mu_0$	$T < -t_{1-\alpha/2, (n-1)}$ eða $T > t_{1-\alpha/2, (n-1)}$

Öryggisbil og tilgátupróf fyrir mismun meðaltals tveggja þýða ($\mu_1 - \mu_2$): svipuð dreifni

Neðra mark $1 - \alpha$ öryggisbils er:

$$\bar{x}_1 - \bar{x}_2 - t_{1-\alpha/2, (n_1+n_2-2)} \cdot s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Efra mark $1 - \alpha$ öryggisbils er:

$$\bar{x}_1 - \bar{x}_2 + t_{1-\alpha/2, (n_1+n_2-2)} \cdot s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Prófstærðin er:

$$T = \frac{\bar{X}_1 - \bar{X}_2 - \delta}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Höfnunarsvæðin eru:

Gagntilgáta	Hafna H_0 ef:
$H_1 : \mu_1 - \mu_2 < \delta$	$T < -t_{1-\alpha, (n_1+n_2-2)}$
$H_1 : \mu_1 - \mu_2 > \delta$	$T > t_{1-\alpha, (n_1+n_2-2)}$
$H_1 : \mu_1 - \mu_2 \neq \delta$	$T < -t_{1-\alpha/2, (n_1+n_2-2)}$ eða $T > t_{1-\alpha/2, (n_1+n_2-2)}$

Vegin dreifni, s_p^2 :

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}.$$

Ef mælingarnar í þýðunum tveimur eru jafn margar, þ.e.a.s. ef $n_1 = n_2$ er

$$s_p^2 = \frac{s_1^2 + s_2^2}{2}.$$

Öryggisbil og tilgátupróf fyrir mismun meðaltals tveggja þýða ($\mu_1 - \mu_2$): ólík dreifni

Neðra mark $1 - \alpha$ öryggisbils er:

$$\bar{x}_1 - \bar{x}_2 - t_{1-\alpha/2,(\nu)} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Efra mark $1 - \alpha$ öryggisbils er:

$$\bar{x}_1 - \bar{x}_2 + t_{1-\alpha/2,(\nu)} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Prófstærðin er:

$$T = \frac{\bar{X}_1 - \bar{X}_2 - \delta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Höfnunarsvæðin eru:

Gagntilgáta	Hafna H_0 ef:
$H_1 : \mu < \mu_0$	$T < -t_{1-\alpha,(\nu)}$
$H_1 : \mu > \mu_0$	$T > t_{1-\alpha,(\nu)}$
$H_1 : \mu \neq \mu_0$	$T < -t_{1-\alpha/2,(\nu)}$ eða $T > t_{1-\alpha/2,(\nu)}$

Fjöldi frígráða ν :

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$$

Parað t-próf

Prófstærðin er:

$$T = \frac{\bar{D} - \mu_{D,0}}{S_D/\sqrt{n}}$$

Höfnunarsvæðin eru:

Gagntilgáta	Hafna H_0 ef:
$H_1 : \mu_D < \mu_{D,0}$	$T < -t_{1-\alpha,(n-1)}$
$H_1 : \mu_D > \mu_{D,0}$	$T > t_{1-\alpha,(n-1)}$
$H_1 : \mu_D \neq \mu_{D,0}$	$T < -t_{1-\alpha/2,(n-1)}$ eða $T > t_{1-\alpha/2,(n-1)}$

Fervikagreining

$$SS_T = SS_{Tr} + SS_E.$$

Fervikasummur	Frígráður	Meðalfervikasummur
SS_{Tr}	$a - 1$	$MS_{Tr} = \frac{SS_{Tr}}{a-1}$
SS_E	$N - a$	$MS_E = \frac{SS_E}{N-a}$
SS_T	$N - 1$	

Prófstærðin er

$$F = \frac{SS_{Tr}/(a-1)}{SS_E/(N-a)} = \frac{MS_{Tr}}{MS_E}.$$

Hafna skal H_0 ef $F > F_{1-\alpha,(a-1,N-a)}$

Línuleg aðhvarfsgreining

$$Y = \beta_0 + \beta_1 x + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2)$$

$$b_1 = r \frac{s_y}{s_x}$$

$$b_0 = \bar{y} - b_1 \bar{x}.$$

$$\hat{y} = b_0 + b_1 x$$

Leifar

$$e_i = y_i - \hat{y}_i.$$

Mat á σ^2 í einföldu línulegu aðhvarfi táknum við með s_e^2 og reiknum með

$$s_e^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 2}.$$

Öryggisbil fyrir β_0 og β_1

Neðra mark $1 - \alpha$ öryggisbils fyrir β_0 er:

$$b_0 - t_{1-\alpha/2, (n-2)} \cdot s_e \sqrt{\frac{1}{n} + \frac{(\bar{x})^2}{s_x^2 \cdot (n-1)}}$$

Efra mark $1 - \alpha$ öryggisbils er:

$$b_0 + t_{1-\alpha/2, (n-2)} \cdot s_e \sqrt{\frac{1}{n} + \frac{(\bar{x})^2}{s_x^2 \cdot (n-1)}}$$

Neðra mark $1 - \alpha$ öryggisbils fyrir β_1 er:

$$b_1 - t_{1-\alpha/2, (n-2)} \cdot s_e \frac{1}{\sqrt{s_x^2 \cdot (n-1)}}$$

Efra mark $1 - \alpha$ öryggisbils er:

$$b_1 + t_{1-\alpha/2, (n-2)} \cdot s_e \frac{1}{\sqrt{s_x^2 \cdot (n-1)}}$$

Spábil fyrir framtíðarmælingar

Neðra mark $1 - \alpha$ spábils fyrir framtíðarmælingu á Y :

$$(b_0 + b_1 x_0) - t_{1-\alpha/2, (n-2)} \cdot s_e \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{s_x^2 (n-1)}}$$

Efra mark $1 - \alpha$ spábils er:

$$(b_0 + b_1 x_0) + t_{1-\alpha/2, (n-2)} \cdot s_e \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{s_x^2 (n-1)}}$$

Próf á fylgnistuðli

Prófstærðin er:

$$T = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Höfnunarsvæðin:

Gagntilgáta	Hafna H_0 ef:
$H_1 : \rho < 0$	$T < -t_{1-\alpha, (n-2)}$
$H_1 : \rho > 0$	$T > t_{1-\alpha, (n-2)}$
$H_1 : \rho \neq 0$	$T < -t_{1-\alpha/2, (n-2)}$ eða $T > t_{1-\alpha/2, (n-2)}$

R-skipanir

Lýsistærðir

```
sum() # leggur saman          var() # dreifni
mean() # meðaltal            quantile() # fjordungamork
median() # midgildi         cor() # fylgni
sd() # staðalfravik
```

Ef við viljum reikna út gildin á lýsistærð talnabreytu fyrir hvern flokk flokkabreytu sem tilheyrir sömu gagnatöflu notum við `tapply()` aðferðina. Við mötum `tapply()` með nafni á talnabreytu, nafni á flokkabreytu og nafninu á aðferðinni sem við viljum beita.

Líkindadreifingar

`dbinom()`

`dbinom()` aðferðin reiknar líkur á forminu $P(X = k)$ þar sem X er slembistærð sem fylgir tví-kostadreifingunni með stíkana n og p :
`dbinom(k,n,p)`

`pbinom()`

`pbinom()` aðferðin reiknar líkur á forminu $P(X \leq k)$ þar sem X er slembistærð sem fylgir tví-kostadreifingunni með stíkana n og p :
`pbinom(k,n,p)`

`dpois()`

`dpois()` aðferðin reiknar líkur á forminu $P(X = k)$ þar sem X er slembistærð sem fylgir Poisson dreifingunni með stíkann λ :
`dpois(k,lambda)`

`ppois()`

`ppois()` aðferðin reiknar líkur á forminu $P(X \leq k)$ þar sem X er slembistærð sem fylgir Poisson dreifingunni með stíkann λ :
`ppois(k,lambda)`

`pnorm()`

Við mötum skipunina `pnorm()` á tilteknu viðmiðunargildi x en hún reiknar líkurnar á því að slembistærð sem fylgir normaldreifingu taki gildi minna en gefna viðmiðunargildið. Þ.e.a.s. reiknar $P(X < x)$ þegar X fylgir normaldreifingu:
`pnorm(x, mu, sigma)`

`qnorm()`

Við mötum skipunina `qnorm()` á tilteknum líkum en hún finnur það viðmiðunargildi x sem er þannig að slembistærð sem fylgir normaldreifingu hefur þær tilteknu líkur á að taka gildi sem er minna en viðmiðunargildið. Þ.e.a.s. finnur það x sem er þannig að $P(X < x)$ er jafnt tilteknu líkunum, p :
`qnorm(p, mu, sigma)`

`rnorm()`

`rnorm()` aðferðin býr til (hermir) gildi sem fylgja normaldreifingu. Við mötum aðferðina með hversu mörg gildi við viljum (`n`), meðaltali (`mean`) og staðalfráviki (`sd`) normaldreifingarinnar:
`rnorm(n, mu, sigma)`