

Mat-1.2620 Sovellettu todennäköisyyslaskenta B**8. harjoitukset / Ratkaisut****Aiheet: Otos ja otosjakaumat****Avainsanat:**

Aritmeettinen keskiarvo, Bernoulli-jakauma, Bernoulli-koe, χ^2 -jakauma, Frekvenssi, Frekvenssijakauma, Keskeinen raja-arvolause, Normaalijakauma, Odotusarvo, Otos, Otoshajonta, Otosjakauma, Otoskoko, Otosvarianssi, Riippumattomuus, Standardoitu normaalijakauma, Suhteellinen frekvenssi, Suhteellinen osuus, Todennäköisyys, Yksinkertainen satunnaisotos, Varianssi

Otos ja otosjakaumat**Yksinkertainen satunnaisotos**

Olkoon

$$X_1, X_2, \dots, X_n$$

yksinkertainen satunnaisotos jakaumasta, jonka *pistetodennäköisyys-* tai *tiheysfunktio* on $f(x)$.

Tällöin havainnot $X_1, X_2, \dots, X_n$ ovat *riippumattomia, identtisesti jakautuneita satunnaismuuttujia*, joilla on *sama pistetodennäköisyys-* tai *tiheysfunktio* $f(x)$:

$$\begin{aligned} X_1, X_2, \dots, X_n &\perp \\ X_i &\sim f(x), i = 1, 2, \dots, n \end{aligned}$$

Otostunnusluku

Olkoon

$$X_1, X_2, \dots, X_n$$

yksinkertainen satunnaisotos jakaumasta, jonka *pistetodennäköisyys-* tai *tiheysfunktio* on $f(x)$.

Olkoon

$$T = g(X_1, X_2, \dots, X_n)$$

jokin *satunnaismuuttujien* $X_1, X_2, \dots, X_n$ (mitallinen) *funktio*. *Satunnaismuuttujaa* T kutsutaan (*otos-*) **tunnusluvuksi**.

Oletetaan, että otoksen poimimisen jälkeen satunnaismuuttujat $X_1, X_2, \dots, X_n$ saavat havaituiksi arvoikseen *havaintoarvot* $x_1, x_2, \dots, x_n$:

$$X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$$

Tällöin tunnusluku

$$T = g(X_1, X_2, \dots, X_n)$$

saa havaituksi arvokseen t funktion g arvon pisteessä $(x_1, x_2, \dots, x_n)$:

$$t = g(x_1, x_2, \dots, x_n)$$

Otosjakauma

Oletetaan, että havainnot

$$X_1, X_2, \dots, X_n$$

muodostavat yksinkertaisen satunnaisotoksen jakaumasta $f(x)$ ja olkoon

$$T = g(X_1, X_2, \dots, X_n)$$

jokin otostunnusluku. Koska tunnusluku T on satunnaismuuttuja, sillä on todennäköisyysjakauma, jota kutsutaan tunnusluvun T **otosjakaumaksi**. Tunnusluvun T otosjakauma muodostaa tilastollisen mallin eli todennäköisyysmallin tunnusluvun T arvojen satunnaisvaihtelulle otoksesta toiseen.

Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat

Aritmeettinen keskiarvo ja otosvarianssi

Oletetaan, että havainnot

$$X_1, X_2, \dots, X_n$$

muodostavat yksinkertaisen satunnaisotoksen jakaumasta, jonka odotusarvo on μ ja varianssi on σ^2 . Tällöin havainnot $X_1, X_2, \dots, X_n$ ovat riippumattomia satunnaismuuttujia, joilla kaikilla on sama odotusarvo ja varianssi:

$$X_1, X_2, \dots, X_n \perp$$

$$E(X_i) = \mu, i = 1, 2, \dots, n$$

$$\text{Var}(X_i) = D^2(X_i) = \sigma^2, i = 1, 2, \dots, n$$

Otoksen ominaisuuksia voidaan kuvata havaintoarvojen aritmeettisellä keskiarvolla ja varianssilla.

Määritellään havaintojen $X_1, X_2, \dots, X_n$ aritmeettinen keskiarvo kaavalla

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Määritellään havaintojen $X_1, X_2, \dots, X_n$ otosvarianssi kaavalla

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

Huomaa, että sekä aritmeettinen keskiarvo $\bar{X}$ että otosvarianssi s^2 ovat havaintojen $X_1, X_2, \dots, X_n$ funktioina satunnaismuuttujia, joiden saamat arvot vaihtelevat satunnaisesti otoksesta toiseen.

Aritmeettisen keskiarvon odotusarvo ja varianssi

Havaintojen $X_1, X_2, \dots, X_n$ aritmeettisellä keskiarvolla $\bar{X}$ on em. oletusten pätiessä seuraava odotusarvo ja varianssi:

$$E(\bar{X}) = \mu$$

$$\text{Var}(\bar{X}) = D^2(\bar{X}) = \frac{\sigma^2}{n}$$

Huomaa, että havaintojen $X_1, X_2, \dots, X_n$ aritmeettisen keskiarvon $\bar{X}$ varianssi otoksessa on aina pienempi kuin havaintojen varianssi, jos otoskoko $n > 1$. Lisäksi aritmeettisen keskiarvon varianssi $\bar{X}$ pienenee, jos otoskoon n annetaan kasvaa.

Aritmeettisen keskiarvon $\bar{X}$ standardipoikkeamaa

$$D(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

kutsutaan tavallisesti keskiarvon keskivirheeksi ja se kuvaa aritmeettisen keskiarvon otosvaihtelua oman odotusarvonsa μ ympärillä.

Otosvarianssin odotusarvo

Havaintojen $X_1, X_2, \dots, X_n$ otosvarianssilla s^2 on em. oletusten pätiessä seuraava odotusarvo:

$$E(s^2) = \sigma^2$$

Aritmeettisen keskiarvon otosjakauma

Oletetaan, että havainnot

$$X_1, X_2, \dots, X_n$$

muodostavat yksinkertaisen satunnaisotoksen normaalijakaumasta $N(\mu, \sigma^2)$. Tällöin havainnot $X_1, X_2, \dots, X_n$ ovat riippumattomia satunnaismuuttujia, jotka noudattavat samaa normaalijakaumaa $N(\mu, \sigma^2)$:

$$X_1, X_2, \dots, X_n \perp$$

$$X_i \sim N(\mu, \sigma^2), i = 1, 2, \dots, n$$

Havaintojen $X_1, X_2, \dots, X_n$ aritmeettinen keskiarvo $\bar{X}$ noudattaa em. oletusten pätiessä *normaalijakaumaa* parametrein μ ja σ^2/n :

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

Erityisesti

$$E(\bar{X}) = \mu$$

$$\text{Var}(\bar{X}) = D^2(\bar{X}) = \frac{\sigma^2}{n}$$

mikä pätee myös ilman normaalisuusoletusta.

Aritmeettisen keskiarvon approksimatiivinen otosjakauma

Oletetaan, että havainnot

$$X_1, X_2, \dots, X_n$$

muodostavat yksinkertaisen satunnaisotoksen jakaumasta, jonka odotusarvo on μ ja varianssi on σ^2 .

Tällöin *keskeisestä raja-arvolauseesta* seuraa, että havaintojen aritmeettinen keskiarvo $\bar{X}$ noudattaa suurissa otoksissa *approksimatiivisesti (asymptoottisesti) normaalijakaumaa* parametrein μ ja σ^2/n :

$$\bar{X} \sim_a N\left(\mu, \frac{\sigma^2}{n}\right)$$

Otosvarianssin otosjakauma

Oletetaan, että havainnot

$$X_1, X_2, \dots, X_n$$

muodostavat yksinkertaisen satunnaisotoksen normaalijakaumasta $N(\mu, \sigma^2)$. Tällöin havainnot $X_1, X_2, \dots, X_n$ ovat riippumattomia satunnaismuuttujia, jotka noudattavat samaa normaali-jakaumaa $N(\mu, \sigma^2)$:

$$X_1, X_2, \dots, X_n \perp$$

$$X_i \sim N(\mu, \sigma^2), i = 1, 2, \dots, n$$

Olko s^2 havaintojen $X_1, X_2, \dots, X_n$ otosvarianssi. Satunnaismuuttuja $(n-1)s^2/\sigma^2$ noudattaa em. oletusten pätiessä χ^2 -jakaumaa vapausastein $(n-1)$:

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

Lisäksi voidaan osoittaa, että aritmeettinen keskiarvo $\bar{X}$ ja otosvarianssi s^2 ovat satunnaismuuttujina riippumattomia:

$$\bar{X} \perp s^2$$

Siten suoraan Studentin t -jakauman määritelmän mukaan

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t(n-1)$$

em. oletusten pätiessä.

Suhteellisen frekvenssin otosjakauma

Frekvenssi ja suhteellinen frekvenssi

Olkoon $A \subset S$ jokin otosavaruuden S tapahtuma ja olkoon

$$p = \Pr(A)$$

$$q = 1 - \Pr(A) = 1 - p$$

Poimitaan otosavaruudesta S yksinkertainen satunnaisotos, jonka koko on n . Olkoon f A -tyyppisten alkioiden *frekvenssi* eli lukumäärä otoksessa ja

$$\hat{p} = \frac{f}{n}$$

vastaava *suhteellinen frekvenssi* eli *osuus*.

Huomaa, että sekä frekvenssi f että suhteellinen frekvenssi $\hat{p} = f/n$ ovat *satunnaismuuttujia*, joiden saamat arvot vaihtelevat satunnaisesti otoksesta toiseen.

Frekvenssin odotusarvo, varianssi ja otosjakauma

Frekvenssin f odotusarvo ja varianssi:

$$E(f) = np$$

$$\text{Var}(f) = D^2(f) = npq$$

jossa $q = 1 - p$.

Frekvenssi f noudattaa otoksessa *binomijakaumaa* parametrein n ja $\Pr(A) = p$:

$$f \sim \text{Bin}(n, p)$$

Suhteellisen frekvenssin odotusarvo ja varianssi

Suhteellisen frekvenssin $\hat{p} = f/n$ odotusarvo ja varianssi:

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

$$\text{Var}(\hat{p}) = D^2(\hat{p}) = \frac{pq}{n}$$

jossa $q = 1 - p$. Huomaa, että suhteellisen frekvenssin $\hat{p}$ varianssi *pienenee*, jos otoskoon n annetaan kasvaa.

Suhteellisen frekvenssin $\hat{p} = f/n$ *standardipoikkeamaa*

$$D(\hat{p}) = \sqrt{\frac{pq}{n}}$$

kutsutaan tavallisesti *suhteellisen frekvenssin keskivirheeksi* ja se kuvaa suhteellisen frekvenssin otosvaihtelua oman odotusarvonsa p ympärillä.

Suhteellisen frekvenssin otosjakauma

Keskeisestä raja-arvolauseesta seuraa, että suhteellinen frekvenssi $\hat{p}$ otoksessa noudattaa em. oletusten pätiessä suurissa otoksissa *approksimatiivisesti normaalijakaumaa*:

$$\hat{p} \sim_a N\left(p, \frac{pq}{n}\right)$$

Tehtävä 8.1.

- (a) Kone valmistaa kuulalaakerin kuulia, joiden halkaisijat vaihtelevat satunnaisesti noudattaen normaalijakaumaa parametrein

$$\mu = 10 \text{ mm}, \sigma = 0.01 \text{ mm}$$

Poimitaan kuulien joukosta yksinkertainen satunnaisotos, jonka koko $n = 10$.

Olkoot $\bar{X}$ ja s^2 kuulien halkaisijoiden aritmeettinen keskiarvo ja otosvarianssi otoksessa.

Mitkä ovat aritmeettisen keskiarvon $\bar{X}$ ja otosvarianssin s^2 muunnoksen $(n-1)s^2/\sigma^2$ jakaumat otoksessa?

- (b) Äänestäjistä 25 % kannattaa puoluetta ABC. Poimitaan äänestäjien joukosta yksinkertainen satunnaisotos, jonka koko $n = 1000$.

Mikä on puolueen ABC kannattajien suhteellisen osuuden f/n approksimatiivinen jakauma otoksessa?

Tehtävä 8.1. – Mitä opimme?

Tehtävän (a)-kohdassa tarkastellaan *aritmeettisen keskiarvon ja otosvarianssin otosjakaumia*.

Tehtävän (b)-kohdassa tarkastellaan *suhteellisen osuuden (approksimatiivista) otosjakaumaa*.

Tehtävä 8.1. – Ratkaisu:

- (a) Oletuksen mukaan havainnot $X_1, X_2, \dots, X_n$ muodostavat *yksinkertaisen satunnaisotoksen* normaalijakaumasta $N(\mu, \sigma^2)$, jossa

$$n = 10$$

$$\mu = 10 \text{ mm}$$

$$\sigma^2 = 0.01^2 \text{ mm} = 0.0001 \text{ mm}^2$$

Siten kuulien halkaisijoiden aritmeettinen keskiarvo $\bar{X}$ noudattaa otoksessa *normaalijakaumaa* $N(\mu, \sigma^2/n)$, jossa

$$\mu = E(\bar{X}) = 10 \text{ mm}$$

$$\frac{\sigma^2}{n} = \text{Var}(\bar{X}) = D^2(\bar{X}) = \frac{0.0001}{10} = 0.00001 \text{ mm}^2$$

Olkoon s^2 kuulien halkaisijoiden varianssi otoksessa. Tällöin satunnaismuuttuja $(n-1)s^2/\sigma^2$ noudattaa otoksessa χ^2 -jakaumaa vapausastein

$$n-1 = 10-1 = 9$$

(b) Olkoon

A = satunnaisesti valittu äänestäjä kannattaa puoluetta ABC

Oletuksen mukaan

$$\Pr(A) = p = 0.25$$

Poimitaan äänestäjien joukosta *yksinkertainen satunnaisotos*, jonka koko on $n = 1000$.

Puoluetta ABC kannattavien äänestäjien suhteellinen frekvenssi $\hat{p} = f/n$ otoksessa noudattaa *suurissa otoksissa approksimatiivisesti normaalijakaumaa*:

$$\hat{p} \sim_a N\left(p, \frac{pq}{n}\right)$$

jossa siis

$$p = \Pr(A) = 0.25$$

$$q = \Pr(A^c) = 1 - \Pr(A) = 1 - p = 0.75$$

Siten puolueen ABC kannattajien suhteellinen frekvenssi $\hat{p} = f/n$ otoksessa noudattaa *suurissa otoksissa approksimatiivisesti normaalijakaumaa* parametrein

$$E(\hat{p}) = p = 0.25$$

$$\text{Var}(\hat{p}) = D^2(\hat{p}) = \frac{pq}{n} = \frac{0.25 \times 0.75}{1000} = \frac{0.1875}{1000} = 0.0001875$$

Tehtävä 8.2.

- (a) Miesten pituus eräässä maassa vaihtelee satunnaisesti noudattaen normaalijakaumaa parametrein

$$\mu = 180 \text{ cm}, \sigma = 5 \text{ cm}$$

Poimitaan miesten joukosta yksinkertainen satunnaisotos, jonka koko $n = 100$.

Olkoot $\bar{X}$ ja s^2 pituuksien aritmeettinen keskiarvo ja otosvarianssi otoksessa.

Mitkä ovat aritmeettisen keskiarvon $\bar{X}$ ja otosvarianssin s^2 muunnoksen $(n-1)s^2/\sigma^2$ jakaumat otoksessa?

- (b) Koneen valmistamista muttereista 5 % on viallisia. Poimitaan muttereiden joukosta yksinkertainen satunnaisotos, jonka koko $n = 100$.

Mikä on viallisten muttereiden suhteellisen osuuden f/n approksimatiivinen jakauma otoksessa?

Tehtävä 8.2. – Mitä opimme?

Tehtävän (a)-kohdassa tarkastellaan *aritmeettisen keskiarvon ja otosvarianssin otosjakaumia*.

Tehtävän (b)-kohdassa tarkastellaan *suhteellisen osuuden (approksimatiivista) otosjakaumaa*.

Tehtävä 8.2. – Ratkaisu:

- (a) Oletuksen mukaan havainnot $X_1, X_2, \dots, X_n$ muodostavat *yksinkertaisen satunnaisotoksen* normaalijakaumasta $N(\mu, \sigma^2)$, jossa

$$n = 100$$

$$\mu = 185 \text{ cm}$$

$$\sigma^2 = 5^2 \text{ cm} = 25 \text{ cm}^2$$

Siten miesten pituuksien aritmeettinen keskiarvo $\bar{X}$ noudattaa otoksessa *normaalijakaumaa* $N(\mu, \sigma^2/n)$, jossa

$$\mu = E(\bar{X}) = 185 \text{ cm}$$

$$\frac{\sigma^2}{n} = \text{Var}(\bar{X}) = D^2(\bar{X}) = \frac{25}{100} = 0.25 \text{ cm}^2$$

Olkoon s^2 miesten pituuksien varianssi otoksessa. Tällöin satunnaismuuttuja $(n-1)s^2/\sigma^2$ noudattaa otoksessa χ^2 -jakaumaa vapausastein

$$n-1 = 100-1 = 99$$

(b) Olkoon

A = satunnaisesti valittu mutteri on viallinen

Oletuksen mukaan

$$\Pr(A) = p = 0.05$$

Poimitaan muttereiden joukosta *yksinkertainen satunnaisotos*, jonka koko on $n = 100$.

Viallisten muttereiden suhteellinen frekvenssi $\hat{p} = f / n$ otoksessa noudattaa *suurissa otoksissa approksimatiivisesti normaalijakaumaa*:

$$\hat{p} \sim_a N\left(p, \frac{pq}{n}\right)$$

jossa siis

$$p = \Pr(A) = 0.05$$

$$q = \Pr(A^c) = 1 - \Pr(A) = 1 - p = 0.95$$

Siten viallisten muttereiden suhteellinen frekvenssi $\hat{p} = f / n$ otoksessa noudattaa *suurissa otoksissa approksimatiivisesti normaalijakaumaa* parametrein

$$E(\hat{p}) = p = 0.05$$

$$\text{Var}(\hat{p}) = D^2(\hat{p}) = \frac{pq}{n} = \frac{0.05 \times 0.95}{100} = \frac{0.0475}{100} = 0.000475$$

kasvaa.

Huomautuksia tehtäviin 8.1. ja 8.2.:

- (1) Tehtävien 8.1. ja 8.2. ideana on kertoa siitä, *millaisia ovat tavanomaisten havainnoista laskettavien otostunnuslukujen jakaumat perusjoukossa, jos havaintojen jakauma perusjoukossa tunnetaan.*
- (2) Otostunnuslukujen jakaumia koskevat tulokset ovat kuitenkin *epäoperationaalisia*, koska *jakaumien parametreja ei yleensä tunneta.*
- (3) Jos havaintojen jakauman parametreja ei tunneta, ne voidaan pyrkiä *estimoimaan* eli arvioimaan otoksesta saatujen tietojen perusteella; ks. lukua **Tilastollisten mallien parametrien estimointi**.
- (4) Perusjoukon parametrien arvoista tehtyjä oletuksia voidaan pyrkiä *testaamaan tilastollisesti* otoksesta saatujen tietojen perusteella; ks. lukua **Tilastollisten hypoteesien testaus**.
- (5) Myös perusjoukon jakauman tyyppiä koskevia oletuksia voidaan pyrkiä *testaamaan tilastollisesti* otoksesta saatujen tietojen perusteella; ks. lukua **Yhteensopivuuden, homogeneisuuden ja riippumattomuuden testaaminen**.

Tehtävä 8.3.

Olkoot

$$X_i, i = 1, 2, \dots, n$$

riippumattomia normaalijakautuneita satunnaismuuttujia, joiden odotusarvo $E(X_i) = \mu$ ja varianssi $\text{Var}(X_i) = \sigma^2$.

Tarkastellaan seuraavia todennäköisyyksiä:

- (1) $\Pr(X_i > \mu + \sigma)$
- (2) $\Pr(X_1 + X_2 + \dots + X_n > n(\mu + \sigma))$
- (3) $\Pr(\bar{X} > \mu + \sigma)$

Tehtävät:

- (a) Määrittää todennäköisyys (1).
- (b) Todista, että todennäköisyys (2) on pienempi kuin todennäköisyys (1), jos $n > 1$.
- (c) Todista, että todennäköisyys (2) pienenee, kun $n \rightarrow +\infty$.
- (d) Todista, että todennäköisyys (3) on sama kuin todennäköisyys (2).
- (e) Määrittää todennäköisyys (2), kun $n = 10$.

Tehtävä 8.3. – Mitä opimme?

Tehtävässä tarkastellaan riippumattomia satunnaismuuttujia, jotka noudattavat samaa normaalijakaumaa ja vertaillaan yksittäisen muuttujan, muuttujien summan ja muuttujien aritmeettisen keskiarvon jakaumia.

Tehtävä 8.3. – Ratkaisu:

Oletuksen mukaan

$$X_1, X_2, \dots, X_n \perp$$

$$X_i \sim N(\mu, \sigma^2), i = 1, 2, \dots, n$$

- (a) Helposti nähdään, että

$$\Pr(X_i > \mu + \sigma) = \Pr\left(\frac{X_i - \mu}{\sigma} > 1\right) = \Pr(Z > 1)$$

jossa *standardoitu satunnaismuuttuja*

$$Z = \frac{X_i - \mu}{\sigma}$$

noudattaa *standardoitua normaalijakaumaa*:

$$Z \sim N(0, 1)$$

Normaalijakauman taulukoiden mukaan

$$\Pr(Z \leq 1) = 0.8413$$

joten komplementtitapahtuman todennäköisyyden kaavan mukaan kysytty todennäköisyys on

$$\Pr(Z > 1) = 1 - \Pr(Z \leq 1) = 0.1587$$

(b)&(c)

Olkoon

$$Y = \sum_{i=1}^n X_i$$

Tällöin

$$E(Y) = n\mu$$

Koska satunnaismuuttujat X_i , $i = 1, 2, \dots, n$ on lisäksi oletettu *riippumattomiksi*, niin

$$\text{Var}(Y) = n\sigma^2$$

Siten kaikille $n > 1$ pätee

$$\Pr(Y > n(\mu + \sigma)) = \Pr\left(\frac{Y - n\mu}{\sigma\sqrt{n}} > \sqrt{n}\right) = \Pr(Z > \sqrt{n}) < \Pr(Z > 1)$$

jossa *standardoitu satunnaismuuttuja*

$$Z = \frac{X_i - \mu}{\sigma\sqrt{n}}$$

noudattaa *standardoitua normaalijakaumaa*:

$$Z \sim N(0,1)$$

Todennäköisyydet

$$\Pr(Z > \sqrt{n})$$

muodostavat *aidosti vähenevän lukujonon*, jos

$$n \rightarrow +\infty$$

(d) Tulos on triviaalisti sama kuin kohdassa (c), koska

$$\Pr(\bar{X} > \mu + \sigma) = \Pr(Y > n(\mu + \sigma))$$

(e) Jos $n = 10$, niin (c)-kohdasta seuraa, että

$$\begin{aligned}\Pr(X_1 + X_2 + \dots + X_n > n(\mu + \sigma)) \\&= \Pr(X_1 + X_2 + \dots + X_{10} > 10(\mu + \sigma)) \\&= \Pr(Z > \sqrt{10}) = \Pr(Z > 3.16)\end{aligned}$$

jossa *standardoitu satunnaismuuttuja*

$$Z = \frac{X_i - n\mu}{\sigma\sqrt{n}} = \frac{X_i - 10\mu}{\sigma\sqrt{10}}$$

noudattaa *standardoitua normaalijakaumaa*:

$$Z \sim N(0,1)$$

Normaalijakauman taulukoiden mukaan

$$\Pr(Z \leq 3.16) = 0.9992$$

joten komplementtitapahtuman todennäköisyyden kaavan mukaan kysytty todennäköisyys on

$$\Pr(Z > 3.16) = 1 - \Pr(Z \leq 3.16) = 0.0008$$

Tehtävä 8.4.

Oletetaan, että havainnot $X_i, i = 1, 2, \dots, 100$ muodostavat yksinkertaisen satunnaisotoksen normaalijakaumasta $N(1,4)$. Määrittää todennäköisyys, että havaintojen aritmeettinen keskiarvo $\bar{X}$ saa suurempia arvoja kuin 1.1.

Tehtävä 8.4. – Mitä opimme?

Tehtävässä tarkastellaan *aritmeettisen keskiarvon otosjakaumaa*.

Tehtävä 8.4. – Ratkaisu:

Oletetaan, että havainnot

$$X_i, i = 1, 2, \dots, 100$$

muodostavat yksinkertaisen satunnaisotoksen normaalijakaumasta $N(\mu, \sigma^2)$, jossa

$$\mu = 1$$

$$\sigma^2 = 4$$

Olkoon havaintojen $X_i, i = 1, 2, \dots, 100$ aritmeettinen keskiarvo

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{100} \sum_{i=1}^{100} X_i$$

Oletuksista seuraa, että satunnaismuuttuja $\bar{X}$ noudattaa normaalijakaumaa parametrein μ ja σ^2/n :

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

jossa siis

$$\mu = 1$$

$$\frac{\sigma^2}{n} = \frac{4}{100} = \frac{1}{25} = 0.04$$

Tehtävänä on määrätä todennäköisyys

$$\Pr(\bar{X} > 1.1)$$

Selvästi

$$\begin{aligned}\Pr(\bar{X} > 1.1) &= \Pr\left(\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} > \frac{1.1 - \mu}{\sigma / \sqrt{n}}\right) \\ &= \Pr\left(Z > \frac{1.1 - 1}{2 / \sqrt{100}}\right) \\ &= \Pr(Z > 0.5)\end{aligned}$$

jossa *standardoitu satunnaismuuttuja*

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

noudattaa *standardoitua normaalijakaumaa*:

$$Z \sim N(0,1)$$

Normaalijakauman taulukoiden mukaan

$$\Pr(Z \leq 0.5) = 0.6915$$

joten komplementtitapahtuman todennäköisyyden kaavan mukaan kysytty todennäköisyys on

$$\begin{aligned}\Pr(\bar{X} > 1.1) &= \Pr(Z > 0.5) \\ &= 1 - \Pr(Z \leq 0.5) \\ &= 1 - 0.6915 \\ &= 0.3085\end{aligned}$$

Tehtävä 8.5.

Oletetaan, että suomalaisten miesten pituus on normaalijakautunut parametrein $\mu = 175$ cm ja $\sigma = 5$ cm. Poimitaan miesten joukosta yksinkertainen satunnaisotos, jonka koko on 100. Määrittää lukuarvo, jota suurempia arvoja havaintojen aritmeettinen keskiarvo saa todennäköisyydellä 0.01.

Tehtävä 8.5. – Mitä opimme?

Tehtävässä tarkastellaan *aritmeettisen keskiarvon otosjakaumaa*.

Tehtävä 8.5. – Ratkaisu:

Oletetaan, että havainnot

$$X_i, i = 1, 2, \dots, 100$$

muodostavat yksinkertaisen satunnaisotoksen normaalijakaumasta $N(\mu, \sigma^2)$, jossa

$$\mu = 175$$

$$\sigma^2 = 25$$

Olkoon havaintojen $X_i, i = 1, 2, \dots, 100$ *aritmeettinen keskiarvo*

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{100} \sum_{i=1}^{100} X_i$$

Oletuksista seuraa, että satunnaismuuttuja $\bar{X}$ *noudattaa normaalijakaumaa* parametrein μ ja σ^2/n :

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

jossa siis

$$\mu = 175$$

$$\frac{\sigma^2}{n} = \frac{25}{100} = \frac{1}{4} = 0.25$$

Tehtävänä on määrätä lukuarvo, jota *suurempia* arvoja havaintojen aritmeettinen keskiarvo $\bar{X}$ saa todennäköisyydellä 0.01.

Koska

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

niin *standardoitu satunnaismuuttuja*

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

noudattaa *standardoitua normaalijakaumaa*:

$$Z \sim N(0,1)$$

Normaalijakauman taulukoista näemme, että

$$\Pr(Z \leq 2.33) = 0.9901 \approx 0.99$$

Komplementtitapahtuman todennäköisyyden kaavan mukaan

$$\Pr(Z > 2.33) = 1 - \Pr(Z \leq 2.33) \approx 1 - 0.99 = 0.01$$

Saamme siten epäyhtälön

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} > 2.33$$

josta aritmeettiselle keskiarvolle saadaan ehto

$$\bar{X} > \mu + \frac{\sigma}{\sqrt{n}} \times 2.33 = 175 + \frac{5}{\sqrt{100}} \times 2.33 = 176.165$$

Siten

$$\Pr(\bar{X} \geq 176.165) = 0.01$$

Tehtävä 8.6.

Oletetaan, että havainnot $X_i, i = 1, 2, \dots, 101$ muodostavat yksinkertaisen satunnaisotoksen normaalijakaumasta $N(1,4)$. Määrittää lukuarvo, jota pienempiä arvoja havaintojen otosvarianssi saa todennäköisyydellä 0.01.

Tehtävä 8.6. – Mitä opimme?

Tehtävässä tarkastellaan *otosvarianssin otosjakaumaa*.

Tehtävä 8.6. – Ratkaisu:

Oletetaan, että havainnot

$$X_i, i = 1, 2, \dots, 101$$

muodostavat yksinkertaisen satunnaisotoksen normaalijakaumasta $N(\mu, \sigma^2)$, jossa

$$\mu = 1$$

$$\sigma^2 = 4$$

Olkoon havaintojen $X_i, i = 1, 2, \dots, 101$ *aritmeettinen keskiarvo*

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{101} \sum_{i=1}^{101} X_i$$

ja *otosvarianssi*

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{100} \sum_{i=1}^{101} (X_i - \bar{X})^2$$

Oletuksista seuraa, että satunnaismuuttuja

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

jossa

$$\sigma^2 = 4$$

$$n = 101$$

noudattaa χ^2 -jakaumaa vapausastein $(n-1)$:

$$V \sim \chi^2(100)$$

Tehtävänä on määrätä lukuarvo, joka erottaa χ^2 -jakauman *vasemmalle hänälle* todennäköisyysmassan, jonka koko on 0.01.

χ^2 -jakauman taulukoista nähdään suoraan, että

$$\Pr(V \leq 70.065) = 0.01$$

kun

$$V \sim \chi^2(100)$$

Koska

$$V = \frac{(n-1)s^2}{\sigma^2} = \frac{100s^2}{4} = 25s^2$$

saamme epäyhtälön

$$25s^2 \leq 70.065$$

josta otosvarianssille s^2 saadaan ehto

$$s^2 \leq 2.803$$

Siten

$$\Pr(s^2 \leq 2.803) = 0.01$$

Tehtävä 8.7.

Oletetaan, että suomalaisten miesten pituus on normaalijakautunut parametrein $\mu = 175$ cm ja $\sigma = 5$ cm. Poimitaan miesten joukosta yksinkertainen satunnaisotos, jonka koko on 101. Määrittää lukuarvo, jota suurempia arvoja otosvarianssi saa todennäköisyydellä 0.01.

Tehtävä 8.7. – Mitä opimme?

Tehtävässä tarkastellaan *otosvarianssin otosjakaumaa*.

Tehtävä 8.7. – Ratkaisu:

Oletetaan, että havainnot

$$X_i, i = 1, 2, \dots, 101$$

muodostavat *yksinkertaisen satunnaisotoksen* normaalijakaumasta $N(\mu, \sigma^2)$, jossa

$$\mu = 175$$

$$\sigma^2 = 25$$

Olkoon havaintojen $X_i, i = 1, 2, \dots, 101$ *aritmeettinen keskiarvo*

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{101} \sum_{i=1}^{101} X_i$$

ja *otosvarianssi*

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{100} \sum_{i=1}^{101} (X_i - \bar{X})^2$$

Oletuksista seuraa, että satunnaismuuttuja

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

jossa

$$\sigma^2 = 25$$

$$n = 101$$

noudattaa χ^2 -jakaumaa vapausastein $(n-1)$:

$$V \sim \chi^2(100)$$

Tehtävänä on määrätä lukuarvo, joka erottaa χ^2 -jakauman *oikealle hännälle* todennäköisyysmassan, jonka koko on 0.01:

χ^2 -jakauman taulukoista nähdään suoraan, että

$$\Pr(V \geq 135.807) = 0.01$$

kun

$$V \sim \chi^2(100)$$

Koska

$$V = \frac{(n-1)s^2}{\sigma^2} = \frac{100s^2}{25} = 4s^2$$

saamme epäyhtälön

$$4s^2 \geq 135.807$$

josta otosvarianssille s^2 saadaan ehto

$$s^2 \geq 33.9518$$

Siten

$$\Pr(s^2 \geq 33.9518) = 0.01$$

Tehtävä 8.8.

Oletetaan, että teemme 100 toisistaan riippumatonta Bernoulli-koetta, jossa kiinnostuksen kohteena olevan tapahtuman A todennäköisyys on 0.2. Määrittää todennäköisyys, että tapahtuman A suhteellinen frekvenssi toistojen joukossa on suurempi kuin 10.

Tehtävä 8.8. – Mitä opimme?

Tehtävässä tarkastellaan *suhteellisen frekvenssin (approksimatiivista) otosjakaumaa*.

Tehtävä 8.8. – Ratkaisu:

Olkoon

f = tapahtuman A *frekvenssi* toistojen joukossa

$\hat{p} = \frac{f}{n}$ = tapahtumien A *suhteellinen frekvenssi* toistojen joukossa

n = toistojen lukumäärä

Koska toistojen lukumäärä $n = 100$ on melko suuri, voimme melko hyvin approksimoida suhteellisen frekvenssin $\hat{p}$ otantajakaumaa normaalijakaumalla:

$$\hat{p} \sim_a N\left(p, \frac{pq}{n}\right)$$

jossa

$$p = 0.2$$

$$q = 1 - p = 0.8$$

$$n = 100$$

Siten *standardoitu satunnaismuuttuja*

$$Z = \frac{\hat{p} - p}{\sqrt{pq/n}}$$

noudattaa *approksimatiivisesti standardoitua normaalijakaumaa*:

$$Z \sim_a N(0,1)$$

Koska

$$\frac{10}{100} = \frac{1}{10} = 0.1$$

tehtävänä on määrätä todennäköisyys

$$\Pr(\hat{p} > 0.1)$$

Selvästi

$$\begin{aligned}\Pr(\hat{p} > 0.1) &= \Pr\left(\frac{\hat{p} - p}{\sqrt{pq/n}} > \frac{0.1 - p}{\sqrt{pq/n}}\right) \\ &= \Pr\left(Z > \frac{0.1 - 0.2}{\sqrt{0.2 \times 0.8/100}}\right) \\ &= \Pr(Z > -2.5)\end{aligned}$$

jossa *standardoitu satunnaismuuttuja*

$$Z = \frac{\hat{p} - p}{\sqrt{pq/n}}$$

noudattaa *approksimatiivisesti standardoitua normaalijakaumaa*:

$$Z \sim_a N(0,1)$$

Normaalijakauman taulukoiden mukaan

$$\Pr(Z \leq -2.5) = 0.0062$$

joten komplementtitapahtuman todennäköisyyden kaavan mukaan kysytty todennäköisyys on

$$\begin{aligned}\Pr(\hat{p} > 0.1) &= \Pr(Z > -2.5) \\ &= 1 - \Pr(Z \leq -2.5) \\ &= 1 - 0.0062 \\ &= 0.9938\end{aligned}$$

Tehtävä 8.9.

Oletetaan, että 30 % suomalaisista kannattaa NATO:on liittymistä. Poimitaan suomalaisten joukosta yksinkertainen satunnaisotos, jonka koko on 100. Määrittää todennäköisyys, että NATO:n kannattajien suhteellinen osuus otoksessa on pienempi kuin 20 %.

Tehtävä 8.9. – Mitä opimme?

Tehtävässä tarkastellaan *suhteellisen frekvenssin* (approksimatiivista) *otosjakaumaa*.

Tehtävä 8.9. – Ratkaisu:

Olkoon

f = NATO:n kannattajien *frekvenssi* otoksessa

$\hat{p} = \frac{f}{n}$ = NATO:n kannattajien *suhteellinen frekvenssi* otoksessa

n = otoskoko

Koska otoskoko $n = 100$ on melko suuri, voimme melko hyvin approksimoida suhteellisen frekvenssin $\hat{p}$ otosjakaumaa normaalijakaumalla:

$$\hat{p} \sim_a N\left(p, \frac{pq}{n}\right)$$

jossa

$$p = 0.3$$

$$q = 1 - p = 0.7$$

$$n = 100$$

Siten *standardoitu satunnaismuuttuja*

$$Z = \frac{\hat{p} - p}{\sqrt{pq/n}}$$

noudattaa *approksimatiivisesti standardoitua normaalijakaumaa*:

$$Z \sim_a N(0,1)$$

Tehtävänä on määrätä todennäköisyys

$$\Pr(\hat{p} < 0.20)$$

Selvästi

$$\begin{aligned}\Pr(\hat{p} < 0.20) &= \Pr\left(\frac{\hat{p} - p}{\sqrt{pq/n}} < \frac{0.20 - p}{\sqrt{pq/n}}\right) \\ &= \Pr\left(Z < \frac{0.2 - 0.3}{\sqrt{0.3 \times 0.7 / 100}}\right) \\ &= \Pr(Z < -2.18)\end{aligned}$$

jossa *standardoitu satunnaismuuttuja*

$$Z = \frac{\hat{p} - p}{\sqrt{pq/n}}$$

noudattaa *approksimatiivisesti standardoitua normaalijakaumaa*:

$$Z \sim_a N(0,1)$$

Normaalijakauman taulukoiden mukaan kysytty todennäköisyys on

$$\Pr(Z \leq -2.18) = 0.0146$$