

---

Johdatus tilastotieteeseen  
**Otos ja otosjakaumat**

# Otos ja otosjakaumat

---

**Yksinkertainen satunnaisotos**

**Otostunnusluvut ja otosjakaumat**

**Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat**

**Suhteellisen frekvenssin otosjakauma**

# Otos ja otosjakaumat:

## Mitä opimme? – 1/3

---

- **Tilastollinen aineisto** koostuu tutkimuksen kohteita kuvaavien muuttujien *havaituista arvoista*.
- Tilastollisissa tutkimusasetelmissä ajatellaan, että jokin **satunnais-ilmiö** *on generoinut havaintoarvot*.
- Siten tilastollisissa tutkimusasetelmissä havaintoarvoihin liittyy aina *epävarmuutta ja satunnaisuutta*.
- Havaintoaineiston **tilastollisella mallilla** tarkoitetaan *aineiston generoiden satunnaismuuttujien todennäköisyysjakaumaa*.
- Yksinkertaisissa tilastollisissa tutkimusasetelmissä havaintoaineiston keräämisessä käytetään **yksinkertaista satunnaisotantaa**.

# Otos ja otosjakaumat:

## Mitä opimme? – 2/3

---

- Koska tilastollisen aineiston havaintoarvot ovat jonkin satunnais-  
ilmion generoimia, myös *kaikki* havainnoista laskettavat suureet  
ovat satunnaisia.
- Tämä merkitsee sitä, että esimerkiksi *havaintoaineistoa kuvaavat*  
*tunnusluvut vaihtelevat satunnaisesti otoksesta toiseen.*
- Tunnusluvun *otosvaihtelua* eli tunnusluvun satunnaista vaihtelua  
otoksesta toiseen voidaan kuvata tunnusluvun **otosjakaumalla**.
- *Otosjakaumien teoria* muodostaa teoreettisen perustan sekä toden-  
näköisyysjakaumien *parametrien estimaattoreiden* ominaisuuksia että  
parametreja koskevien hypoteesien testauksessa käytettävien *testi-*  
*suureiden* ominaisuuksia koskevalle tilastolliselle tutkimukselle.

# Otos ja otosjakaumat:

## Mitä opimme? – 3/3

---

- Esimerkkeinä tarkastellaan seuraavien otostunnuslukujen otosjakaumia:
  - **Aritmeettinen keskiarvo**
  - **Otosvariassi**
  - **Suhteellinen frekvenssi**
- Tarkastelun kohteena ovat sekä ym. otostunnuslukujen **odotusarvot** ja **variassit** että niiden **otosjakaumat**.

# Otos ja otosjakaumat: Esitiedot

---

- Esitiedot: ks. seuraavia lukuja:
    - Tilastollisten aineistojen kerääminen ja mittaaminen**
    - Tilastollisten aineistojen kuvaaminen**
  - Tarvitset esitietoja myös seuraavista kalvokokoomelman **Johdatus todennäköisyyslaskentaan** luvuista:
    - Satunnaismuuttujat ja todennäköisyysjakaumat**
    - Jakaumien tunnusluvut**
    - Diskreettejä jakaumia**
    - Jatkuvia jakaumia**
    - Normaalijakaumasta johdettuja jakaumia**
    - Moniulotteiset satunnaismuuttujat ja todennäköisyysjakaumat**
- Ks. myös lukuja:
- Satunnaismuuttujien muunnokset ja niiden jakaumat**
  - Konvergenssikäsitteet ja raja-arvolauseet**

# Otos ja otosjakaumat: Lisätiedot

---

- Todennäköisyysjakaumien parametrien *estimointia* käsitellään luvuissa
  - Estimointi**
  - Estimointimenetelmät**
  - Väliestimointi**
- Todennäköisyysjakaumien parametreja koskevien tilastollisten hypoteesien *testaamista* käsitellään yleisesti luvussa
  - Tilastolliset testit**

# Otos ja otosjakaumat

---

>> Yksinkertainen satunnaisotos

Otostunnusluvut ja otosjakaumat

Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat

Suhteellisen frekvenssin otosjakauma



# Yksinkertainen satunnaisotos

---

## *Avainsanat*

Havainto

Havaintoarvo

Tilastollinen aineisto

Tilastollinen malli

Todennäköisyysjakauma

Yksinkertainen satunnaisotos

## Yksinkertainen satunnaisotos

# Tilastollinen aineisto

---

- **Tilastollinen aineisto** koostuu tutkimuksen kohteita kuvaavien muuttujien *havaituista arvoista*.
- Tilastollisissa tutkimusasetelmissä havaintoarvoihin liittyy aina *epävarmuutta* ja *satunnaisuutta*.
- Seurauksia:
  - (i) Tilastollisissa tutkimusasetelmissä ajatellaan, että *havaintoarvot on generoinut ilmiö, joka on luonteeltaan satunnainen*.
  - (ii) Tilastollisen tutkimuksen kohteita kuvaavat muuttujat tulkitaan tilastollisissa tutkimusasetelmissä *satunnaismuuttujiksi* ja havaintoarvot tulkitaan näiden *satunnaismuuttujien realisoituneiksi arvoiksi*.

# Yksinkertainen satunnaisotos

## Tilastollinen malli

---

- **Tilastollisella mallilla** tarkoitetaan tutkimuksen kohteita kuvaavien satunnaismuuttujien *todennäköisyysjakaumaa*, jonka ajatellaan *generoineen ko. satunnaismuuttujien havaitut arvot*.
- Havaintoarvojen ajatellaan syntyneen *arpomalla* tilastollisena mallina käytetystä todennäköisyysjakaumasta saatavin todennäköisyyksin.
- Huomautus:

Todennäköisyysjakaumat riippuvat tavallisesti *parametreista* eli vakioista, joiden arvoja ei yleensä tunneta.

# Yksinkertainen satunnaisotos

## Tilastolliset mallit ja tilastollinen päättely

---

- Kun tilastollista mallia sovelletaan jotakin reaalimaailman ilmiötä kuvaavan havaintoaineiston analysointiin, kohdataan tavallisesti seuraavat mallin **parametreja** koskevat ongelmat:
  - (i) Parametrien arvoja *ei tunneta* ja ne on **estimoitava** eli *arvioitava* havaintoaineistosta.
  - (ii) Parametrien arvoista on esitetty *oletuksia* tai *väitteitä*, joita halutaan **testata** eli asettaa koetteelle havaintoaineistosta saatua informaatiota vastaan.
- Tilastollisten mallien parametrien estimointi ja testaus muodostavat keskeisen osan **tilastollista päättelyä**.

## Satunnaisotos ja satunnaisotanta

---

- **Satunnaisotos** poimitaan *arpomalla* havaintoyksiköt perusjoukosta otokseen.
- Arpomisessa käytettävää menetelmää kutsutaan **satunnaisotannaksi**.
- Satunnaisotannassa *sattuma* määrää mitkä perusjoukon alkioista tulevat otokseen.

## Satunnaisotanta: Kommentteja

---

- Jos havaintoyksiköt poimitaan perusjoukosta satunnaisotannalla, pätee seuraava:
  - (i) **Havaintoyksiköitä kuvaavien muuttujien *havaitut arvot ovat satunnaisia* siinä mielessä, että ne vaihtelevat satunnaisesti otoksesta toiseen.**
  - (ii) ***Kaikki* havaintoyksiköitä kuvaavien muuttujien *havaituista arvoista lasketut tunnusluvut ovat satunnaisia* siinä mielessä, että ne vaihtelevat satunnaisesti otoksesta toiseen.**

# Yksinkertainen satunnaisotos

---

- Olkoot

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

*riippumattomia, identtisesti jakautuneita* satunnaismuuttujia, joilla on *sama* pistetodennäköisyys- tai tiheysfunktio

$$f(x)$$

- Tällöin satunnaismuuttujat

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

muodostavat **yksinkertaisen satunnaisotoksen** jakaumasta  $f(x)$ .

# Yksinkertainen satunnaisotos

## Havainnot ja havaintoarvot

---

- Olkoon

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

*yksinkertainen satunnaisotos* jakaumasta  $f(x)$ .

- Kutsumme satunnaismuuttujia  $X_1, X_2, \dots, X_n$  tavallisesti **havainnoiksi**.
- *Otoksen poimimisen jälkeen* satunnaismuuttujat  $X_1, X_2, \dots, X_n$  saavat havaituiksi arvoikseen **havaintoarvot**

$$x_1, x_2, \dots, x_n$$

- Merkitään:

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



## Yksinkertainen satunnaisotos: Kommentteja 1/2

---

- Olkoon

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

*yksinkertainen satunnaisotos jakaumasta  $f(x)$ .*

- Tällöin havaintoarvot

$$x_1, x_2, \dots, x_n$$

*on saatu toistamalla arvontaa toisistaan riippumattomin toistoin  $n$  kertaa samoin, jakaumasta  $f(x)$  saatavin todennäköisyyksin.*

- Havaintoarvot  $x_1, x_2, \dots, x_n$  ovat kiinteitä eli *ei-satunnaisia*, mutta ne vaihtelevat toisistaan riippumatta ja *satunnaisesti otoksesta toiseen.*

## Yksinkertainen satunnaisotos: Kommentteja 2/2

---

- Satunnaisuus liittyy yksinkertaisessa satunnaisotannassa siihen, että *havaintoarvot vaihtelevat toisistaan riippumatta ja satunnaisesti otoksesta toiseen.*
- **Satunnaisuus ei siis liity otannan tuloksena saatuihin havaintoarvoihin, vaan otoksen poimintatapaan.**

# Tilastollinen malli yksinkertaiselle satunnaisotokselle 1/2

---

- Olkoon

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

*yksinkertainen satunnaisotos jakaumasta  $f(x)$ .*

- Satunnaismuuttujien  $X_1, X_2, \dots, X_n$  yhteisjakauma muodostaa **tilastollisen mallin** *havaintoarvojen satunnaiselle vaihtelulle otoksesta toiseen.*

# Tilastollinen malli yksinkertaiselle satunnaisotokselle 2/2

---

- Koska satunnaismuuttujat

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

on oletettu riippumattomiksi, niin satunnaismuuttujien  $X_1, X_2, \dots, X_n$  yhteisjakauma on muotoa

$$f(x_1, x_2, \dots, x_n) = f(x_1) \times f(x_2) \times \dots \times f(x_n)$$

jossa

$$X_i \sim f(x_i), i = 1, 2, \dots, n$$

# Otos ja otosjakaumat

---

**Yksinkertainen satunnaisotos**

**>> Otostunnusluvut ja otosjakaumat**

**Aritmeettisen keskiarvon ja otosvariانسsin otosjakaumat**

**Suhteellisen frekvenssin otosjakauma**

# Otostunnusluvut ja otosjakaumat

---

## *Avainsanat*

Otosjakauma

Otostunnusluvut

Tilastollinen malli

Yksinkertainen satunnaisotos

## Otostunnusluvut 1/3

---

- 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*  $f(x)$ :

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

$$X_i \sim f(x), i = 1, 2, \dots, n$$

## Otostunnusluvut ja otosjakaumat

# Otostunnusluvut 2/3

---

- 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**.



## Otostunnusluvut 3/3

---

- 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

$(x_1, x_2, \dots, x_n)$ :

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

## Otostunnusluvut ja otosjakaumat

# Otosjakauma

---

- Oletetaan, että satunnaismuuttujat

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

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

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

jokin otostunnusluku.

- Tunnusluvun  $T$  jakaumaa kutsutaan *tunnusluvun  $T$  otosjakaumaksi*.
- Tunnusluvun  $T$  otosjakauma muodostaa **tilastollisen mallin** *tunnusluvun  $T$  arvojen satunnaiselle vaihtelulle otoksesta toiseen*.

# Eräiden tavallisten tunnuslukujen otosjakaumat

---

- Olkoon

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

*yksinkertainen satunnaisotos* jakaumasta  $f(x)$ .

- Jatkossa tarkastellaan seuraavien tunnuslukujen (ks. lukua Tilastollisten aineistojen kuvaaminen) otosjakaumia:
  - **Aritmeettinen keskiarvo**
  - **Otosvarianssi**
  - **Suhteellinen frekvenssi**

# Otos ja otosjakaumat

---

**Yksinkertainen satunnaisotos**

**Otostunnusluvut ja otosjakaumat**

**>> Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat**

**Suhteellisen frekvenssin otosjakauma**

# Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat

---

## *Avainsanat*

Aritmeettisen keskiarvon otosjakauma

Keskeinen raja-arvolause

$\chi^2$ -jakauma

Normaalijakauma

Otosjakauma

Otostunnusluvut

Otosvarianssin otosjakauma

Riippumattomuus

Suhteellisen frekvenssin otosjakauma

Tilastollinen malli

$t$ -jakauma

Yksinkertainen satunnaisotos

## Aritmeettinen keskiarvo:

### Määritelmä 1/2

---

- Oletetaan, että havainnot

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

muodostavat *yksinkertaisen satunnaisotoksen* satunnaismuuttujan  $X$  jakaumasta, jonka *odotusarvo* ja *variانسsi* ovat

$$E(X) = \mu$$

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

- Tällöin kaikilla satunnaismuuttujilla  $X_i$ ,  $i = 1, 2, \dots, n$  on *sama* odotusarvo  $\mu$  ja *sama* variانسsi  $\sigma^2$ .

## Aritmeettinen keskiarvo:

### Määritelmä 2/2

---

- Olkoon

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{X_1 + X_2 + \dots + X_n}{n}$$

havaintojen  $X_1, X_2, \dots, X_n$  **aritmeettinen keskiarvo**.

- Aritmeettinen keskiarvo  $\bar{X}$  kuvaa havaintojen *keskimääräistä arvoa*.
- Aritmeettinen keskiarvo  $\bar{X}$  on satunnaismuuttuja, jonka saamat *arvot vaihtelevat satunnaisesti otoksesta toiseen*.

## Aritmeettisen keskiarvon otosjakauma: Odotusarvo ja varianssi

---

- Aritmeettisen keskiarvon  $\bar{X}$  odotusarvo ja varianssi:

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

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

- Aritmeettisen keskiarvon  $\bar{X}$  *standardipoikkeamaa*

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

kutsutaan tavallisesti **keskiarvon keskivirheeksi** ja se kuvaa aritmeettisen keskiarvon otosvaihtelua oman odotusarvonsa  $\mu$  ympärillä.



# Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat

## Aritmeettisen keskiarvon otosjakauma: Odotusarvon johto

---

- Olkoot  $X_1, X_2, \dots, X_n$  riippumattomia satunnaismuuttujia, joille

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

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

- Odotusarvon yleisten ominaisuuksien perusteella pätee (myös *ilman riippumattomuusoletusta*):

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right)$$

$$= \frac{1}{n} \sum_{i=1}^n E(X_i)$$

$$= \frac{1}{n} \sum_{i=1}^n \mu$$

$$= \frac{1}{n} n\mu = \mu$$

# Aritmeettisen keskiarvon ja otosvariانسsin otosjakaumat

## Aritmeettisen keskiarvon otosjakauma:

### Varianssin johto

---

- Olkoot  $X_1, X_2, \dots, X_n$  riippumattomia satunnaismuuttujia, joille
$$E(X_i) = \mu, i = 1, 2, \dots, n$$
$$\text{Var}(X_i) = \sigma^2, i = 1, 2, \dots, n$$
- Varianssin yleisten ominaisuuksien perusteella pätee (koska satunnaismuuttujat  $X_1, X_2, \dots, X_n$  on oletettu riippumattomiksi):

$$\begin{aligned}\text{Var}(\bar{X}) &= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) \\ &= \frac{1}{n^2} \sum_{i=1}^n \sigma^2 \\ &= \frac{1}{n^2} n \sigma^2 = \frac{\sigma^2}{n}\end{aligned}$$

# Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat

## Aritmeettisen keskiarvon otosjakauma:

### Jakauman käyttäytyminen otoskoon kasvaessa

---

- Koska aritmeettisen keskiarvon  $\bar{X}$  odotusarvo on

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

ja varianssi on

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

niin aritmeettisen keskiarvon otosjakauma *keskittyy yhä voimakkaammin havaintojen yhteisen odotusarvon  $\mu$  ympärille, kun otoskoko  $n$  kasvaa.*

# Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat

## Aritmeettisen keskiarvon otosjakauma: Normaalijakautunut otos

---

- Oletetaan, että havainnot

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

muodostavat *yksinkertaisen satunnaisotoksen normaali-jakaumasta*

$$N(\mu, \sigma^2)$$

- Tällöin havaintojen *aritmeettinen keskiarvo*  $\bar{X}$  *noudattaa eksaktisti* (eli myös äärellisissä otoksissa) **normaalijakaumaa**:

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

# Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat

## Aritmeettisen keskiarvon otosjakauma:

### Perustelu, kun otos on normaalijakautunut 1/2

---

- Olkoon

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

*yksinkertainen satunnaisotos normaalijakaumasta*

$$N(\mu, \sigma^2)$$

- Koska oletuksen mukaan havainnot  $X_1, X_2, \dots, X_n$  ovat *riippumattomia*, niin

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

ja

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

# Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat

## Aritmeettisen keskiarvon otosjakauma: Perustelu, kun otos on normaalijakautunut 2/2

---

- Perustelu:

Ks. todistusta normaalijakautuneen otoksen aritmeettisen keskiarvon  $\bar{X}$  ja otosvarianssin  $s^2$  riippumattomuudelle  $>$  sekä kalvokokoelman **Johdatus todennäköisyyslaskentaan** lukuja **Jatkuvia jakaumia, Moniulotteiset satunnaismuuttujat ja todennäköisyysjakaumat** sekä **Satunnaismuuttujien muunnokset ja niiden jakaumat**.

# Aritmeettisen keskiarvon ja otosvariانسsin otosjakaumat

## Aritmeettisen keskiarvon otosjakauma: Asymptoottinen jakauma

---

- Oletetaan, että havainnot

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

muodostavat *yksinkertaisen satunnaisotoksen* satunnaismuuttujan  $X$  jakaumasta, jonka *odotusarvo* on  $\mu$  ja *variانسsi* on  $\sigma^2$ .

- Tällöin havaintojen *aritmeettinen keskiarvo*  $\bar{X}$  *noudattaa suurissa otoksissa approksimatiivisesti normaalijakaumaa* jonka odotusarvo on  $\mu$  ja variانسsi on  $\sigma^2 / n$  :

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

# Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat

## Aritmeettisen keskiarvon otosjakauma:

### Kommentteja 1/2

---

- *Oletukset havaintojen riippumattomuudesta, samasta jakaumasta ja normaalisuudesta ovat välttämättömiä aritmeettisen keskiarvon eksaktia eli tarkkaa otosjakaumaa koskevalle tulokselle.*
- Aritmeettisen keskiarvon otosjakaumaa koskeva *asymptoottinen* tulos seuraa **keskeisestä raja-arvolauseesta**; ks. kalvokokoelman Johdatus todennäköisyyslaskentaan lukua Jatkuvia jakaumia tai lukua Konvergenssikäsitteet ja raja-arvolauseet.



## Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat

# Aritmeettisen keskiarvon otosjakauma:

### Kommentteja 2/2

---

- Aritmeettisen keskiarvon otosjakaumaa koskeva asymptoottinen tulos *pätee tietyin lisäehdoin* myös monissa sellaisissa tilanteissa, joissa *havaintojen riippumattomuutta ja samaa jakaumaa koskevat oletukset eivät päde*.

# Aritmeettisen keskiarvon ja otosvariانسsin otosjakaumat

## Standardoidun aritmeettisen keskiarvon otosjakauma: Odotusarvo ja varianssi

---

- Koska

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

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

niin *standardoidun* satunnaismuuttujan

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

*odotusarvo ja varianssi* ovat

$$E(Z) = 0$$

$$\text{Var}(Z) = 1$$

# Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat

## Standardoidun aritmeettisen keskiarvon otosjakauma: Normaalijakautunut otos

---

- Oletetaan, että havainnot

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

muodostavat yksinkertaisen satunnaisotoksen normaali-jakaumasta

$$N(\mu, \sigma^2)$$

- Tällöin *standardoitu satunnaismuuttuja*

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

*noudattaa eksaktisti* (eli myös äärellisissä otoksissa)

**standardoitua normaalijakaumaa:**

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

# Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat

## Standardoidun aritmeettisen keskiarvon otosjakauma: Asymptoottinen jakauma

---

- Oletetaan, että havainnot

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

muodostavat *yksinkertaisen satunnaisotoksen* satunnaismuuttujan  $X$  jakaumasta, jonka *odotusarvo* on  $\mu$  ja *varianssi* on  $\sigma^2$ .

- Tällöin *standardoitu satunnaismuuttuja*

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

*noudattaa suurissa otoksissa approksimatiivisesti standardoitua normaalijakaumaa:*

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

## Otosvariانسsi:

### Määritelmä 1/2

---

- Oletetaan, että havainnot

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

muodostavat *yksinkertaisen satunnaisotoksen* satunnaismuuttujan  $X$  jakaumasta, jonka *odotusarvo* ja *variانسsi* ovat

$$E(X) = \mu$$

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

- Tällöin kaikilla satunnaismuuttujilla  $X_i$ ,  $i = 1, 2, \dots, n$  on *sama* odotusarvo  $\mu$  ja *sama* variانسsi  $\sigma^2$ .

## Otosvariانسsi:

### Määritelmä 2/2

---

- Olkoon

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

havaintojen  $X_1, X_2, \dots, X_n$  **otosvariانسsi**, jossa

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

on havaintojen  $X_1, X_2, \dots, X_n$  *aritmeettinen keskiarvo*.

- Otosvariانسsi  $s^2$  kuvaa havaintoarvojen *vaihtelua niiden aritmeettisen keskiarvon ympärillä*.
- Otosvariانسsi  $s^2$  on satunnaismuuttuja, jonka saamat *arvot vaihtelevat satunnaisesti otoksesta toiseen*.

## Otosvariانسsin otosjakauma: Odotusarvo ja variانسsi

---

- **Otosvariانسsin  $s^2$  odotusarvo:**

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

- Jos lisäksi voidaan olettaa, että havainnot  $X_1, X_2, \dots, X_n$  noudattavat normaalijakaumaa  $N(\mu, \sigma^2)$ , niin **otosvariانسsin  $s^2$  variانسsi** on

$$\text{Var}(s^2) = D^2(s^2) = \frac{2\sigma^4}{n-1}$$

- Siten otosvariانسsin  $s^2$  *standardipoikkeama* on normaalisen otoksen tapauksessa

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

## Otosvarianssin otosjakauma: Normaalijakautunut otos 1/2

---

- Oletetaan, että havainnot

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

muodostavat yksinkertaisen satunnaisotoksen normaali-jakaumasta

$$N(\mu, \sigma^2)$$

- Tällöin satunnaismuuttuja

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

noudattaa  **$\chi^2$ -jakaumaa** vapausastein  $n$ :

$$Y \sim \chi^2(n)$$



## Otosvarianssin otosjakauma: Normaalijakautunut otos 2/2

---

- Oletetaan, että havainnot

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

muodostavat yksinkertaisen satunnaisotoksen normaali-jakaumasta

$$N(\mu, \sigma^2)$$

- Tällöin satunnaismuuttuja

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

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

$$V \sim \chi^2(n-1)$$

## Otosvariانسsin otosjakauma:

### Perustelu, kun otos on normaalijakautunut 1/6

---

- Olkoon

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

*yksinkertainen satunnaisotos normaalijakaumasta*

$$N(\mu, \sigma^2)$$

- Olkoon

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

*havaintojen  $X_1, X_2, \dots, X_n$  aritmeettinen keskiarvo ja*

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

*havaintojen  $X_1, X_2, \dots, X_n$  (harhaton) otosvariانسsi.*

## Otosvarianssin otosjakauma:

### Perustelu, kun otos on normaalijakautunut 2/6

---

- Määritellään satunnaismuuttuja  $Y$  kaavalla

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

- Koska havainnot  $X_1, X_2, \dots, X_n$  ovat *riippumattomia* ja noudattavat normaalijakaumaa  $N(\mu, \sigma^2)$ :

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

niin *standardoidut* satunnaismuuttujat

$$Y_i = \frac{X_i - \mu}{\sigma}, i = 1, 2, \dots, n$$

ovat *riippumattomia* ja noudattavat standardoitua normaalijakaumaa  $N(0, 1)$ :

$$Y_i \sim N(0, 1), i = 1, 2, \dots, n$$

## Otosvarianssin otosjakauma:

### Perustelu, kun otos on normaalijakautunut 3/6

---

- Edellä esitetystä seuraa, että satunnaismuuttuja  $Y$  on *riippumattomien, standardoitua normaalijakaumaa*  $N(0, 1)$  *noudattavien satunnaismuuttujien*  $Y_i$ ,  $i = 1, 2, \dots, n$  *neliösumma*:

$$Y = \sum_{i=1}^n Y_i^2$$

- Suoraan  $\chi^2$ -jakauman määritelmästä seuraa, että satunnaismuuttuja  $Y$  noudattaa  $\chi^2$ -jakaumaa vapausastein  $n$ :

$$Y \sim \chi^2(n)$$

Ks. kalvokokoelman **Johdatus todennäköisyyslaskentaan** lukua **Normaalijakaumasta johdettuja jakaumia**.

## Otosvarianssin otosjakauma:

### Perustelu, kun otos on normaalijakautunut 4/6

---

- Määritellään nyt satunnaismuuttuja  $V$  kaavalla

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

- Satunnaismuuttuja  $V$  saadaan satunnaismuuttujasta

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

korvaamalla odotusarvo  $\mu$  harhattomalla estimaattorillaan  $\bar{X}$ .

- Satunnaismuuttujan  $V$  määritelmässä esiintyvän summan termit

$$U_i = \frac{X_i - \bar{X}}{\sigma}, i = 1, 2, \dots, n$$

*eivät ole riippumattomia.*

## Otosvariانسsin otosjakauma:

### Perustelu, kun otos on normaalijakautunut 5/6

---

- Voidaan kuitenkin osoittaa, että  $V$  voidaan esittää *riippumattomien, standardoitua normaalijakaumaa*  $N(0, 1)$  *noudattavien satunnaismuuttujien*  $V_i, i = 1, 2, \dots, n - 1$  *neliösummana* (ks. todistusta normaalijakautuneen otoksen aritmeettisen keskiarvon  $\bar{X}$  ja otosvariانسsin  $s^2$  riippumattomuudelle >):

$$V = \sum_{i=1}^{n-1} V_i^2$$

- Siten suoraan  $\chi^2$ -jakauman määritelmästä seuraa, että satunnaismuuttuja  $Y$  noudattaa  $\chi^2$ -jakaumaa vapausastein  $(n - 1)$ :

$$V \sim \chi^2(n - 1)$$

Ks. kalvokokoelman **Johdatus todennäköisyyslaskentaan** lukua **Normaalijakaumasta johdettuja jakaumia**.

## Otosvariانسsin otosjakauma:

### Perustelu, kun otos on normaalijakautunut 6/6

---

- Huomautuksia:
  - (i) Satunnaismuuttuja  $Y$  noudattaa  $\chi^2$ -jakaumaa, jonka vapausasteiden lukumäärä on sama kuin havaintojen lukumäärä  $n$ .
  - (ii) Kun satunnaismuuttujasta  $Y$  siirrytään satunnaismuuttujaan  $V$  menetetään yksi vapausaste.
  - (iii) Yhden vapausasteen menetys on seurausta siitä, että parametrin  $\mu$  korvaaminen estimaattorillaan  $\bar{X}$  riippumattomissa satunnaismuuttujissa

$$Y_i = \frac{X_i - \mu}{\sigma}, i = 1, 2, \dots, n$$

luo yhden (lineaarisen) side-ehdon satunnaismuuttujien

$$U_i = \frac{X_i - \bar{X}}{\sigma}, i = 1, 2, \dots, n$$

välille.

## Otosvariانسsin otosjakauma: Kommentteja

---

- *Oletukset havaintojen riippumattomuudesta ja samasta jakaumasta ovat välttämättömiä otosvariانسsin eksaktia eli tarkkaa otosjakaumaa koskevalle tulokselle.*



# Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat

## Aritmeettisen keskiarvon ja otosvarianssin riippumattomuus 1/2

---

- Oletetaan, että havainnot

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

muodostavat yksinkertaisen satunnaisotoksen normaali-jakaumasta

$$N(\mu, \sigma^2)$$

- Olkoon

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

havaintojen  $X_1, X_2, \dots, X_n$  aritmeettinen keskiarvo ja

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

havaintojen  $X_1, X_2, \dots, X_n$  (harhaton) otosvarianssi.

# Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat

## Aritmeettisen keskiarvon ja otosvarianssin riippumattomuus 2/2

---

- Tällöin  $\bar{X}$  ja  $s^2$  ovat *riippumattomia*:

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

- Lisäksi

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

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

# Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat

## Aritmeettisen keskiarvon ja otosvarianssin riippumattomuus: Perustelu 1/8

---

- Oletetaan, että havainnot

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

muodostavat yksinkertaisen satunnaisotoksen normaalijakaumasta

$$N(\mu, \sigma^2)$$

- Olkoon

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

havaintojen  $X_1, X_2, \dots, X_n$  aritmeettinen keskiarvo ja

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

havaintojen  $X_1, X_2, \dots, X_n$  (harhaton) otosvarianssi.

# Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat

## Aritmeettisen keskiarvon ja otosvarianssin riippumattomuus: Perustelu 2/8

---

- Otoksen yhteisjakauman tiheysfunktio voidaan kirjoittaa havaintojen riippumattomuuden ja normaalisuuden takia seuraavaan muotoon:

$$f(x_1, x_2, \dots, x_n) = (2\pi)^{-\frac{1}{2}n} \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right\}$$

- Määritellään *lineaarinen* muunnos

$$\begin{cases} Y_1 = \frac{1}{\sqrt{n}} X_1 + \frac{1}{\sqrt{n}} X_2 + \frac{1}{\sqrt{n}} X_3 + \dots + \frac{1}{\sqrt{n}} X_n \\ Y_2 = \frac{1}{\sqrt{2}} X_1 - \frac{1}{\sqrt{2}} X_2 \\ Y_3 = \frac{1}{\sqrt{6}} X_1 + \frac{1}{\sqrt{6}} X_2 - \frac{2}{\sqrt{6}} X_3 \\ \vdots \\ Y_n = \frac{1}{\sqrt{n(n-1)}} X_1 + \frac{1}{\sqrt{n(n-1)}} X_2 + \frac{1}{\sqrt{n(n-1)}} X_3 + \dots - \frac{n-1}{\sqrt{n(n-1)}} X_n \end{cases}$$

# Aritmeettisen keskiarvon ja otosvariانسsin otosjakaumat

## Aritmeettisen keskiarvon ja otosvariانسsin riippumattomuus: Perustelu 3/8

---

- Muunnos voidaan esittää *matriisein* muodossa

$$\mathbf{Y} = \mathbf{B}\mathbf{X}$$

jossa

$$\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)$$

$$\mathbf{X} = (X_1, X_2, \dots, X_n)$$

ja  $n \times n$ -matriisi

$$\mathbf{B} = \begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & \cdots & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \cdots & -\frac{n-1}{\sqrt{n(n-1)}} \end{bmatrix}$$

on ortogonaalinen ( $\mathbf{B}'\mathbf{B} = \mathbf{B}\mathbf{B}' = \mathbf{I}$ ).

# Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat

## Aritmeettisen keskiarvon ja otosvarianssin riippumattomuus: Perustelu 4/8

---

- Matriisi **B** nähdään ortogonaaliseksi alla esitettävällä tavalla.
- Määritellään  $n \times n$ -matriisi

$$\mathbf{C} = \begin{bmatrix} 1 & 1 & 1 & 1 & \dots & 1 & 1 \\ 1 & -1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 1 & -2 & 0 & \dots & 0 & 0 \\ 1 & 1 & 1 & -3 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 1 & 1 & 1 & \dots & -(n-2) & 0 \\ 1 & 1 & 1 & 1 & \dots & 1 & -(n-1) \end{bmatrix}$$

- On helppo nähdä, että matriisin **C** rivit ovat *kohtisuorassa* toisiaan vastaan.
- Matriisi **B** saadaan matriisista **C** *normeeraamalla* sen rivit niin, että niiden pituudeksi tulee 1.

# Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat

## Aritmeettisen keskiarvon ja otosvarianssin riippumattomuus: Perustelu 5/8

---

- Koska muunnos

$$\mathbf{Y} = \mathbf{B}\mathbf{X}$$

on ortogonaalinen, niin muunnosta vastaavan *Jacobin determinantin* itseisarvo = 1.

- Koska

$$Y_1 = \frac{1}{\sqrt{n}}(X_1 + X_2 + \dots + X_n) = \sqrt{n}\bar{X}$$

ja

$$\begin{aligned} Y_1^2 + Y_2^2 + \dots + Y_n^2 &= \mathbf{Y}'\mathbf{Y} = \mathbf{X}'\mathbf{B}'\mathbf{B}\mathbf{X} = \mathbf{X}'\mathbf{X} = X_1^2 + X_2^2 + \dots + X_n^2 \\ &= \sum_{i=1}^n (X_i - \bar{X})^2 + n\bar{X}^2 \end{aligned}$$

niin

$$Y_2^2 + \dots + Y_n^2 = \sum_{i=1}^n (X_i - \bar{X})^2 = (n-1)s^2$$

# Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat

## Aritmeettisen keskiarvon ja otosvarianssin riippumattomuus: Perustelu 6/8

---

- Koska

$$\begin{aligned}\sum_{i=1}^n (X_i - \mu)^2 &= \sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^n (\bar{X} - \mu)^2 \\ &= \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2 \\ &= Y_2^2 + \dots + Y_n^2 + (Y_1 - \sqrt{n}\mu)^2\end{aligned}$$

niin satunnaismuuttujien  $Y_1, Y_2, \dots, Y_n$  yhteisjakauman tiheysfunktiksi saadaan

$$\begin{aligned}f(y_1, y_2, \dots, y_n) &= \frac{1}{(2\pi)^{\frac{1}{2}n} \sigma^n} e^{-\frac{1}{2\sigma^2} [(Y_1 - \sqrt{n}\mu)^2 + Y_2^2 + \dots + Y_n^2]} \\ &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2} (Y_1 - \sqrt{n}\mu)^2} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2} Y_2^2} \dots \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2} Y_n^2}\end{aligned}$$



# Aritmeettisen keskiarvon ja otosvariانسsin otosjakaumat

## Aritmeettisen keskiarvon ja otosvariانسsin riippumattomuus: Perustelu 7/8

---

- Edellä esitetystä seuraa, että satunnaismuuttujat

$$Y_1, Y_2, \dots, Y_n$$

ovat riippumattomia ja normaalijakautuneita:

$$Y_1, Y_2, \dots, Y_n \perp$$

$$Y_1 = \sqrt{n}\bar{X} \sim N(\sqrt{n}\mu, \sigma^2)$$

$$Y_i \sim N(0, \sigma^2), i = 2, \dots, n$$

- Lisäksi

$$s^2 = \frac{1}{n-1}(Y_2^2 + \dots + Y_n^2) = \frac{\sigma^2}{n-1} \left[ \left( \frac{Y_2}{\sigma} \right)^2 + \dots + \left( \frac{Y_n}{\sigma} \right)^2 \right]$$

jossa

$$\frac{Y_i}{\sigma} \sim N(0, 1), i = 2, \dots, n$$

# Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat

## Aritmeettisen keskiarvon ja otosvarianssin riippumattomuus: Perustelu 8/8

---

- Siten olemme todistaneet, että

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

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

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

- Huomautus:

Todistuksessa on sovellettu kalvokokoelman **Johdatus todennäköisyytlaskentaan** luvun **Satunnaismuuttujien muunnokset ja niiden jakaumat** teoriaa sekä luvussa **Normaalijakaumasta johdettuja jakaumia** esitettyä  $\chi^2$ -jakauman määritelmää.

# Aritmeettisen keskiarvon ja otosvariانسsin otosjakaumat

## Aritmeettisen keskiarvon ja otosvariانسsin otosjakaumat: Seuraus 1/2

---

- Oletetaan, että havainnot

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

muodostavat yksinkertaisen satunnaisotoksen normaali-jakaumasta

$$N(\mu, \sigma^2)$$

- Olkoon

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

havaintojen  $X_1, X_2, \dots, X_n$  aritmeettinen keskiarvo ja

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

havaintojen  $X_1, X_2, \dots, X_n$  (harhaton) otosvariانسsi.

# Aritmeettisen keskiarvon ja otosvariانسsin otosjakaumat

## Aritmeettisen keskiarvon ja otosvariانسsin otosjakaumat: Seuraus 2/2

---

- Tällöin

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

# Aritmeettisen keskiarvon ja otosvariانسsin otosjakaumat

## Aritmeettisen keskiarvon ja otosvariانسsin otosjakaumat: Todistus 1/3

---

- Oletetaan, että havainnot

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

muodostavat yksinkertaisen satunnaisotoksen normaalijakaumasta

$$N(\mu, \sigma^2)$$

- Olkoon

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

havaintojen  $X_1, X_2, \dots, X_n$  aritmeettinen keskiarvo ja

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

havaintojen  $X_1, X_2, \dots, X_n$  (harhaton) otosvariانسsi.

# Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat

## Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat: Todistus 2/3

---

- Aikaisemmin on todettu, että

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

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

ja lisäksi

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

- Aritmeettista keskiarvoa  $\bar{X}$  koskevasta jakaumatuloksesta seuraa, että

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

# Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat

## Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat: Todistus 3/3

---

- Siten suoraan *t*-jakauman määritelmästä seuraa, että

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{1}{n-1} \left( \frac{(n-1)s^2}{\sigma^2} \right)}} \sim t(n-1)$$

Ks. kalvokokoelman **Johdatus todennäköisyyslaskentaan** lukua **Normaalijakaumasta johdettuja jakaumia**.

# Otos ja otosjakaumat

---

**Yksinkertainen satunnaisotos**

**Otostunnusluvut ja otosjakaumat**

**Aritmeettisen keskiarvon ja otosvariانسsin otosjakaumat**

**>> Suhteellisen frekvenssin otosjakauma**



# Suhteellisen frekvenssi otosjakauma

---

## *Avainsanat*

**Keskeinen raja-arvolause**

**Normaalijakauma**

**Otosjakauma**

**Otostunnusluvut**

**Suhteellisen frekvenssin otosjakauma**

**Tilastollinen malli**

**Yksinkertainen satunnaisotos**

## Frekvenssi ja suhteellinen frekvenssi: Määritelmät 1/3

---

- Olkoon  $P$  jokin otosavaruuden  $S$  alkioiden *ominaisuus*.
- Jos otosavaruuden  $S$  alkiolla  $x$  on ominaisuus  $P$ , merkitään

$$P(x)$$

- Olkoon

$$A = \{x \in S \mid P(x)\}$$

niiden otosavaruuden  $S$  alkioiden *osajoukko*, joilla on ominaisuus  $P$ .

- Oletetaan, että **tapahtuman  $A$  todennäköisyys** on

$$\Pr(A) = p$$

## Frekvenssi ja suhteellinen frekvenssi: Määritelmät 2/3

---

- Poimitaan otosavaruudesta  $S$  yksinkertainen satunnaisotos, jonka *koko* on  $n$ .
- Olkoon

$f$

niiden havaintoyksiköiden **frekvenssi**, joilla on ominaisuus  $P$  ja olkoon

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

vastaava **suhteellinen frekvenssi**.

## Frekvenssi ja suhteellinen frekvenssi: Määritelmät 3/3

---

- Frekvenssi

$$f$$

kuvaa *A*-tyyppisten alkioiden **lukumäärää** otoksessa ja vastaava suhteellinen frekvenssi

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

kuvaa *A*-tyyppisten alkioiden **suhteellista osuutta** otoksessa.

- Frekvenssi  $f$  ja vastaava suhteellinen frekvenssi  $\hat{p}$  ovat *satunnaismuuttujia*, joiden saamat *arvot vaihtelevat satunnaisesti otoksesta toiseen*.

## Frekvenssi:

### Odotusarvo, varianssi ja jakauma 1/2

---

- Olkoon  $A$  jokin otosavaruuden  $S$  tapahtuma:

$$A \subset S$$

- Poimitaan otosavaruudesta  $S$  yksinkertainen satunnaisotos, jonka koko on  $n$ .
- Olkoon

$$f$$

$A$ -tyyppisten alkioiden lukumäärä eli *frekvenssi* otoksessa.

## Frekvenssi:

## Odotusarvo, varianssi ja jakauma 2/2

---

- **Frekvenssin  $f$  odotusarvo ja varianssi:**

$$E(f) = np$$

$$\text{Var}(f) = npq$$

jossa  $q = 1 - p$ .

- *Frekvenssi  $f$  noudattaa eksaktisti **binomijakaumaa** parametrein  $n$  ja  $\text{Pr}(A) = p$ :*

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

Ks. kalvokokoelman **Johdatus todennäköisyyslaskentaan** lukua **Diskreettejä jakaumia** tai lukua **Satunnaismuuttujien muunnokset ja niiden jakaumat**.

## Suhteellisen frekvenssin otosjakauma

# Suhteellinen frekvenssi: Odotusarvo ja varianssi 1/2

---

- Olkoon  $A$  jokin otosavaruuden  $S$  *tapahtuma*:

$$A \subset S$$

- Poimitaan otosavaruudesta  $S$  *yksinkertainen satunnaisotos*, jonka *koko* on  $n$ .
- Olkoon

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

$A$ -tyyppisten alkioiden *suhteellinen osuus* eli *frekvenssi* otoksessa.

## Suhteellinen frekvenssi: Odotusarvo ja varianssi 2/2

---

- **Suhteellisen frekvenssin  $\hat{p}$  odotusarvo ja varianssi:**

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

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

jossa  $q = 1 - p$ .

- Suhteellisen frekvenssin  $\hat{p}$  *standardipoikkeamaa*

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

kutsutaan tavallisesti **suhteellisen frekvenssin keski-  
virheeksi** ja se kuvaa suhteellisen frekvenssin  $f/n$  otos-  
vaihtelua oman odotusarvonsa  $p$  ympärillä.



## Suhteellisen frekvenssin otosjakauma: Jakauman käyttäytyminen otoskoon kasvaessa

---

- Koska suhteellisen frekvenssin  $\hat{p}$  odotusarvo

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

ja varianssi on

$$\text{Var}(\hat{p}) = pq/n, \quad q = 1 - p$$

niin suhteellisen frekvenssin otosjakauma *keskittyy yhä voimakkaammin tapahtuman A todennäköisyyden p ympärille, kun otoskoko n kasvaa.*

## Suhteellisen frekvenssin otosjakauma: Asymptoottinen jakauma

---

- *Suhteellinen frekvenssi  $\hat{p}$  noudattaa suurissa otoksissa approksimatiivisesti normaalijakaumaa:*

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

- *Siten standardoitu satunnaismuuttuja*

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

*noudattaa suurissa otoksissa approksimatiivisesti standardoitua normaalijakaumaa:*

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

## Suhteellisen frekvenssin otosjakauma: Kommentti

---

- Suhteellisen frekvenssin otosjakaumaa koskeva asymptoottinen tulos seuraa **keskeisestä raja-arvolauseesta**; ks. kalvokokoelman Johdatus todennäköisyyslaskentaan lukua Jatkuvia jakaumia tai lukua Konvergenssikäsitteet ja raja-arvolauseet.