

# 6 Kertaus

## 6.1 Kurssin keskeiset asiat

1.

a) Muuttujan  $X$  arvot luetaan vaaka-akselilta. Muuttuja  $X$  saa arvot 0, 1, 2, 3, 4 tai 5.

b) Pistetodennäköisyys  $p(1)$  luetaan arvoa  $X = 1$  vastaavan pylvään korkeudesta.

$$p(1) \approx 0,08$$

c) Pistetodennäköisyys  $P(X = 2)$  luetaan arvoa  $X = 2$  vastaavan pylvään korkeudesta.

$$P(X = 2) \approx 0,23$$

d) Tapahtuma  $X < 2$  tarkoittaa tapahtumaa  $X \leq 1$  eli " $X = 0$  tai  $X = 1$ ". Kertymätodennäköisyys  $P(X \leq 1)$  on pistetodennäköisyyksien  $p(0)$  ja  $p(1)$  summa.

$$P(X \leq 1) = p(0) + p(1) \approx 0,01 + 0,08 = 0,09$$

e) Tapahtuma  $0 < X \leq 2$  tarkoittaa tapahtumaa  $1 \leq X \leq 2$  eli " $X = 1$  tai  $X = 2$ ".

$$P(1 \leq X \leq 2) = p(1) + p(2) \approx 0,08 + 0,23 = 0,31$$

f) Todennäköisin arvo on se arvo, jonka pistetodennäköisyys on suurin eli jota vastaava pylväs on korkein. Todennäköisin arvo on  $X = 3$ .

2.

a) Muuttujan  $X$  mahdolliset arvot ovat 0, 1, 2, 3 tai 4. Arvoja vastaavat pistetodennäköisyydet luetaan  $y$ -akselilta.

Muuttujan arvo	Pistetodennäköisyys
0	0,24
1	0,41
2	0,26
3	0,08
4	0,01

Kannattaa tarkistaa, että pistetodennäköisyyksien summaksi tulee 1.

$$0,24 + 0,41 + 0,26 + 0,08 + 0,01 = 1$$

b)

Muuttujan arvo	Kertymätodennäköisyys
0	0,24
1	$0,24 + 0,41 = 0,65$
2	$0,65 + 0,26 = 0,91$
3	$0,91 + 0,08 = 0,99$
4	$0,99 + 0,01 = 1$

3.

Binomijakautunut satunnaismuuttuja  $Y =$  "onnistuneiden heittojen lukumäärä".

$$Y \sim \text{Bin}(7; 0,49)$$

Toistojen lukumäärä on  $n = 7$ .

Onnistumistodennäköisyys yhdessä toistossa on  $p = 0,49$ .

a) Muuttujan  $Y$  mahdolliset arvot ovat 0, 1, 2, 3, 4, 5, 6 tai 7.

b) Binomijakautuneen satunnaismuuttujan odotusarvo on

$$E(Y) = n \cdot p = 7 \cdot 0,49 = 3,43$$

Jos 7 heiton heittosarjaa toistetaan useita kertoja, onnistuneiden heittojen lukumäärän keskiarvo on yhdessä sarjassa 3,43.

c) Binomijakautuneen satunnaismuuttujan keskihajonta on

$$\begin{aligned} D(Y) &= \sqrt{n \cdot p \cdot (1 - p)} \\ &= \sqrt{7 \cdot 0,49 \cdot 0,51} \\ &= 1,3226 \dots \\ &\approx 1,32 \end{aligned}$$

Kaava löytyy taulukkokirjasta.

$$\begin{aligned} D(Y) &= \sqrt{n \cdot p \cdot (1 - p)} \\ p &= 0,49 \\ 1 - p &= 1 - 0,49 = 0,51 \end{aligned}$$

#### 4.

Havaintojen kokonaismäärä nähdään viimeisen luokan summafrekvenssistä ( $sf$ ).

Havaintoja on yhteensä  $N = 25$ .

$x$	Tod. alaraja	Tod. yläraja	$f$	$f\%$	$sf$	$sf\%$
1-4	0,5	4,5	5	$\frac{5}{25} = 0,2$ = 20 %	5	20
5-8	4,5	8,5	$15 - 5$ = 10	$\frac{10}{25} = 0,4$ = 40 %	15	$20 + 40$ = 60
9-12	8,5	12,5	$18 - 15$ = 3	$\frac{3}{25} = 0,12$ = 12 %	$25 - 7$ = 18	$60 + 12$ = 72
13-16	12,5	16,5	7	$\frac{7}{25} = 0,28$ = 28 %	25	$72 + 28$ = 100

Suhteellinen summafrekvenssi ( $sf\%$ )  
saadaan myös summafrekvenssin ( $sf$ )  
prosenttiosuutena kaikista havainnoista.

5.

a) Vastaajien lukumäärä nähdään viimeisen luokan summafrekvenssistä ( $sf\%$ ).

Vastaajia oli yhteensä  $N = 34$ .

b) Luokitus on tasavälinen. Luokkavälin pituus saadaan todellisten luokkarajojen avulla, erotuksena.

Esimerkiksi ensimmäisen luokan todellinen alaraja on 159,5 (cm) ja todellinen yläraja on 164,5 (cm), joten luokkavälin pituus on

$$164,5 - 159,5 = 5 \text{ (cm)}.$$

Moodiluokka on luokka, jonka frekvenssi ( $f$  tai  $f\%$ ) on suurin.

Moodiluokka on siis 165–169 (cm).

Mediaaniluokka on luokka, jonka suhteellinen summafrekvenssi ( $sf\%$ ) ensimmäisen kerran ylittää arvon 50 (%). Mediaaniluokka on siis 170–174 (cm).

c) Vähintään 174,5 cm pituiset vastaajat kuuluvat kolmeen ylimpään luokkaan. Näitä vastaajia oli yhteensä

$$9 + 4 + 2 = 15.$$

d) Alle 169,5 cm pituiset vastaajat kuuluvat kahteen alimpaan luokkaan. Näitä vastaajia oli yhteensä

$$2 + 10 = 12$$

eli prosentteina

$$\frac{12}{34} = 0,3529 \dots \approx 35 \%$$

e) Vähintään 169,5 cm mutta alle 179,5 cm pituiset vastaajat kuuluvat luokkiin 170–174 (cm) ja 175–179 (cm). Näitä vastaajia oli yhteensä

$$7 + 9 = 16$$

eli prosentteina

$$\frac{16}{34} = 0,4705 \dots \approx 47 \%$$

## 6.

Todelliset luokkarajat näkyvät vaak akselilta, pylväiden reunojen kohdalta. Frekvenssitaulukossa ilmoitetaan pyöristetyt luokkarajat.

Luokkakeskukset lasketaan todellisten luokkarajojen avulla, keskiarvona.

Frekvenssit ( $f$ ) luetaan pysty akselilta, pylväiden korkeuksien kohdalta.

Massa (kg)	Luokkakeskus	$f$	$f\%$	$sf$	$sf\%$
45-54	$\frac{44,5 + 54,5}{2}$ = 49,5	8	$\frac{8}{22} = 0,363 \dots$ = 36 %	8	$\frac{8}{22} = 0,363 \dots$ = 36 %
55-64	$\frac{54,5 + 64,5}{2}$ = 59,5	5	$\frac{5}{22} = 0,227 \dots$ = 23 %	8 + 5 = 13	$\frac{13}{22} = 0,590 \dots$ = 59 %
65-74	$\frac{64,5 + 74,5}{2}$ = 69,5	4	$\frac{4}{22} = 0,181 \dots$ = 18 %	13 + 4 = 17	$\frac{17}{22} = 0,772 \dots$ = 77 %
75-84	$\frac{74,5 + 84,5}{2}$ = 79,5	3	$\frac{3}{22} = 0,136 \dots$ = 14 %	17 + 3 = 20	$\frac{20}{22} = 0,909 \dots$ = 91 %
85-94	$\frac{84,5 + 94,5}{2}$ = 89,5	2	$\frac{2}{22} = 0,090 \dots$ = 9 %	20 + 2 = 22	$\frac{22}{22} = 1 = 100\%$

↑  
Frekvenssi ( $f$ ) luetaan histogrammista pylvään korkeuden kohdalta.



b) Luokan 45–54 (kg) frekvenssi on suurin. Luokan luokkakeskus on 49,5 (kg), joten tyypillisin lampaan massa oli 49,5 kg.

c) Massojen jakauma on oikealle vino, eli vaihteluvälillä 45 kg – 94 kg pienet massat ovat yleisempiä kuin suuret massat.

Luokitellun aineiston tunnusluvut lasketaan luokkakeskusten avulla. Massan keskiarvo on

$$\begin{aligned}\bar{x} &= \frac{49,5 \cdot 8 + 59,5 \cdot 5 + 69,5 \cdot 4 + 79,5 \cdot 3 + 89,5 \cdot 2}{22} = \frac{1389}{22} \\ &= 63,136 \dots \approx 63 \text{ (kg)}.\end{aligned}$$

---

*Huomautus:* Keskiarvon käyttöä tunnuslukuna tulisi välttää, jos jakauma on voimakkaasti vino. Tällöin mediaani on parempi keskiluku keskimääräisen arvon kuvailuun.

Mediaaniluokka on se luokka, jonka suhteellinen summafrekvenssi ensimmäisen kerran ylittää arvon 50 (%). Lampaiden massan mediaaniluokka on 55–64 (kg) ja luokan luokkakeskus antaa likiarvon mediaanille. Siis  $M_d \approx 59,5 \approx 60$  (kg).

7.

a) Alakvartiili on pituus, jota pienempiä arvoja on 25 %.

Piirretään kuvaan vaakasuora  $y = 0,25$  ja katsotaan, missä kohdassa suora leikkaa summafrequenssijakauman kuvaajan.

Alakvartiili on  $Q_1 \approx 166$  (cm).

b) Mediaani on pituus, jota pienempiä arvoja on 50 %.

Piirretään kuvaan vaakasuora  $y = 0,50$  ja katsotaan, missä kohdassa suora leikkaa summafrequenssijakauman kuvaajan.

Mediaani on  $M_d \approx 172$  (cm).

c) Pituus, jota pidempiä oli 25 % opiskelijoista on pituus, jota lyhyempiä on  $100\% - 25\% = 75\%$  opiskelijoista. Tämä pituus on pituuden yläkvartiili.

Piirretään kuvaan vaakasuora  $y = 0,75$  ja katsotaan, missä kohdassa suora leikkaa summafrequenssijakauman kuvaajan.

Yläkvartiili on  $Q_3 \approx 177$  (cm).

8.

a)

$X =$  "nuoren naisen pituus (cm)"

$$X \sim N(167, 5)$$

$$\mu = 167 \text{ ja } \sigma = 5$$

Arvon 175 (cm) normitettu arvo on

$$\frac{175 - \mu}{\sigma} = \frac{175 - 167}{5} = \frac{8}{5} = 1,6$$

$Y =$  "nuoren miehen pituus (cm)"

$$Y \sim N(181, 6)$$

$$\mu = 181 \text{ ja } \sigma = 6$$

Arvon 175 (cm) normitettu arvo on

$$\frac{175 - \mu}{\sigma} = \frac{175 - 181}{6} = \frac{-6}{6} = -1$$

b) 175 cm:n pituinen nainen on pidempi kuin naiset keskimäärin.

- normitettu arvo 1,6 on positiivinen

175 cm:n pituinen mies on lyhyempi kuin miehet keskimäärin

- normitettu arvo  $-1$  on negatiivinen

175 cm:n pituinen nainen on naisten pituusjakaumassa poikkeavampi arvo kuin samanpituinen mies miesten jakaumassa.

- normitettujen arvojen itseisarvot ovat  $|1,6| = 1,6$  ja  $|-1| = 1$
- $1,6 > 1$

Siis 175 cm pituinen nainen on suhteellisesti pidempi kuin samanpituinen mies.

9.

a) Arvo  $X = 22$  on yhtä kaukana odotusarvosta 20 kuin arvo  $X = 18$ .

Normaalijakauman symmetrisyyden perusteella

$$P(X > 22) = P(X < 18) = 0,02.$$


$$22 - 20 = 2$$

$$20 - 18 = 2$$

b) Komplementtisäännön perusteella

$$P(X \geq 18) = 1 - P(X < 18) = 1 - 0,02 = 0,98.$$

c) a-kohdan perusteella  $P(X > 22) = P(X \geq 22) = 0,02$  joten komplementtisäännön mukaan

$$P(X < 22) = 1 - 0,02 = 0,98.$$

Puolet arvoista (50 %) on odotusarvoa 20 pienempiä, joten

$$P(20 \leq X < 22) = 0,98 - 0,50 = 0,48.$$

Siis

$$P(19 < X < 22) = 0,34 + 0,48 = 0,82.$$

## 10.

a) Kertymätodennäköisyys luetaan kertymätaulukosta.

$$P(X < 5) = F(5) \approx 0,20$$

b) Kysytty todennäköisyys saadaan kertymätodennäköisyyden komplementtina.

$$\begin{aligned} P(X \geq 8) &= 1 - P(X < 8) \\ &= 1 - F(8) \\ &\approx 1 - 0,95 \\ &= 0,05 \end{aligned}$$

c) Arvoa  $X = 7$  ei löydy taulukosta. Huomataan, että arvo  $X = 7$  on yhtä kaukana jakauman odotusarvosta  $\mu = 6$  kuin arvo  $X = 5$ .  
Normaalijakauman symmetrisyyden perusteella

$$P(X > 7) = P(X < 5) = F(5) \approx 0,20.$$

d) c-kohdan ja komplementtisäännön perusteella,

$$P(X \leq 7) \approx 1 - 0,20 = 0,80.$$

Puolet arvoista (50 %) on odotusarvoa  $\mu = 6$  suurempia, joten

$$P(6 < X \leq 7) \approx 0,80 - 0,50 = 0,30.$$

## 11.


Kannatusprosentin luottamusväli on 13,5 % – 17,7 %.

a) Otoksesta laskettu prosenttiluku on luottamusvälin keskipiste.

$$\frac{17,7 + 13,5}{2} = \frac{31,2}{2} = 15,6$$

Otoksesta laskettu kannatusprosentti oli 15,6 %.

b) Virhemarginaali on  $17,7 - 15,6 = 2,1$  (%-yksikköä).


$$15,6 - 13,5 = 2,1 \text{ (%-yks.)}$$

Virhemarginaali on tapana ilmaista plus-miinus-merkillä muodossa  $\pm 2,1$  % -yksikköä.

---

*Huomautus:* Luottamusväli ulottuu virhemarginaalin (2,1 %-yksikköä) verran otoksesta lasketusta tunnusluvusta (15,6 %) molempiin suuntiin.

Luottamusväli voidaan siis ilmaista muodossa  $15,6 \% \pm 2,1 \%$ .

c) Riskitaso on 5 % eli luottamustaso on  $100 \% - 5 \% = 95 \%$ .

Suhteellisen osuuden 95 %:n virhemarginaalin laskukaava on:

$$1,96 \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

Laskukaava löytyy taulukkokirjasta.

Virhemarginaalin suuruuteen vaikuttavat:

- otoksesta laskettu prosenttiosuus  $\hat{p}$
- otoksen koko  $n$
- otoskoko  $n$  on nimittäjässä, joten mitä suurempi on otoskoko, sitä pienempi on virhemarginaali.

Valittu riskitaso (tai luottamustaso) vaikuttaa virhemarginaalin kaavassa olevaan **kertoimeen** eli kriittiseen arvoon.

Kertoimet löytyvät taulukkokirjasta.

- 5 %:n riskitasolla eli 95 %:n luottamustasolla kerroin on **1,96**
- 1 %:n riskitasolla 99 %:n luottamustasolla kerroin on **2,58**
- 0,1 %:n riskitasolla 99,9 %:n luottamustasolla kerroin on **3,29**

Siis mitä pienempi riskitaso (eli mitä suurempi luottamustaso), sitä suurempi kerroin ja sitä suurempi virhemarginaali.

Valittu riskitaso vaikuttaa siis virhemarginaalin suuruuteen.



## 12.

a) Yhteensä 60 kokeessa testi toimi onnistuneesti 48 kokeessa. Otoksen perusteella estimaatti eli arvio onnistumistodennäköisyydelle on

$$\frac{48}{60} = 0,8 = 80 \%$$

b) Lasketaan virhemarginaali. Suhteellisen osuuden 95 %:n virhemarginaalin laskukaava on:

$$1,96 \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

Laskukaava löytyy taulukkokirjasta.

$$n = 60$$

$$\hat{p} = 0,80$$

$$1 - \hat{p} = 1 - 0,80 = 0,20$$

Virhemarginaaliksi saadaan

$$1,96 \cdot \sqrt{\frac{0,80 \cdot 0,20}{60}} = 0,1012 \dots \approx 10 \text{ (\% - yks.)}$$

Onnistumistodennäköisyyden 95 %:n luottamusväliksi saadaan 80 % ± 10 %-yksikköä eli 70 % – 90 %.

$$\begin{aligned} 80 - 10 &= 70 \\ 80 + 10 &= 90 \end{aligned}$$

# 6 Kertaus

## 6.2 Matemaattisia malleja

13.

Joukkue B voittaa, jos joukkue saa vähintään 2 pistettä eli jos kolmesta vapaaheitosta vähintään kaksi onnistuu.

**Tapa 1:**

Mallinnetaan tilannetta binomijakaumalla.

$X =$  ”onnistuneiden heittojen lukumäärä kolmen heiton sarjassa”

$$n = 3$$

$$p = P(\text{heitto onnistuu}) = 0,75$$

$$X \sim \text{Bin}(3; 0,75)$$

Lasketaan tapahtuman  $X \geq 2$  todennäköisyys.

$$P(\text{B voittaa}) = P(X \geq 2) = 0,84375 \approx 0,84$$

Laskinohjelmiston binomijakaumatoimintoon syötetään:  $n = 3$  ja  $p = 0,75$ , alaraja 2 ja yläraja 3.

## Tapa 2:

Lasketaan voiton todennäköisyys yhteen- ja kertolaskusääntöjen avulla. Joukkue B voittaa, jos kaikki kolme heittoa onnistuu tai jos kaksi heitoista onnistuu ja yksi ei onnistu. Yksi epäonnistunut heitto voi sijaita kolmen heiton sarjassa kolmessa eri kohdassa.

$$P(\text{heitto onnistuu}) = 0,75$$

$$P(\text{heitto ei onnistu}) = 1 - 0,75 = 0,25$$

$$\begin{aligned} P(\text{B voittaa}) &= P(3 \text{ heittoa onnistuu}) + P(2 \text{ heittoa onnistuu ja yksi ei}) \\ &= 0,75 \cdot 0,75 \cdot 0,75 + 0,75 \cdot 0,75 \cdot 0,25 + 0,75 \cdot 0,25 \\ &\quad \cdot 0,75 + 0,25 \cdot 0,75 \cdot 0,75 \\ &= 0,84375 \approx 0,84 \end{aligned}$$

## 14.

Satunnaismuuttuja  $X =$  ”onnistuneiden heittojen lukumäärä 8 heiton sarjassa”

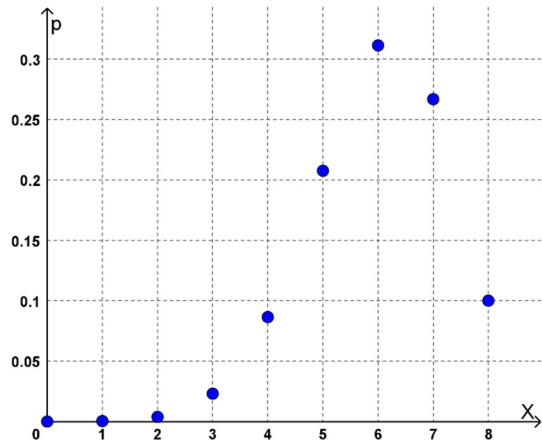
$$n = 8$$

$$p = P(\text{heitto onnistuu}) = 0,75$$

$$X \sim \text{Bin}(8; 0,75)$$

a) Muuttujan  $X$  mahdolliset arvot ovat 0–8. Määritetään pistetodennäköisyydet sopivan ohjelmiston avulla ja kootaan ne taulukkoon. Havainnollistetaan jakaumaa pistekuvaajalla, jossa muuttujan arvot ovat vaaka-akselilla ja niitä vastaavat pistetodennäköisyydet pystyakselilla.

$k$	$P(X = k)$
0	0,0000152... $\approx$ 0,000015
1	0,000366... $\approx$ 0,00037
2	0,00384... $\approx$ 0,0038
3	0,0230... $\approx$ 0,023
4	0,0865... $\approx$ 0,087
5	0,207... $\approx$ 0,21
6	0,311... $\approx$ 0,31
7	0,266... $\approx$ 0,27
8	0,100... $\approx$ 0,10



b) Lasketaan tapahtuman  $X \geq 6$  todennäköisyys.

$$P(X \geq 6) = 0,678 \dots \approx 0,68$$

Laskinohjelmiston binomijakaumatoimintoon syötetään:  $n = 8$  ja  $p = 0,75$ , alaraja 6 ja yläraja 8.

c) Binomijakaumaa noudattavan satunnaismuuttujan  $X$

- odotusarvo on  $E(X) = n \cdot p = 8 \cdot 0,75 = 6$
- keskihajonta on

$$D(X) = \sqrt{n \cdot p \cdot (1 - p)} = \sqrt{8 \cdot 0,75 \cdot 0,25} = 1,224 \dots \approx 1,2$$

Odotusarvo ilmaisee, että jos 8 heiton sarjoja toistetaan useita kertoja, onnistuneiden heittojen lukumäärä on keskimäärin 6 onnistunutta heitto/sarja.

Huomaa: Odotusarvo ei ole sama asia kuin todennäköisin arvo. Tässä tapauksessa kuitenkin suurin pistetodennäköisyys on arvolla  $X = 6$ , joten onnistuneita heittoja on 8 heiton sarjassa todennäköisimmin 6.

Keskihajonta ilmaisee onnistuneiden heittojen lukumäärässä esiintyvää vaihtelua, kun 8 heiton sarjaa toistetaan useita kertoja.

## 15.

a) Satunnaismuuttuja  $X =$  ”kymppiin heitettyjen tikkojen lukumäärä 5 heiton sarjassa”

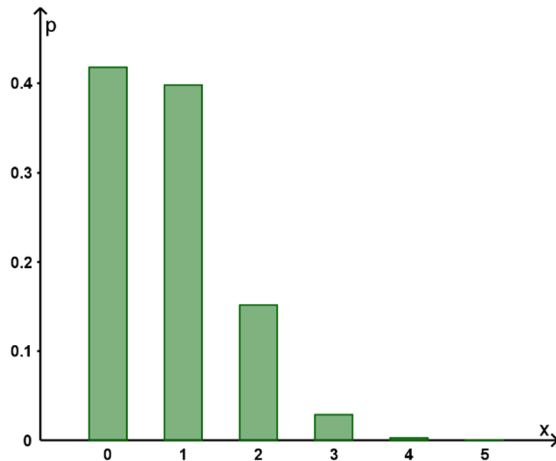
Toistojen lukumäärä on  $n = 5$ .

Onnistumisen todennäköisyys yhdessä heitossa on  $p = P(\text{heitto onnistuu}) = 0,16$ .

Muuttuja  $X$  noudattaa binomijakaumaa:  $X \sim \text{Bin}(5; 0,16)$ .

b) Muuttujan  $X$  mahdolliset arvot ovat 0–5. Määritetään pistetodennäköisyydet sopivan ohjelmiston avulla ja kootaan ne taulukkoon. Havainnollistetaan jakaumaa esimerkiksi pylväskuvaajalla.

$k$	$P(X = k)$
0	0,4182... $\approx 0,42$
1	0,3982... $\approx 0,40$
2	0,1517... $\approx 0,15$
3	0,0289... $\approx 0,029$
4	0,00275... $\approx 0,0028$
5	0,000104... $\approx 0,00010$



c) Lasketaan tapahtuman ” $X = 2$  tai  $X = 3$ ” eli tapahtuman  $2 \leq X \leq 3$  todennäköisyys.

$$P(2 \leq X \leq 3) = 0,1806 \dots \approx 0,18$$

Laskinohjelmiston binomijakaumatoimintoon syötetään:  $n = 5$  ja  $p = 0,16$ , alaraja 2 ja yläaraja 3.

d) Lasketaan tapahtuman  $X > 3$  eli tapahtuman  $X \geq 4$  todennäköisyys.

$$P(X \geq 4) = 0,00285 \dots \approx 0,0029$$

Laskinohjelmiston  
binomijakaumatoimintoon  
syötetään:  $n = 5$  ja  $p = 0,16$ ,  
alaraja 4 ja yläraja 5.

e) b-kohdan jakaumataulukosta nähdään, että suurin pistetodennäköisyys on arvolla  $X = 0$ . Todennäköisin kymppien lukumäärä viiden heiton sarjassa on siis nolla (eli ei yhtään kymppiä).

## 16.

a) Satunnaismuuttuja  $X =$  ”sadepäivien lukumäärä viikon aikana”

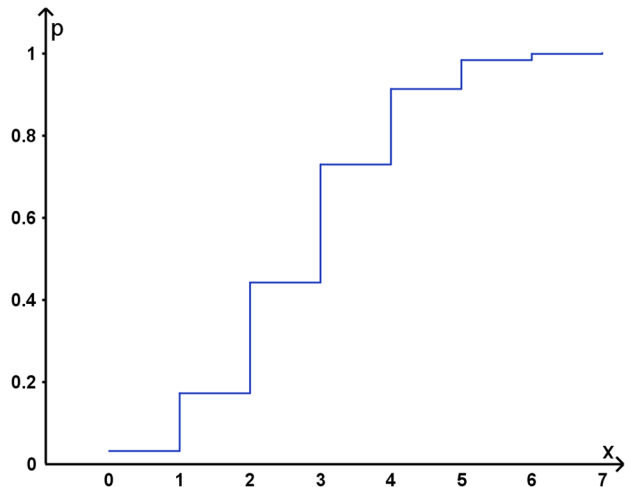
Viikossa on 7 päivää, eli  $n = 7$ .

Sateen todennäköisyys yhtenä päivänä on  $p = 0,39$ .

Muuttuja  $X$  noudattaa binomijakaumaa:  $X \sim \text{Bin}(7; 0,39)$ .

Muuttujan  $X$  mahdolliset arvot ovat 0–7. Määritetään kertymätodennäköisyydet sopivan ohjelmiston avulla ja kootaan ne taulukkoon. Havainnollistetaan jakaumaa porraskuvaajalla.

$k$	$P(X \leq k)$
0	0,03142... $\approx 0,03$
1	0,1720... $\approx 0,17$
2	0,4418... $\approx 0,44$
3	0,7293... $\approx 0,73$
4	0,9131... $\approx 0,91$
5	0,9836... $\approx 0,98$
6	0,9986... $\approx 0,999$
7	1



b) Binomijakaumaa noudattavan satunnaismuuttujan  $X$

- odotusarvo on  $\mu = E(X) = n \cdot p = 7 \cdot 0,39 = 2,73 \approx 2,7$
- keskihajonta on

$$\begin{aligned}\sigma &= D(X) = \sqrt{n \cdot p \cdot (1 - p)} \\ &= \sqrt{7 \cdot 0,39 \cdot 0,61} = 1,290 \dots \approx 1,3\end{aligned}$$



c) Lasketaan muuttujan arvon  $X = 5$  normitettu arvo.

$$\frac{5 - \mu}{\sigma} = \frac{5 - 2,73}{1,290 \dots} = 1,759 \dots$$

Normitetun arvon itseisarvo on  $|1,759 \dots| = 1,759 \dots < 2$ , joten arvo  $X = 5$  ei poikkea merkitsevästi odotusarvosta.

## 17.

a) Havaintoyksiköitä (rivejä) on yhteensä 243. Siis, aineiston koko on  $N = 243$ .

Muuttujia (sarakkeita) ovat: sukupuoli, ikä, pituus, opintotuki (€/kk), opiskelu (h/päivä), kokonaistyytyväisyys (1–5) sekä vastausaika (s). Muuttujia on yhteensä 7.

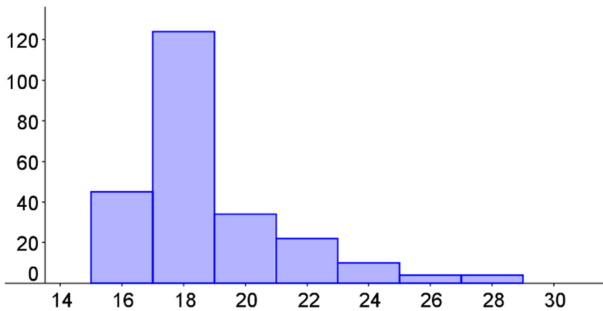
b) Määritetään ikä-sarakkeen pienin arvo tilastotoiminnoilla.

Min = 15. Nuorin vastaaja oli siis 15-vuotias.

c) Tehdään luokittelu esimerkiksi matematiikkaohjelmiston taulukkosovelluksessa. Kopioidaan iät laskentataulukkoon ja tehdään yhden muuttujan analyysi. Asetetaan luokat käsin: ensimmäisen luokan todellinen alaraja on 15 ja luokkavälin leveys 2.

Muodostetaan frekvenssijakaumat esimerkiksi laskentataulukossa. Pyöristetään frekvenssit kokonaisluvun tarkkuuteen.

Ikä (vuotta)	$f$	$sf$	$f\%$	$sf\%$
15–16	45	45	19	19
17–18	124	169	51	70
19–20	34	203	14	84
21–22	22	225	9	93
23–24	10	235	4	97
25–26	4	239	2	98
27–29	4	243	2	100



d) Luokan 17–18 frekvenssi ( $f$  ja  $f\%$ ) on suurin, eli suurin osa vastaajista on 17–18-vuotiaita.

e) Alle 19-vuotiaat vastaajat kuuluvat kahteen alimpaan luokkaan. Näiden luokkien frekvenssien summa on

$$45 + 124 = 169$$

eli prosentteina

$$\frac{169}{243} = 0,6954 \dots \approx 69,5 \%$$

Vastaajista alle 19-vuotiaita oli 69,5 %.

f) *Huomautus:* Tehtävänannossa ilmaus ”yli 22-vuotias” on tarkoitus tulkita: ”23-vuotta täyttänyt”.

23-vuotta täyttäneet kuuluvat kolmeen ylimpään luokkaan. Näiden luokkien frekvenssien summa on

$$10 + 4 + 4 = 18.$$

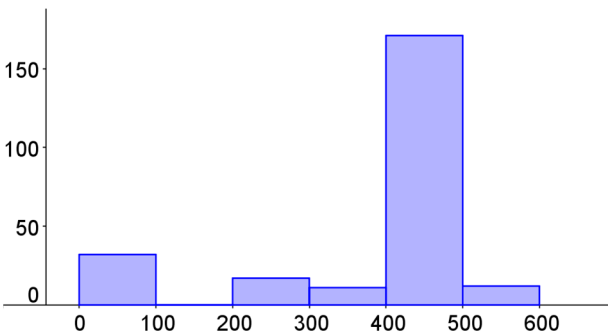
Vastaajista yli 22-vuotiaita (vähintään 23-vuotiaita) oli 18 vastaajaa.

## 18.

a) Tehdään luokittelu esimerkiksi matematiikkaohjelmiston taulukkosovelluksessa. Kopioidaan opintotuet laskentataulukkoon ja tehdään yhden muuttujan analyysi.

Esimerkiksi tilastojen yhteenvetotaulukosta nähdään, että opintotuen pienin arvo on 0 (€/kk).

Asetetaan luokat käsin: valitaan ensimmäisen luokan todelliseksi alarajaksi 0 ja luokkavälin leveydeksi 100.



Opintotuki (€/kk)	$f$
0-99	32
100-199	0
200-299	17
300-399	11
400-499	171
500-599	12

Tunnusluvut nähdään tilastojen yhteenvetotaulukosta. Vastaajien ( $N = 243$ ) saaman opintotuen määrän (€/kk)

- minimi on 0 €/kk ja maksimi 550 €/kk,
- mediaani on 460 €/kk eli puolet vastaajista sai opintotukea vähintään tämän määrän ja puolet korkeintaan tämän määrän,
- kvartiiliväli on  $[Q_1, Q_3] = [400 \text{ €}, 470 \text{ €}]$  eli puolet vastaajista sai opintotukea 400–470 €/kk,
- moodiluokka on 400–499 €/kk eli suurimmalla osalla vastaajista opintotuen määrä on tällä välillä. Näitä opiskelijoita oli yhteensä 171 eli prosentteina

$$\frac{171}{243} = 0,703 \dots \approx 70 \text{ \%}.$$

Lisäksi on mielekästä ilmoittaa, kuinka suuri osuus vastaajista ei saanut opintotukea, eli tuen määrä oli 0 €/kk. Näitä vastaajia oli 32 (lukumäärä nähdään esimerkiksi jakauman pylväskaaviosta) eli prosentteina

$$\frac{32}{243} = 0,131 \dots \approx 13 \text{ \%}.$$

b) i) Määritetään tilastotoiminnoilla vastausaikojen minimi ja maksimi.

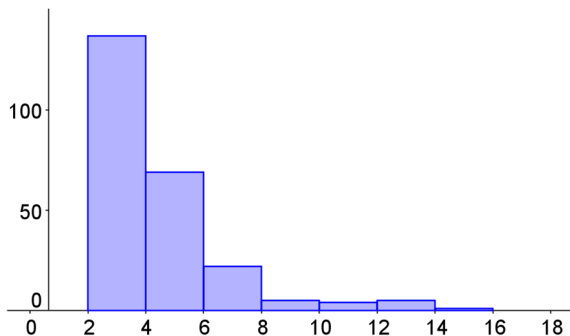
- $\min = 2$  ja  $\max = 14,38 \approx 14$

Vastausaika vaihteli siis välillä 2 min – 14 min

ii) Määritetään tilastotoiminnoilla vastausaikojen keskiarvo ja mediaani.

- $\bar{x} = 4,42 \dots \approx 4,4$  (min)
- $Md = 3,72 \approx 3,7$  (min)

Havainnollistetaan vastausajan jakaumaa graafisesti histogrammilla. Havaitaan, että jakauma ei ole symmetrinen vaan oikealle vino. Jakauman ”häntä” eli pieni määrä suuria arvoja siirtää keskiarvoa (4,4 min) mediaanista (3,7 min) katsoen oikealle.



Keskiarvo ja mediaani ovat yhtä suuria vain symmetrisissä jakaumissa.

iii) Keskiarvon käyttöä tunnuslukuna tulisi välttää, jos jakauma on vino. Mediaani soveltuu tällöin keskiarvoa paremmin kuvaamaan muuttujan keskimääräistä arvoa. Tässä tapauksessa keskiarvo ja mediaani ovat kuitenkin pyydetyllä pyöristystarkkuudella yhtä suuria ( $\bar{x} \approx 4$  ja  $Md \approx 4$ ).

Vastaajat käyttivät kyselyyn keskimäärin 4 minuuttia.

## 19.

a) Määritetään sukupuoli-sarakkeesta arvon 1 (mies) lukumäärä.

Miehiä oli yhteensä 114.

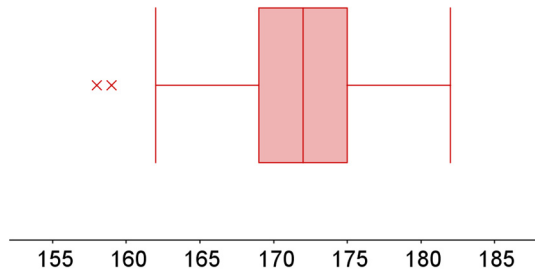
Naisia oli tällöin  $243 - 114 = 129$ .

Miehet (sukupuoli = 1) ovat havaintomatriisissa ensimmäisenä. Viimeinen havaintoyksikkö, jonka sukupuoli = 1 on 114. havaintoyksikkö. Siis, vastaajista 114 oli miehiä.

b) Määritetään tyttöjen pituuksien tunnusluvut sopivalla ohjelmistolla tilastotoiminnoilla. Tyttöjen pituuksien

- min = 158 ja max = 182 eli pituuksien vaihteluväli on 158 cm – 182 cm
- keskiarvo on  $\bar{x} = 171,70 \dots \approx 171,7$  (cm)
- keskihajonta on  $\sigma = 4,77 \dots \approx 4,8$  (cm)

Havainnollistetaan tyttöjen pituusjakaumaa laatikkokuvailla. Poikkeavat arvot on merkitty kuvioon rastilla (X). Kuvaajasta nähdään, että kaksi pienintä arvoa poikkeaa merkittävästi keskiarvosta.



Kaksi pienintä arvoa ovat 158 cm ja 159 cm. Lasketaan näiden arvojen normitetut arvot.

$$\frac{158 - \bar{x}}{\sigma} = \frac{158 - 171,70 \dots}{4,77 \dots} = -2,867 \dots$$

$$\frac{159 - \bar{x}}{\sigma} = \frac{159 - 171,70 \dots}{4,77 \dots} = -2,658 \dots$$

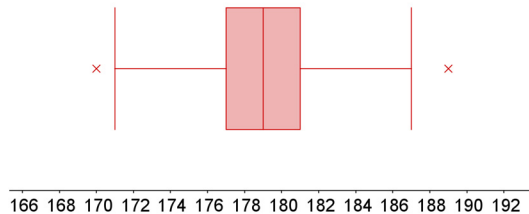
Molempien arvojen normitettujen arvojen itseisarvo on suurempi kuin 2, joten pituudet 158 cm ja 159 cm poikkeavat merkitsevästi tyttöjen pituuksien keskiarvosta.

Tyttöjen pituuksissa on kaksi keskiarvoa merkitsevästi pienempää arvoa.

c) Määritetään poikien pituuksien tunnusluvut sopivalla ohjelmistolla tilastotoiminnoilla. Poikien pituuksien

- min = 170 ja max = 189 eli pituuksien vaihteluväli on 170 cm – 189 cm
- keskiarvo on  $\bar{x} = 179,27 \dots \approx 179,3$  (cm)
- keskihajonta on  $\sigma = 3,71 \dots \approx 3,7$  (cm)

Havainnollistetaan poikien pituusjakaumaa laatikkokuviolla. Poikkeavat arvot on merkitty kuvioon rastilla (X). Kuvaajasta nähdään, että pienin ja suurin arvo poikkeavat merkitsevästi keskiarvosta.





Pienin arvo on 170 cm ja suurin arvo 189 cm. Lasketaan näiden arvojen normitetut arvot.

$$\frac{170 - \bar{x}}{\sigma} = \frac{170 - 179,27 \dots}{3,71 \dots} = -2,493 \dots$$

$$\frac{189 - \bar{x}}{\sigma} = \frac{189 - 179,27 \dots}{3,71 \dots} = 2,615 \dots$$

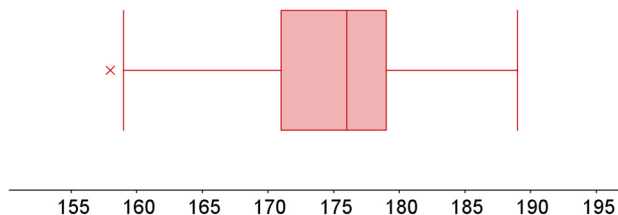
Molempien arvojen normitetujen arvojen itseisarvo on suurempi kuin 2, joten pituudet 170 cm ja 189 cm poikkeavat merkittävästi poikien pituuksien keskiarvosta.

Poikien pituuksissa on yksi keskiarvoa merkittävästi pienempi arvo ja yksi keskiarvoa merkittävästi suurempi arvo.

d) Määritetään kaikkien vastaajien pituuksien tunnusluvut sopivalla ohjelmistolla tilastotoiminnoilla. Vastaajien pituuksien

- min = 158 ja max = 189 eli pituuksien vaihteluväli on 158 cm - 189 cm
- keskiarvo on  $\bar{x} = 175,25 \dots \approx 175,3$  (cm)
- keskihajonta on  $\sigma = 5,73 \dots \approx 5,7$  (cm)

Havainnollistetaan vastaajien pituusjakaumaa laatikkokuvioilla. Poikkeavat arvot on merkitty kuvioon rastilla (X). Kuvaajasta nähdään, että pienin arvo poikkeaa merkittävästi keskiarvosta.



Pienin arvo on 158 cm. Lasketaan normitettu arvo.

$$\frac{158 - \bar{x}}{\sigma} = \frac{158 - 175,25 \dots}{5,73 \dots} = -3,009 \dots$$

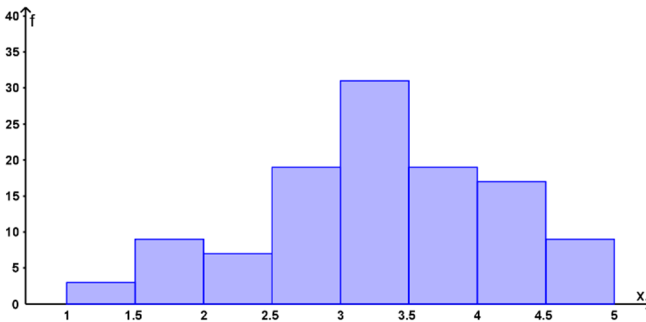
Normitetun arvon itseisarvo on suurempi kuin 2, joten pituus 158 cm poikkeaa merkitsevästi vastaajien keskiarvosta.

Vastaajien pituuksissa on yksi keskiarvoa merkitsevästi pienempi arvo.

## 20.

Määritetään poikien kokonaisuustyytyväisyys-muuttujan tunnusluvut sopivalla ohjelmistolla tilastotoiminnoilla. Poikien tyytyväisyyden

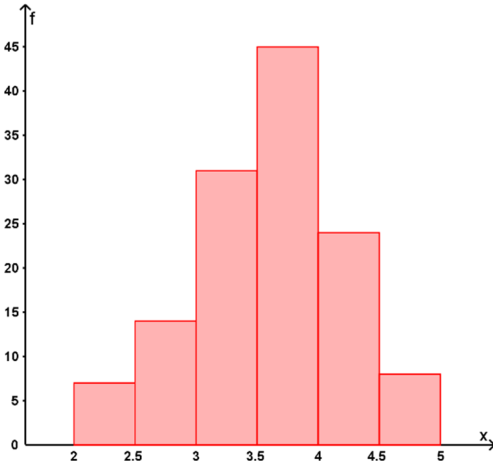
- $N = 114$
- $\min = 1,1$  ja  $\max = 5$  eli kokonaistyytyväisyys vaihtelee välillä 1,1–5
- vaihteluvälin pituus on  $5 - 1,1 = 3,9$
- keskiarvo on  $\bar{x} = 3,236 \dots \approx 3,24$
- keskihajonta on  $\sigma = 0,877 \dots \approx 0,88$



Havainnollistetaan poikien tyytyväisyyttä histogrammilla. Ensimmäisen luokan todelliseksi alarajaksi on valittu arvo 1 ja luokkavälin leveydeksi 0,5.

Määritetään tyttöjen kokonaisuustyytyväisyys-muuttujan tunnusluvut sopivalla ohjelmistolla tilastotoiminnoilla. Tyttöjen tyytyväisyyden

- $N = 129$
- $\min = 2$  ja  $\max = 5$  eli kokonaistyytyväisyys vaihtelee välillä 2–5
- vaihteluvälin pituus on  $5 - 2 = 3$
- keskiarvo on  $\bar{x} = 3,548 \dots \approx 3,55$
- keskihajonta on  $\sigma = 0,614 \dots \approx 0,61$



Havainnollistetaan tyttöjen tyytyväisyyttä histogrammilla.

Ensimmäisen luokan todelliseksi alarajaksi on valittu pienin arvo 2 ja luokkavälin leveydeksi 0,5.

Histogrammeihin on valittu sama luokittelu vertailun helpottamiseksi.

Sekä histogrammeista että tunnusluvuista nähdään, että poikien kokonaistyytyväisyydessä esiintyy suurempaa vaihtelua kuin tyttöjen tyytyväisyydessä: vaihteluvälin pituus ( $3,9 > 3$ ) sekä keskihajonta ( $0,88 > 0,61$ ) ovat pojilla suurempia kuin tytöillä.

Poikien tyytyväisyyden keskiarvo on hieman tyttöjen keskiarvo pienempi ( $3,24 < 3,55$ ).

Voidaan siis todeta, että tytöt ovat keskimäärin poikia tyytyväisempiä oppilaitoksen palveluihin.

## 21.

Satunnaismuuttuja  $X =$  ”korjausta vaativan vian ilmenemiseen kuluva aika (a)”

$$X \sim N(3; 1,2)$$

a) Lasketaan tapahtuman  $X > 4$  todennäköisyys.

$$P(X > 4) = 0,2023 \dots \approx 20 \%$$

Laskinohjelmiston normaali-jakaumatoimintoon syötetään: alaraja 4, yläraja  $\infty$ ,  $\mu = 3$  ja  $\sigma = 1,2$ .

Todennäköisyys ilmaisee myös yli neljä vuotta ilman korjausta toimivien pakastimien prosenttiosuuden. Näitä pakastimia on siis 20 %.

b) Merkitään takuu-aikaa kirjaimella  $a$  (vuotta). Pakastimelle tehdään takuukorjaus, jos  $X \leq a$ . Takuu-aika tehdään enintään 5 % pakastimista, kun

$$P(X \leq a) \leq 0,05$$
$$a \leq 1,026 \dots \approx 1,0$$

Laskinohjelmiston käänteiseen normaalijakaumatoimintoon syötetään: pinta-ala 0,05,  $\mu = 3$  ja  $\sigma = 1,2$ .

Pakastimille voidaan siis myöntää vuoden takuu.

c) Merkitään keskihajontaa kirjaimella  $\sigma$ .

$$X \sim N(3, \sigma)$$

$P(X > 2) \geq 0,9$ , joten kertymätodennäköisyys on

$$P(X \leq 2) < 1 - 0,9 = 0,1.$$

Määritetään kertymätodennäköisyyttä 0,1 vastaava normitettu arvo.  
Normitetuksi arvoksi saadaan

$$Z = -1,281 \dots$$

Laskinohjelmiston käänteiseen normaalijakaumatoimintoon syötetään: pinta-ala 0,1,  $\mu = 0$  ja  $\sigma = 1$ .

Kertymätodennäköisyys  $P(X \leq 2)$  on pienempi kuin 0,1, joten arvon 2 normitettu arvo on pienempi kuin  $-1,281 \dots$

Muodostetaan epäyhtälö ja ratkaistaan  $\sigma$ .

$$\frac{2 - 3}{\sigma} < -1,281 \dots$$

$$0 < \sigma < 0,7803 \dots$$

Suurin  $\sigma$ , joka kahden desimaalin tarkkuudella toteuttaa ehdon  $0 < \sigma < 0,7803 \dots$  on 0,78.

22.

Satunnaismuuttuja  $X =$  "kirkasvalolampun kestoikä (h)"

$$X \sim N(3000, 400)$$

a) Lasketaan tapahtuman  $X < 2300$  todennäköisyys.

$$P(X < 2300) = 0,0400 \dots \approx 4,0 \%$$

Laskinohjelmiston normaali-jakaumatoimintoon syötetään: yläraja 2300,  $\mu = 3000$  ja  $\sigma = 400$ .

Todennäköisyys ilmaisee myös niiden lamppujen prosenttiosuuden, joiden kestoikä on alle 2300 tuntia. Näitä lampuja on 4,0 % eli tehdas vaihtaa 4,0 % lampuista.

b) Merkitään takuuaikaa kirjaimella  $a$  (tuntia). Tehdas vaihtaa lampun, jos  $X \leq a$ . Tehdas vaihtaa 2 % lampuista, kun

$$P(X \leq a) = 0,02$$

$$a = 2178,50 \dots \approx 2200 \text{ (h)}$$

Laskinohjelmiston käänteiseen normaalijakaumatoimintoon syötetään: pinta-ala 0,02,  $\mu = 3000$  ja  $\sigma = 400$ .

Kahden prosentin vaihtomäärää vastaava takuuaika on siis 2200 tuntia.

## 23.

Satunnaismuuttuja  $X = \text{"joulutortun paino (g)"}$

$$X \sim N(75, 4)$$

Määritetään joulutorttujen painon mediaani.

**Tapa 1:** Merkitään mediaania kirjaimella  $a$  (g). Mediaani on muuttujan arvo, joita pienempiä arvoja on 50 % eli

$$P(X < a) = 0,50$$

$$a = 75 \text{ (g)}$$

Laskinohjelmiston käänteiseen normaalijakaumatoimintoon syötetään: pinta-ala 0,50,  $\mu = 75$  ja  $\sigma = 4$ .

Mediaani on siis  $Md = 75$  g. Puolet joulutorttuista painavat alle 75 g ja puolet yli 75 g.

**Tapa 2:** Mediaani voidaan päätellä myös normaalijakauman symmetrian perusteella. Joulutorttujen odotusarvo on  $\mu = 75$ . Koska normaalijakauma on symmetrinen, puolet arvoista on odotusarvoa pienempiä. Siis, mediaani on  $Md = 75$  g.

Määritetään joulutorttujen painon kvartiilit.

Merkitään alakvartiilia kirjaimella  $Q_1$  (g). Alakvartiili on muuttujan arvo, joita pienempiä arvoja on 25 % eli

$$P(X < Q_1) = 0,25$$

$$Q_1 = 72,3020 \dots$$

$$Q_1 \approx 72 \text{ (g)}$$

Laskinohjelmiston käänteiseen normaalijakaumatoimintoon syötetään: pinta-ala 0,25,  $\mu = 75$  ja  $\sigma = 4$ .



Kevyin neljännes (25 %) joulutortuista painaa alle 72 g.

Merkitään yläkvartiilia kirjaimella  $Q_3$  (g). Yläkvartiili on muuttujan arvo, joita pienempiä arvoja on 75 % eli

$$P(X < Q_3) = 0,75$$

$$Q_3 = 77,6979 \dots$$

$$Q_3 \approx 78 \text{ (g)}$$

Laskinohjelmiston käänteiseen normaalijakaumatoimintoon syötetään: pinta-ala 0,75,  $\mu = 75$  ja  $\sigma = 4$ .

Painavin neljännes (25 %) joulutortuista painaa yli 78 g.

24.

Satunnaismuuttuja  $X =$  "yhden laskiaispullan paino (g)"

$$X \sim N(50, 9)$$

a) Lasketaan tapahtuman  $40 \leq X \leq 60$  todennäköisyys.

$$P(40 \leq X \leq 60) = 0,7334 \dots \approx 0,73 = 73 \%$$

Laskinohjelmiston normaali-  
jakaumatoimintoon syötetään:  
alaraja 40, yläraja 60,  $\mu = 50$  ja  $\sigma = 9$

b) Satunnaismuuttuja  $Y =$  "neljän pullan yhteispaino (g)"

$Y \sim N(\mu, \sigma)$ , jossa

- odotusarvo on  $\mu = 4 \cdot 50 = 200$  (g)
- keskihajonta on  $\sigma = 9 \cdot \sqrt{4} = 18$  (g)

Lasketaan tapahtuman  $180 \leq Y \leq 220$  todennäköisyys.

$$P(180 \leq Y \leq 220) = 0,7334 \dots \approx 0,73 = 73 \%$$

Laskinohjelmiston normaali-  
jakaumatoimintoon syötetään: alaraja  
180, yläraja 220,  $\mu = 200$  ja  $\sigma = 18$

c) Satunnaismuuttuja  $\bar{X}$  = "neljän pullan keskipaino (g)"

$\bar{X} \sim N(\mu, \sigma)$ , missä

- odotusarvo on  $\mu = 50$  (g)
- keskihajonta on  $\sigma = \frac{9}{\sqrt{4}} = 4,5$  (g)

Lasketaan tapahtuman  $45 \leq \bar{X} \leq 55$  todennäköisyys.

$$P(45 \leq \bar{X} \leq 55) = 0,7334 \dots \approx 0,73 = 73 \%$$

Laskinohjelmiston normaali-  
jakaumatoimintoon syötetään: alaraja  
45, yläraja 55,  $\mu = 50$  ja  $\sigma = 4,5$

d) Lasketaan todennäköisyys, että yksi pulla painaa vähintään 50 g, eli  $X \geq 50$ . Pullan painon odotusarvo on 50 g, joten

$$P(X \geq 50) = 0,50.$$

Laskinohjelmiston normaali-  
jakaumatoiminnolla  
todennäköisyydeksi voi tulla  
 $P(X \geq 50) = 0,4999 \dots$

Kertolaskusäännön mukaan, neljästä pullasta jokainen painaa vähintään 50 g todennäköisyydellä

$$(0,50)^4 = 0,0625 \approx 6,3 \%$$

Jos käytät laskimesta saatua  
todennäköisyyttä tapahtumalle  $X \geq 50$ ,  
todennäköisyydeksi tulee  
 $(0,4999 \dots)^4 = 0,06249 \dots \approx 6,2 \%$

25.

Otoksessa:

$$\bar{x} = 158 \text{ cm}$$

$$s = 3,9 \text{ cm}$$

$$n = 61$$

Valittu luottamustaso on  $C = 0,95$ .

Perusjoukon keskihajonta ei ole tiedossa, joten kriittisenä arvona käytetään t-jakaumasta katsottua kerrointa.

Määritetään luottamusväli sopivan ohjelmiston avulla.

Luottamusvälin

- alaraja on  $157,001... \approx 157 \text{ (cm)}$
- yläraja on  $158,998... \approx 159 \text{ (cm)}$

Luottamusväli on siis  $157 \text{ cm} - 159 \text{ cm}$ .

Esimerkiksi laskinohjelmistossa luottamusväli-toiminto voi olla ”t-väli”.

Tiedonsyöttömenetelmäksi valitaan ”tilastot”.

Luottamusväli ilmaisee, millä välillä koko perusjoukon (kaikkien suomenhevosorien) säkäkorkeuden **keskiarvo** on valitulla luottamustasolla 95 %. Luottamusvälin perusteella ei voida päätellä, mikä säkäkorkeus vähintään on.

## 26.

a) Perusjoukon keskihajonta ei ole tiedossa, joten kriittisenä arvona käytetään t-jakaumasta katsottua kerrointa.

Määritetään luottamusväli sopivan ohjelmiston avulla.

Valittu luottamustaso on  $C = 0,95$ .

Syötetään ajoajat ohjelmiston taulukkosovellukseen.

Luottamusvälin

- alaraja on  $48,8517... \approx 48,9$  (min)
- yläraja on  $53,3482... \approx 53,3$  (min)

Luottamusväli on siis  $48,9$  min –  $53,3$  min.

Esimerkiksi laskinohjelmistossa luottamusväli-toiminto voi olla ”t-väli”.

Tiedonsyöttömenetelmäksi valitaan ”tiedot”.

Matematiikkaohjelmiston taulukkosovelluksessa määritetään ajoaikojen keskiarvo ja otoskeskihajonta. Tämän jälkeen todennäköisyys-sovelluksessa määritetään keskiarvon T-Estimaatti. Sovellukseen syötetään:

luottamustaso:  $0,95$

keskiarvo:  $51,1$

otoskeskihajonta  $s = 3,1428...$

otoskoko  $N = 10$

b) a-kohdassa määritetty luottamusväli ilmaisee, että ajoaikojen keskiarvo on 95 % varmuudella välillä 48,9 min – 53,3 min.

Liikennöitsijän ilmoittama ajoaika 55 min ei ole luottamusvälillä, joten luottamusvälin perusteella liikennöitsijän ilmoitusta on syytä epäillä.

27.

a) Otoksessa:

$$\bar{x} = 15,0 \%$$

$$s = 3,6 \%$$

$$n = 9$$

Valittu luottamustaso on  $C = 0,95$ .

Perusjoukon keskihajonta ei ole tiedossa, joten kriittisenä arvona käytetään t-jakaumasta katsottua kerrointa.

Määritetään luottamusväli sopivan ohjelmiston avulla.

Luottamusvälin

- alaraja on  $12,232... \approx 12,2$  (%)
- yläraja on  $17,767... \approx 17,8$  (%)

Luottamusväli on siis  $12,2 \%$  –  $17,8 \%$ .

Esimerkiksi laskinohjelmistossa luottamusväli-toiminto voi olla ”t-väli”.

Tiedonsyöttömenetelmäksi valitaan ”tilastot”.

b) Luottamusväli ilmaisee, millä välillä koko perusjoukon (puutavaraerän) kosteuden **keskiarvo** on valitulla luottamustasolla 95 %. Sen perusteella ei voida tehdä päätelmiä yksittäisistä havainnoista, eli yksittäisistä kosteusmittauksista.

Siis ei voida sanoa, että 95 % kosteusmittauksista antaa tuloksen, joka on luottamusvälillä.

## 28.

Aikaisempi markkinaosuus: 14,0 %

Otoksessa:

$$\bar{x} = 14,42 \%$$

$$s = 1,12 \%$$

$$n = 40$$

Perusjoukon keskihajonta ei ole tiedossa, joten kriittisenä arvona käytetään t-jakaumasta katsottua kerrointa.

Määritetään luottamusväli sopivan ohjelmiston avulla.

Valitaan ensin luottamustaso  $C = 0,95$ .

95 % luottamusvälin

- alaraja on 14,061...  $\approx$  14,06 (%)
- yläraja on 14,778...  $\approx$  14,78 (%)

Esimerkiksi laskinohjelmistossa luottamusväli-toiminto voi olla ”t-väli”.

Tiedonsyöttömenetelmäksi valitaan ”tilastot”.

95 % luottamusväli on siis 14,06 % – 14,78 % eli hajuvesimerkin markkinaosuus on 95 % varmuudella tällä välillä.

Valitaan sitten luottamustaso  $C = 0,99$ .

99 % luottamusvälin

- alaraja on 13,940...  $\approx$  13,94 (%)
- yläraja on 14,899...  $\approx$  14,90 (%)

Esimerkiksi matematiikka-ohjelmistossa luottamusväli-toiminto on todennäköisyyssovelluksessa nimellä ”keskiarvon t-estimaatti”.



99 % luottamusväli on siis 13,94 % – 14,90 % eli hajuvesimerkin markkinaosuus on 99 % varmuudella tällä välillä.

Aikaisempi markkinaosuus, 14,0 %, ei sisälly 95 % luottamusväliin, joten tämän perusteella on perusteltua otaksua, että markkinaosuus olisi muuttunut. Ero aiemman markkinaosuuden (14,0 %) ja luottamusvälin alarajan (14,06 %) välillä on kuitenkin pieni.

Sen sijaan 99 % luottamusvälille aikaisempi markkinaosuus sisältyy, joten tällä luottamustasolla ei voida sanoa, että markkinaosuus olisi muuttunut, ja tulkintaan jää epävarmuutta.

## 6.3 Monivalintatehtävät

1.

$$P(\text{itää}) = 0,72$$

$$P(\text{ei idä}) = 1 - 0,72 = 0,28$$

Yksi sipuli itää ja muut  $5 - 1 = 4$  sipulia eivät idä. Itävä sipuli voi olla mikä vaan viidestä vaihtoehdosta. Yhteen- ja kertolaskusääntöjen perusteella

$$\begin{aligned} P(\text{vain yksi itää}) &= 5 \cdot 0,72 \cdot 0,28 \cdot 0,28 \cdot 0,28 \cdot 0,28 \\ &= 0,0221 \dots \approx 0,022 \end{aligned}$$

Vaihtoehdoista b on oikein.

**2.**

Pistetodennäköisyyksien summa on

$$p(3) + p(4) \approx 0,30 + 0,19 = 0,49.$$

Vaihtoehdoista b on oikein.

**3.**

Kuvaajan perusteella kertymätodennäköisyydet ovat

$$P(X \leq 4) = 1 \quad \text{ja} \quad P(X \leq 3) = 0,70$$

Siis

$$P(X = 4) = 1 - 0,70 = 0,30$$

Vaihtoehtoista a on oikein.

4.

Moodiluokan frekvenssi ( $f$ ) on suurin, joten luokka 7–9 on moodiluokka.

Moodina voidaan pitää moodiluokan luokkakeskusta, joka on

$$\frac{7 + 9}{2} = \frac{16}{2} = 8.$$

Vaihtoehdoista b on oikein.

5.

Mediaaniluokka on se luokka, jonka suhteellinen summafrekvenssi ( $sf\%$ ) ensimmäisen kerran ylittää arvon 50 (%), joten pituuden mediaaniluokka on 16–21 (cm).

Vaihtoehdoista c on oikein.

**6.**

Ikä pyöristetään alaspäin, lähimpään kokonaislukuun. Ikäluokkaan 0–2-vuotiaat kuuluvat ne lapset, jotka eivät ole täyttäneet kolmea vuotta. Luokan todellinen yläraja on siis 3 vuotta.

Vaihtoehdoista a on oikein.

7.

Laatikkokuvaajan perusteella:

- $\min = 100 \text{ g}$
- $Q_1 = 150 \text{ g}$
- $Md = 165 \text{ g}$
- $Q_3 = 195 \text{ g}$
- $\max = 250 \text{ g}$

Laatikkokuvaaja havainnollistaa viittä tunnuslukua:

- **min** eli minimiarvo (pienin arvo), vasemmanpuoleisen viiksen alku
- **Q<sub>1</sub>** eli alakvartiili (arvo, jota pienempiä arvoja on 25 %), laatikko-osan vasen reuna
- **Md** eli mediaani, laatikon sisällä oleva pystyviiva
- **Q<sub>3</sub>** eli yläkvartiili (arvo, jota pienempiä arvoja on 75 %), laatikko-osan oikea reuna
- **max** eli maksimiarvo (suurin arvo), oikeanpuoleisen viiksen loppu

Laatikko-osa muodostaa kvartiilivälin  $[Q_1, Q_3] = [150 \text{ g}, 195 \text{ g}]$  jolla on puolet havainnoista

Vaihtoehtoista c on oikein.



8.

Laatikkokuvaajan perusteella:

- $\min = 100 \text{ g}$
- $Q_1 = 150 \text{ g}$
- $Md = 165 \text{ g}$
- $Q_3 = 195 \text{ g}$
- $\max = 250 \text{ g}$

Mediaani (165 g) on arvo, jota pienempiä arvoja on puolet.

Vaihtoehdoista b on oikein.

**9.**

Vastausajan alakvartiili  $Q_1 = 31$  min ilmaisee, että neljännes (25 %) opiskelijoista täytti kyselyn alle 31 minuutissa. Siis, nopein neljännes käytti kyselyyn alle 31 min.

Vaihtoehtoista b on oikein.

**10.**

Symmetrisessä jakaumassa keskiarvo ja mediaani ovat yhtä suuret.

Keskiarvo ja mediaani ovat lähinnä toisiaan jakaumassa a ( $\bar{x} = 80$  € ja  $M_d = 79$  €).

Vaihtoehdoista a on oikein.

**11.**

$$X \sim \text{Bin}(10; 0,45)$$

$$n = 10$$

$$p = 0,45$$

Binomijakaumaa noudattavalle satunnaismuuttujalle odotusarvo on

$$E(X) = n \cdot p = 10 \cdot 0,45 = 4,5.$$

Vaihtoehdoista b on oikein.

## 12.

Myytyjä lippuja on 82 eli  $n = 82$ .

Lipun ostanut saapuu todennäköisyydellä  $100 \% - 2 \% = 98 \%$  eli  $p = 0,98$ .

Satunnaismuuttuja  $X =$  "saapuvien matkustajien lukumäärä".

$X \sim \text{Bin}(82; 0,98)$

Binomijakaumaa noudattavalle satunnaismuuttujalle odotusarvo on

$$E(X) = n \cdot p = 82 \cdot 0,98 = 80,36 \approx 80.$$

Vaihtoehtoista b on oikein.

**13.**

$$X \sim \text{Bin}(8; 0,2)$$

Koska onnistumistodennäköisyys  $p = 0,2 < 0,5$ , onnistuminen on epätodennäköisempää eli muuttujan pienet arvot ovat yleisempiä: jakauma on painottunut keskikohdan vasemmalle puolelle.

Vaihtoehdoista b on oikein.

**14.**

$$X \sim N(10; 0,45)$$

Odotusarvo on  $\mu = 10$ .

Keskihajonta on  $\sigma = 0,45$ .

Normaalijakauma on symmetrinen, joten mediaani ja odotusarvo ovat yhtä suuria. Siis  $Md = 10$ .

Väite  $c$  ei pidä paikkaansa

Vaihtoehdoista  $c$  on oikein.

**15.**

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

Odotusarvo on  $\mu = 0$ .

Keskihajonta on  $\sigma = 1$ .

Normaalijakauman prosenttisäännön mukaan, noin 95 % muuttujan arvoista on korkeintaan kahden keskihajonnan päässä odotusarvosta, eli välillä  $[-2, 2]$ .

Vaihtoehtoista b on oikein.



16.

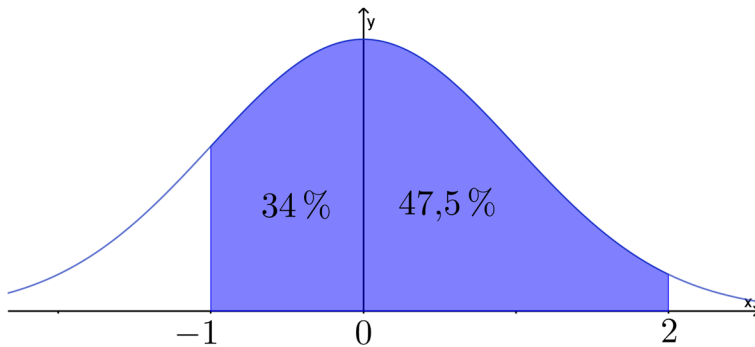
$$X \sim N(0, 1)$$

Odotusarvo on  $\mu = 0$ .

Keskihajonta on  $\sigma = 1$ .

Normaalijakauman prosenttisäännön mukaan, noin 68 % muuttujan arvoista on korkeintaan yhden keskihajonnan päässä odotusarvosta, eli välillä  $[-1, 1]$ . Näistä arvoista puolet, eli  $\frac{68\%}{2} = 34\%$  on symmetrian perusteella odotusarvon vasemmalla puolella, eli välillä  $[-1, 0]$ .

Prosenttisäännön mukaan, noin 95 % muuttujan arvoista on korkeintaan kahden keskihajonnan päässä odotusarvosta, eli välillä  $[-2, 2]$ . Näistä arvoista puolet, eli  $\frac{95\%}{2} = 47,5\%$  on symmetrian perusteella odotusarvon oikealla puolella, eli välillä  $[0, 2]$ .



$$\text{Siis } P(-1 \leq X \leq 2) = 0,34 + 0,475 = 0,815 \approx 0,82.$$

Vaihtoehdoista b on oikein.

17.

$$X \sim N(5, 3)$$

Odotusarvo on  $\mu = 5$ .

Keskihajonta on  $\sigma = 3$ .

Normaalijakauman prosenttisäännön mukaan, noin 68 % muuttujan arvoista on korkeintaan yhden keskihajonnan päässä odotusarvosta.

- $5 - 3 = 2$
- $5 + 3 = 8$

Siis noin 68 % arvoista on välillä  $[2, 8]$ .

Vaihtoehdoista c on oikein.

**18.**

Satunnaismuuttuja  $X =$  "verhotangon pituus (cm)".

$$X \sim N(50, 2)$$

Odotusarvo on  $\mu = 50$ .

Keskihajonta on  $\sigma = 2$ .

Normitetaan arvo 53.

$$\frac{53 - \mu}{\sigma} = \frac{53 - 50}{2} = \frac{3}{2} = 1,5$$

Vaihtoehtoista b on oikein.

## 19.

Satunnaismuuttuja  $X =$  "verhotangon pituus (cm)".

$$X \sim N(50, 2)$$

Odotusarvo on  $\mu = 50$ .

Keskihajonta on  $\sigma = 2$ .

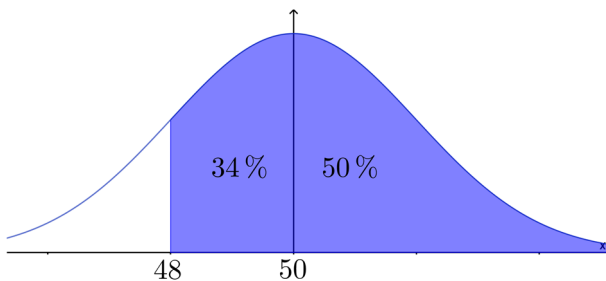
Normaalijakauman prosenttisäännön mukaan, noin 68 % muuttujan arvoista on korkeintaan yhden keskihajonnan päässä odotusarvosta.

- $50 - 2 = 48$
- $50 + 2 = 52$

Siis noin 68 % arvoista on välillä  $[48, 52]$ .

Symmetrian perusteella puolet, eli  $\frac{68\%}{2} = 34\%$  arvoista on odotusarvon vasemmalla puolella, eli välillä  $[48, 50]$ .

Lisäksi puolet (50 %) arvoista on odotusarvoa 50 suurempia.



Siis  $P(X \geq 48) = 0,34 + 0,50 = 0,84$  eli 84 % verhotangoista on vähintään 48 cm pituisia.

Vaihtoehdoista c on oikein.

---

*Huomautus:* Oikean vaihtoehdon voi päätellä myös ilman prosenttisääntöä.

Tehtävässä lasketaan tapahtuman  $X \geq 48$  todennäköisyyttä.

- Arvo 48 on odotusarvon vasemmalla puolella, joten todennäköisyys on yli 0,50.
- Arvo 48 poikkeaa odotusarvosta yhden keskihajonnan verran ( $50 - 48 = 2$ ).
- Tapahtuman  $X \geq 48$  todennäköisyys on siis lähempänä arvoa 1 kuin arvoa 0,50.

20.

Satunnaismuuttuja  $X =$  "verhotangon pituus (cm)".

$$X \sim N(50, 2)$$

Odotusarvo on  $\mu = 50$ .

Keskihajonta on  $\sigma = 2$ .

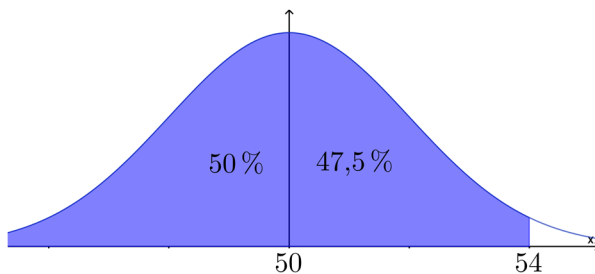
Normaalijakauman prosenttisäännön mukaan, noin 95 % muuttujan arvoista on korkeintaan kahden keskihajonnan päässä odotusarvosta.

- $50 - 2 \cdot 2 = 46$
- $50 + 2 \cdot 2 = 54$

Siis noin 95 % arvoista on välillä  $[46, 54]$ .

Symmetrian perusteella puolet, eli  $\frac{95\%}{2} = 47,5\%$  arvoista on odotusarvon oikealla puolella, eli välillä  $[50, 54]$ .

Lisäksi puolet (50 %) arvoista on odotusarvoa 50 pienempiä.



Siis,  $P(X \leq 54) = 0,50 + 0,475 = 0,975 \approx 0,98$ .

Vaihtoehtoista a on oikein.

---

*Huomautus:* Oikean vaihtoehdon voi päätellä myös ilman prosenttisääntöä.

Tehtävässä lasketaan tapahtuman  $X \leq 54$  todennäköisyyttä.

- Arvo 54 on odotusarvon vasemmalla puolella, joten todennäköisyys on yli 0,50.
- Arvo 54 poikkeaa odotusarvosta kahden keskihajonnan verran ( $54 - 50 = 4$ ).
- Yli kaksi keskihajontaa poikkeavat arvot ovat normaalijakaumassa harvinaisia.
- Tapahtuman  $X \leq 54$  todennäköisyys on siis lähellä arvoa 1.

## 21.

Riskitasoa 5 % vastaa luottamustaso  $100 \% - 5 \% = 95 \%$ .

Suhteellisen osuuden 95 %:n virhemarginaalin laskukaava on:

$$1,96 \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

Laskukaava löytyy taulukkokirjasta.

$$n = 1000$$

$$\hat{p} = 0,16$$

$$1 - \hat{p} = 1 - 0,16 = 0,84$$

Virhemarginaaliksi saadaan

$$1,96 \cdot \sqrt{\frac{0,16 \cdot 0,84}{1000}} = 0,0227 \dots \approx 2,3 \text{ (prosenttiyksikköä)}$$

Vaihtoehdoista b on oikein.



## 22.

Keskiarvon luottamusväli ilmaisee rajat, joiden sisällä perusjoukon (kaikkien suomalaisten) keskiarvo on.

Voidaan sanoa, että suomalaisten saunomiskertojen keskiarvo on jokin luku välillä 6,0–6,4 kertaa/kk.

Luottamusväliin liittyy otoksesta johtuvaa epävarmuutta, mikä ilmaistaan luottamustasolla. Siis saatuun arvioon voidaan luottaa 95 %:n varmuudella.

Vaihtoehtoista c on oikein.

### 23.

Otoksen perusteella tehtyihin johtopäätöksiin liittyy aina epävarmuutta, joka ilmaistaan tutkimuksen virhemarginaalina eli muodostamalla otoksesta lasketulle tunnusluvulle luottamusväli. Virhemarginaalin leveyteen vaikuttaa valittu riskitaso.

c-kohdassa on ilmaistu:  $\alpha = 0,01$  eli riskitaso on 1 %.

Valittu luottamustaso on siis  $100 \% - 1 \% = 99 \%$ .

Suhteellisen osuuden 99 %:n virhemarginaalin laskukaava on:

$$2,58 \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

Laskukaava löytyy taulukkokirjasta.

$$n = 200$$

$$\hat{p} = \frac{30}{200} = 0,15$$

Punatukkaisten osuus otoksessa on 15 %.

$$1 - \hat{p} = 1 - 0,15 = 0,85$$

Virhemarginaaliksi saadaan

$$2,58 \cdot \sqrt{\frac{0,15 \cdot 0,85}{200}} = 0,0651 \dots \approx 7 \text{ (prosenttiyksikköä)}$$

Otoksen perusteella punatukkaisten osuus on siis  $15 \% \pm 7 \%$ .

Vaihtoehdoista c on oikein.

24.

Riskitaso on 5 %.

Valittu luottamustaso on siis  $100 \% - 5 \% = 95 \%$ .

Suhteellisen osuuden 95 %:n virhemarginaalin laskukaava on:

$$1,96 \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

Laskukaava löytyy taulukkokirjasta. ←

$$n = 1000$$

$$\hat{p} = 0,30$$

$$1 - \hat{p} = 1 - 0,30 = 0,70$$

← Kannattajien osuus otoksessa on 30 %.

Virhemarginaaliksi saadaan

$$1,96 \cdot \sqrt{\frac{0,30 \cdot 0,70}{1000}} = 0,028 \dots \approx 3 \text{ (prosenttiyksikköä)}$$

Ehdokkaan kannatuksen luottamusväliksi saadaan  $30 \% \pm 3 \%$ .

Vaihtoehdoista a on oikein.

25.

Luottamusväli ilmaisee, millä välillä *perusjoukon* keskiarvo on tietyllä varmuudella (siis: vaihtoehto b on väärin).

Luottamustason (usein 95 %) päättää tutkija (siis: vaihtoehto c on väärin).

Keskiarvon keskivirheen laskukaava on:

$$\frac{s}{\sqrt{n}}$$

Laskukaava löytyy taulukkokirjasta.

Otoskoko  $n$  on nimittäjässä, joten otoskoon kasvaessa keskivirhe pienenee.

Vaihtoehtoista a on oikein.