

806109P TILASTOTIETEEN PERUSMENETELMÄT I

1. välikoe 11.3.2011 (Jari Päckilä)

VALITSE VIIDESTÄ TEHTÄVÄSTÄ NELJÄ JA VASTAA VAIN NIIHIN!

1. Valitse kohdissa A-F oikea (vain yksi) vaihtoehto. Oikeasta vastauksesta saat +1 pistettä, väärästä et menetä pisteitä.

A) Mikä seuraavista otantamenetelmistä ei ole ns. todennäköisyysotantaan perustuva otantamenetelmä, vaan harkintaotantaa, joka johtaa ”mukavuusotokseen”?

- a1) yksinkertainen satunnaisotanta palauttaen,
- a2) yksinkertainen satunnaisotanta palauttamatta,
- a3) ryväotanta,
- a4) systemaattinen otanta,
- a5) kiintiöpoiminta,
- a6) ositettu otanta suhteellisella kiintiöinnillä.

B) Ylipeitto otantatutkimuksessa tarkoittaa sitä, että

- b1) käytettäessä yksinkertaista satunnaisotantaa palauttamatta otoskoko tulee liian suureksi,
- b2) kehikkopopulaatio sisältää sellaisia havaintoyksiköitä, jotka eivät kuulu kohdepopulaatioon,
- b3) kohdepopulaatio sisältää sellaisia havaintoyksiköitä, jotka eivät ole mukana kehikkopopulaatiossa,
- b4) ositettua otantaa käytettäessä jokin ositteista tulee yliedustetuksi ositekohtaisten otoskokojen laskennassa tapahtuvien pyöristysvirheiden takia,
- b5) haastattelututkimukseen osallistuva henkilö valitsee tarjolla olevista vastausvaihtoehdoista useamman kuin yhden,
- b6) samaan perusjoukkoon kohdistuu samanaikaisesti useampia tutkimuksia.

C) Välimatka-asteikon muuttujasta x on saatu kolme havaintoa: 0, 3 ja 6. Havaintoarvon 0 standardoitu arvo on

- c1) -1, c2) $-1/3$, c3) 0, c4) $1/3$, c5) 1, c6) 3.

D) Havaintoaineistosta on laskettu erilaisia tilastollisia tunnuslukuja. Seuraavista tuloksista ainoa mahdollinen (lasketun arvon perusteella) on

- d1) $C = 2.0$, d2) $s = -2.0$, d3) $Q_1 = Q_3 = 10.0$,
d4) vaihteluvälin pituus $w = -3$, d5) $OR = -1$, d6) variaatiokerroin $V = -0.5$.

E) Luokitteluasteikollisen muuttujan x mahdolliset arvot ovat A, B, C ja D. Sata havaintoa sisältävässä havaintoaineistossa arvo A esiintyy 25 kertaa, arvo B 35 kertaa, arvo C 20 kertaa ja arvo D 20 kertaa. Muuttujan x

- e1) jakauma voidaan esittää pistekuviona,
- e2) jakauman moodi on 35,
- e3) hajontaa voidaan kuvailla vaihteluvälin avulla,
- e4) summajakauma on mielekäs muodostaa,
- e5) havaintoarvot voidaan standardoida.
- e6) Mikään edellä esitetyistä kohdista e1)–e5) ei pidä paikkaansa.

F) Arvottujen lohkojen koejärjestelyssä

- f1) lohkot jaetaan satunnaisesti eri käsittelyille,
- f2) koeyksiköt jaetaan satunnaisesti eri lohkoihin,
- f3) koeyksiköt valitaan satunnaisotannalla tarjolla olevasta perusjoukosta,
- f4) lohkoja on aina yhtä monta kuin käsittelyjä,
- f5) jokaisessa lohossa koeyksiköt jaetaan satunnaisesti eri käsittelyille,
- f6) kunkin lohkon sisällä koeyksiköt ovat mahdollisimman heterogeenisiä (erilaisia) sellaisten ominaisuuksien suhteen, joilla oletetaan olevan vaikutusta vastemuuttujaan.

2. Eräessä vuonna 2003 tehdyssä haastattelututkimuksessa ($n = 218$) tutkittiin opiskelijoiden terveystietoisuutta ja terveyttä. Haastattelulomakkeen yhdessä kysymyksessä vastaaja pyydettiin arvioimaan omaa terveydentilaansa ($= x$). Vastausten jakauma oli seuraava:

Arvio omasta terveydentilasta	frekvenssi
huono	1
melko huono	3
keskitasoinen	32
melko hyvä	84
hyvä	98
Yhteensä	218

- a) Määrää muuttujan x :n mitta-asteikko ja se, onko muuttuja jatkuva vai diskreetti. (1 p)
- b) Esitä x :n jakauma graafisesti. (1 p)
- c) Määrää (ja laske) x :n jakaumalle kaksi sijaintilukua ja kaksi hajontalukua. (2.5 p)
- d) Alla oleva R-ohjelman tulostus liittyy vastaajan ikää (vuosina) kuvaavaan muuttujaan. Esitä iän jakauma laatikko-jana -kuviona. (1.5 p)

```
> numSummary(terveys[, "IKÄ"], statistics=c("mean", "sd", "quantiles"),
quantiles=c( 0, .25, .5, .75, 1 ))
```

```
   mean    sd 0% 25% 50% 75% 100%  n
20.505  3.422 16  17  20  22   30 218
```

3. Eräässä perunanviljelykokeessa tutkittiin Fambo ja Van Gogh-lajikkeita.

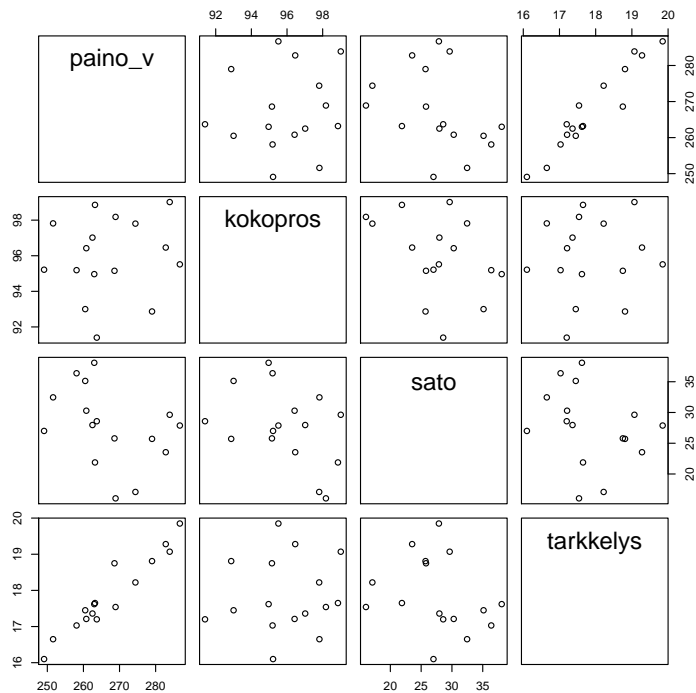
- a) Kahdeksalla koealalla viljeltiin Fambo-lajikkeen perunaa. Koealoittain mukuloiden lukumäärän havaittiin olevan

46, 52, 50, 46, 56, 52, 48 54.

a1) Esitä havainnot pistekuviona. (1 p)

a2) Laske mukuloiden lukumäärän keskiarvo ja keskihajonta. (2 p)

- b) Toisessa kokeessa kerättiin koealoittain ($n = 16$) tiedot mm. perunoiden yhteispainosta (g, `paino_v`), tavoitekoon onnistumisesta (prosentteina tavoitekoosta, `kokopros`), sato­määrästä (tonnia/hehtaari, `sato`) ja tärkkelyspitoisuudesta (prosentteina, `tarkkelys`). Koealat olivat pinta-aloiltaan yhtä suuria. Alla aineistoon liittyvää R-tulostusta:



```
> cor(potut)      # korrelaatiomatriisi
      paino_v kokopros  sato tarkkelys
paino_v  1.000    0.111 -0.364      C
kokopros  0.111      B -0.403    0.074
sato     -0.364  -0.403  1.000      D
tarkkelys  A      0.074 -0.263    1.000
```

```
> cov(potut)      # kovarianssimatriisi
      paino_v kokopros  sato tarkkelys
paino_v 127.059  2.773 -25.545  11.158
kokopros  2.773  4.879 -5.539  0.170
sato    -25.545 -5.539 38.685 -1.689
tarkkelys 11.158  0.170 -1.689  1.066
```

Täydennä korrelaatiomatriisi (eli määrää korrelaatiokerroimet A, B, C ja D) ja tulkitse korrelaatiokerroin A. (3 p)

4. Peikkokulman Supervalinta ilmoittaa punaisten ja sinisten karkkien alennusmyynneistä. Mörrimöykky pitää erityisesti punaisista karkeista ja siniset ovat aivan kamalan ihania (tarttuvat ikävästi hampaisiin ja aiheuttavat karvaita röyhtäyksiä). Tänä kesänä Mörrimöykyn hampaita on vihlonut ikävästi, mutta silti Mörrimöykky ryntää tohkeissaan orava-kortin kera karkkiostoksille. Syvällä mielessä kuitenkin pyörii se, miten hän on saanut hammassärkyä syötyään tietyn värisiä karkkeja. Syötyjen karkkien värin ja karkkien syönnin jälkeisen hammassärlyn määrän välisen yhteyden tutkimiseen Mörrimöykkyllä on käytössä seuraava ristiintaulukko:

Hammassärlyn määrä karkkien syönnin jälkeen	Syötyjen karkkien väri		Yhteensä
	punainen	sininen	
särkee paljon	43	29	72
särkee vähän	323	298	621
ei särje lainkaan	108	174	282
Yhteensä	474	501	975

Tutki muuttujien välistä riippuvuutta

- a) ehdollisten prosenttijakaumien avulla, (2 p)
- b) tilanteeseen sopivan tunnusluvun avulla. (4 p)

5. Tutkimuksessa selvitettiin sitä, miten runsas lannoitteen määrä (= muuttuja **lannoite**, 100 kg/ha) vaikuttaa tietystä kasvusta saatavan sadon määrään (= muuttuja **sato**, 100 kg/ha). Havaintoaineisto muodostui kymmeneltä koelajalta kerätyistä tiedoista. Muuttujien lannoite ja sato välinen kovarianssi on -9.44 ja lisäksi ko. muuttujista on laskettu R-ohjelmalla seuraavat tunnuslukujen arvot:

```
> numSummary(satokoe, statistics=c("mean", "sd", "quantiles"),
  quantiles=c( 0,.25,.5,.75,1 ))
```

```
      mean      sd   0%  25%  50%  75% 100%  n
lannoite  7.5  0.745  6.5  7.0  7.5  8.0  8.5  10
sato     65.6 13.023 45.0 55.2 69.0 77.2 80.0 10
```

- a) Sovita aineistoon regressiosuora $y = a + bx$, missä vastemuuttujana on sadon määrä ja selittävänä muuttujana lannoitteen määrä. Tulkitse regressiokertoimet a ja b selväkielisesti. Määrää myös regressioyhtälön determinaatikerroin eli selitysaste ja tulkitse se. (4 p)
- b) Mikä on regressioyhtälön antama ennustearvo sadon määräksi, jos käytetty lannoitemäärä on 725 kg/ha? (1 p)
- c) Lisää regressiosuoran kuvaaja liitteen 1 sirontakuviioon. **Palauta liite 1 nimelläsi varustettuna vastauspaperisi mukana!** (1 p)

Liite 1

Nimi:

