

806109P TILASTOTIETEEN PERUSMENETELMÄT I
Loppukoe 7.1.2008 (Hanna Heikkinen ja Marjatta Mankinen)

VASTAA VIITEEN TEHTÄVÄÄN!

1. Valitse kohdissa A-F oikea vaihtoehto. Perusteluja ei vaadita, mutta ne on kuitenkin hyvä laittaa näkyviin. Jokaisessa kohdassa saat oikeasta vastauksesta pisteen, väärästä vastauksesta et menetä pisteitä.

A) Järjestysasteikkoa olevalle muuttujalle on sallittua

- a1) esittää frekvenssijakauma janakuviona,
- a2) esittää summajakauma graafisesti,
- a3) laittaa luokat frekvenssien suuruusjärjestyksen mukaisesti,
- a4) muodostaa laatikko-jana-kuvio,
- a5) laskea vinousmitta,
- a6) määrätä kvartiiliväli.

B) Satunnaistus kokeellisessa tutkimuksessa tarkoittaa sitä, että

- b1) koeyksiköt valitaan yksinkertaisella satunnaisotannalla,
- b2) koeyksiköt valitaan satunnaistettua harkintaotantaa käyttäen,
- b3) tutkittavat käsittelyt valitaan satunnaisesti tarjolla olevista käsittelyistä,
- b4) koeyksiköt jaetaan eri käsittelyille satunnaisesti,
- b5) koeyksiköt jaetaan eri käsittelyille satunnaistettua kiintiöintiä käyttäen,
- b6) kyseessä on ns. kaksoissokkokoe eli ne, jotka voisivat vaikuttaa tuloksiin (esim. koehenkilöt, käsittelyn antajat) ja ne, jotka arvioivat tuloksia (esim. hoitavat lääkärit), eivät tiedä, minkä käsittelyn kukin koeyksikkö on saanut.

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

- c1) $s=-2.1$, c2) $\rho=+1.3$,
- c3) $C=1.2$, c4) $Q_1 = Q_3 = -3.4$,
- c5) $w=-8.0$, c6) $r = -2.7$.

Kohdat D-F liittyvät seuraavaan, erään liikkeen asiakkaita koskevasta kyselyaineistosta saatuun ristiintaulukkoon:

Tyytyväisyys asiakaspalveluun	Sukupuoli		Yhteensä
	mies (=M)	nainen (=N)	
erittäin tyytyväinen (=ET)	30	30	60
melko tyytyväinen (=MT)	50	100	150
melko tyytymätön (=MTT)	20	70	90
Yht.	100	200	300

D) Jos tyytyväisyys asiakaspalveluun -muuttujan frekvenssijakauma esitettäisiin graafisesti, sopiva kuvio olisi

- d1) histogrammi, d2) frekvenssimonikulmio,
d3) pylväskuvio, d4) janakuvio,
d5) porraskuvio, d6) pistekuvio.

E) Taulukosta laskettu suhteellinen kontingenssikerroin on

- e1) 0.200 e2) 0.283 e3) 0.452 e4) 0.585 e5) 0.639 e6) ei mikään edellisistä.

F) Valitaan satunnaisesti yksi asiakkaista. Mikä seuraavista väitteistä on epätosi?

- f1) $P(ET) = 0.200$ f2) $P(MTT|M) = 0.200$
f3) $P(N|MTT^c) = 0.619$ f4) $P(N \cap MTT) = 0.200$
f5) $P(ET \cup MT) = 0.700$ f6) $P(M \cup MT) = 0.667$.

2. Kolmentoista Suomessa toimivan autokaupan yrityksen sijoitetun pääoman tuottoprosentit olivat v. 2006 seuraavat:

24, 34, 18, 17, 19, 6, 23, 32, 18, 12, 23, 14, 8

a) Muodosta sijoitetun pääoman tuottoprosentin

- a1) runko-lehti-kuvio,
a2) pistekuvio,
a3) laatikko-jana-kuvio. (4 p)

b) Laske sijoitetun pääoman tuottoprosentille kaksi sopivaa keskilukua ja kaksi sopivaa hajontalukua. Perustele valintasi! (2 p)

Huom. Sijoitetun pääoman tuottoprosentti, lyhyesti sipo, on välimatka-asteikkoa oleva muuttuja, joka kuvaa yrityksen kannattavuutta; mitä suurempi sipo, sitä parempi kannattavuus.

3. Liitteessä on R:llä saatuja tuloksia aineistosta, jonka muodostaa 20 miehen tiedot seuraavista muuttujista: paino (kg), vyötärön ympärysmitta (cm) ja rasvaprosentti.

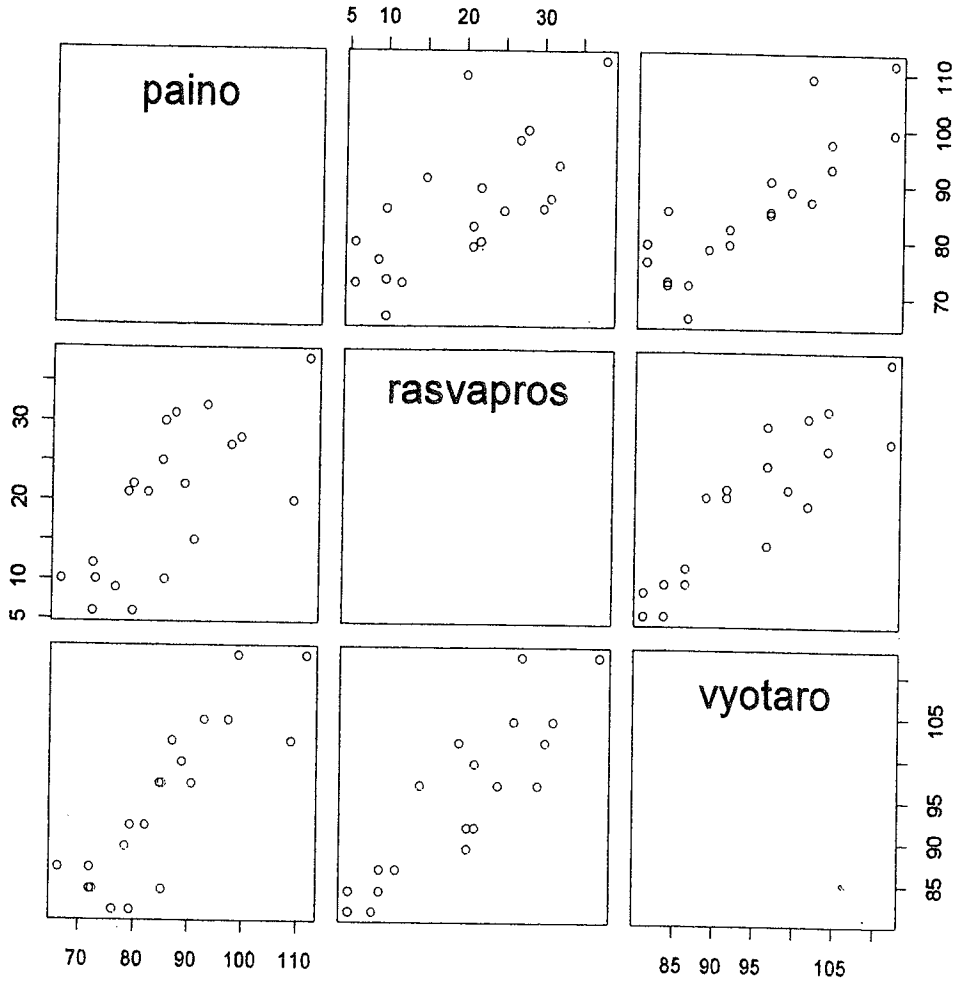
a) Tulkitse painon ja vyötärön ympärysmittan välistä riippuvuutta tulostuksesta löytyvän a1) kuvion, a2) riippuvuusluvun arvon perusteella. (1 p)

b) Kumpi muuttujista paino vai vyötärön ympärysmitta on parempi selittäjä rasvaprosentille? Perustele valintasi! (0.5 p)

c) Määrää b)-kohdassa valitsemasi muuttuja selittäjänä regressioyhtälö rasvaprosentille ja tulkitse saadun yhtälön kertoimet. Määrää myös regressioyhtälön determinatiokerroin (= selitysaste) ja tulkitse se. (4 p)

d) Ennusta b)-kohdan yhtälöä käyttäen sellaisen miehen rasvaprosentti, joka painaa 78 kg ja jonka vyötärön ympärysmitta on 90 cm. (0.5 p)

4. Pitkän ajan tilastoista selviää, että vuotuinen sademäärä Oulussa noudattaa likimain normaalijakaumaa keskiarvona 520 mm ja keskihajontana 60 mm. Eri vuosien sademäärät oletetaan toisistaan riippumattomiksi.
- a) Mikä on todennäköisyys, että Oulussa sataa satunnaisesti valitun vuoden aikana yli 600 mm? (1 p)
 - b) Mikä on todennäköisyys, että seuraavista kahdeksasta vuodesta kahtena vuotena sataa Oulussa yli 600 mm? (2 p)
 - c) Juuso opiskelee Oulun yliopistossa viisi vuotta. Merkitään Juuson opiskeluvuosien sademäärien aritmeettista keskiarvoa \bar{X} .
 - c1) Mitä jakaumaa \bar{X} noudattaa? (1 p)
 - c2) Mikä on todennäköisyys, että niiden viiden vuoden sademäärien keskiarvo on välillä 500-600 mm? (2 p)
5. Terveystoimittaja halusi tutkia, missä iässä vauvat oppivat ryömimään. Ryömimisen oppimisiän voidaan olettaa noudattavan normaalijakaumaa keskihajontana 7.2 viikkoa.
- a) Viidenkymmenen satunnaisesti poimitun vauvan ryömimisen oppimisikäien keskiarvo oli 32.4 viikkoa.
 - a1) Määrää keskimääräiselle ryömimisen oppimisikälle piste-estimaatti ja 90 % :n luottamusväli. Tulkitse saamasi väli. (3 p)
 - a2) Eräässä lastenhoito-oppaassa kerrotaan, että vauvat oppivat ryömimään keskimäärin 40 viikon iässä. Kommentoi tätä väitettä a1)-kohdassa laskemasi luottamusvälin perusteella. (1 p)
 - b) Jos keskimääräisen ryömimisen oppimisiän 95 %:n luottamusvälin pituudeksi haluttaisiin korkeintaan 2 viikkoa, montako vauvaa otokseen pitäisi ottaa? (2 p)
6. Oy Suklaa Ab valmistaa 200 gramman suklaakonvehtipusseja, joissa samankokoiset konvehdit on täytetty usealla erilaisella täytteellä. Tuoteselosteen mukaan konvehdeista 10 % on vaniljatäytteisiä, 20 % mansikkatäytteisiä, 20 % pähkinätäytteisiä, 10 % tryffelitäytteisiä ja loput ovat kokonaan suklaata. Tehtaan laaduntarkkailija ottaa satunnaisesti kauhallisen pussitettavaksi meneviä konvehdeja tarkasteluun. Lajittelussa selviää, että kauhalliseen tulleista 200 konvehdistä vaniljatäytteisiä on 22, mansikkatäytteisiä 36, pähkinätäytteisiä 41, tryffelitäytteisiä 14 ja kokonaan suklaisia 87 kappaletta. Tutki sopivan merkitsevyydestin avulla, pitääkö tuoteselostus paikkansa. Merkitse testauksen kaikki vaiheet näkyviin. (6 p)



```
> numSummary(rasva[,c("paino", "rasvapros", "vyotaro")], statistics=c("mean",
"sd"))
```

	mean	sd	n
paino	85.550	12.097477	20
rasvapros	19.750	9.563500	20
vyotaro	94.105	9.699456	20

```
> cor(rasva[,c("paino", "rasvapros", "vyotaro")], use="complete.obs")
```

	paino	rasvapros	vyotaro
paino	1.0000000	0.6965958	0.8522118
rasvapros	0.6965958	1.0000000	0.8863919
vyotaro	0.8522118	0.8863919	1.0000000

```
> cov(rasva[,c("paino", "rasvapros", "vyotaro")], use="complete.obs")
```

	paino	rasvapros	vyotaro
paino	146.34895	80.59211	99.99763
rasvapros	80.59211	91.46053	82.22237
vyotaro	99.99763	82.22237	94.07945