

TILASTOTTEEN PERUSMENETLMÄT I

Loppukoe 1.10.2007 (Marjatta Mankinen ja Jari Päkkiä)

VALITSE KUUDESTA TEHTÄVÄSTÄ VIISI JA VASTAA VAIN NIINIIN!

1. Vastaa lyhyesti seuraaviin kysymyksiin. Jokaisessa kohdassa a)-f) oikealla vastauksella saa yhden pisteen (vääristä vastauksesta ei menetä pisteitä).

- a) 12000 henkilön populaatiosta on poimittu ositetulla otannalla suhteellista kiintöimintää käyttäen 1000 henkilön otos. Osituksen perustana on käytetty henkilön ikää luokitusk-sella: 0-14 -vuotiaat, 15-64 -vuotiaat ja väh. 65-vuotiaat. Otoksessa 15-64 -vuotiaita on 555. Kuinka monta 15-64 -vuotiaasta on populaatiossa?

- b) Luettele vertailevien kokeiden suunnittelun kolme tärkeintä periaatetta.

- c) Eräällä palkkakunnalla mitattiin tammikuun jokaisena päivänä ulkoilman lämpötilaa aa-mulla kello 9. Havainnoista ($n = 31$) laskettuja tunnustuksia on esitetty alla olevassa R:n tulostuksessa.

> summary(lampotila)

```

Min. 1st Qu. Median      Mean 3rd Qu.      Max.
 6.90  10.10  14.40  13.44  16.05  23.60

```

Tunnustuskujen laskemisen jälkeen huomattiin tietojenkäsittelyssä tapahtunut virhe: kaik-kien havaintoarvojen edestä oli puuttunut miinusmerkki (havainnot olivat siis todellisuus-nessa pakkaslukemia!). Lisäksi lämpömittari huomattiin virheelliseksi siten, että mittari näytti systemaattisesti kaksi astetta liian suurta pakkaslukemia. Esimerkiksi todellinen lämpötila -14.0 näkyi mittarivirheen ja tallennusvirheen takia aineistossa lukemana 16.0. Muuna R:n tulostuksessa esiintyvät tunnustuskujen arvot vastaamaan todellisia havain-toja.

- d) Havaintoaineistossa on neljä havaintoa: 3, 5, 5 ja 7. Laske havaintoarvon 3 stardardoitu arvo.

- e) Suhdeasteikollisen muuttujan x mitaamiseen ollaan kehittälemässä uutta mittaria. Ky-seistä ominaisuutta mitataan uudella mittarilla samalta tilastoyksiköltä viisi kertaa pe-räkkäin. Mittaustilanteesta tiedetään se, että muuttujan x todellinen arvo jokaisella mit-tauskerralla on 10. Havainnollista pistekuvioiden avulla viidestä mittauksesta saatavaa aineistoa, kun tiedetään, että uusi mittari on

e1) validi, mutta konsistenssiltään huono,

e2) konsistenssiltään hyvä, mutta se ei ole validi.

- e) Erästä havaintoaineistosta sukupuolen ja tupakoinnin väliseksi ristintaulukoksi saatiin seuraavaa:

	Tupakointi		Yhteensä
Sukupuoli	Kyllä	Ei	
mies	20	60	80
nainen	30	90	120
Yhteensä	50	150	200

Määrää kyseiseen ristintaulukkaan liittyvän χ^2 -tunnusluvun arvo.

4. a) Automerkin AA mallin 320i hiilidioksidipäästöjen määrän on todettu noudattavan normaalijakaumaa odotusarvona 146 g/km ja keskihajontana 10 g/km.

a1) Monellako prosentilla näistä autoista hiilidioksidipäästöjen määrä ylittää 160 g/km?

a2) Mikä on kyseisten autojen hiilidioksidipäästöjen määrän alakvartiili?

a3) Mikä on todennäköisyys, että kahden satunnaisesti valitun mallia 320i olevan AA-auton yhteenlasketut hiilidioksidipäästöt jäävät alle 280 g/km? (4.5 p)

b) Olkoon $(X_1, X_2, \dots, X_{100})$ satunnaisotos jakaumasta $\text{Exp}(0.5)$. Mitä jakaumaa noudattaa $\bar{X} = \frac{1}{100} \sum_{i=1}^{100} X_i$? Perustelee vastauksesi! (1.5 p)

5. Alueen A lähiöstölle suunnitellaan rakennettavaksi ydinvoimala. Alueen kaikki asukkaat eivät kuitenkaan suhtaudu myötämielesisti hankkeeseen ja alueen aikuisväestöstä otetusta satunnaisotoksesta ydinvoimalan rakentamista vastustavien 95%:n luottamusväliksi saatiin (67.0%, 70.6%).

a) Tulkitse saatu väli.

b) Mikä on saadun otoksen antama piste-estimaatti ydinvoimalan rakentamista vastustavien prosenttiosuudelle ja mikä on ns. virhemarginaali?

c) Määrää luottamusvälin laskemisessa käytetty otoskoko.

d) Mitä luottamusväliksi olisi saatu, jos luottamustasona olisi ollut d1) 0.99 d2) 0.90? (6 p)

6. Eräessä tutkimuksessa (Strayer and Johnston, 2001) haluttiin selvittää, onko ajonakaisella matkapuhelimeen puhumisella vaikutusta kuljettajan reaktioaikaan. Koehenkilöinä oli 63 yliopisto-opiskelijaa, jotka jaettiin satunnaisesti koeryhmään (31) ja kontrolliryhmään (32). Kooneessa, joka simuloi ajotilanteita, välähti epätasaisin välein punainen tai vihreä valo. Koehenkilöitä pyydettiin painamaan "jarrunäppäintä" niin pian kuin mahdollista, kun he havaitsivat punaisen valon. Koetta suorittaessaan kontrolliryhmään kuuluvat kuuntelivat radiota, kun taas koeryhmäläiset keskustelivat matkapuhelimen välityksellä poliittisesta aiheesta eri huoneessa olevan henkilön kanssa. Jokaisen koehenkilön keskimääräinen reaktioaika (millisekunteina) mitattiin ja näistä saatiin seuraavat ryhmäkohtaiset tulokset:

koeryhmä	keskihajonta
573.1	58.9
533.7	65.3

Näyttääkö näiden tulosten perusteella matkapuhelimeen puhuminen ajon aikana lisäävän reaktioaika? Suorita merkittävyystestaus. Kirjoita kaikki testauksen vaiheet (myös oletukset) näkyviin. (6 p)

2. Kyliäkaupan teettämässä asiakaskyselyssä selvitettiin mm. asiakkaiden tyytyväisyyttä kaupan asiakaspalveluun. Kyselyyn vastanneista asiakaspalveluun erittäin tyytyväisiä oli 15, melko tyytyväisiä 8, melko tyytymättömiä 6 ja erittäin tyytymättömiä 3 asiakasta.

- a) Määrää tarkasteltavan muuttujan mita-asteikko ja esitä ko. jakama graafisesti. (2 p)
- b) Määrää (ja laske) tyytyväisyysmuuttujan jakumalle kaksi sijaintia ja kaksi hajontaa ku-
vaavaa tunnuslukua. (2 p)

c) Kyselyn lopussa vastaaja pyydettiin antamaan yleisarvosana kaupan palveluista piste-
asteikolla 0-100. Kymmenen ensimmäistä palautettua lomaketta käsitteävässä aineistossa
yleisarvosanojen aritmeettinen keskiarvo oli 70 pistettä ja keskihajonta 5 pistettä. Näiden
tunnuslukujen laskeamisen jälkeen kauppaan palautettiin vielä kaksi vastauslomaketta,
joissa yleisarvosanat olivat 85 ja 87 pistettä. Määrää täydennetty aineiston ($n = 12$)
yleisarvosanan aritmeettinen keskiarvo ja keskihajonta. (2 p)

3. Erään tutkimuksen tarkoituksena oli selvittää sitä, miten erään suolan liukeneminen ve-
teen riippuu veden lämpötilasta. Tutkimusaineistossa on kahdeksan havaintoa ja kaksi muut-
tujaa: veden lämpötila (=lampotila, °C) ja veteen liuenneen suolan paino (=suolan.paino,
grammoina). Muuttuja suolan.paino kertoo, kuinka paljon suola liukeni 100 grammaan tietyn
lämpöistä vettä. Alla on esitetty ko. aineistoon liittyvää R-ohjelman tulostusta:

```
> summary(suolan.paino)
  Min. 1st Qu.  Median  Mean 3rd Qu.  Max.
 53.50  63.78  73.05  72.50  81.53  90.00
> sd(suolan.paino)
[1] 12.70793
> summary(lampotila)
  Min. 1st Qu.  Median  Mean 3rd Qu.  Max.
  0.0   17.5   35.0   35.0   52.5   70.0
> sd(lampotila)
[1] 24.49490
cov(suolan.paino, lampotila)
[1] 311
```

- a) Laske tarkasteltavien muuttujien välinen korrelaatiokerroin r ja tulkitse tulos lyhyesti.
(2 p)
- b) Sovita aineistoon regressioyhtälö $y = a + bx$, jossa lämpötilan avulla selitetään veteen
liuenneen suolan määrää. Tulkitse regressioyhtälön kertoimet selväkielisesti. (2 p)
- c) Määrää b)-kohdan regressioyhtälön determinaatikerroin (=selitysaste) ja tulkitse tulos.
(1 p)

d) Yhdellä havaintoaineiston tilastoyksiköllä muuttujien arvot olivat: lampotila=30 °C ja
suolan.paino=70.6 g. Määrää kyseiselle tilastoyksikölle b)-kohdan regressioyhtälön an-
tama ennustearvo ja residuaali. (1 p)