

806109P TILASTOTIETEEN PERUSMENETELMÄT I  
Loppukoe 20.6.2011 (Marjatta Mankinen ja Jari Päckilä)

VALITSE KUUESTA TEHTÄVÄSTÄ VIISI JA VASTAA VAIN NIIHIN!

1. Rukiinviljelykokeessa oli käytössä kahdeksan samankokoista koealaa. Kokeen päättyessä koealoilta saadut ruissadot (kg/hehtaari) olivat:

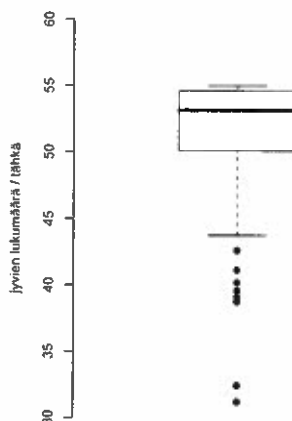
3300 3200 3800 3150 3550 3250 3600 3100

a) Laske aineiston perusteella seuraavien tunnuslukujen arvot

- a1) aritmeettinen keskiarvo,                      a2) mediaani,                      a3) varianssi,  
a4) kvartiiliväli.                                      a5) vaihteluvälin pituus,  
a6) havaintoarvon 3800 standardoitu arvo.

(3 p)

b) Sadasta satunnaisesti valitusta rukiin tähkästä laskettiin tähkässä olevien jyvien lukumäärä. Alla oleva R-ohjelman tulostus liittyy jyvien lukumäärän jakaumaan.



```
# jyvien lukumäärä/tähkä talletettu muuttujaan jyva
> summary(jyva)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  31.0   50.0   54.0   51.28  54.0   55.0
> sd(jyva)
[1] 4.92
```

Määrää tulostuksen perusteella jyvien lukumäärän jakaumalle sopiva sijainti- ja hajontaluku. Perustele valintasi lyhyesti. (1.5 p)

c) Jyvien painoa punnittiin puntarilla, joka oli validi ja konsistenssiltaan huono. Selitä lyhyesti millainen puntari oli. (1.5 p)

2. Osakan ympäristössä Japanissa tehtiin 1990-luvun alussa eräs haastattelututkimus, johon valittiin mukaan 18-79 -vuotiaita ko. alueella asuneita henkilöitä. Haastattelujen perusteella saatiin muodostettua mm. seuraava ristiintaulukko:

| Sukupuoli | Suhtautuminen ydinvoimaan |                  |                  |            | Yhteensä |
|-----------|---------------------------|------------------|------------------|------------|----------|
|           | myönteinen                | melko myönteinen | melko kielteinen | kielteinen |          |
| mies      | 335                       | 1298             | 330              | 204        | 2167     |
| nainen    | 210                       | 1317             | 496              | 336        | 2359     |
| Yhteensä  | 545                       | 2615             | 826              | 540        | 4526     |

- Mitä mitta-asteikkoa ristiintaulukon muuttujat ovat? Ilmoita kummankin muuttujan osalta myös se, onko kyseessä jatkuva vai diskreetti muuttuja. (2 p)
- Esitä kummankin muuttujan jakauma graafisesti (kaksi eri kuvaa). (2 p)
- Tutki (ja tulkitse) sukupuolen ja ydinvoimaan suhtautumisen välistä riippuvuutta ehdollisten prosenttijakaumien avulla. (2 p)

3. Kymmenestä kuusesta mitattiin puun korkeus (metreissä) sekä rungon halkaisija (senttimetreissä) 1.5 metrin korkeudelta. Mittaustulokset olivat seuraavat:

| Puun numero       | 1    | 2    | 3   | 4    | 5    | 6    | 7    | 8    | 9    | 10   |
|-------------------|------|------|-----|------|------|------|------|------|------|------|
| Rungon halkaisija | 13.7 | 10.1 | 5.8 | 20.7 | 17.8 | 14.8 | 14.4 | 13.4 | 17.8 | 20.7 |
| Puun korkeus      | 16.3 | 12.1 | 8.0 | 17.4 | 18.4 | 16.1 | 16.6 | 12.9 | 17.5 | 19.4 |

- Tutki muuttujien välistä riippuvuutta sopivan kuvion avulla ja tulkitse tulos lyhyesti. (1 p)
- Tutki muuttujien välistä riippuvuutta sopivan riippuvuustunnusluvun avulla ja tulkitse tulos lyhyesti. Voit käyttää laskelmissasi apuna tehtävän lopussa esitettyä R-ohjelman tulostusta. (2 p)
- Tulkitse seuraavalla sivulla esitetyn R-ohjelmalla tehdyn regressioanalyysin tulokset (määrittele regressioyhtälö, kertoimien selväkielinen tulkinta, determinaatiokerroin eli selityssaste ja sen tulkinta). (3 p)

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

```
      mean  sd  0%   25%   50%   75% 100%  n
halkaisija 14.92 4.642 5.8 13.475 14.60 17.800 20.7 10
korkeus    15.47 3.464 8.0 13.700 16.45 17.475 19.4 10
```

```
> malli.1 <- lm(korkeus~halkaisija, data=puu)
```

```
> summary(malli.1)
```

```
Call:
```

```
lm(formula = korkeus ~ halkaisija, data = puu)
```

```
Residuals:
```

```
      Min       1Q   Median       3Q      Max
-2.087049 -0.870523  0.004141  0.874669  1.677889
```

```
Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.10073    1.49607   3.409  0.00923
halkaisija   0.69499    0.09617   7.227  9e-05
```

```
Residual standard error: 1.339 on 8 degrees of freedom
```

```
Multiple R-squared:  0.8672, Adjusted R-squared:  0.8506
```

```
F-statistic: 52.22 on 1 and 8 DF,  p-value: 9.009e-05
```

4. Merkkiä A olevien autojen hiilimonoksidipäästöjen ( $=X$ ) on todettu noudattavan normaalijakaumaa odotusarvolla 1.8 g/km ja keskihajonnalla 0.25 g/km.

- a) Mikä on todennäköisyys, että satunnaisesti valitun A-merkkisen auton hiilimonoksidipäästöt ylittävät 2.2 g/km? (2 p)
- b) Eräällä yrityksellä on A-merkkisiä autoja 40 (= satunnaisotos A-merkkisistä autoista) ja olkoon  $\bar{X}$  näiden autojen hiilimonoksidipäästöjen keskiarvo.
  - b1) Mitä jakaumaa  $\bar{X}$  noudattaa? (1 p)
  - b2) Mikä on todennäköisyys, että  $\bar{X}$  on välillä 1.75 g/km – 1.82 g/km? (2 p)
- c) Jos ei tiedettäisi, että  $X$  noudattaa normaalijakaumaa, voitaisiinko
  - i) a)-kohdan, ii) b2) -kohdan todennäköisyys silti laskea? Perustele vastauksesi. (1 p)

5. a) Pankissa A uskotaan aiemman kokemuksen perusteella, että 7% henkilöistä, jotka saavat lainan A-pankista, ei maksa lainaansa takaisin ennalta sovitun aikataulun mukaisesti. A-pankki on vastikään myöntänyt lainan 200 henkilölle (= satunnaisotos A-pankin myöntämistä lainoista).

- a1) Mitä jakaumaa tässä joukossa noudattaa  $X$  = niiden lukumäärä, jotka eivät maksa lainaansa ennalta sovitun aikataulun mukaisesti. Määrää myös  $X$ :n odotusarvo ja varianssi. (1.5 p)
- a2) Mikä on todennäköisyys, että kaksikymmentä 200:sta ei maksa lainaansa takaisin ennalta sovitun aikataulun mukaisesti? Voit laskea kysytyyn todennäköisyyden tarkasti tai sopivaa sopivaa approksimaatiota käyttäen. (1.5 p)

b) Biologit tutkivat metsästäjien tietyltä alueelta USA:ssa saalistamaa 153 hirveä ja löysivät näistä 32:sta Lymen sairautta aiheuttavia punkkeja. Tee tarvittavat oletukset ja määrää 95%:n luottamusväli niiden hirvien suhteelliselle osuudelle kyseisellä alueella, jotka kantavat Lymen sairautta aiheuttavia punkkeja. Tulkitse saamasi väli. (3 p)

6. Tutkija halusi selvittää, jakautuvatko kuolemaan johtavat polkupyöräonnettomuudet tasaisesti eri viikonpäiville. Hän valitsi satunnaisesti 200 pyöräilijän kuolemaan johtanutta onnettomuutta, kirjasi viikonpäivän, jona onnettomuus tapahtui ja sai seuraavat tulokset:

| viikonpäivä | frekvenssi |
|-------------|------------|
| sunnuntai   | 16         |
| maanantai   | 35         |
| tiistai     | 16         |
| keskiviikko | 28         |
| torstai     | 34         |
| perjantai   | 41         |
| lauantai    | 30         |
| yhteensä    | 200        |

Näyttääkö kuolemaan johtavat polkupyöräonnettomuudet jakautuvan tasaisesti eri viikonpäiville? Analysoi aineistoa sopivalla merkitsevyydestillä. Kirjoita vastaukseesi kaikki merkitsevyydestin vaiheet. (6 p)