

## 806114P TILASTOTIETEEN PERUSTEET B

Loppukoe 18.10.2010

Hanna Heikkinen

### VASTAA KAIKKIIN TEHTÄVIIN!

1. Tutkimusryhmän tavoitteena oli tutkia erilaisten häiriötekijöiden vaikutusta erään tarkkaavaisuustestin tuloksiin. Tutkimuksen aikana tarkkaavaisuustesti esitettiin jokaiselle koehenkilölle kaksi kertaa: ilman häiriötä (olosuhde A) ja voimakkaasti puheella häiriten (olosuhde B). Tutkimukseen ilmoittautui mukaan 32 vapaaehtoista koehenkilöä, joista naisia oli 11 ja loput 21 miehiä. Vapaaehtoisten joukosta kokeeseen valittiin mukaan 8 koehenkilöä.

- a) Otantakehikossa naiset oli sijoitettu kehikon alkuun ennen miehiä. Millä todennäköisyydellä koehenkilöiden joukossa on täsmälleen kolme naista, kun otanta tehtiin
- a1) yksinkertaisella satunnaisotannalla palauttamatta? (1 p)
- a2) systemaattisella otannalla? (1 p)
- b) Tarkkaavaisuustestin pistemäärät olosuhteissa A ja B olivat koehenkilöiden osalta seuraavat:

koehenkilö	1	2	3	4	5	6	7	8
olosuhde A	28	17	26	20	16	32	26	18
olosuhde B	21	15	25	20	14	23	22	21

Tutki muuttujien välistä riippuvuutta sekä graafisesti että Spearmanin järjestykskorrelaatiokertoimen avulla. Tulkitse tulokset lyhyesti. (2 p)

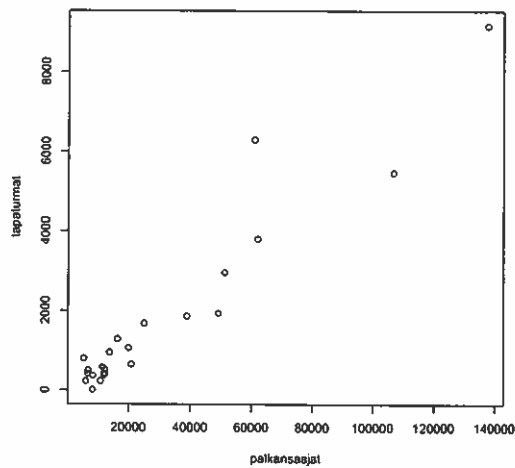
+b) Olisiko mahdollista käyttää muita riippuvuustunnuslukuja muuttujien välisen riippuvuuden tarkasteluun? Perustele. (+ 1 p)

- c) Kokeen seuraavassa vaiheessa kokeeseen lisätään uutena häiriötekijänä valonvälähdyksiä. Määritellään  $X$  siten, että se kuvaa kahden peräkkäisen valonvälähdyksen välistä aikaa sekunteina ja oletetaan, että testaustilanteessa  $X \sim \text{Exp}(0.2)$ . Laske  $X$ :n odotusarvo ja todennäköisyys sille, että kahden peräkkäisen valonvälähdyksen välinen aika on  $X$ :n odotusarvon ja mediaanin välissä. (2 p)

2. a) Täydennä seuraavan sivun taulukko (kopioi vastauspaperiisi) siten, että muuttujien  $x$  ja  $y$  välillä
- a1) ei ole lainkaan riippuvuutta. (1 p)
- a2) on suurin mahdollinen riippuvuus. (1 p)

y	x		yhteensä
	0	1	
0			30
1			30
yhteensä	30	30	60

- b) Suomalaisten työpaikkatapaturmien määrän ja palkansaajien määrän välistä yhteyttä on tarkasteltu 24 eri ammattialalla vuonna 2008 (kuviota alla, Lähde: Tilastokeskus). Muuttujien välille on sovitettu lineaarinen regressiomalli ja sen R-ohjelman tulostus on kuvion alla.



```
> malli<-lm(tapaturmat~palkansaajat)
> summary(malli)
```

```
Call:
lm(formula = tapaturmat ~ palkansaajat)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-1241.82 -255.18  -41.61   216.95  2530.89
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.634e+02  1.931e+02  -0.846   0.407
palkansaajat  6.443e-02  4.344e-03  14.830 6.18e-13 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 703.3 on 22 degrees of freedom
Multiple R-squared:  0.9091,    Adjusted R-squared:  0.9049
F-statistic: 219.9 on 1 and 22 DF,  p-value: 6.177e-13
```

- b1) Kirjoita analyysin regressioyhtälö. Tulkitse yhtälön kertoimet ja determinatiokertoimen arvo selväkielisesti. (2 p)
- b2) Metsätyössä palkansaaajia on 12 300 ja työpaikkatapaturmia on tapahtunut 507. Mikä on metsätyön työpaikkatapaturmien ennustearvo ja residuaali? (1 p)
- c) Jatkoa b)-kohtaan. Kohdassa b) oli otettu mukaan vain alat, joilla on alle 200 000 palkansaaajaa. Alla on taulukoituna myös kuuden suurimman ammattialan havainnot. Sopiiko b)-kohdassa sovitettu regressiomalli myös suurten alojen palkansaaajien ja työpaikkatapaturmien määrän välisen yhteyden kuvaamiseen? Perustelee. (1 p)

Ammattiala	Tapaturmat	Palkansaaajat
Tekninen, tieteellinen, lainopillinen, humanistinen ja taiteellinen työ	1 800	388 600
Terveystenhoito, sosiaalialan työt	4 676	334 700
Hallinto- ja toimistotyö	1 788	400 800
Kaupallinen työ	2 192	251 300
Teollinen työ, koneenhoito, kaivos- ja louhintatyö sekä rakennustyö	28 782	454 100
Palvelutyö ym.	7 009	225 600

3. a) Kontrolloitujen kokeiden suorittamiseen liittyy olennaisena osana satunnaistus, toistaminen ja lohkojen muodostaminen.
- a1) Selitä, mitä nämä tarkoittavat ja mihin niillä pyritään. (3 p)
- a2) Kerro, miten nämä huomioidaan kaltaistettujen parien koejärjestelyssä. (1 p)
- b) Eräessä tutkimuksessa selvitettiin sitä, onko poikien ja tyttöjen keskimääräisissä syntymäpainoissa eroa. Tutkimukseen osallistui 204 satunnaisesti valittua vastasyntyntä. Seuraavalla sivulla on esitetty tutkimusongelmaan liittyvää R-ohjelman tulostusta.
- Millaisiin mallioletuksiin käytetyt tilastomenetelmät perustuvat? Mitkä ovat tutkimusongelmaan liittyvät tilastollisen päättelyn johtopäätökset? Perustelee johtopäätöksesi lyhyesti. (2 p)

```

> #poikien syntymäpainot (grammoina) vektorissa pojat, n=96
> #tyttöjen syntymäpainot (grammoina) vektorissa tytöt, n=108
> t.test(pojat, tytöt, mu=0, var.equal=TRUE, conf.level=0.95)

```

```

Two Sample t-test
data: pojat and tytöt
t = 1.1511, df = 202, p-value = 0.2510
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -52.43632 199.54761
sample estimates:
mean of x mean of y
 3641.763  3568.207

```

4. Yrityksen johto epäili, että työntekijät olivat alkaneet venyttää viikonlopun viettoja ilmoittautumalla sairaaksi perjantaisin ja maanantaisin. Asiaa tutkittiin neljän viikon aikana rekisteröimällä poissaolojen lukumäärät eri työpäivinä. Seuraavassa taulukossa on esitetty kyseisenä seuranta-aikana havaittujen poissaolojen jakauma.

viikonpäivä	ma	ti	ke	to	pe	Yht.
poissaolojen lkm	51	32	34	39	44	200

Tutki sopivan merkitsevyydestin avulla, jakautuvatko työntekijöiden poissaolot tasan eri työpäiville vai tukeeko saatu aineisto johdon epäilyjä. Merkitse vastaukseesi näkyviin kaikki merkitsevyydestauksen vaiheet. (6 p)

**LIITE:** Lisäyksiä kaavakokoelmaan:

$$s_{xy} = \frac{1}{n-1} \left( \sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} \right)$$

$$X \sim \text{Geom}(p), \quad P(X = k) = (1-p)^{k-1} p$$

$$E(X) = \frac{1}{p} \quad \text{ja} \quad D^2(X) = \frac{1-p}{p^2}$$

$$X \sim \text{HypGeom}(N, M, n), \quad P(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$$

$$E(X) = n \frac{M}{N} \quad \text{ja} \quad D^2(X) = n \frac{M}{N} \left( 1 - \frac{M}{N} \right) \frac{N-n}{N-1}$$