

806109 TILASTOTIETEEN PERUSMENETELMÄT I
Harjoitus 7, viikko 9, kevät 2013
(Muut kuin taloustieteiden tiedekunnan opiskelijat)

MUISTA MIKROLUOKKAHARJOITUKSET VIIKOLLA 9!

Ilmoittaudu Weboodissa 4.3.2013 klo 8.00-12.00 (sali L4) pidettävään 1. välikokeeseen!

1. Jatkoa harjoituksen 6 tehtävään 4: Sähkölämmitteisen loma-asunnon sähkön kulutusta ja ulkoilman lämpötilaa seurattiin seitsemän vuorokautta. Tällöin saatiin seuraavat havainnot:

vuorokausi:	1	2	3	4	5	6	7
Sähkön kulutus (kWh):	32	28	23	21	30	28	22
Ulkoilman lämpötila (°C):	5	8	12	10	-1	3	7

- Sovita aineistoon regressiosuora $y = a + bx$, missä y = sähkön kulutus ja x = ulkoilman lämpötila. Tulkitse regressiokertoimet a ja b selväkielisesti. Määrää myös regressioyhtälön determinaatikerroin eli selitysaste ja tulkitse se.
- Paljonko regressioyhtälö ennustaa loma-asunnon lämmönkulutuksen olevan (=ennustearvo), jos ulkoilman lämpötila on 9 astetta?

2. Pienen ala-asteen kakkosluokan liikuntaryhmän pojat ($n=11$) ottelivat neliottelun, jossa lajeina olivat 60 metrin juoksu, 100 metrin juoksu, pallonheitto ja pituushyppy.

- Pallonheitossa lopputulokset (metreinä) olivat

18.7 23.4 15.9 19.6 21.0 19.7 18.5 19.7 21.8 18.4 20.1

Laske pallonheiton lopputulosten aritmeettinen keskiarvo ja keskihajonta.

- Liitteessä 1 on esitetty tarkasteltavasta aineistosta saatu R:n tulostus. Kommentoi 60 metrin juoksun ja 100 metrin juoksun välistä riippuvuutta korrelaatiokertoimen r perusteella. Missä korrelaatiodiagrammin kuvassa/kuvissa (1-12) kuvataan 60 metrin juoksun ja 100 metrin juoksun välistä riippuvuutta?
- Määrää regressioyhtälö $y = a + bx$ ja tulkitse kertoimet, kun vastemuuttujana on 100 metrin juoksu ja selittävänä muuttujana 60 metrin juoksu. Määrää lisäksi regressioyhtälön determinaatikerroin (eli selitysaste) ja tulkitse se.
- Matin tulokset neliottelussa olivat: 60 metrin juoksu = 12.5 sekuntia, 100 metrin juoksu = 23.2 sekuntia, pallonheitto = 19.6 metriä ja pituushyppy 2.20 metriä. Ennusta Matille 100 metrin juoksun loppuaika c)-kohdan regressioyhtälön avulla. Laske ennustearvon ja todellisen arvon erotus eli jäännöstermi (eli residuaali).

3. Kalabiologi Pekka Brofelt keräsi Längelmävedellä vuonna 1916 aineistoa tutkimuksiinsa eri kalojen ruumiinmitoista ja niiden yhteyksistä toisiinsa. Tässä tehtävässä tarkastelemme erityisesti sitä, miten lahnoilla (*Abramis brama*) ruumiin paino (grammoina) riippuu maksimipituudesta (muuttuja *pituus* senttimetreinä mitattuna suusta pyrstön päähän) ja korkeudesta (*korkeus*, senttimetreinä kalan korkeimmalta kohdalta.)

Aineiston lähde: Brofelt, P. Bidrag till kännedom om fiskbeståndet in våra sjöar. Längelmävesi. Kirjassa Järvi, T.H. Finlands Fiskeriet Band 4, Meddelanden utgivna av fiskeriföreningen i Finland. Helsingfors 1917.

Liitteessä 2 on esitetty otteita R-ohjelman tulostuksesta. Käytä tulostuksen tietoja apunasi vastatessasi seuraaviin kysymyksiin.

- a) Mallissa 1 vastemuuttujana on lahnan paino ja selittäjänä lahnan korkeus. Miksi lahnan korkeus näyttäisi olevan hieman pituutta parempi selittäjä lahnan painolle?
- b) Tulkitse regressioanalyysin tulokset, jotka liittyvät malliin 2 (määrää regressioyhtälö, kertoimien selväkielinen tulkinta, determinaatikertoimen arvo ja sen tulkinta).
- c) Erään lahnan paino on 500 grammaa, pituus 36.4 cm ja korkeus 13.8 cm. Ennusta vastemuuttujan arvo ko. tilastoyksikölle
 - c1) regressioyhtälön 1 (*malli1*) avulla,
 - c2) regressioyhtälön 2 (*malli2*) avulla.
 - c3) Laske vasteen todellisen arvon ja ennustearvon erotus eli jäännöstermi (eli residuaali) kohdissa c1) ja c2).

4. Liitteessä 3 on esitetty neljä erilaista hajontakuviota muuttujien x ja y välillä. Kunkin pisteparven keskelle on sovitettu regressiosuora $y = a + bx$. Sovitusten tuloksena on saatu seuraavat regressiosuorat:

A: $\hat{y} = 15.7 - 0.2x$

B: $\hat{y} = 14.7 - 0.7x$

C: $\hat{y} = 10.3 + 0.3x$

D: $\hat{y} = 3.1 + 1.3x$

Mihin kuvioon kukin sovitteista kuuluu?

5. a) Mikä on todennäköisyys, että satunnaisesti valittu viikonpäivä on

a1) sunnuntai a2) lauantai tai sunnuntai?

b) Herätyskello pysähtyy pariston loputtua. Millä todennäköisyydellä kello pysähtyy

b1) klo 01.00 ja 06.00 välisenä aikana,

b2) lauantaina tai sunnuntaina klo 01.00 ja 06.00 välisenä aikana?

6. Paperisilppurin jäljiltä paperiarkki on silputtu 264 palaan, joista jokaisessa on korkeintaan yksi kirjain. Kirjaimia on seuraavasti:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	R	S	T	U	V	Ä	Ö
12	1	0	3	8	0	0	5	15	4	7	2	6	12	8	6	6	4	6	12	2	3	1

Valitset satunnaisesti yhden lapun. Mikä on todennäköisyys, että

- a) se on tyhjä, b) siinä on vokaali, c) siinä on konsonantti,
- d) siinä on kirjain, jota ei ole missään muussa lapussa,
- e) siinä on F,
- f) siinä on joku sanassa "tilastotiede"oleva kirjain?

7. Alla olevassa taulukossa on esitetty nykyisten kansanedustajien lukumäärät sukupuolen ja syntymävuoden mukaan:

Sukupuoli	Syntymävuosi					Yhteensä
	1940–49	1950–59	1960–69	1970–79	1980–89	
Miehet	21	35	35	16	3	110
Naiset	6	16	36	25	7	90
Yhteensä	27	51	71	41	10	200

Valitaan satunnaisesti (umpimähkään) yksi kansanedustaja. Mikä on todennäköisyys, että valittu kansanedustaja on

- a) mies,
- b) syntynyt 1960-luvulla,
- c) mies ja syntynyt 1960-luvulla,
- d) mies tai syntynyt 1960-luvulla,
- e) nainen, kun tiedetään, että hän on syntynyt 1940-luvulla,
- f) syntynyt 1940-luvulla, kun tiedetään, että hän on nainen,
- g) syntynyt 1940- tai 1980-luvulla,
- h) syntynyt 1940- tai 1980-luvulla, kun tiedetään, että hän on mies?

Huom.: Mikroluokkaharjoitusten harjoitusryhmät viikolla 9 ovat:

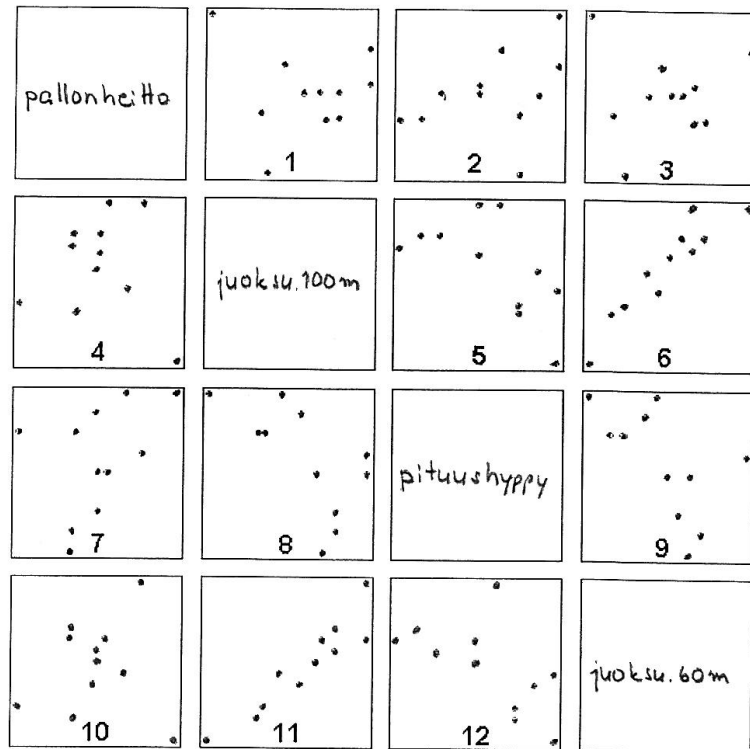
MA KLO 12.15–13.45 (M304) (ryhmä suunnattu biologeille)
TI KLO 8.30–10.00 (M304)
TI KLO 12.15–13.45 (M302)
KE KLO 10.15–11.45 (M302)
KE KLO 14.30–16.00 (M302)
TO KLO 12.15–13.45 (M302)
TO KLO 12.15–13.45 (M304)
PE KLO 10.15–11.45 (M302) (ryhmä suunnattu biologeille)
PE KLO 14.15–15.45 (M304) (ryhmä suunnattu biologeille)

Vastauksia tehtäviin:

1. a) 30.6 -0.68 0.49 b) 24.5
2. a) 19.71 ja 1.96 c) $a \approx -13.74$, $b \approx 2.93$ ja $\hat{R}^2 \approx 0.87$ d) 22.8 ja 0.4
3. c1) 482 g c2) 501 g c3) 18 g ja -1 g
5. a1) 0.1429 a2) 0.2857 b1) 0.2083 b2) 0.0595
6. a) 0.5341 b) 0.2235 c) 0.2424 d) 0.0076 e) 0 f) 0.2197
7. a) 0.5500 b) 0.3550 c) 0.1750 d) 0.7300 e) 0.2222 f) 0.0667
g) 0.1850 h) 0.2182

LIITE 1:

LIITE 1



#keskiarvot:

```
> mean(juoksu.100m)
[1] 23.3182
```

```
> mean(juoksu.60m)
[1] 12.6636
```

```
> mean(pituushyppy)
[1] 2.1318
```

varianssit:

```
> var(juoksu.100m)
[1] 1.5796
```

```
> var(juoksu.60m)
[1] 0.1605
```

```
> var(pituushyppy)
[1] 0.0186
```

#korrelaatiomatriisi

```
> cor(neliottelu, use="complete.obs")
```

	juoksu.100m	juoksu.60m	pallonheitto	pituushyppy
juoksu.100m	1.0000	0.9328	-0.1067	-0.6302
juoksu.60m	0.9328	1.0000	0.0322	-0.5983
pallonheitto	-0.1067	0.0322	1.0000	0.3907
pituushyppy	-0.6302	-0.5983	0.3907	1.0000



LIITE 2:

```
>cor(lahna) #korrelaatomatriisi
      paino  pituus korkeus
paino  1.0000  0.9639  0.9707
pituus 0.9639  1.0000  0.9540
korkeus 0.9707  0.9540  1.0000
```

```
> malli1 <- lm(paino~korkeus)
> summary(malli1)
```

Call:

```
lm(formula = paino ~ korkeus)
```

Residuals:

Min	1Q	Median	3Q	Max
-62.8	-37.4	-9.9	19.0	106.7

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-913.94	67.91	-13.5	1e-14
korkeus	101.16	4.42	22.9	<2e-16

Residual standard error: 50.4 on 32 degrees of freedom
Multiple R-squared: 0.942, Adjusted R-squared: 0.941
F-statistic: 523 on 1 and 32 DF, p-value: <2e-16

>

```
> malli2 <- lm(paino~korkeus+pituus)
> summary(malli2)
```

Call:

```
lm(formula = paino ~ korkeus + pituus)
```

Residuals:

Min	1Q	Median	3Q	Max
-66.7	-27.2	-7.2	15.4	94.9

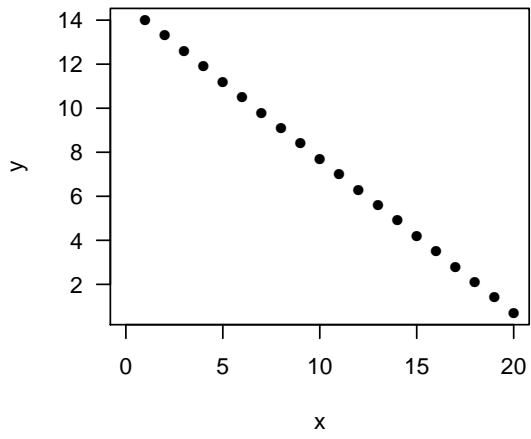
Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1068.84	73.91	-14.46	2.5e-15
korkeus	59.27	12.74	4.65	5.8e-05
pituus	20.65	5.99	3.45	0.0017

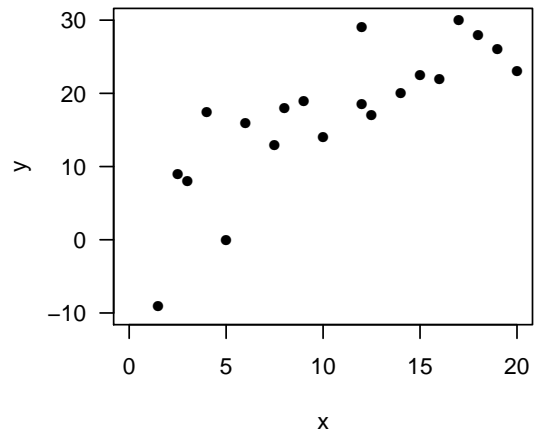
Residual standard error: 43.5 on 31 degrees of freedom
Multiple R-squared: 0.958, Adjusted R-squared: 0.956
F-statistic: 356 on 2 and 31 DF, p-value: <2e-16

LIITE 3:

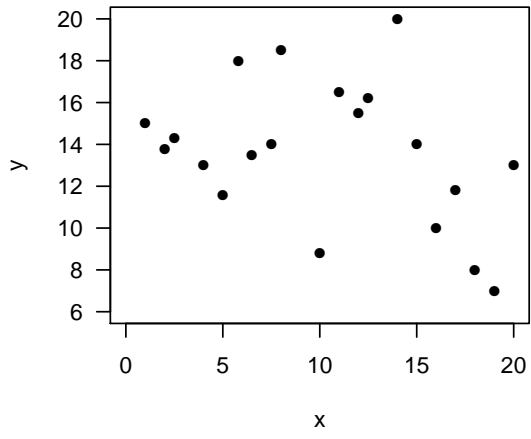
KUVA 1



KUVA 2



KUVA 3



KUVA 4

