

806109 TILASTOTIETEEN PERUSMENETELMÄT I
Harjoitus 7, viikko 9, kevät 2012
(Muut kuin taloustieteiden tiedekunnan opiskelijat)

MUISTA MIKROLUOKKAHARJOITUKSET VIIKOILLA 8 JA 9!

1. Jatkoa harjoituksen 5 tehtävään 7: Selvitettäessä tentissä viipymisen (aika minuuteissa) ja tenttituloksen välistä riippuvuutta saatiin seuraavat mittaustulokset:

Opiskelija	1	2	3	4	5	6	7	8	9	10
viipymisaika	200	150	100	180	230	160	170	175	210	190
pistemäärä	24	15	5	22	23	18	19	19	20	21

Sovita aineistoon regressiosuora $y = a + bx$, missä y = pistemäärä ja x = viipymisaika. Tulkitse regressiokertoimet a ja b selväkielisesti. Määrää myös regressioyhtälön determinaatikerroin eli selitysaste ja tulkitse se.

2. Kevään 2010 välikoetehtävä: Erään pohjoisen luonnonkasvin ekofysiologiaa selvittävän kasvatuskokeen tulosten analysoinnissa saatiin R-ohjelmalla liitteessä 1 esitetty tulostus (osa tuloksista peitetty merkinnällä xx.xxx). Kokeessa mitatut muuttujat olivat kasvin sisältämä tyyppi (muuttuja N , milligrammoina) ja kasvin biomassa (muuttuja biomass, milligrammoina). Aineistosta laskettu Pearsonin tulomomenttikorrelaatiokerroin oli 0.85. Muuttujan N varianssi oli 0.0840 ja muuttujan biomass varianssi 40.4469.

- Onko Pearsonin tulomomenttikorrelaatiokerroimen käyttäminen mielekästä/luvallista tälle aineistolle? Jos on, niin mistä syystä? Kumpi aineiston muuttujista on valittu regressioyhtälössä $y = a + bx$ vasteeksi ja kumpi selittäväksi muuttujaksi?
- Määrää regressioyhtälön kertoimet a ja b ja tulkitse ne lyhyesti.
- Määrää regressioyhtälön determinaatikerroimen arvo (selitysaste) ja tulkitse se lyhyesti.
- Laske mallin mukainen ennustearvo, kun kasvin biomassa on 40 milligrammaa ja typen määrä 1.8 milligrammaa.

3. Liitteessä 2 on esitetty neljä erilaista hajontakuviota muuttujien x ja y välillä. Kunkin pisteparven keskelle on sovitettu regressiosuora $y = a + bx$.

- Regressiosuorien kulmakertoimien b arvot näissä kuvioissa olivat -0.70, 0.25, -0.37 ja 1.48. Mihin kuvioon kukin näistä luvuista kuuluu?
- Regressiosuorien vakiokerroimien a arvot näissä kuvioissa olivat 3.5, 9.9, 15.0 ja 18.6. Mihin kuvioon kukin näistä luvuista kuuluu?
- Korrelaatiokerroimen r arvot näissä kuvioissa olivat 0.34, 0.90, -1.00, -0.63. Mihin kuvioon kukin näistä luvuista kuuluu?

4. Palataan harjoituksen 2 tehtävässä 7 kerättyyn irtokarkkidataan. Tuolloin tehtävänä oli arvioida 100 irtokarkin yhteispainoa neljän satunnaisesti valitun irtokarkin yhteispainon avulla. R:n aineistoon ko. muuttujat on talletettu seuraavilla nimillä:

yhtpaino = 100 irtokarkin yhteispaino

paino = neljän satunnaisesti valitun irtokarkin yhteispaino

Liitteessä 3 on esitetty otteita R-ohjelman tulostuksesta. Käytä tulostuksen tietoja apunasi vastatessasi seuraaviin kysymyksiin.

- a) Tulkitse regressioanalyysin tulokset, jotka liittyvät tulostuksen regressiomalliin (määrää regressioyhtälö, kertoimien selväkielinen tulkinta, determinaatikertoimen arvo ja sen tulkinta).
- b) Ennen R-ohjelmalla tehtyä regressiomallin $y = \beta_0 + \beta_1 x$ regressiokertoimien arviointia olisi voitu tehdä vaikkapa seuraavanlaisia teoreettisia pohdintoja:
 - Koska arviointitilanteessa opiskelija sai arpoa punnitukseen neljä irtokarkkia eli yhden kahdeskymmenesviidesosan karkkien kokonaismäärästä, kannattaisi opiskelijan arvioida 100 irtokarkin yhteispaino kertomalla neljän punnitsemansa irtokarkin yhteispaino luvulla 25. Tämän päättelyn perusteella regressiomallissa $y = \beta_0 + \beta_1 x$ regressiokertoimen β_1 estimaatti b tulisi olemaan 25.
 - Jos opiskelijan punnitsemien neljän irtokarkin yhteispaino olisi ollut teoreettiset nolla grammaa, olisi ollut loogista olettaa, että myös 100 irtokarkin yhteispaino olisi samaiset 0 grammaa (karkit olisivat siis ”tyhjää täynnä”!). Tämän päättelyn perusteella regressiomallissa $y = \beta_0 + \beta_1 x$ regressiokertoimen β_0 estimaatti a tulisi olemaan 0.

Vertaile a)-kohdan tuloksista löytyvien kertoimien a ja b arvoja edellä esitetyn teoreettisen pohdiskelun mukaisiin a:n ja b:n arvoihin. Näyttäisivätkö opiskelijat toimineen teoreettisen pohdiskelun mukaisesti yhteispainon arvioinnissaan? Jos eivät, keksitkö selityksiä pohdinnasta poikkeavalle menettelylle?

Huomautettakoon, että liitteen 3 lopussa tarkasteltavien muuttujien väliseen korrelaatio-diagrammiin (eli hajontakuviin) on lisätty kaksi suoraa: yhtenäisellä viivalla on kuvattu regressioanalyysin tulosten perusteella piirretty regressiosuora $y = 122.869 + 21.158x$ ja katkoviivalla on kuvattu se teoreettinen tilanne, jossa päättely olisi tehty edellä esitellyn logiikan mukaisesti (eli $a=0$ ja $b=25$ ja täten $y = 0 + 25x$)!

- c) Eräessä vastauslomakkeessa neljän satunnaisesti valitun irtokarkin yhteispainoksi oli ilmoitettu 27 grammaa ja 100 irtokarkin yhteispainoksi 605 grammaa. Ennusta R-tulostuksen (malli1) avulla vastemuuttujan arvo ko. tilastoyksikölle. Laske myös vasteen todellisen arvon ja ennustetun arvon erotus eli residuaali.

5. Ville on päättänyt uusia talonsa ulkomaalaukset. Hän valitsee arpomalla sekä seinän värin että ikkunanpuitteiden värin. Seinän väriksi on tarjolla kolme vaihtoehtoa: rosa, sininen ja lila. Ikkunanpuitteiden väriksi tarjolla on joko valkoinen tai keltainen väri. Valittava ikkunanpuitteen väri ei riipu millään tavalla seinän väristä. Millä todennäköisyydellä talon

- a) seinäväriksi tulee sininen?
- b) seinäväriksi tulee rosa ja ikkunanpuitteiden väriksi valkoinen?
- c) seinäväriksi tulee sininen tai ikkunanpuitteiden väriksi keltainen?

6. Ihmiset kuuluvat johonkin neljästä pääveriryhmästä A, B, AB ja O. Suomalaisten veriryhmäjakauma on seuraava: A 44%, B 17%, AB 8% ja O 31 %. Millä todennäköisyydellä suomalaisen avioparin aviopuolisot kuuluvat

- a) A-veriryhmään,
- b) samaan veriryhmään,
- c) eri veriryhmiin?

7. Seuraavassa on esitetty tehtävässä 4 analysoidusta irtokarkki-aineistosta muodostettu sukupuolen ja 100 irtokarkin yhteispainoarvion välinen ristiintaulukko:

Sukupuoli	Irtokarkkien yhteispaino (grammoja)				Yhteensä
	500 – 749	750 – 999	1000 – 1249	1250 – 1499	
Miehet	7	8	5	3	23
Naiset	12	13	8	3	36
Yhteensä	19	21	13	6	59

Valitaan aineistosta satunnaisesti (umpimähkään) yksi irtokarkkien yhteispainoa arvioinut opiskelija. Mikä on todennäköisyys, että valittu opiskelija on

- a) mies,
- b) arvioinut yhteispainon välille 750 – 999 grammaa,
- c) mies ja arvioinut yhteispainon välille 750 – 999 grammaa,
- d) mies tai arvioinut yhteispainon välille 750 – 999 grammaa,
- e) nainen, kun tiedetään, että hän on arvioinut yhteispainon välille 500 – 749 grammaa,
- f) arvioinut yhteispainon välille 500 – 749 grammaa, kun tiedetään, että hän on nainen,
- g) arvioinut yhteispainon välille 500 – 749 grammaa tai välille 1250 – 1499 grammaa,
- h) arvioinut yhteispainon välille 500 – 749 grammaa tai välille 1250 – 1499 grammaa, kun tiedetään, että hän on mies?

- Mikroluokkaharjoitusten harjoitusryhmät viikolle 9 ovat:

MA KLO 12.15–13.45 (M302)

MA KLO 14.30–16.00 (M302)

TI KLO 8.30–10.00 (M302) (ryhmä suunnattu biologeille)

TI KLO 14.30–16.00 (M302)

KE KLO 10.00–11.30 (M304)

KE KLO 12.15–13.45 (M302)

TO KLO 8.30–10.00 (M302) (ryhmä suunnattu biologeille)

TO KLO 10.15–11.45 (M304)

TO KLO 14.30–16.00 (M304)

PE KLO 10.15–11.45 (M302) (ryhmä suunnattu biologeille)

- Ilmoittautuminen perjantaina 9.3. klo 12–16 pidettävään 1. välikokeeseen on avattua WebOodiin.
- Välikoeviikolla perjantain klo 12–16 luokkaharjoitus siirtyy pidettäväksi tiistaina klo 14–16 ja torstaina klo 14–16. Harjoitussalit ilmoitetaan myöhemmin. Keskiviikon harjoitusryhmä pidetään normaalin aikataulun mukaisesti.

Huom.: Ilmoittautuminen perjantaina 9.3. klo 12–16 pidettävään 1. välikokeeseen on avattua WebOodiin.

Vastauksia tehtäviin:

1. a) -5.68 0.14 0.82

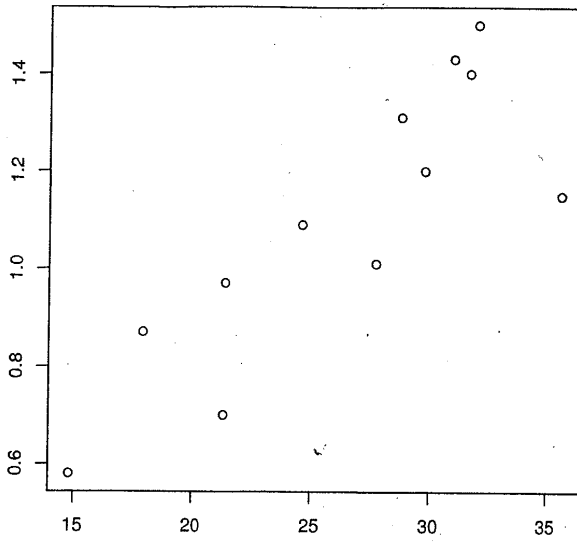
2. b) 0.078982 0.0387 c) 0.723 d) 1.6

4. c) 694.1 ja -89.1

5. a) 0.3333 b) 0.1667 c) 0.6667

6. a) 0.1936 b) 0.3250 c) 0.6750

LIITE 1:



```
> summary(model)
```

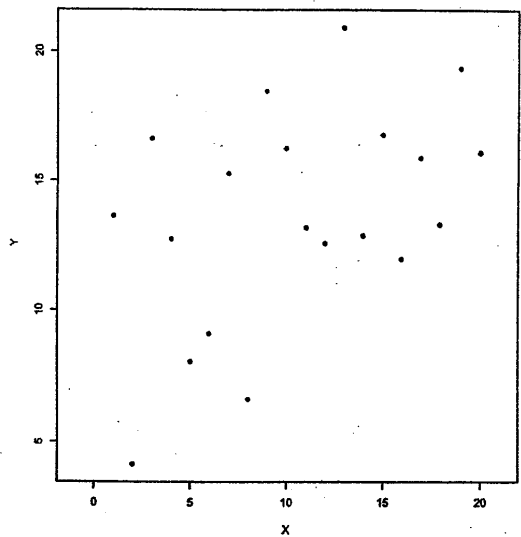
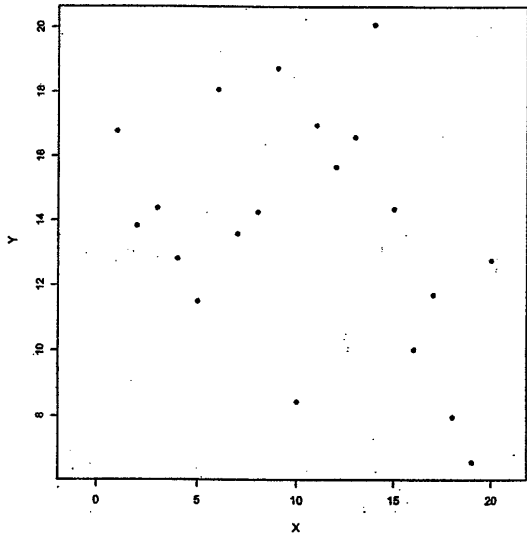
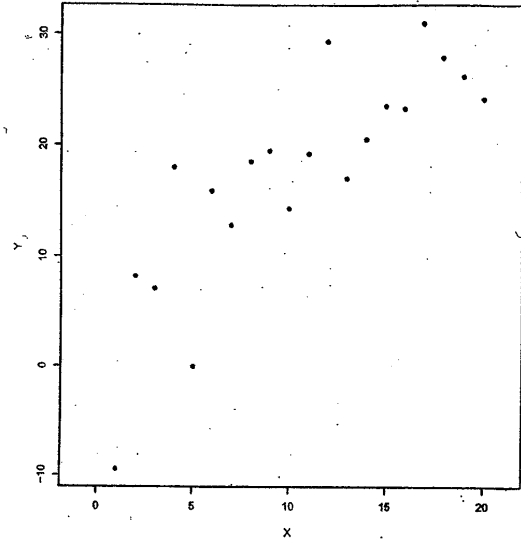
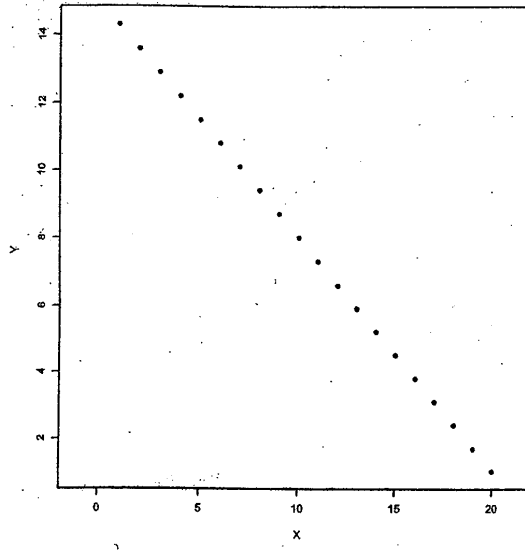
```
Call:  
lm(formula = N ~ biomass)
```

```
Residuals:  
    Min       1Q   Median       3Q      Max  
-0.30731 -0.09022  0.06011  0.10151  0.18157
```

```
Coefficients:  
              Estimate Std. Error t value Pr(>|t|)  
(Intercept) 0.078982   0.206457   0.383 0.710052  
biomass      xx.xxx    0.007616   5.079 0.000478 ***
```

```
Residual standard error: 0.1606 on 10 degrees of freedom  
Multiple R-squared:  xx.xxx , Adjusted R-squared: xx.xxx  
F-statistic: 25.8 on 1 and 10 DF, p-value: 0.0004783
```

LIITE 2:



LIITE 3:

```
> summary(malli1)
```

Call:

```
lm(formula = yhtpaino ~ paino, data = karkkidata)
```

Residuals:

Min	1Q	Median	3Q	Max
-464.23	-15.29	3.18	40.21	410.87

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	122.869	61.419	2.001	0.0502 .
paino	21.158	1.633	12.956	<2e-16 ***

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

Residual standard error: 120 on 57 degrees of freedom

Multiple R-squared: 0.7465, Adjusted R-squared: 0.7421

F-statistic: 167.9 on 1 and 57 DF, p-value: < 2.2e-16

