

**Tehtävät – Tulee valita/tehdä 4 tehtävää 5:stä, ei enempää!**

**1. Periaatteet/käsitteet.** Vastaa lyhyesti mutta selkeästi seuraaviin kysymyksiin/tehtäviin.

(Pisteitys 1 p / kohta eli yhteensä 5 p)

- Mitä kaikkea viiksilaatikkokuvaaja (*box-and-whiskers plot*) voi esittää? Vastauksen tulee sisältää esimerkki monipuolisesta kahden ryhmän viiksilaatikkokuvaajasta selityksineen (piirrä kuvaaja ja selitä ihan kaikki sen osat jopa akseleita myöten).
- Mikä on  $p$ -arvon tarkka määritelmä, minkälaisia arvoja se voi saada, mihin sen arvoa verrataan ja miten se tulee esittää tutkimusartikkelissa kokeen tulosta esitettäessä?
- Mitä tarkoittaa kausaalisuus? Minkälaisella tutkimustavalla kausaalisuutta voidaan tutkia? Anna esimerkki kausaalisuutta selvittävästä tutkimuksesta.
- Mitä tarkoittaa vastesuhde (*response ratio*) ja minkälaisessa tilanteessa sitä käytetään?
- Mitä profilikuvaajalla (*profile plot*) pyritään esittämään ja missä tilanteessa sellaisia käytetään? Anna kaksi erilaista esimerkkiä tällaisesta kuvaajasta selittäen samalla, minkälaista tilannetta ne kuvaavat.

**2. Koesuunnittelun/tilastollinen testauksen periaatteet.**

Olet suunnittelemassa kokeellista tutkimusta, jolla pitäisi selvittää: 1) vaikuttaako valittu vierasaine valitun kasvin biomassaan (massoissa yli 100 gramman ero) ja versojen pituuteen (pituudessa yli 5 cm:n ero), 2) kuinka suuri annos vierasainetta tarvitaan, jotta tilastollisesti merkitsevää vaikutusta havaitaan. Tarkoituksena on annostella vierasaine kasvien kasteluveden joukkoon. Käytössäsi on 40 yksilöä valitun kasvin lajiketta A ja 40 yksilöä kasvin lajiketta B. Tarkoituksena on tehdä vain yksi koe, mutta tulosten tulisi päteä molemmille lajikkeille.

Mitä kaikkia vaiheita sinun tulee tehdä koetta suunnitellessasi, koetta käytännössä toteuttaessasi ja kokeessa kerättyjä mittaustuloksia analysoidessasi (tilastollinen merkitsevyytestaus), jotta kokeesta saataisiin mahdollisimman tarkat ja edellä esitettyihin tavoitteisiin nähden mielekkäät tulokset? Kerrothan kokeen suunnittelun, toteutuksen ja tulosten analysoinnin kaikki työvaiheet järjestyksessä numeroituja listoja käyttäen. Kiitos! (5 p)

**3. Koesuunnittelutehtävä.**

Elinympäristönsä tutkimuksesta kiinnostunut eläkeläinen nimeltään Järvinen oli huolestunut ihan kesämökkinsä vierestä menevällä laivareitillä kulkevien veneiden ja laivojen aiheuttamien päästöjen vaikutuksesta järiveden puhtauteen. Hän pohti mm. sitä, voiko järvivettä käyttää vaimon useita kukkalajeja sisältävän kukkapenkin kasteluun vai tulisiko sen sijaan käyttää puhtaampaa kaivovettä.

Hän päätti tutkia kasteluveden vaikutusta kasvien kasvuun perustamalla oman kukkapenkin. Hän laittoi viiteentoista 20 cm ja korkeudeltaan 25 cm ruukkuun samanlaista kukkamultaa yhtä suuret määrät ja jakoi ne kolmeen 5 ruukun ryhmään ja merkitsi ryhmät kirjaimilla A–C. Ensimmäiselle 5 ruukun ryhmälle A annettaisiin pelkkää kaivovettä, toiselle 5 ruukun ryhmälle B puoli järvi- ja puoli kaivovettä ja kolmannelle 5 ruukun ryhmälle C pelkkää järvivettä.

Kasviksi hän valitsi vaimon suosikkikukan, orvokin, koska hän ajatteli, että ainakin niiden tulisi menestyä. Hän istutti kuhunkin ruukkuun kolme orvokinsiementä ja valitsi kasvin kannalta suotuisan kasvupaikan. Paikan olosuhteet vastasivat vaimon kukkapenkkiä. Ruukut hän sijoitti ryhmittäin kasvupaikalle.

Hän kasteli ruukut säännöllisesti noudattaen kiinnitettyä kastelusuunnitelmaa. Kasvukauden aikana hän tutki itäneiden kasvien lukumäärän ruukuittain, ja kauden keskellä hän laski kukkien ja kellastuneiden lehtien määrät kasveittain. Kukkimiskauden päättyessä hän vielä poisti kasvit ruukuista ja punnitsi kuivattujen kasvien kokonaismassat. (Pisteitys 1 p / kohta eli yhteensä 5 p)

- Mitä koe tarkasti ottaen tutkii? Määrittele sille myös mielekkäät tilastolliset hypoteesit.
- Miten tässä kokeessa määritellään koeyksilö? Luettele myös kokeen käsittely(t), tasot ja vasteet!
- Montako lolikoa ja toistoa kokeessa on? Onko koe balansoitu?
- Millä tilastomenetelmillä tilastollisia hypoteeseja voidaan testata ja miten vastearvot (määrä- ja punnitus-tulokset) tulee muodostaa ja käsitellä niitä varten?
- Huomioitaessa Järvisen alkuperäinen huolenaihe (sopiiko järvivesi vaimon kukkapenkin kasteluun?), mikä suoritettussa kokeessa on hyvää ja mitä asioita tulisi ehdottomasti parantaa?

c) Olet tekemässä tutkimusdatan analyysiä ja olet saanut yhtenä tuloksena oikeisen taulukon.

ANOVA oksa ryhmien välillä on

Dependent Variable: pituus keskiarvo-  
neliö

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	847,047	5	169,409	14,370	,000
Within Groups	282,928	24	11,789		
Total	1129,975	29			

*5 ryhmä.  
24 kpl /ryhmä.*

Mitä taulukon numerot tarkoittavat? Kerro, jos voit päätellä niistä jotakin suoritetusta kokeesta, esim. montako ryhmää siinä on ollut ja ryhmien koko, jos ryhmät ovat samankokoisia. Miten tulokset saamasi tuloksen ja tulisiko data-analyysiä jatkaa jollakin tavalla? (2 p)

d) Olet tehnyt kovarianssianalyysin tilanteessa, jossa sinulla on ollut yksi käsittely (Treatment) ja yksi kovariaatti (x) ja saanut tuloksena taulukot:

#### Tests of Between-Subjects Effects

Dependent Variable: y

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	53,507 <sup>a</sup>	3	17,836	29,432	,000
Intercept	58,054	1	58,054	95,798	,000
x	16,555	1	16,555	27,319	,000
Treatment	16,932	2	8,466	13,970	,000
Error	10,302	17	,606		
Total	881,000	21			
Corrected Total	63,810	20			

a. R Squared = ,839 (Adjusted R Squared = ,810)

#### Parameter Estimates

Dependent Variable: y

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Intercept	4,698	,485	9,683	,000	3,674	5,722
x	,743	,142	5,227	,000	,443	1,043
[Treatment=15]	-1,861	,424	-4,390	,000	-2,756	-,967
[Treatment=20]	,327	,428	,762	,456	-,577	1,230
[Treatment=25]	0 <sup>a</sup>	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

Mitä kovariaatin avulla tässä analyysissä pyritään tekemään (vastaathan havainnollisesti, jos pystyt)? Miten sen rooli eroaa esim. käsittely(i)stä? Mitä voit saatujen tulostaulukoiden perusteella päätellä käsittelyn (Treatment) ja kovariaatin (x) vaikutuksesta vasteeseen (y)? Minkä käsittelyryhmän vastearvot ovat olleet suurimpia? Tulisiko tämän aineiston analysointia jatkaa tulosten perusteella jotenkin? (1 p)

## YMPA238 Kokeellisen ja yhteiskunnallisen ympäristöntutkimuksen tilastolliset menetelmät (2012)

Tentti 12.10.2012, tentaattori Anssi Lensu

**Tehtävät – Tulee valita/tehdä 4 tehtävää 5:stä, ei enempää!**

### 1. Periaatteet/käsitteet. Vastaa lyhyesti mutta selkeästi seuraaviin kysymyksiin/tehtäviin.

(Pisteytys 1 p / kohta)

- Mitä kaikkea viiksilaatikkokuvaaja voi esittää? Vastauksen tulee sisältää esimerkki monipuolisesta viiksilaatikkokuvaajasta selityksineen.
- Mikä on  $p$ -arvo, miten se tulee tulkita ja miten se tulee esittää tutkimusartikkelissa?
- Äärellisen populaation korjaus. Milloin korjausta tarvitaan ja mitä tunnuslukua laskettaessa sitä käytetään?
- Miten parametrittomat testit eroavat parametrisista testeistä ja milloin parametrittomia tulee käyttää?
- Mitä profiilikuvaujalla (*profile plot*) pyritään esittämään? Anna kaksi erilaista esimerkkiä tällaisesta kuvaujasta selittäen samalla, minkälaisia tilannetta ne kuvaavat.

### 2. Tutkimusmenetelmät. Vastaathan tässäkin lyhyesti mutta kattavasti.

- Miten kokeellinen tutkimus tulee suorittaa ja (erityisesti) miten se eroaa passiivisesta empiirisestä seurannasta? (1 p)
- Minkä vuoksi kokeellinen tutkimus antaa parempia tai varmempia tuloksia erityisesti luonnontieteiden piirissä kuin passiivinen empiirinen seuranta? Mitä lisäarvoa sillä saavutetaan? (1 p)
- Minkälaisissa tilanteissa/minkä seikkojen vuoksi kokeellista tutkimusta ei aina voida käyttää? (1 p)
- Luettele ja selitä lyhyesti hyvän kokeen suunnittelun kolme tärkeintä periaatetta. (2 p)

### 3. Koesuunnitelman pohdintatehtävä.

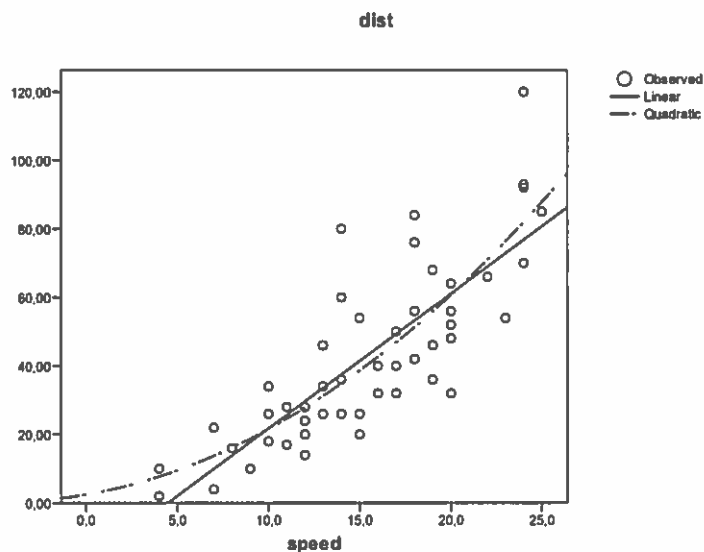
Metsämaan omistaja nimeltään Pennonen oli kiinnostunut puuntaimien kasvun nopeuttamisesta lannoittamalla ja tarjolla olevien kahden lannoiteaineen kustannustehokkuudesta. Tavoitteena hänellä oli selvittää, antaako kaliumi lannoiteaine A paremman kasvutuloksen puuntaimille kuin edullisempi aine B, tai onko aineilla ylipäättään taimien kasvua edistävää vaikutusta.

Hän päätti tutkia asiaa istuttaessaan uusia männyntaimia aiemmin hakatulle alueelle. Hän valitsi sellaisen alueen kasvupaikaksi, jossa maaperä oli tasalaatuisia isommalla alueella ja jota eivät lähialueiden puut varjostaneet. Hän jätti yhden metrin etäisyydet istutuskolojen väliin ja pystytti kasvualueen ympärille aidan pitämään eläimet poissa taimikon luota.

Koetta varten hän sai 37 suunnilleen samanlaista taimia samalta taimitarhalla ja päätti alueelliselta metsänhoitajalta saamansa ohjeen mukaisesti sijoittaa lannoiteaineet istutuskoloihin laitettavan tasalaatuisen mullan joukkoon. Tarkoituksena oli mitata taimien korkeudet kokeen alussa heti istutuksen jälkeen, sekä yhden vuoden kuluttua ja tutkia kasvun määrää  $\Delta y = y_{\text{loppu}} - y_{\text{alku}}$ .

(Pisteytys 1 p / kohta)

- Jotta koe vastaisi Pennosen alkuperäiseen kysymykseen mahdollisimman hyvin ja antaisi tarkan tuloksen, minkä kokosiin ryhmiin taimet tulisi jakaa ja minkälaisia käsittelyitä ryhmille tulisi tehdä? Miten taimikohtaiset sijoituspaikat ja käsittelyt tulee valita?
- Minkä tyyppinen koe tässä on kyseessä ja minkälaiset otokset sen tuloksena muodostuu? Mikä on/mitkä ovat mielekkäät tilastolliset nollahypoteesi(t)?
- Montako käsittelyryhmää, lohkoa, faktoria ja faktorikohtaista tasoa siinä on?
- Millä menetelmällä ryhmien välisiä eroja tulee tutkia, jos tavallisimmat parametrinen testausmenetelmien ennako-oletukset täyttyvät? Jos tilastollisesti merkitseviä eroja todetaan olevan, miten tulosten analyysiä voidaan/tulisi jatkaa?
- Mitä hyvää koejärjestelyssä on ja miltä osin kokeen voisi tehdä paremmin?



Kumpi malli sopii paremmin dataan ja minkä asioiden vuoksi? Paljonko nopeuden yhden yksikön lisäys vaikuttaa jarrutusmatkaan? (1 p)

c) Olet tekemässä tutkimusdatan analyysiä ja olet saanut yhtenä tuloksena oheisen taulukon.

#### ANOVA

Dependent Variable: pituus

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	87,047	5	17,409	1,477	,234
Within Groups	282,928	24	11,789		
Total	369,975	29			

Mitä taulukon numerot tarkoittavat? Kerro, jos voit päätellä niistä jotakin suoritetusta kokeesta, esim. montako ryhmää siinä on ollut ja ryhmien koko, jos ryhmät ovat samankokoisia. Miten tulkitset saamasi tuloksen ja tulisiko data-analyysiä jatkaa jollakin tavalla? (2 p)

d) Olet tutkimassa tilannetta, jossa sinulla on kontrolliryhmä (1) ja kaksi käsittelyryhmää (2, 3), ja todennut jonkin ryhmän poikkeavan jostakin toisesta tilastollisesti merkitsevästi. Olet edelleen tehnyt Dunnettin testin selvittääksesi sen, mikä/mitkä käsittelyryhmät poikkeavat kontrollista ja saanut oheisen taulukon:

#### Multiple Comparisons

Measure: MEASURE\_1

Dunnett t (2-sided)<sup>a</sup>

(I) Käsitely	(J) Käsitely	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
2	1	,5328	,56872	,546	-,7406	1,8062
3	1	-3,3557*	,56872	,000	-4,6291	-2,0823

Based on observed means.

\*. The mean difference is significant at the ,05 level.

a. Dunnett t-tests treat one group as a control, and compare all other groups against it.

Minkä ryhmien erot ovat tilastollisesti merkitseviä ja kumpaan suuntaan käsittelyryhmät eroavat kontrollista? Eron suunta tulee kertoa siinäkin tapauksessa, vaikka ero ei olisi tilastollisesti merkitsevä. (1 p)

## YMPA238 Kokeellisen ja yhteiskunnallisen ympäristöntutkimuksen tilastolliset menetelmät (2011)

Tentti 14.10.2011, tentaattori Anssi Lensu

**Tehtävät – Tulee valita/tehdä 4 tehtävää 5:stä, ei enempää!**

1. **Periaatteet/käsitteet.** Vastaa lyhyesti mutta selkeästi seuraaviin kysymyksiin/tehtäviin.  
(Pisteytys 1 p / kohta)
  - a) Määrittele tilastollinen nollahypoteesi. Kerro myös, millä tavoin se tulee määritellä? Anna esimerkki nollahypoteesista.
  - b) Mitä tarkoittaa keskiarvon (esim. 95 %:n) luottamusväli? Kerro sen tulkinta.
  - c) Miten parametrittomat testit eroavat parametrisista testeistä ja milloin parametrittomia tulee käyttää?
  - d) Mikä on kokeellisen tutkimuksen faktori? Anna esimerkki faktorista.
  - e) Milloin tukija sortuu näennäistoistojen käyttöön? Anna esimerkki tällaisesta tilanteesta.
2. **Tutkimusmenetelmät.** Vastaathan tässäkin lyhyesti mutta kattavasti.
  - a) Miten kokeellinen tutkimus tulee suorittaa ja (erityisesti) miten se eroaa passiivisesta empiirisestä seurannasta? (2 p)
  - b) Minkä vuoksi kokeellinen tutkimus antaa parempia tai varmempia tuloksia erityisesti luonnontieteiden piirissä kuin passiivinen empiirinen seuranta? Mitä lisäarvoa sillä saavutetaan? (1 p)
  - c) Minkälaisissa tilanteissa/minkä seikkojen vuoksi kokeellista tutkimusta ei aina voida käyttää? (1 p)
  - d) Luettele hyvän kokeen suunnittelun kolme tärkeintä periaatetta. (1 p)
3. **Koesuunnittelutehtävä.**

Elinympäristönsä tutkimuksesta kiinnostunut eläkeläinen nimeltään Järvinen oli huolestunut ihan kesämökkkinsä vierestä menevällä laivareitillä kulkevien veneiden ja laivojen aiheuttamien päästöjen vaikutuksesta järveden puhtauteen. Hän poltti mm. sitä, voiko järvivettä käyttää vaimon kukkapenkin kasvien kasteluun vai tulisiko sen sijaan käyttää puhtaampaa kaivovettä.

Hän päätti tutkia kasteluveden vaikutusta kasvien kasvuun perustamalla oman kukkapenkin. Hän laittoi viiteentoista  $\varnothing$ 20 cm ja korkeudeltaan 25 cm ruukkuun samanlaista kukkamultaa yhtä suuret määrät. Sen jälkeen hän numeroi ruukut numeroilla 1:stä 15:een ja päätti valita satunnaisesti sen, minkälaista kasteluvettä kukin ruukku saisi. Ensimmäiselle 5 ruukun ryhmälle annettaisiin pelkkää kaivovettä, toiselle 5 ruukun ryhmälle puoliksi järvi- ja puoliksi kaivovettä ja kolmannelle 5 ruukun ryhmälle pelkkää järvivettä. Ryhmien valinta tehtiin paperilappu- ja hatusta vetämällä (1 – 15 numeroidut samanlaiset laput, viisi ensimmäistä sokkona nostettua ensimmäiseen ryhmään, viisi seuraavaa toiseen ja loput kolmanteen).

Kasviksi hän valitsi vaimon suosikkikukan, orvokin, koska hän ajatteli, että ainakin niiden tulisi menestyä. Hän istutti kuluunkin ruukkuun kolme orvokinsientä ja valitsi kasvin kannalta suotuisan kasvupaikan. Paikan olosuhteet vastasivat vaimon kukkapenkkiä. Ruukut hän sijoitti systematiikkaa vältellen kasvupaikalle ja vaihtoi niiden paikkoja viikoittain (ruukkujen paikkoja vaihdeltiin systematiikkaa vältellen).

Hän kasteli ruukut säännöllisesti noudattaen kiinnitettyä kastelusuunnitelmaa. Kasvukauden aikana hän tutki itäneiden kasvien lukumäärän ruukuittain, ja kauden keskellä hän laski kukkien ja kellastuneiden lehtien määrät kasveittain. Kukkimiskauden päättyessä hän vielä poisti kasvit ruukuista ja punnitsi kuivattujen kasvien kokonaismassat.

(Pisteytys 1 p / kohta)

  - a) Mitä koe tarkasti ottaen tutkii? Määrittele sille myös mielekkäät tilastolliset hypoteesit.
  - b) Miten tässä kokeessa määritellään koeyksilö? Luettele myös kokeen käsittely(t), tasot ja vasteet!
  - c) Minkälaisia satunnaistamisia kokeeseen liittyy? Montako lohkoa ja toistoa kokeessa on? Onko koe balansoitu?
  - d) Millä tilastomenetelmillä nollahypoteesia voidaan testata ja miten vastearvot (määrä- ja punnitustulokset) tulee tuottaa/käsitellä niitä varten?
  - e) Huomioitaessa Järvisen alkuperäinen huolenaihe (sopiiko järvivesi vaimon kukkapenkin kasteluun?), mikä suoritettussa kokeessa on hyvää ja mitä asioita tulisi parantaa?

Mitä sinun tulisi tehdä, jotta voisit käyttää tavallista Studentin *t*-testiä ryhmien tulotietojen vertaamiseen? Selitäthän yksityiskohtaisesti toimintatavan (ohjelman toimintoja ei kuitenkaan tarvitse esittää). (1 p)

c) Olet tekemässä tutkimusdatan analyysiiä ja olet saanut yhtenä tuloksena oikeisen taulukon.

#### ANOVA

Dependent Variable: pituus

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	847,047	5	169,409	14,370	,000
Within Groups	282,928	24	11,789		
Total	1129,975	29			

Mitä taulukon numerot tarkoittavat? Kerro, jos voit päätellä niistä jotakin suoritettusta kokeesta! Miten tulokset saamasi tuloksen ja tulisiko data-analyysiiä jatkaa jollakin tavalla? (2 p)

d) Olet tehnyt kovarianssianalyysin tilanteessa, jossa sinulla on ollut yksi käsittely (Treatment) ja yksi kovariaatti (*x*). Mitä voit oikeisen tulostaulukon perusteella päätellä käsittelyn ja kovariaatin merkityksestä vasteelle? Mitä kovariaatin avulla tässä analyysissä pyritään tekemään? Vastaathan havainnollisesti. (1 p)

#### Tests of Between-Subjects Effects

Dependent Variable: *y*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	53,507 <sup>a</sup>	3	17,836	29,432	,000
Intercept	58,054	1	58,054	95,798	,000
<i>x</i>	16,555	1	16,555	27,319	,000
Treatment	16,932	2	8,466	13,970	,000
Error	10,302	17	,606		
Total	881,000	21			
Corrected Total	63,810	20			

<sup>a</sup>. R Squared = ,839 (Adjusted R Squared = ,810)

## YMPA238 Kokeellisen ja yhteiskunnallisen ympäristöntutkimuksen tilastolliset menetelmät (2013)

Tentti 8.11.2013, tentaattori Anssi Lensu

**Tehtävät – Tulee valita/tehdä 4 tehtävää 5:stä, ei enempää!**

1. **Periaatteet/käsitteet.** Vastaa lyhyesti mutta selkeästi seuraaviin kysymyksiin/tehtäviin.  
(Pisteytys 1 p / kohta)

- Mitä kaikkea sirontakuvaajalla pyritään saamaan / voidaan saada selville? Miten havainnollistettavat muutujat tulee sijoittaa sirontakuvaajassa akseleille? Miten tulisi toimia esim. säädettävän veden lämpötilan ja havaitun liukenemisajan tapauksessa?
- Mitä tarkoittaa keskiarvon (esim. 95 %:n) luottamusväli? Kerro sen tarkka tulkinta.
- Mikä on kokeellisen tutkimuksen faktori? Anna esimerkki vähintään kaksi faktoria sisältävästä kokeesta.
- Olet tekemässä kyselytutkimuksen vastaajien valintaa ja päätät käyttää systemaattista otantaa (*systematic sampling*) vastaajien valintaan. Käytössäsi on kaupungin (asukkaita 50 027) aakkosjärjestetty vaaliluettelo ja haluat poimia siitä 500 henkilöä. Selitä, miten systemaattinen otanta tällöin käytännössä tehdään?
- Mitä profiilikuvaajalla (*profile plot*) pyritään esittämään? Anna kaksi erilaista esimerkkiä tällaisesta kuvaajasta selittäen samalla, minkälaista tilannetta ne kuvaavat.

2. **Koesuunnittelun/tilastollinen testauksen periaatteet.** Vastaathan tässäkin lyhyesti (esim. numeroitua listaa käyttäen) mutta selkeästi.

Olet suunnittelemassa kokeellista tutkimusta, jossa on tarkoitus tutkia, onko tietyllä vierasaineella vaikutusta valitun kasvin taimien kasvuun ja muuttuuko vaikutus, jos vierasainetta annostellaan vähän tai enemmän. Miten kokeen toteutat käytännössä, jos haluat saada siitä vastaukset em. kysymyksiin. Kerrothan tarkasti, mitä vaiheita sinun tulee tehdä koetta suunnitellessasi, koetta käytännössä toteuttaessasi ja kokeessa kerättyjä mittaustuloksia analysoidessasi (tilastollinen merkitsevyytestaus), jotta koe tulisi toteutettua mahdollisimman hyvin? Kerrothan suunnittelun ja tulosten analysoinnin kaikki työvaiheet järjestyksessä. Kiitos! (5 p)

3. **Koesuunnittelutehtävä.**

Elinympäristönsä tutkimuksesta kiinnostunut eläkeläinen nimeltään Järvinen oli huolestunut ihan kesämökkkinsä vierestä menevällä laivareitillä kulkevien veneiden ja laivojen aiheuttamien päästöjen vaikutuksesta järveden puhtauteen. Hän pohti mm. sitä, voiko järvivettä käyttää vaimon kukkapenkin kasvien kasteluun vai tulisiko sen sijaan käyttää puhtaampaa kaivovettä.

Hän päätti tutkia kasteluveden vaikutusta kasvien kasvuun perustamalla oman kukkapenkin. Hän laittoi viiteentoista  $\varnothing 20$  cm ja korkeudeltaan 25 cm ruukkuun samanlaista kukkamultaa yhtä suuret määrät. Sen jälkeen hän numeroi ruukut numeroilla 1:stä 15:een ja päätti valita satunnaisesti sen, minkälaista kasteluvettä kukin ruukku saisi. Ensimmäiselle 5 ruukun ryhmälle annettaisiin pelkkää kaivovettä, toiselle 5 ruukun ryhmälle puoliksi järvi- ja puoliksi kaivovettä ja kolmannelle 5 ruukun ryhmälle pelkkää järvivettä. Ryhmien valinta tehtiin paperilappuja hatusta vetämällä (1 – 15 numeroidut samanlaiset laput, viisi ensimmäistä sokkona nostettua ensimmäiseen ryhmään, viisi seuraavaa toiseen ja loput kolmanteen).

Kasviksi hän valitsi vaimon suosikkikukan, orvokin, koska hän ajatteli, että ainakin niiden tulisi menestyä. Hän istutti kuhunkin ruukkuun kolme orvokinsiemettä ja valitsi kasvin kannalta suotuisan kasvupaikan. Paikan olosuhteet vastasivat vaimon kukkapenkkiä. Ruukut hän sijoitti systematiikkaa vältellen kasvupaikalle ja vaihtoi niiden paikkoja viikoittain (ruukkujen paikkoja vaihdeltiin systematiikkaa vältellen).

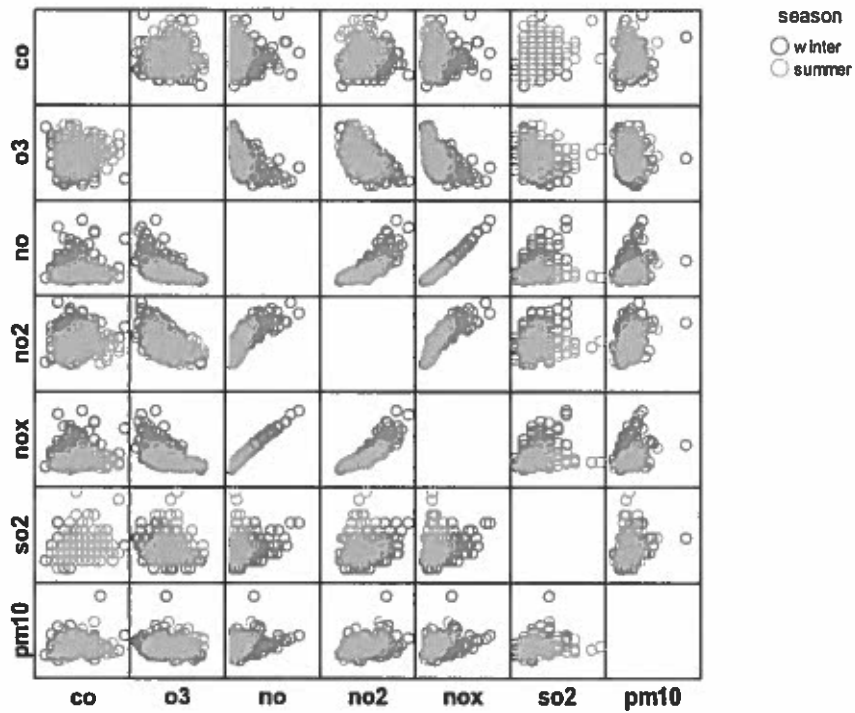
Hän kasteli ruukut säännöllisesti noudattaen kiinnitettyä kastelusuunnitelmaa. Kasvukauden aikana hän tutki itäneiden kasvien lukumäärän ruukuittain, ja kauden keskellä hän laski kukkien ja kellastuneiden lehtien määrät kasveittain. Kukkimiskauden päättyessä hän vielä poisti kasvit ruukuista ja punnitsi kuivattujen kasvien kokonaismassat.

(Pisteytys 1 p / kohta)

- Mitä koe tarkasti ottaen tutkii? Määrittele sille myös mielekkäät tilastolliset hypoteesit.
- Miten tässä kokeessa määritellään koeyksilö? Luettele myös kokeen käsittely(t), tasot ja vasteet!
- Minkälaisia satunnaistamisia kokeeseen liittyy? Montako lohkoa ja toistoa kokeessa on? Onko koe balansoitu?
- Millä tilastomenetelmillä nollahypoteesia voidaan testata ja miten vastearvot (määrä- ja punnitustulokset) tulee tuottaa/käsitellä niitä varten?
- Huomioitaessa Järvisen alkuperäinen huolenaihe (sopiiko järvi- ja kaivovettä vaimon kukkapenkin kasteluun?), mikä suoritettuna kokeessa on hyvää ja mitä asioita tulisi parantaa?

Mitä sinun tulisi tehdä, jotta voisit käyttää tavallista Studentin *t*-testiä ryhmien tulotietojen vertaamiseen? Selitäthän yksityiskohtaisesti toimintatavan (ohjelman toimintoja ei kuitenkaan tarvitse esittää). Mitä testiä käytät, jos Studentin *t*-testi ei tule mahdolliseksi? (1 p)

- c) Olet tutkimassa ilmanlaatuudataa, joka sisältää usean yhdisteen mitatut pitoisuusarvot. Tavoitteesi on vähentää tutkittavien muuttujien määrää tutkimalla muuttujien välisiä riippuvuuksia. Jos saat yhtenä osatuloksena oheisen sirontakuvaajien matriisin, mitä voit sen perusteella päätellä (minkä muuttujien välinen riippuvuus on selkeä ja minkä välinen kenties heikompi)? Miten voit määrittää riippuvuuksien tilastollisen merkitsevyyden? (1 p)



- d) Olet tekemässä tutkimusdatan analyysiä ja olet saanut yllteenä tuloksena oheisen taulukon.

ANOVA

Dependent Variable: pituus

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	847,047	5	169,409	14,370	,000
Within Groups	282,928	24	11,789		
Total	1129,975	29			

Mitä taulukon numerot tarkoittavat? Kerro, jos voit päätellä niistä jotakin suoritetusta kokeesta, esim. montako ryhmää siinä on ollut mukana ja montako havaintoa kussakin ryhmässä, jos ryhmät ovat samankokoisia. Miten tulkitset saamasi tuloksen ja tulisiko data-analyysiä jatkaa jollakin tavalla? (2 p)