Tilastotiede käytännön tutkimuksessa 17.12.2008 [KOKO KURSSI]

1. Barometri tilastojen täydentäjänä

   - barometrien yleiset muodostamisperiaatteet ja ominaisuudet
   - kuluttajabarometri verrattuna elinkustannusindeksiin.

2. Tilastollisen mallintamisen työkierto

a) Kuvaile mallintamisen yleisperiaatteet

b) Havainnollista asiaa oheisen havaintoaineiston tapauksessa, jos aineistoon sovitettaisiin lineaarista mallia. Oheen on liitetty myös graafinen havainnollistus vastaavasta residuaalianalyysistä.

3. Empiirisen tutkimuksen yleisperiaate uuden tieteellisen teorian todentamisen näkökulmasta

   - esittele periaatteellisella tasolla ja yksinkertaisen esimerkin avulla tieteellistä hypoteesia ja siitä johdettua tilastollista hypoteesia
   - miten kommentoisit Mendelin risteytyskokeiden tuloksia teorian ja empiirisen evidenssin yhteensopivuuden näkökulmasta (ks. liite) ?

4. Tilastollinen merkitsevyys ja satunnaisvaihtelun määrä kahden populaation vertailussa

a) Satunnaisvaihtelun huomioon otto testattaessa kahden populaation keskimääräisen tason eroa – yleisperiaate.

b) Satunnaisvaihtelun yhteys tilastolliseen luotettavuuteen – miten havainnollistaisit asiaa tilastotiedettä ei tunteville ?

5. Kokeellinen tiedonkeruu

a) Yleisperiaatteet, miksi kokeellista tiedonkeruuta tarvitaan

b) Havainnollista koesuunnittelun periaatteita ja ongelmia maanviljelykokeiden tapauksessa (eri lajikkeiden satoisuus / eri lannoitusmenetelmien hyödyllisyys jne).

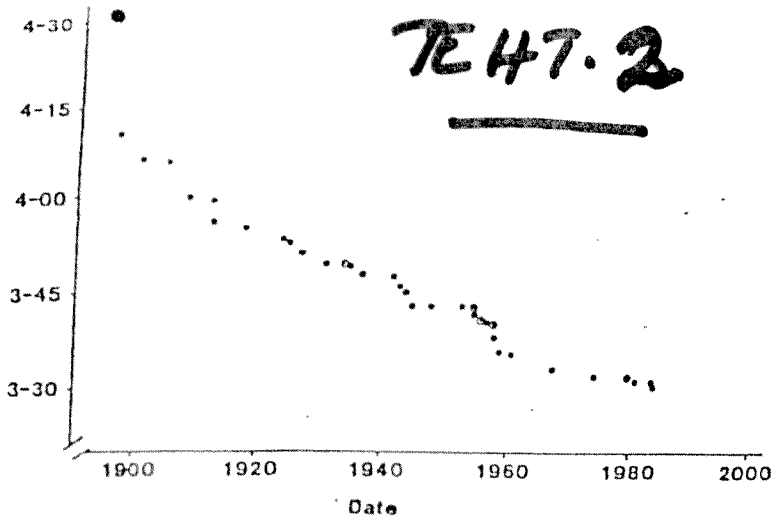


FIGURE 1 World best times for the 1500 m.

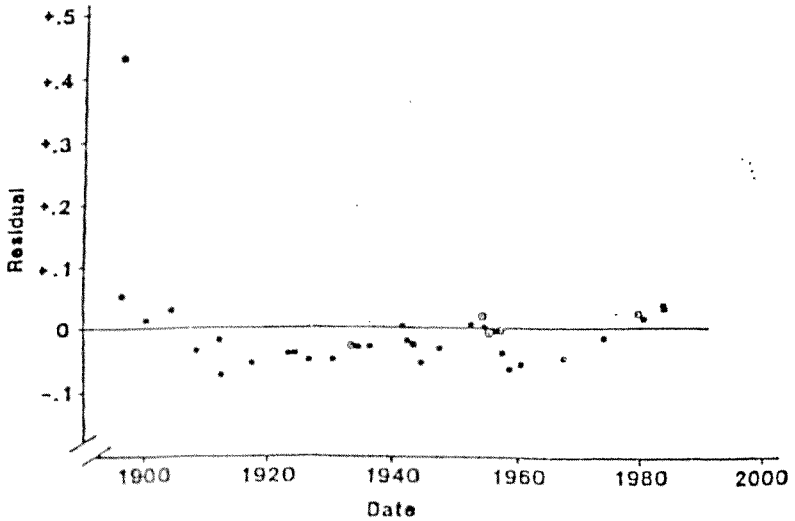


FIGURE 2 Residuals from straight-line fit.

Table 2.3.2 : χ^2 value of deviation from expected and probability ($\chi^2 \geqslant$ observed value) for each group of experiments conducted by Mendel (Source : R. A. Fisher, *Annals of Science*, 1, 1936)

Experiments to test hypothesis	degrees of freedom	χ^2_0 (observed)	$P(\chi^2 > \chi^2_0)$
3 : 1 ratios	7	2.1389	0.95
2 : 1 ratios	8	5.1733	0.74
bifactorial	8	2.8110	0.94
gametic ratios	15	3.6730	0.9987
trifactorial	26	15.3224	0.95
Total	64	29.1186	0.99987
Illustrations of plant variation	20	12.4870	0.90
Total	84	41.6056	0.99993

We see that the probabilities are extremely high in each case indicating that "data are probably faked to show a remarkably close agreement with theory". The overall probability of such good agreement is

$$1 - .99993 = \frac{7}{100000}$$

TILASTOTIEDE KÄYTÄNNÖN TUTKIMUKSESSA

KOKO KURSSI 22.1.09

1. Esittele varianssianalyysin käyttötarkoitus ja testisuureen muodostamisperiaate. Havainnollista varianssianalyysin soveltamista (ilman laskutoimituksia) 4 laboratorion mittaustarkkuuden vertailussa (ks. oheismateriaali)

2. Elinkustannusindeksin laadinta

   - yleisperiaatteet
   - miten elinkustannusindeksi poikkeaa eri aikoina tehdyn "ruokakassin" hintojen vertailusta. Miksi "ruokakassien" hintavertailun ei katsota olevan riittävää ?

3. Käsitteen määrittelyn rooli ja vaikutus tutkimustuloksiin. Esittele tätä työttömyystilastointia käyttäen, ts. käsitteenä on "työtön".
Vertaile yhteiskuntatiellisten käsitteiden määrittelyn vaikeutta/luonnetta teollisuuden/teknisten alojen käsitteen määrittelyyn ja mittaamiseen. Käytä esimerkkinä jousen kovuuden määrittelyä ja mittaamista.

4. Haastattelumenetelmän vaikutus tulosten luotettavuuteen

Esittele esimerkkien valossa posti- ja puhelinhaastattelujen käyttökelpoisuutta sensitiivisten asioiden tutkimisessa.

5. Lisäinformaation käyttö otantatutkimuksissa

- yleisperiaatteet, mitä lisäinformaation käytöllä tavoitellaan
- esittele ja vertaile tästä näkökulmasta kahta keskeistä otantamenetelmää: Ositettu otanta ja ryväsotanta. Havainnollista ko. menetelmien käyttöä sopivin esimerkein.

Tehtävä 1

ESIM. LABORATORIOIDEN VERTAILU

Neljä laboratoriota A, B, C ja D

A & B "Internal Laboratories"
C & D "External Laboratories"

Onko laboratorioissa eroja ???

Testi: Samasta näytteestä valmistettuja koe-eriä lähetetään kaikkien laboratorioiden analysoitavaksi.

Tulokset

A	B	C	D
55.9	58.7	60.7	62.7
56.1	61.4	60.3	64.5
57.3	60.9	60.9	63.1
55.2	59.1	61.4	59.2
58.1	58.2	62.3	60.3

	$\overline{X}$	s^2
A	56.52	1.352
B	59.66	1.983
C	61.12	0.592
D	61.96	4.668

EHT. 2

E S I M E R K K I :

I N F L A A T I O N M I T T A A M I N E N

O H E I S M A T E R I A A L I :

1. **Hyvinvointikatsaus 4/2001**

   Elinkustannusindeksi 50 v.

2. **Tilastokeskuksen www–tiedote**

3. **SCB www-materiaalia**

Y L E I S P E R I A A T E :

Elinkustannusindeksin laskenta perustuu

a) ajoittain päivitettävään tuotekoriin, jossa eri tuoteryhmien painot määräytyvät kulutustutkimusten perusteella, ja

b) kuukausittain tehtävään hintatietojen keruuseen

KIEMI/SYKSY 2008/50

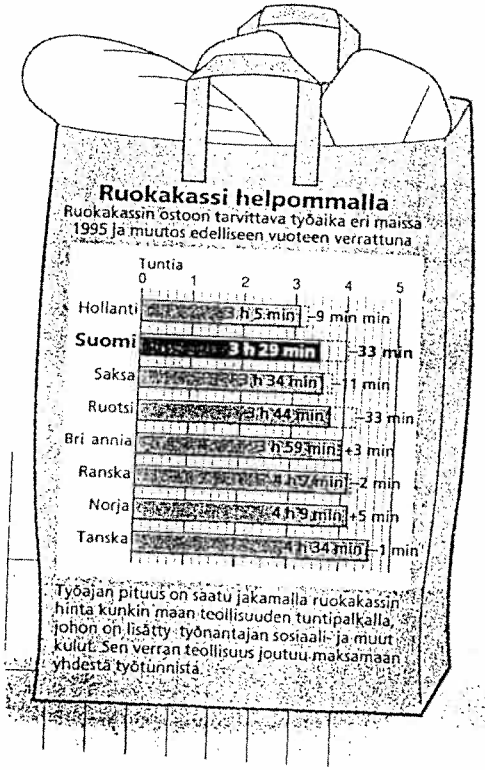
N SANOMAT
TTAJA

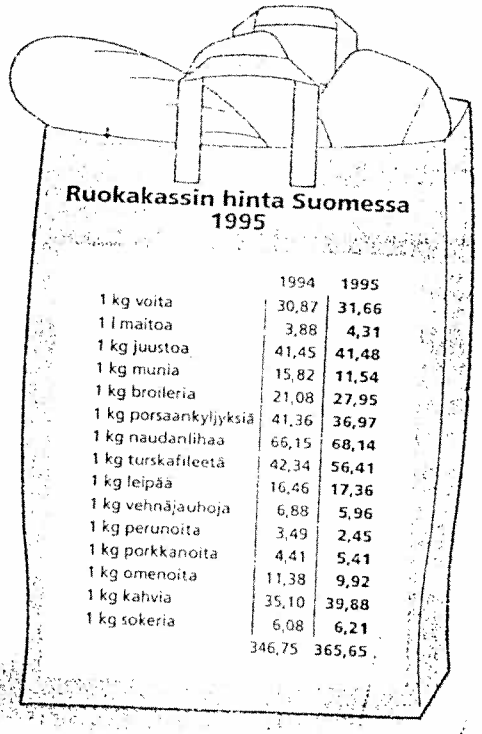
D

15. marraskuuta 1995

Sivut 1 – 12

D 1





	1994	1995
1 kg voita	30,87	31,66
1 l maitoa	3,88	4,31
1 kg juustoa	41,45	41,48
1 kg munia	15,82	11,54
1 kg broileria	21,08	27,95
1 kg porsaankyljyksiä	41,36	36,97
1 kg naudanlihaa	66,15	68,14
1 kg turskafileetä	42,34	56,41
1 kg leipää	16,46	17,36
1 kg vehnäjauhoja	6,88	5,96
1 kg perunoita	3,49	2,45
1 kg porkkanoita	4,41	5,41
1 kg omenoita	11,38	9,92
1 kg kahvia	35,10	39,88
1 kg sokeria	6,08	6,21
	346,75	365,65

TEHT-3

ESIMERKKI

Lähde:
KUME, H.: Laadun parantamisen tilastolliset menetelmät (1991)

IDEA: Havaintoaineistopohjainen ongelmanratkaisu ja informatiivinen tiedonkeruu.

Esimerkin tapauksessa:

- tehtaan ongelmana on halkeamat jousissa

Jotta ongelmaa voitaisiin analysoida täytyy löytää mitattavissa oleva suure, joka kuvaa halkeamaherkkyyttä.
Insinöörit: Käytä metallin kovuutta !!!

Halkeamien välttämiseksi

- jousen kovuuden tulisi olla toleranssien sisällä !!!

Tämän toteutumiseksi tuotannon tulisi olla tasalaatuista, ts. kovuuden varianssin tulisi olla pieni.

Ongelmanratkaisu havaintoaineistopohjaisesti:
Jäljitä ne tekijät, jotka aiheuttavat kovuusmittausten vaihtelevuutta !!!

NIEMI/SYKSY 2008/23

ESIM. Prosessin analysointi

CASE: Traktorin lehtijousen valmistusprosessi.

ONGELMA: Halkeamat jousissa.

Merkintöjä:

A_1 = pienet jouset A_2 = suuret jouset

B_1, B_2 työvuoro (2 päivässä)

P_1, P_2 jousen sijainti lämpökasittelyuunissa

P_1 = keskellä, P_2 laidassa

HUOM.
Jousen kovuus toimii hyvänä indikaattorina halkeamista. Mitataan siis kovuutta !!!

Standardi kovuudelle:

Maksimikovuus: 460 Hb
Minimikovuus: 350 Hb

Tilastotiede käytännön tutkimuksessa (koko kurssi) 3.3.2009

1. Haastattelumenetelmän vaikutus tulosten luotettavuuteen

Esittele ja vertaile esimerkkien valossa posti- ja puhelinhaastattelujen käyttökelpoisuutta sensitiivisten asioiden tutkimisessa.

2. Kaksisuuntainen luokitus

a) Luokitusten riippumattomuuden testaus – yleisperiaate
Havainnollista asiaa myös oheisen hiusten väri – silmien väri –aineiston avulla.

b) Mitä tulos "luokitukset ovat riippumattomat" merkitsee/kertoo ilmiötä tutkivalle tutkijalle.

c) Kolmas muuttuja selittäjänä ja ehdollinen riippumattomuus
Yleisperiaate. Havainnollista asiaa myös oheisen Politiikka/Urheilukiinnostus Aineiston avulla, jossa sukupuoli on selittävänä/kolmantena muuttujana.

3. Barometri tilastojen täydentäjänä:

a) barometrien yleiset muodostamisperiaatteet ja ominaisuudet
b) kuluttajabarometri verrattuna elinkustannusindeksiin.

4. Tilastollisen mallin hyvyyden arviointi

a) Tilastollisen mallintamisen työkierto.
b) Tilastollisen mallin hyvyyden arviointi - yleisperiaate.
c) Arvioi lineaarisen mallin soveltuvuutta oheiseen havaintoaineistoon (1500 m juoksun ME:n kehitys)
Käytettävissäsi on residuaalianalyysi, jonka graafinen havainnollistus on liittenä. Arvioi lineaarisen mallin soveltuvuutta myös mallinnettavan ilmiön sisällöllisin perustein.

5. Miten havainnollistaisit ei-tilastotieteilijälle seuraavia käsitteitä:

a) tilastollisesti merkitsevä tulos
b) tilastollinen riippumattomuus

NIEMI / SYKSY 2008 / 147

TEHT. 2.

Table B.5 Observed frequencies of people with a particular hair and eye colour

Eye colour	Hair colour				
	Black	Brunette	Red	Blond	Total
Brown	68	119	26	7	220
Blue	20	84	17	94	215
Hazel	15	54	14	10	93
Green	5	29	14	16	64
Total	108	286	71	127	592

ESIM. EHDOLLINEN RIIPPUMATTOMUUS (jatk.)

DATA: 1960 haastateltua
1010 miestä & 950 naista

Kokonaisaineisto:

	U+	U-	Total
P+	549	360	909
P-	511	540	1051
Total	1060	900	1960

Kysymyksessä on frekvenssiaineisto kaksisuuntaisessa luokituksessa.

KYSYMYS: Ovatko luokitukset riippumattomat ?

Entä ehdollistaminen ? Vakioidaan sukupuoli, ts. käytetään sukupuolta luokitusmuuttujana.

MIEHET

	U+	U-	Total
P+	448	201	649
P-	252	109	361
Total	700	310	1010

NAISET

	U+	U-	Total
P+	101	159	260
P-	259	431	690

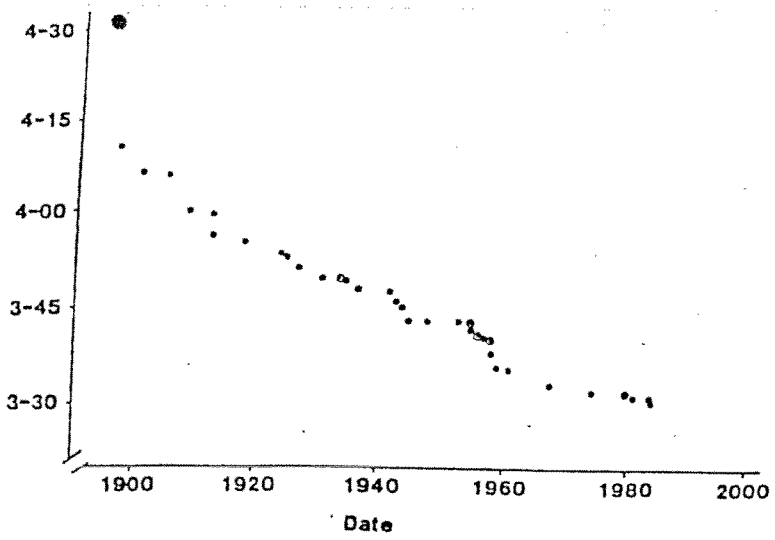


FIGURE 1 World best times for the 1500 m.

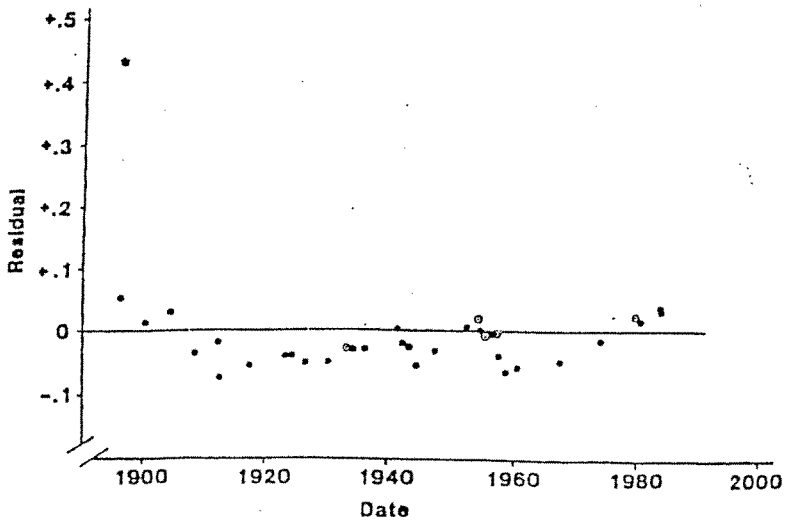


FIGURE 2 Residuals from straight-line fit.

Tilastotiede käytännön tutkimuksessa 19.5.2009

1. Tilastollisen riippumattomuuden käsite

- miten selittäisit ja havainnollistaisit tilastollista riippumattomuutta
- tutkija haluaa käyttää huippujuoksijoita kuvaavassa mallissa pikajuoksunopeutta kuvaamaan kahta muuttujaa: kunkin juoksijan edellisen kesän ennätys a) 100 m juoksussa ja b) 200 m juoksussa.
  Miten kommentoisit ko. muuttujien riippumattomuutta ja toisaalta kahden muuttujan käytön lisäarvoa verrattuna siihen, että käytettäisiin vain esim. 100 m juoksun ennätystä.

2. Kaksisuuntainen luokitus

a) Luokitusten riippumattomuuden testaus – yleisperiaate
Havainnollista asiaa myös oheisen hiusten väri – silmien väri –aineiston avulla.

b) Mitä tulos "luokitukset ovat riippumattomat" merkitsee/kertoo
ilmiötä tutkivalle tutkijalle.

c) Kolmas muuttuja selittäjänä ja ehdollinen riippumattomuus
Yleisperiaate. Havainnollista asiaa myös oheisen Politiikka/Urheilukiinnostus
Aineiston avulla, jossa sukupuoli on selittävänä/kolmantena muuttujana.

3. Empiirisen tutkimuksen yleisperiaate uuden tieteellisen teorian todentamisen näkökulmasta

- esittele periaatteellisella tasolla ja yksinkertaisen esimerkin avulla tieteellistä hypoteesia ja siitä johdettua tilastollista hypoteesia
- tarkastele sopivan testisuureen ja kriittisen alueen määrittämistä sekä satunnaisvaihtelusta johtuvan tilastollisen epävarmuuden suuruuteen vaikuttavia tekijöitä tutkittavan muuttujan keskimääräistä tasoa kosken hypoteesin tapauksessa.

4. Aineiston valikoituneisuus empiiristen tutkimusten virhelähteenä

- arvioi netti-kyselyn luotettavuutta tiedonkeruumenetelmänä, mitä mahdollisia virhelähteitä ko. menetelmään sisältyy ?
- valikoituneisuuden korjausmenetelmät aineiston analysointivaiheessa.

Huom.: Netti-kyselyllä tässä tarkoitetaan jollekin keskustelufoorumille lähetettyä tiedustelua, jossa pyydetään ottamaan kantaa (esim. Kyllä / Ei – vastauksen muodossa) tiettyyn kysymykseen ja pyydetään lähettämään vastaukset sähköpostilla kyselijän sähköpostiosoitteeseen.

5. Haastattelulomakkeen ja -tilanteen suunnittelu

   - vastauskato haastattelututkimusten ongelmana, esittele vastauskadon ongelmaa ja sen vaikutusta tulosten luotettavuuteen yleisesti.

Tarkastele lisäksi

   - kognitiivisen rasitteen vaikutukset vastauskäyttäytymiseen. ja kognitiivisen rasitteen huomioon otto kyselyn suunnitteluvaiheessa; ja
   - arkaluonteisia asioita koskevat kyselyt, esittele ja vertaile esimerkkien avulla eri haastattelumenetelmien toimivuutta.

NIEMI/SYKSY 2008/147

TEHT. 2.

Table B.5 Observed frequencies of people with a particular hair and eye colour

Eye colour	Hair colour				
	Black	Brunette	Red	Blond	Total
Brown	68	119	26	7	220
Blue	20	84	17	94	215
Hazel	15	54	14	10	93
Green	5	29	14	16	64
Total	108	286	71	127	592

ESIM. EHDOLLINEN RIIPPUMATTOMUUS (jatk.)

DATA: 1960 haastateltua
1010 miestä & 950 naista

Kokonaisaineisto:

	U+	U-	Total
P+	549	360	909
P-	511	540	1051
Total	1060	900	1960

Kysymyksessä on frekvenssiaineisto kaksisuuntaisessa luokituksessa.

KYSYMYS: Ovatko luokitukset riippumattomat ?

Entä ehdollistaminen ? Vakioidaan sukupuoli, ts. käytetään sukupuolta luokitusmuuttujana.

MIEHET

	U+	U-	Total
P+	448	201	649
P-	252	109	361
Total	700	310	1010

NAISET

	U+	U-	Total
P+	101	159	260
P-	259	431	690

Tilastotiede käytännön tutkimuksessa / Yleistentti 11.6.2009 8 op

1. Haastattelulomakkeen ja -tilanteen suunnittelu
   - kognitiivisen rasitteen vaikutukset vastauskäyttäytymiseen ja kognitiivisen rasitteen huomioon otto kyselyn suunnitteluvaiheessa
   - arkaluonteisia asioita koskevat kyselyt, esittele ja vertaile esimerkkien avulla eri haastattelumenetelmien toimivuutta.

2. Tilastollinen merkitsevyys ja satunnaisvaihtelun määrä kahden populaation vertailussa
   a) Satunnaisvaihtelun huomioon otto testattaessa kahden populaation keskimääräisen tason eroa – yleisperiaate.
   b) Satunnaisvaihtelun yhteys tilastolliseen luotettavuuteen – miten havainnollistaisit asiaa tilastotiedettä ei-tunteville ?

3. Empiirisen tutkimuksen yleisperiaate uuden tieteellisen teorian todentamisen näkökulmasta
   - esittele periaatteellisella tasolla ja yksinkertaisen esimerkin avulla tieteellistä hypoteesia ja siitä johdettua tilastollista hypoteesia
   - tarkastele sopivan testisuureen ja kriittisen alueen määrittämistä sekä satunnaisvaihtelusta johtuvan tilastollisen epävarmuuden suuruuteen vaikuttavia tekijöitä tutkittavan muuttujan keskimääräistä tasoa kosken hypoteesin tapauksessa.

4. Barometri tilastojen täydentäjänä
   - barometrien yleiset muodostamisperiaatteet ja ominaisuudet
   - kuluttajabarometri verrattuna elinkustannusindeksiin.

5. Faktorianalyysin yleisperiaatteet
   a) Esittele faktoriaanalyysin perusidea
   b) Havainnollista asiaa myös oheisen STRESSI –muuttujan ja siihen liittyvien epäsuorien kysymysten avulla.
   c) Tarkastele myös multikollineaarisuuden ongelmaa, ja usean selittävän muuttujan sisällyttämiseen liittyvää problematiikkaa/periaatteistoa lineaaristen mallien tapauksessa.