

## Tilastotiede käytännön tutkimuksessa

### 2. välikoe

1. a) Esittele ja vertaile ositetun otannan ja ryväsotannan periaatteita ja
- b) arvioi näiden menetelmien soveltuvuutta oheiseen kunta-aineistoon jos tehtävänä olisi estimoida otantapohjaisesti (esim. 8 suuruinen otos) Keski-Suomen läänin kuntien keskiväkiluku.

### 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 Poliitikka/Urheilukiinnostus Aineiston avulla, jossa sukupuoli on selittävänä/kolmantena muuttujana.

### 3. Faktorianalyysin yleisperiaatteet

Havainnollista asiaa myös oheisen STRESSI –muuttujan ja siihen liittyvien epäsuorien kysymysten avulla.

### 4. Tilastollisen mallin hyvyyden arviointi

#### Yleisperiaate

Arvioi lineaarisen mallin soveltuvuutta oheiseen havaintoaineistoon. Käytettävissäsi on residuaalianalyysi, jonka graafinen havainnollistus on liitteenä.

# TEHT. 1

## Otannan perusteikkaat 17

Taulukko 2.3. Lääni-87. Tilastotietoja Keski-Suomen läänin kunnista 31.12.1987. Ennakkotieto väestöstä on tilastotiedotuksesta VÄ.4/88 ja muut tiedot Suomen Tilastollisen Vuosikirjan (1985/86) taulukosta 16.

Järj. nro	Kunta- muoto	Kunnan nimi	Väestö 31.12.1987	Väestö- ja asuntolaskenta 1980 Asunto- kuntia	Elinkeino- indeksi
		Koko lääni	248 347	87 284	134
		Kaupungit	127 455	47 759	035
1	1	Jyväskylä	65 653	25 718	036
2	1	Jämsä	12 477	4 454	144
3	1	Jämsänkoski	8 035	2 960	044
4	1	Keuruu	12 940	4 761	125
5	1	Saarijärvi	10 570	3 575	224
6	1	Suolahti	6 231	2 309	044
7	1	Aänekoski	11 549	3 982	054
		Muut kunnat	120 892	39 525	224
8	0	Hankasalmi	6 093	2 087	313
9	0	Joutsa	4 565	1 714	324
10	0	Jyväskylän mlk.	27 162	8 309	045
11	0	Kannonkoski	2 005	728	413
12	0	Karstula	5 577	1 777	323
13	0	Klonula	2 334	617	422
14	0	Kivijärvi	2 034	627	314
15	0	Konginkangas	1 600	553	322
16	0	Konnevesi	3 483	1 150	413
17	0	Korpilahti	4 932	1 809	323
18	0	Kuhmoinen	3 475	1 467	323
19	0	Kyyjärvi	1 966	609	423
20	0	Laukaa	14 980	4 539	134
21	0	Leivonmäki	1 381	568	413
22	0	Luhanka	1 212	455	512
23	0	Multia	2 464	944	423
24	0	Muurame	5 974	1 622	035
25	0	Petäjävesi	3 730	1 270	323
26	0	Pihltpudas	5 829	1 868	413
27	0	Pylkönmäki	1 348	476	412
28	0	Sumlainen	1 378	473	422
29	0	Säynätsalo	3 322	1 194	053
30	0	Toivakka	2 417	828	314
31	0	Uuralainen	2 835	873	323
32	0	Vittasaari	8 796	2 968	323

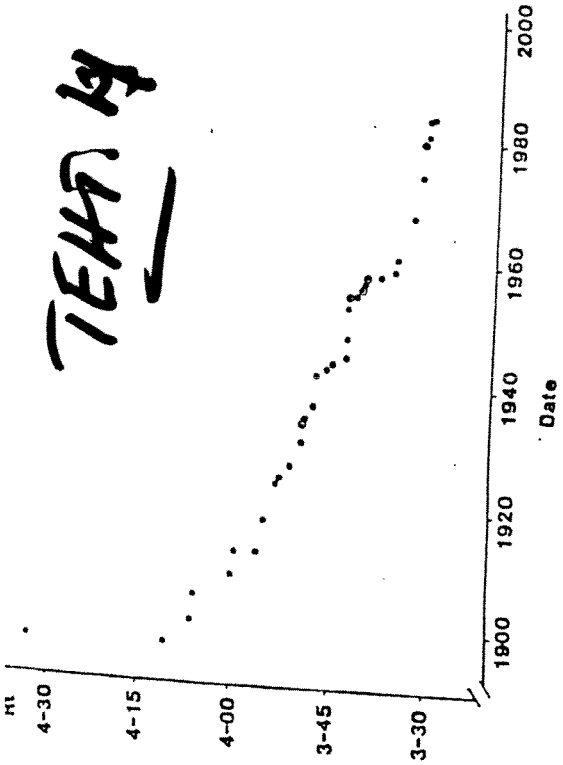


FIGURE 1 World best times for the 1500 m.

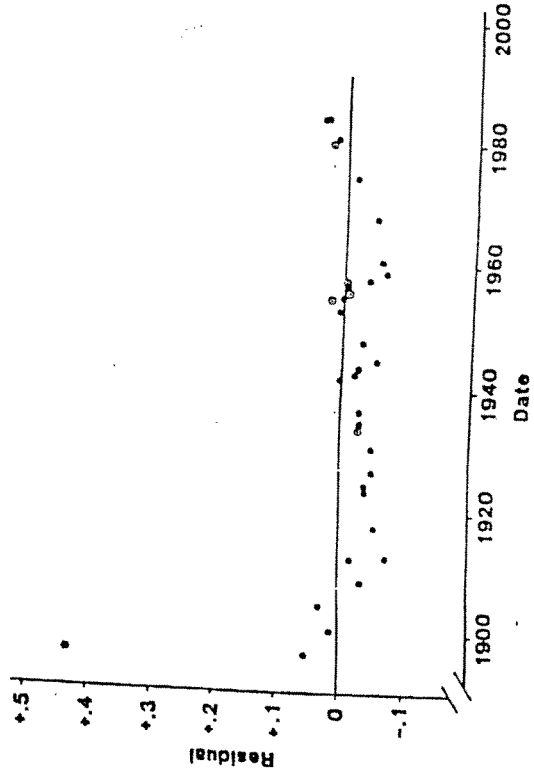


FIGURE 2 Residuals from straight-line fit.

DATA: 1960 haastateltua  
1010 miestä & 950 naista

Kokonaisaineisto:

	U+	U-	Total
P+	549	360	909
P-	511	539	1051
Total	1060	900	1960

Kysymyksessä on frekvenssiaineisto  
kaksisuuntaisessa luokituksessa.

KYSYMYYS: 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

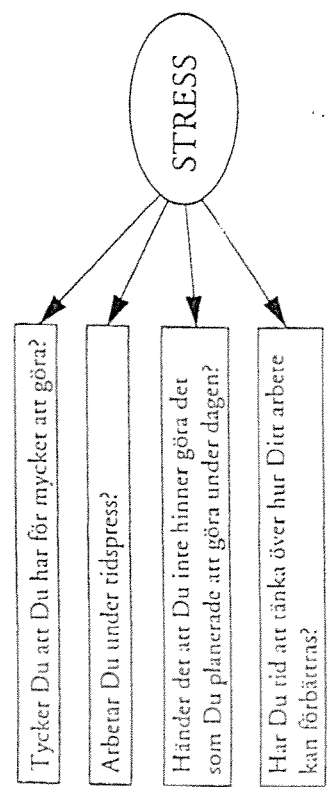
	U+	U-	Total
P+	101	159	260
P-	259	431	690
Total	360	590	950

108 → 108.  
NIEMI / SYKSY 2008 / 147  
TEHT. 2.

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

Eye colour	Hair colour				Total
	Black	Brunette	Red	Blond	
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

TEHT. 3.



## Tilastotiede käytännön tutkimuksessa

Jakson II koe, Seppo Laaksonen

Elokuu 2007

Vastaa kuuteen (6) kysymykseen. Ainakin yksi vastauksista tulee valita kahdesta viimeisestä vaihtoehdosta.

1. Eräs lehti julkaisee aikasarjan ensilumen tulemisesta tietylle paikkakunnalle sadalta vuodelta. Lehden otsikko on ”Ensilumi tulee aina vain myöhemmin.” Et automaattisesti usko väitettä vaan haluat tutkia asiaa. Miten tutkit sitä? Kertooko tuollainen aikasarja jostakin isommasta ilmiöstä?
2. ISI (International Statistical Institute) on luonut tilastoeettiset säännöt vuonna 1985 (löytyvät netistä ISI:n sivulta). Niitä uudistetaan parhaillaan. Kuvaa keskeiset piirteet näistä (engl. DECLARATION ON PROFESSIONAL ETHICS).
3. Köyhyyden mittaamisen menetelmät joita on käytetty Euroopan Unionin maissa? Esitä kaksi selkeästi erilaista menetelmää ja kerro riittävän tarkasti myös miten ne voidaan käytännössä toteuttaa eli miten hankkia aineisto ja estimoida?
4. Irakin kuolleisuustutkimuksen menetelmän pääideat ja sitä kohtaan esitetty kritiikki.
5. Mitä tarkoittavat estimaattori ja estimaatti? Kerro lisäksi kummastakin kaksi päätyyppiä. Esimerkki ja/tai kuva voi auttaa asian havainnollistamisessa.
6. Onnellisuuden mittaaminen ja selittäminen. Sisällytä mm. seuraavia asioita tähän:
  - Miten sitä surveyllä voidaan mitata?
  - Pohdiskele esittelemäsi mittarin laatua
  - Minkälaisia mallittamisen mahdollisuuksia on käytettävissä?
  - Mikä on U-käyrä tässä yhteydessä, ja onko se pätevä?
7. Tilastollisia tutkimusaineistoja (mikroaineistoja) on suhteellisen helppo muodostaa jos on olemassa sopiva kehikko (esim. rekisteri) josta tutkimusyksiköt poimitaan joko otannalla tai kokonaistutkimuksena, mutta näin ei ole aina ja silti halutaan tutkia tärkeitä ilmiöitä. Esitä kolme erilaista tilanne-esimerkkiä tällaisesta sekä selosta näistä yhden perusteet tarkasti.
8. Palkkamallin käyttö sukupuolten palkkaerojen tutkimisessa. Anna huonompia ja parempia metodisia vaihtoehtoja asian selvittämiseksi sekä mallittamisen että mittaamisen kannalta. Alla luentomateriaalista otettu sivu josta on hyötyä vastaamisessa.

## Palkkamalli 6

Lopetamme tämän jälkeen mallittamisen absoluuttiselle muuttujalle **mwage** koska siinä on ongelmia joista on kerrottu aikaisemmin. Siirrymmekin selittämään muuttujia **logmwage** =log(mwage) ja **logextra** ja teemme perä perää malleja joissa on lisätty selittäjiä. Esitämme ohessa tulokset vain miesten estimaatille (plus  $R^2$ : lle) jolloin vertailuryhmänä ovat naiset ja estimaatti heille on =0:

Selittäjät	Selittävänä logmwage	Selittävänä logextra
gender	0.247 (0.060)	0.568 (0.012)
gender age age2	0.269 (0.284)	0.525 (0.018)
gender age age2 urban tenure tenure2 whours	0.177 (0.630)	0.519 (0.023)
gender age age2 urban tenure tenure2 whours occu	0.179 (0.724)	0.522 (0.074)
gender age age2 urban tenure tenure2 whours occu educ	0.145 (0.783)	0.455 (0.091)
gender age age2 urban tenure tenure2 whours occuexact educ	0.068 (0.871)	0.400 (0.280)

12. Seuraavaksi on European Social Surveystä faktorianalyysin tuloksia muuttujista joiden tulkinta näkyy metadatasta. Ne on kysytty siten että pieni arvo tarkoittaa ko. asian olevan tärkeän. Kaksi rotatointua faktoria on tulostettu. Tulkitse jompikumpi näistä, ja sen jälkeen selosta miten Suomen (Country=FI) naiset (Gender=2) ja/tai miehet (Gender=1) eroavat valitsemistasi kahdesta muusta taulukon maasta ja ovatko erot merkitseviä?

Rotated Factor Pattern

		Factor2
IPCRTIV	Important to think new ideas and being creative	-0.19470
IMPRICH	Important to be rich, have money and expensive things	0.06234
IPEQOPT	Important that people are treated equally and have equal opportunities	0.13001
IPSHABT	Important to show abilities and be admired	0.03438
IMPSAFE	Important to live in secure and safe surroundings	0.57902
IMPDIFF	Important to try new and different things in life	-0.04078
IPFRULE	Important to do what is told and follow rules	0.65377
IPUDRST	Important to understand different people	0.17826
IPMODST	Important to be humble and modest, not draw attention	0.60343
IPGDTIM	Important to have a good time	-0.01150
IMPFREE	Important to make own decisions and be free	-0.03753
IPHLPL	Important to help people and care for others well-being	0.26144
IPSUCES	Important to be successful and that people recognize achievements	0.08377
IPSTRGV	Important that government is strong and ensures safety	0.58722
IPADVNT	Important to seek adventures and have an exiting life	-0.18496
IPBHPRP	Important to behave properly	0.71408
IPRSPOT	Important to get respect from others	0.31531
IPLYLFR	Important to be loyal to friends and devote to people close	0.22057
IMPENV	Important to care for nature and environment	0.28398
IMPTRAD	Important to follow traditions and customs	0.62627
IMPFUN	Important to seek fun and things that give pleasure	0.05386

Rotated Factor Pattern

		Factor3
IPCRTIV	Important to think new ideas and being creative	0.28371
IMPRICH	Important to be rich, have money and expensive things	0.28866
IPEQOPT	Important that people are treated equally and have equal opportunities	-0.04558
IPSHABT	Important to show abilities and be admired	0.20933
IMPSAFE	Important to live in secure and safe surroundings	-0.10568
IMPDIFF	Important to try new and different things in life	0.63130
IPFRULE	Important to do what is told and follow rules	-0.03038
IPUDRST	Important to understand different people	0.16140
IPMODST	Important to be humble and modest, not draw attention	0.06957
IPGDTIM	Important to have a good time	0.71142
IMPFREE	Important to make own decisions and be free	0.29281
IPHLPPL	Important to help people and care for others well-being	0.11115
IPSUCES	Important to be successful and that people recognize achievements	0.32327
IPSTRGV	Important that government is strong and ensures safety	-0.01820
IPADVNT	Important to seek adventures and have an exiting life	0.68357
IPBHPRP	Important to behave properly	-0.05777
IPRSPOT	Important to get respect from others	0.06993
IPLYLFR	Important to be loyal to friends and devote to people close	0.12630
IMPENV	Important to care for nature and environment	0.00834
IMPTRAD	Important to follow traditions and customs	-0.02383
IMPFUN	Important to seek fun and things that give pleasure	0.79724

		Factor2				Factor3			
		Mean	rr	P25	P75	Mean	rr	P25	P75
Country	Gender								
AT	1	0.41	0.03	-0.31	1.04	-0.25	0.03	-0.96	0.38
	2	0.26	0.03	-0.45	0.92	-0.19	0.03	-0.93	0.48
BE	1	-0.02	0.03	-0.66	0.56	-0.33	0.03	-0.93	0.29
	2	-0.08	0.03	-0.77	0.55	-0.19	0.03	-0.80	0.38
CH	1	0.26	0.03	-0.43	0.90	-0.30	0.03	-0.97	0.29
	2	0.27	0.03	-0.50	0.96	-0.15	0.03	-0.77	0.41
CZ	1	0.04	0.03	-0.63	0.58	-0.09	0.03	-0.77	0.49
	2	-0.26	0.03	-0.88	0.23	0.24	0.03	-0.46	0.93
DE	1	0.27	0.03	-0.49	0.95	-0.17	0.03	-0.83	0.45
	2	0.14	0.03	-0.57	0.75	0.12	0.03	-0.50	0.74
DK	1	0.31	0.04	-0.43	0.97	-0.29	0.04	-0.94	0.34
	2	0.37	0.04	-0.40	1.14	-0.11	0.04	-0.82	0.55
EE	1	0.19	0.03	-0.45	0.77	0.10	0.03	-0.56	0.76
	2	0.03	0.03	-0.58	0.58	0.34	0.03	-0.34	1.03
ES	1	-0.14	0.03	-0.81	0.47	0.01	0.04	-0.76	0.71
	2	-0.19	0.04	-0.85	0.48	0.09	0.04	-0.72	0.88
FI	1	0.25	0.04	-0.41	0.87	0.00	0.04	-0.67	0.62
	2	0.12	0.03	-0.55	0.72	0.03	0.03	-0.66	0.67
FR	1	0.24	0.04	-0.48	0.93	-0.29	0.03	-0.92	0.39
	2	0.17	0.04	-0.58	0.89	-0.07	0.03	-0.75	0.61
GB	1	0.07	0.03	-0.59	0.73	-0.10	0.03	-0.77	0.50
	2	0.13	0.03	-0.63	0.80	0.15	0.03	-0.55	0.89



Helsingin yliopisto  
Matematiikan ja tilastotieteen laitos

**Tilastotiede käytännön tutkimuksessa, 2. jakso**  
**Loppukoe, 2. jakso, 12.11.2008**

Valitse seuraavista neljä tehtävää.

1. Määrittele (sanallisesti ja/tai kaavojen avulla):

- a) Sisällymistodennäköisyys (*inclusion probability*).
- b) Asetelmakerroin (*design effect*).
- c) Asetelmapaino (*design weight*).

2. Otanta-asetelmat

- a) Selosta yksiasteisen ryväsotannan (*one-stage cluster sampling*) periaate ja työvaiheet.
- b) Selosta implisiittisen osittamisen tarkoitus ja periaate systemaattisessa otannassa.

3. Vastauskato

- a) Selosta käsitteet harmiton (*ignorable*) ja harmillinen (*non-ignorable*) vastauskato.
- b) Selosta tilastollisen imputoinnin tarkoitus ja periaate.

4. Survey-prosessi

- a) Selosta tyypillisen empiirisen kvantitatiivisen tutkimusprosessin eli survey-prosessin päävaiheet ja vaiheiden keskeinen sisältö.
- b) Laadi omalta tieteenalaltasi esimerkki survey-prosessista päävaiheineen.

5. Katso erillinen paperi.

-----  
Kustakin tehtävästä voi saada enintään 6 pistettä (maksimipistemäärä on siten 24 pistettä).  
Alin hyväksytty pistemäärä 2. jaksolta on 12 pistettä.

Helsingin yliopisto  
Matematiikan ja tilastotieteen laitos

**Tilastotiede käytännön tutkimuksessa**  
**Loppukoe, 2. jakso, 12.11.2008**

**Vastauslomake kysymykseen 5**

Nimi: \_\_\_\_\_

5. Merkitse (x) oheiseen taulukkoon mielestäsi oikea vaihtoehto kunkin väitteen osalta (kustakin oikeasta vastauksesta 1 piste, kustakin väärästä vastauksesta  $-\frac{1}{2}$  pistettä, kustakin tyhjästä kohdasta 0 pistettä).

	Väite on...	
	oikein	väärin
a) Ositetussa systemaattisessa otannassa suhteellisella kiintiöinnillä sisällysmistodennäköisyys on sama kaikille perusjoukon alkioille		
b) Kaksiasteisen ryväsotannan tilastollinen tehokkuus on yleensä parempi kuin yksinkertaisen satunnaisotannan tilastollinen tehokkuus		
c) On hyvä ajatus korvata puuttuvat tiedot koko aineiston keskiarvolla		
d) Asetelmakerroin deff on ykköstä suurempi, kun rypäät ovat positiivisesti sisäkorreloituneita		
e) Imputoinnilla pyritään hallitsemaan yksikkökatoa ( <i>unit nonresponse</i> )		
f) Monitasomallien avulla ei voida reagoida havaintojen korreloituneisuuteen rypäissä		

Muista merkitä nimesi (selvästi, kiitos) ja palauttaa vastauslomake!

Helsingin yliopisto  
Matematiikan ja tilastotieteen laitos

**Tilastotiede käytännön tutkimuksessa, 2. jakso**  
**Loppukoe, 2. jakso, 17.12.2008**

Valitse seuraavista neljä tehtävää.

1. Määrittele (sanallisesti ja/tai kaavojen avulla):

- a) Otos (*sample*).
- b) Asetelmakerroin (*design effect*).
- c) Tehokas otoskoko (*efficient sample size*).

2. Otanta-asetelmat

- a) Selosta alkiotasoisien systemaattisen otannan toimintaperiaate ja työvaiheet.
- b) Selosta alkiotasoisien ositetun otannan toimintaperiaate ja työvaiheet, mukaan lukien eri kiintiöintimenetelmät.

3. Vastauskato

- a) Selosta käsitteet harmiton (*ignorable*) ja harmillinen (*non-ignorable*) vastauskato.
- b) Selosta tilastollisen imputoinnin tavoite ja toimintaperiaate.

4. Survey-prosessi

- a) Selosta tyypillisen empiirisen kvantitatiivisen tutkimusprosessin eli survey-prosessin päävaiheet ja vaiheiden keskeinen sisältö.
- b) Laadi omalta tieteenalaltasi esimerkki survey-prosessista päävaiheineen.

5. Katso erillinen paperi.

-----  
Kustakin tehtävästä voi saada enintään 6 pistettä (maksimipistemäärä on siten 24 pistettä).  
Alin hyväksytty pistemäärä 2. jaksolta on 12 pistettä.

Helsingin yliopisto  
Matematiikan ja tilastotieteen laitos

**Tilastotiede käytännön tutkimuksessa, 2. jakso**  
**Loppukoe, 2. jakso, 17.12.2008**

**Vastauslomake kysymykseen 5**

Nimi: \_\_\_\_\_

5. Merkitse (x) oheiseen taulukkoon mielestäsi oikea vaihtoehto kunkin väitteen osalta (kustakin oikeasta vastauksesta 1 piste, kustakin väärästä vastauksesta  $-\frac{1}{2}$  pistettä, kustakin tyhjästä kohdasta 0 pistettä).

	Väite on...	
	oikein	väärin
a) Ositetussa systemaattisessa otannassa tasakiintiöinnillä sisällysmistodennäköisyydet ovat samat kaikille perusjoukon alkioille		
b) Yksiasteisen ryväsotannan tilastollinen tehokkuus on yleensä huonompi kuin yksinkertaisen satunnaisotannan tilastollinen tehokkuus		
c) On hyvä ajatus korvata puuttuvat tiedot koko aineiston keskiarvolla		
d) Asetelmakerroin $d_{eff}$ on ykköstä pienempi, kun rypäät ovat positiivisesti sisäkorreloituneita		
e) Imputoinnilla pyritään hallitsemaan eräkatoa ( <i>item nonresponse</i> )		
f) Monitasomallien avulla ei voida reagoida havaintojen korreloituneisuuteen rypäissä		

Muista merkitä nimesi (selvästi, kiitos) ja palauttaa vastauslomake!

Helsingin yliopisto  
Matematiikan ja tilastotieteen laitos

**Tilastotiede käytännön tutkimuksessa, 2. jakso**  
**2. välikoe, uusinta 19.12.2007**

Valitse seuraavista neljä tehtävää.

1. Määrittele (sanallisesti ja/tai kaavojen avulla):

- a) Tehokas otoskoko (*efficient sample size*).
- b) Asetelmakerroin (*design effect*).
- c) Kohdeperusjoukko (*target population*).

2. Otanta-asetelmat

- a) Selosta alkiotasoisien systemaattisen otannan toimintaperiaate ja työvaiheet.
- b) Selosta suhteellisen kiintiöinnin (*proportional allocation*) ja optimaalisen kiintiöinnin (*optimal allocation*) periaate ositetussa otannassa.

3. Vastauskato

- a) Selosta käsitteet harmiton (*ignorable*) ja harmillinen (*non-ignorable*) vastauskato sekä yksikkökato (*unit nonresponse*) ja eräkato (*item nonresponse*).
- b) Miten vastauskadon eri muotoihin voidaan reagoida tiedonkeruun sekä painotuksen, tilastollisen estimoinnin ja analyysin yhteydessä? Anna esimerkkejä.

4. Survey-prosessi

- a) Selosta tyypillisen empiirisen kvantitatiivisen tutkimusprosessin eli survey-prosessin päävaiheet ja vaiheiden keskeinen sisältö.
- b) Laadi omalta tieteenalaltasi esimerkki survey-prosessista päävaiheineen.

5. Terveys 2000-tutkimus

- a) Perustele ryväotannan käyttö Terveys 2000-tutkimuksen tutkimus- ja otanta-asetelmassa.
- b) Kuvaile otanta-asetelmaan reagoivien asetelmaperusteisten tilastollisten analyysimenetelmien käyttöä Terveys 2000-tutkimuksessa. Anna esimerkkejä.

6. Katso erillinen paperi.

-----  
Kustakin tehtävästä voi saada enintään 6 pistettä (maksimipistemäärä on siten 24 pistettä).  
Alin hyväksytty pistemäärä on 12 pistettä.

Helsingin yliopisto  
Matematiikan ja tilastotieteen laitos

**Tilastotiede käytännön tutkimuksessa, 2. jakso**  
**2. välikoe, uusinta 19.12.2007**

**Vastauslomake kysymykseen 6**

Nimi: \_\_\_\_\_

6. Merkitse (x) ohaiseen taulukkoon mielestäsi oikea vaihtoehto kunkin väitteen osalta (kustakin oikeasta vastauksesta 1 piste, kustakin väärästä vastauksesta  $-\frac{1}{2}$  pistettä, kustakin tyhjästä kohdasta 0 pistettä).

	Väite on...	
	oikein	väärin
a) Systemaattisessa otannassa sisällymisdennäköisyys on sama kaikille perusjoukon alkioille		
b) Ositetussa otannassa suhteellisella kiintiöinnillä sisällymisdennäköisyys voi vaihdella		
c) Regressiomallin sovittamisessa SPSS-ohjelmistolla kannattaa skaalata painot niin, että niiden summa yli otosaineiston on perusjoukon koko $N$		
d) Asetelmakerroin $d_{eff}$ on ykköstä suurempi, kun rypäät ovat positiivisesti sisäkorreloituneita		
e) Imputoinnilla pyritään hallitsemaan eräkatoa ( <i>item nonresponse</i> )		
f) Asetelmaperusteisella ( <i>design-based</i> ) analyysillä ei voida reagoida havaintojen korreloituneisuuteen		

Muista merkata nimesi (selvästi, kiitos) ja palauttaa vastauslomake!

**Tilastotiede käytännön tutkimuksessa, 2. jakso**  
**Loppukoe, 2. jakso, 24.1.2008**

Valitse seuraavista neljä tehtävää.

1. Määrittele (sanallisesti ja/tai kaavojen avulla):

- a) Tehokas otoskoko (*efficient sample size*).
- b) Sisältymistodennäköisyys (*inclusion probability*).
- c) Kehikkoperusjoukko (*frame population*).

2. Otanta-asetelmat

- a) Selosta alkiotasoisien systemaattisen otannan toimintaperiaate ja työvaiheet.
- b) Selosta implisiittisen osittamisen periaate systemaattisessa otannassa.
- c) Selosta suhteellisen kiintiöinnin (*proportional allocation*), tasakiintiöinnin (*equal allocation*) ja optimaalisen kiintiöinnin (*optimal allocation*) periaate ositetussa otannassa.

3. Vastauskato

- a) Selosta käsitteet harmiton (*ignorable*) ja harmillinen (*non-ignorable*) vastauskato.
- b) Selosta käsitteet yksikkökato (*unit nonresponse*) ja eräkato (*item nonresponse*).
- c) Selosta yksikkökadon ja eräkadon tilastolliset hallintamenetelmät.

4. Survey-prosessi

- a) Selosta tyypillisen empiirisen kvantitatiivisen tutkimusprosessin eli survey-prosessin päävaiheet ja vaiheiden keskeinen sisältö.
- b) Laadi omalta tieteenalaltasi esimerkki survey-prosessista päävaiheineen.

5. PISA 2000 -tutkimus

- a) Perustele ryväotannan käyttö PISA-tutkimuksen otanta-asetelmassa.
- b) Perustele lineaarisen kaksitasomallin (sekamallin) käyttö lukutaidon mallintamisessa PISA-tutkimuksessa.

6. Katso erillinen paperi.

-----  
Kustakin tehtävästä voi saada enintään 6 pistettä (maksimipistemäärä on siten 24 pistettä).  
Alin hyväksytyy pistemäärä 2. jaksolta on 12 pistettä.

Helsingin yliopisto  
Matematiikan ja tilastotieteen laitos

**Tilastotiede käytännön tutkimuksessa, 2. jakso**  
**Loppukoe, 2. jakso, 24.1.2008**

**Vastauslomake kysymykseen 6**

Nimi: \_\_\_\_\_

6. Merkitse (x) oheiseen taulukkoon mielestäsi oikea vaihtoehto kunkin väitteen osalta (kustakin oikeasta vastauksesta 1 piste, kustakin väärästä vastauksesta  $-\frac{1}{2}$  pistettä, kustakin tyhjästä kohdasta 0 pistettä).

	Väite on...	
	oikein	väärin
a) Systemaattisessa otannassa sisällymistodennäköisyys on sama kaikille perusjoukon alkioille		
b) Ositetussa otannassa tasakiintiöinnillä sisällymistodennäköisyys on vakio kaikille perusjoukon alkioille		
c) Regressiomallin sovittamisessa SPSS-ohjelmistolla kannattaa skaalata painot niin, että niiden summa yli otosaineiston on perusjoukon koko $N$		
d) Asetelmakerroin $d_{eff}$ on ykköstä pienempi, kun rypäät ovat negatiivisesti sisäkorreloituneita		
e) Uudelleenpainotuksella pyritään hallitsemaan eräkatoa ( <i>item nonresponse</i> )		
f) Monitasomallin avulla voidaan reagoida havaintojen korreloituneisuuteen rypäissä		

Muista merkata nimesi (selvästi, kiitos) ja palauttaa vastauslomake!



Helsingin yliopisto  
Matematiikan ja tilastotieteen laitos

**Tilastotiede käytännön tutkimuksessa, 2. jakso**  
**Loppukoe, 2. jakso, 4.3.2008**

Valitse seuraavista neljä tehtävää.

1. Määrittele (sanallisesti ja/tai kaavojen avulla):

- a) Estimaattori.
- b) Estimaattorin keskivirhe.
- c) Sisäkorrelaatio (*intra-cluster correlation*) ryväsotannan yhteydessä.

2. Painotus ja otanta-asetelmat

- a) Selosta asetelmapainon (*design weight*) ja analyysipainon (*analysis weight*) yhtäläisyydet/erot sekä painojen konstruktioperiaate.
- b) Selosta kaksiasteisen (*two-stage*) otannan periaate ja työvaiheet.
- c) Selosta implisiittisen osittamisen periaate systemaattisessa otannassa.

3. Vastauskato

- a) Selosta käsitteet harmiton (*ignorable*) ja harmillinen (*non-ignorable*) vastauskato.
- b) Selosta käsitteet yksikkökato (*unit nonresponse*) ja eräkato (*item nonresponse*).
- c) Selosta yksikkökadon ja eräkadon tilastolliset hallintamenetelmät.

4. Terveys 2000-tutkimus

- a) Perustele ryväsotannan käyttö Terveys 2000-tutkimuksen tutkimus- ja otanta-asetelmassa.
- b) Kuvaile otanta-asetelmaan reagoivien asetelmaperusteisten tilastollisten analyysimenetelmien käyttöä Terveys 2000-tutkimuksessa. Anna esimerkkejä.

5. Katso erillinen paperi.

-----  
Kustakin tehtävästä voi saada enintään 6 pistettä (maksimipistemäärä on siten 24 pistettä).  
Alin hyväksytty pistemäärä 2. jaksolta on 12 pistettä.

Helsingin yliopisto  
Matematiikan ja tilastotieteen laitos

**Tilastotiede käytännön tutkimuksessa, 2. jakso**  
**Loppukoe, 2. jakso, 4.3.2008**

**Vastauslomake kysymykseen 5**

Nimi: \_\_\_\_\_

5. Merkitse (x) oheiseen taulukkoon mielestäsi oikea vaihtoehto kunkin väitteen osalta (kustakin oikeasta vastauksesta 1 piste, kustakin väärästä vastauksesta  $-\frac{1}{2}$  pistettä, kustakin tyhjästä kohdasta 0 pistettä).

	Väite on...	
	oikein	väärin
a) Ositetussa systemaattisessa otannassa sisällymistodennäköisyys on sama kaikille perusjoukon alkioille		
b) Kokonaismäärän estimoinnissa käytetään asetelmapainoa		
c) Asetelmakerroin $d_{eff}$ on ykköstä pienempi, kun sisäkorrelaatiokerroin on negatiivinen		
d) Uudelleenpainotuksella pyritään hallitsemaan yksikkökatoa ( <i>unit nonresponse</i> )		
e) Monitasomallin avulla ei voida reagoida havaintojen korreloituneisuuteen rypäissä		
f) PPS-otannassa sisällymistodennäköisyys on vakio kaikille perusjoukon alkioille		

Muista merkata nimesi (selvästi, kiitos) ja palauttaa vastauslomake!

Helsingin yliopisto  
Matematiikan ja tilastotieteen laitos

**Tilastotiede käytännön tutkimuksessa**  
**Loppukoe, 2. jakso, 3.4.2008**

Valitse seuraavista neljä tehtävää.

1. Määrittele (sanallisesti ja/tai kaavojen avulla):

- a) Sisällyttämistodennäköisyys (*inclusion probability*).
- b) Asetelmakerroin (*design effect*).
- c) Tehokas otoskoko (*efficient sample size*).

2. Otanta-asetelmat

- a) Selosta alkiotasoisien systemaattisen otannan toimintaperiaate ja työvaiheet.
- b) Selosta alkiotasoisien ositetun otannan toimintaperiaate ja työvaiheet, mukaan lukien eri kiintiöintimenetelmät.

3. Vastauskato

- a) Selosta käsitteet harmiton (*ignorable*) ja harmillinen (*non-ignorable*) vastauskato.
- b) Selosta käsitteet yksikkökato (*unit nonresponse*) ja eräkato (*item nonresponse*).
- c) Selosta yksikkökadon ja eräkadon tilastolliset hallintamenetelmät.

4. Survey-prosessi

- a) Selosta tyypillisen empiirisen kvantitatiivisen tutkimusprosessin eli survey-prosessin päävaiheet ja vaiheiden keskeinen sisältö.
- b) Laadi omalta tieteenalaltasi esimerkki survey-prosessista päävaiheineen.

5. Katso erillinen paperi.

-----  
Kustakin tehtävästä voi saada enintään 6 pistettä (maksimipistemäärä on siten 24 pistettä).  
Alin hyväksytty pistemäärä 2. jaksolta on 12 pistettä.

Helsingin yliopisto  
Matematiikan ja tilastotieteen laitos

**Tilastotiede käytännön tutkimuksessa**  
**Loppukoe, 2. jakso, 3.4.2008**

**Vastauslomake kysymykseen 5**

Nimi: \_\_\_\_\_

5. Merkitse (x) oheiseen taulukkoon mielestäsi oikea vaihtoehto kunkin väitteen osalta (kustakin oikeasta vastauksesta 1 piste, kustakin väärästä vastauksesta  $-\frac{1}{2}$  pistettä, kustakin tyhjästä kohdasta 0 pistettä).

	Väite on...	
	oikein	väärin
a) Ositetussa systemaattisessa otannassa tasakiintiöinnillä sisällymistodennäköisyys on aina sama kaikille perusjoukon alkioille		
b) Ryväotannan tilastollinen tehokkuus on yleensä parempi kuin yksinkertaisen satunnaisotannan tilastollinen tehokkuus		
c) Regressiomallin sovittamisessa SPSS-ohjelmistolla kannattaa skaalata painot niin, että niiden summa yli otosaineiston on otoskoko $n$		
d) Asetelmakerroin deff on ykköstä pienempi, kun rypäät ovat positiivisesti sisäkorreloituneita		
e) Imputoinnilla pyritään hallitsemaan yksikkökatoa ( <i>unit nonresponse</i> )		
f) Monitasomallin avulla voidaan reagoida havaintojen korreloituneisuuteen rypäissä		

Muista merkata nimesi (selvästi, kiitos) ja palauttaa vastauslomake!

Helsingin yliopisto  
Matematiikan ja tilastotieteen laitos

**Tilastotiede käytännön tutkimuksessa, 2. jakso**  
**Loppukoe, 2. jakso, 20.5.2008**

Valitse seuraavista neljä tehtävää.

1. Määrittele (sanallisesti ja/tai kaavojen avulla):

- a) Asetelmakerroin (*design effect*).
- b) Estimaattorin keskivirhe (*standard error*).
- c) Sisällymistodennäköisyys (*inclusion probability*).

2. Otanta-asetelmat

- a) Selosta alkiotasoisien systemaattisen otannan toimintaperiaate ja työvaiheet.
- b) Selosta alkiotasoisien ositetun otannan toimintaperiaate ja työvaiheet, mukaan lukien tärkeimmät kiintiöintimenetelmät.

3. Vastauskato

- a) Selosta käsitteet harmiton (*ignorable*) ja harmillinen (*non-ignorable*) vastauskato.
- b) Selosta käsitteet yksikkökato (*unit nonresponse*) ja eräkato (*item nonresponse*).
- c) Selosta yksikkökadon ja eräkadon tilastolliset hallintamenetelmät.

4. Survey-prosessi

- a) Selosta tyypillisen empiirisen kvantitatiivisen tutkimusprosessin eli survey-prosessin päävaiheet ja vaiheiden keskeinen sisältö.
- b) Laadi omalta tieteenalaltasi esimerkki survey-prosessista päävaiheineen.

5. Katso erillinen paperi.

-----  
Kustakin tehtävästä voi saada enintään 6 pistettä (maksimipistemäärä on siten 24 pistettä).  
Alin hyväksytyy pistemäärä 2. jaksolta on 12 pistettä.

Helsingin yliopisto  
Matematiikan ja tilastotieteen laitos

**Tilastotiede käytännön tutkimuksessa**  
**Loppukoe, 2. jakso, 20.5.2008**

**Vastauslomake kysymykseen 5**

Nimi: \_\_\_\_\_

5. Merkitse (x) oheiseen taulukkoon mielestäsi oikea vaihtoehto kunkin väitteen osalta (kustakin oikeasta vastauksesta 1 piste, kustakin väärästä vastauksesta  $-\frac{1}{2}$  pistettä, kustakin tyhjästä kohdasta 0 pistettä).

	Väite on...	
	oikein	väärin
a) Ositetussa systemaattisessa otannassa suhteellisella kiintiöinnillä sisällymistodennäköisyys on sama kaikille perusjoukon alkioille		
b) Ryväotannan tilastollinen tehokkuus on yleensä huonompi kuin yksinkertaisen satunnaisotannan tilastollinen tehokkuus		
c) Regressiomallin sovittamisessa SPSS-ohjelmistolla kannattaa skaalata painot niin, että niiden summa yli otosaineiston on otoskoko $n$		
d) Asetelmakerroin deff on ykköstä suurempi, kun rypäät ovat negatiivisesti sisäkorreloituneita		
e) Imputoinnilla pyritään hallitsemaan eräkatoa ( <i>item nonresponse</i> )		
f) Monitasomallin avulla ei voida reagoida havaintojen korreloituneisuuteen rypäissä		

Muista merkata nimesi (selvästi, kiitos) ja palauttaa vastauslomake!

Helsingin yliopisto  
Matematiikan ja tilastotieteen laitos

**Tilastotiede käytännön tutkimuksessa, 2. jakso**  
**Loppukoe, 2. jakso, 12.6.2008**

Valitse seuraavista neljä tehtävää.

1. Määrittele (sanallisesti ja/tai kaavojen avulla):

- a) Parametri.
- b) Estimaattori.
- c) Sisällymistodennäköisyys (*inclusion probability*).

2. Painotus ja otanta-asetelmat

- a) Selosta asetelmapainon (*design weight*) ja analyysipainon (*analysis weight*) yhtäläisyydet/erot sekä painojen konstruktioperiaate.
- b) Selosta kaksiaasteisen (*two-stage*) otannan periaate ja työvaiheet.
- c) Selosta implisiittisen osittamisen periaate systemaattisessa otannassa.

3. Vastauskato

- a) Selosta käsitteet harmiton (*ignorable*) ja harmillinen (*non-ignorable*) vastauskato.
- b) Selosta käsitteet yksikkökato (*unit nonresponse*) ja eräkato (*item nonresponse*).
- c) Selosta yksikkökadon ja eräkadon tilastolliset hallintamenetelmät.

4. Survey-prosessi

- a) Selosta tyypillisen empiirisen kvantitatiivisen tutkimusprosessin eli survey-prosessin päävaiheet ja vaiheiden keskeinen sisältö.
- b) Laadi omalta tieteenalaltasi esimerkki survey-prosessista päävaiheineen.

5. Katso erillinen paperi.

-----  
Kustakin tehtävästä voi saada enintään 6 pistettä (maksimipistemäärä on siten 24 pistettä).  
Alin hyväksytty pistemäärä 2. jaksolta on 12 pistettä.

Helsingin yliopisto  
Matematiikan ja tilastotieteen laitos

**Tilastotiede käytännön tutkimuksessa**  
**Loppukoe, 2. jakso, 12.6.2008**

**Vastauslomake kysymykseen 5**

Nimi: \_\_\_\_\_

5. Merkitse (x) oheiseen taulukkoon mielestäsi oikea vaihtoehto kunkin väitteen osalta (kustakin oikeasta vastauksesta 1 piste, kustakin väärästä vastauksesta  $-\frac{1}{2}$  pistettä, kustakin tyhjästä kohdasta 0 pistettä).

	Väite on...	
	oikein	väärin
a) Ositetussa systemaattisessa otannassa tasakiintiöinnillä sisällyttämistodennäköisyys on sama kaikille perusjoukon alkioille		
b) Ryväotannan tilastollinen tehokkuus on yleensä parempi kuin yksinkertaisen satunnaisotannan tilastollinen tehokkuus		
c) Regressiomallin sovittamisessa SPSS-ohjelmistolla kannattaa skaalata painot niin, että niiden summa yli otosaineiston on otoskoko $n$		
d) Asetelmakerroin $d_{eff}$ on ykköstä pienempi, kun rypäät ovat positiivisesti sisäkorreloituneita		
e) Imputoinnilla pyritään hallitsemaan eräkatoa ( <i>item nonresponse</i> )		
f) Monitasomallin avulla voidaan reagoida havaintojen korreloituneisuuteen rypäissä		

Muista merkata nimesi (selvästi, kiitos) ja palauttaa vastauslomake!



Helsingin yliopisto  
Matematiikan ja tilastotieteen laitos

**Tilastotiede käytännön tutkimuksessa, 2. jakso**  
**Loppukoe, 2. jakso, 14.8.2008**

Valitse seuraavista neljä tehtävää.

1. Määrittele (sanallisesti ja/tai kaavojen avulla):

- a) Asetelmakerroin (*design effect*).
- b) Estimaattorin keskivirhe (*standard error*).
- c) Asetelmapaino (*design weight*).

2. Otanta-asetelmat

- a) Selosta systemaattisen otannan periaate ja työvaiheet.
- b) Selosta implisiittisen osittamisen periaate systemaattisessa otannassa.

3. Vastauskato

- a) Selosta käsitteet harmiton (*ignorable*) ja harmillinen (*non-ignorable*) vastauskato.
- b) Selosta käsitteet yksikkökato (*unit nonresponse*) ja eräkato (*item nonresponse*).
- c) Selosta yksikkökadon ja eräkadon tilastolliset hallintamenetelmät.

4. Survey-prosessi

- a) Selosta tyypillisen empiirisen kvantitatiivisen tutkimusprosessin eli survey-prosessin päävaiheet ja vaiheiden keskeinen sisältö.
- b) Laadi omalta tieteenalaltasi esimerkki survey-prosessista päävaiheineen.

5. Katso erillinen paperi.

-----  
Kustakin tehtävästä voi saada enintään 6 pistettä (maksimipistemäärä on siten 24 pistettä).  
Alin hyväksytty pistemäärä 2. jaksolta on 12 pistettä.

Helsingin yliopisto  
Matematiikan ja tilastotieteen laitos

**Tilastotiede käytännön tutkimuksessa**  
**Loppukoe, 2. jakso, 14.8.2008**

**Vastauslomake kysymykseen 5**

Nimi: \_\_\_\_\_

5. Merkitse (x) oheiseen taulukkoon mielestäsi oikea vaihtoehto kunkin väitteen osalta (kustakin oikeasta vastauksesta 1 piste, kustakin väärästä vastauksesta  $-\frac{1}{2}$  pistettä, kustakin tyhjästä kohdasta 0 pistettä).

	Väite on...	
	oikein	väärin
a) Ositetussa systemaattisessa otannassa tasakiintiöinnillä sisällymistodennäköisyys on sama kaikille perusjoukon alkioille		
b) Ryväotannan tilastollinen tehokkuus on yleensä huonompi kuin yksinkertaisen satunnaisotannan tilastollinen tehokkuus		
c) Puuttuvat tiedot kannattaa korvata koko aineiston keskiarvolla		
d) Asetelmakerroin $deff$ on ykköstä pienempi, kun rypäät ovat positiivisesti sisäkorreloituneita		
e) Imputoinnilla pyritään hallitsemaan eräkatoa ( <i>item nonresponse</i> )		
f) Monitasomallien avulla voidaan reagoida havaintojen korreloituneisuuteen rypäissä		

Muista merkitä nimesi (selvästi, kiitos) ja palauttaa vastauslomake!