

Helsingin yliopisto
Sosiaalitieteiden laitos
Seppo Tammikuu 2013
Surveymetodiikan koe

Päätä viimeistään silloin kun jätät vastauksesi, kuinka moneen opintopisteeseen pyrit. Jos haluat saavuttaa perusmäärän, vastaa kahdeksaan (8) kysymykseen, jos vähemmän niin vastaavasti vähempään, minimissään neljään (4). Huomaa kuitenkin että harjoitustöiden ja tentin laajuuden pitää olla oikeassa balanssissa. Laita posti ainakin yhteen paperiin ellei weboodin posti ole paras, ja jokaiseen nimesi ja opiskelijanumerosi. Kerron tuloksesta siis postilla.

1. Eräs melko kokematon 'tutkija' teki pienimuotoisen kyselyn tietylle rajatulle tavoiteperusjoukolle. Hän jakeli kyselylomakkeet suoraan taloihin joista toivoi omistajan tai pääasujan täyttävän sen. Hän saikin yli 500 täytettyä lomaketta paperisina, ollen siis tyytyväinen. Nyt hänelle tuli ongelma, mitä tehdä seuraavaksi, koska hän halusi innolla analysoida tuloksia, minkä asiantuntija hän ei vielä myöskään ole. Yksi tehtävä hänelle on muodostaa lomakkeen vastauksista muuttujat ja antaa näille arvot (koodit) ja tallentaa kaikki asianmukaisesti tiedostoon. Liitteenä on lomakkeesta kolme kysymystä. Tee kunkin niistä osalta malli tallennusta varten.
2. Kehitysmaissa tehdään paljonkin surveytutkimuksia, joita rahoitetaan vaikkapa Suomen kehitysyhteistyörahoista. Toki Maailmanpankki ja myös ehkä EU ovat merkittävämpiä rahoittajia, ja konsultoijina on usein korkean tason asiantuntijoita. Tästä syystä kehitysmaiden surveyasetelmat ovat yleensä hyviä, joiltakin osin jopa parempia kuin kehittyneissä maissa. Tätä edistää muun muassa halvahko työvoima. Kerro Etiopian otanta-asetelman keskeiset ominaisuudet siltä osin kuin niistä kerroin kurssilla ja sen nettisivulla. Keskity siis otanta-asetelmaan.

3. Surveyn synonyymina tietyissä tilanteissa käytetään nimeä 'gallup.' Hyväksyn sen, vaikkei se olekaan tyylikäs. Mistähän nimi tulee? Oikeasta vastauksesta tuohon saat pienen bonuksen, mutta olennaisinta on selostaa antamani selostuksen perusteella:

- Mitä tiedon keruumenetelmää puoluekannatusgallupeissa nykyisin Suomessa käytetään?
- Mikä on alkuperäisen otospainotuksen (peruspainon muodostamisen) idea?
- Miten tuon otospainon avulla saatua kannatusmittausta edelleen parannetaan jotta saataisiin mahdollisimman luotettava tulos eli siis tuotetaan uusi parannettu paino? Mihin se siis perustuu?

4. Surveyestimaatin (piste-estimaatin kuten keskiarvon ja frekvenssin) hyvä osuvuus on toki erinomaisen tärkeä, mutta sille pitäisi saada myös estimaatin epävarmuutta kuvaava tunnusluku. Kerro mitä tällaisia tunnuslukuja on kurssilla käyty läpi ja selosta myös lyhyesti mitä mikäkin tarkoittaa. Sano myös jotain epävarmuuden todennäköisyydestä.

5. Alla on viisi käsitettä. Selosta niistä ainakin neljä mahdollisimman napakasti ja oikein:

- (a) Ensisijainen poimintayksikkö (psu).
- (b) Yksiasteinen otanta
- (c) Kaksivaiheinen otanta
- (d) Otoksen allokonti (kiintiöinti)
- (e) Pitkittäisaineisto.

6. Otantatiedosto on erittäin tärkeä jonka muodostaminen tapahtuu useassa vaiheessa. Kerro näistä vaiheista. Ja vielä tärkeämpänä, selosta minkälainen on hyvä otantatiedosto?

7. Sivulla 4 on ESS:stä kaksi kysymystä. Valitan että ne ovat englanniksi. Ei tässä ihmeellistä kielitaitoa tosin tarvita, kun pyydän tekemään noista kahdesta sellaisen 'yhdistemuuttujan' joka kertoo myötämielisestä suhtautumisesta immigrantteja (maahan muuttaneita) kohtaan siten että hyvin kielteinen saa arvon=0 ja hyvin myönteinen arvon=100. Jos mahdollista, esitä asia laskulausekkeena eli sellaisena jonka saa helposti SAS:iin, SPSS:ään ym ohjelmistoon. Anna muuttujalle myös osuva nimi.

8. Ota vaikkapa edellinen tehtävä pohjaksi ja kerro erilaisista mahdollisuuksista asetella asteikkoja asenteita koskeviin kysymyksiin. Ota kriittinen asennekin mukaan kun pohdit asteikkoja.

9. Nettisurveyden (nettikyselyjen) toteuttamisessa huomioon otettavia seikkoja. Selosta myös mitä ongelmia näissä kyselyissä voi tulla vastaan, ja miten ongelmista voitaisiin selvittää kohtuudella?

10. Tilastollisen editoinnin tarkoitus ja keskeiset tehtävät.

11. Viimeisellä sivulla on otanta-asetelman huomioon ottaneella regressiomallilla selitetty äidinkielellä ilmaistun tekstin ymmärtämistä = READING (PISA-testi) muutamilla muuttujilla joita on ollut myös harjoituksissa. Annapa lyhyt ja napakka tulkinta tulokselle kertoen kaikista selittäjistäkin jotain. Olennaista on myös kertoa mitä mikäkin estimaatti tarkoittaa 'tilastollisesti'. Huomioi myös epävarmuus.

12. Otanta-asetelmana on kaksiasteinen ryväotanta siten että on käytetty alueellisia (pieniä) ryppäitä. Tästä syystä alueiden sisällä ovat vastaukset jossain määrin samankaltaisia eli ryppäät ovat sisäkorreloituneita. Edellisistä tiedusteluista on päätelty että sisäkorrelaatio on 0,025 (laskettu usean muuttujan mediaanina). Päätettiin poimia pps:llä ryppäitä yhteensä 250 ja kustakin 10 poimintayksikköä eli kehikon mukaan tavoiteperusjoukon jäsentä. Siten brutto-otoskoko on 2500. Todellisuudessa kävi niin että vastauksia saatiin 60 prosentilta. Lisäksi havaittiin isohko 10 prosentin ylipeitto johtuen kuolemista tai muuttamisesta tavoiteperusjoukon ulkopuolelle ja myös kehikon virheistä. Lopulta oli myös niin, että otoksen poiminta ei ollut täysin satunnainen vaan tiettyjä ryhmiä otettiin mukaan suhteellisesti enemmän kuin toisia. Tällöin voitiin laskea että asetelmavaikutus johtuen erisuurista todennäköisyyksistä DEFFp = 1,2. Laske näillä perusteilla efektiivinen otoskoko.

How about people of a different race or ethnic group from most [country] people?

Instruction(s): Pre: STILL CARD 14

Post: Still use this card

Variable name and label: **IMDFETN** Allow many/few immigrants of different race/ethnic group from majority

Values and categories

1 Allow many to come and live here

2 Allow some

3 Allow a few

4 Allow none

7 Refusal

8 Don't know

Would you say it is generally bad or good for [country]'s economy that people come to live here from other countries?

Instruction(s): Pre: CARD 15

Post: Please use this card

Variable name and label: **IMBGECO** Immigration bad or good for country's economy

Values and categories

00 Bad for the economy

01 1

02 2

03 3

04 4

05 5

06 6

07 7

08 8

09 9

10 Good for the economy

77 Refusal

88 Don't know

Liite tehtävään 1:

Kysymys 3. Oliko siirtolapuutarhatoiminta teille tuttua jo ennen oman palstan vuokraamista?

_____ Kyllä _____ Ei

Kysymys 4. Mikäli vastasitte edelliseen kysymykseen kyllä, niin millä tavoin olitte tutustunut toimintaan?

- _____ Vanhemmilla on/oli siirtopuutarhapalsta
_____ Muilla sukulaisillani on/oli siirtolapuutarhapalsta
_____ Tuttavillani oli/on siirtolapuutarhapalsta
_____ Toiminta oli minulle muulla tavoin tuttua, miten _____
-

Kysymys 7. Mihin ikäluokkaan kuuluvat perheenjäsenenne, jotka käyttävät palstaa lisäksenne

Merkitkää lukumäärä kyseisen ikäluokan kohdalle

Ikäluokka	a) Vuokraaja itse	b) Muut perheenjäsenet
0- 7	_____	_____
8-16	_____	_____
17-24	_____	_____
25-34	_____	_____
35-44	_____	_____
45-54	_____	_____
55-64	_____	_____
65-74	_____	_____
75+	_____	_____

Tests of Model Effects SELITETTÄVÄNÄ READING eli tekstin ymmärtäminen

Effect	Num DF	F Value	Pr > F
Intercept	1	2.97E7	<.0001
ST04Q01	1	168241	<.0001
CNT	9	1.515E9	<.0001
JOYREAD (Lukemisen määrä nautinnolla)	1	1963038	<.0001
IC02Q03 (Onko internet ja jos niin käyttääkö?)	2	525897	<.0001
IMMIG	2	143711	<.0001
FAMSTRUC	2	150952	<.0001

R-square 0.3043

Estimated Regression Coefficients				
Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	459.368929	0.10538746	4358.86	<.0001
ST04Q01 Female	18.957657	0.04621878	410.17	<.0001
ST04Q01 Male	0.000000	0.00000000	.	.
CNT Belgium	19.504312	1.24224357	15.70	<.0001
CNT Denmark	-5.393639	0.00436523	-1235.6	<.0001
CNT Estonia	-0.848748	0.01403510	-60.47	<.0001
CNT Finland	28.600972	0.00735594	3888.15	<.0001
CNT Germany	1.124552	0.01532890	73.36	<.0001
CNT Japan	9.303739	0.01838094	506.16	<.0001
CNT Korea	28.509005	0.01185313	2405.19	<.0001
CNT Netherlands	18.265993	0.00670833	2722.88	<.0001
CNT Norway	5.766525	0.00507325	1136.65	<.0001
CNT Russian Federation	-41.674812	0.01764202	-2362.2	<.0001
CNT Sweden	0.000000	0.00000000	.	.
JOYREAD	35.182123	0.02511064	1401.08	<.0001
IC02Q03 No	-37.809599	0.03694231	-1023.5	<.0001
IC02Q03 Yes, I use it	-8.307090	0.04368077	-190.18	<.0001
IC02Q03 Yes, but I don't use it	0.000000	0.00000000	.	.
IMMIG First-Generation	0.383632	0.10880263	3.53	0.0004
IMMIG Native	38.131355	0.07281776	523.65	<.0001
IMMIG Second-Generation	0.000000	0.00000000	.	.
FAMSTRUC Mixed	-23.205060	0.07649796	-303.34	<.0001
FAMSTRUC Nuclear family	13.378338	0.03565884	375.18	<.0001
FAMSTRUC Single parent family	0.000000	0.00000000	.	.