

Otantamenetelmät / Syksy 2008

Loppuentti 16.12.2008

AINEOPINNOT

Valitse seuraavista neljä tehtävää.

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

- a) Tehokas otoskoko (*Effective sample size*).
- b) Imputointi (*Imputation*).
- c) Ryväs (*Cluster*).

2. Systemaattinen otanta (*Systematic sampling*)

- a) Selosta Bernoulli-tyyppisen systemaattisen otannan poimintaproseduuri. Mitä erityispiirteitä Bernoulli-otantaan sisältyy ”tavanomaiseen” systemaattiseen otantaan verrattuna?
- b) Selosta implisiittisen osittamisen menetelmä systemaattisessa otannassa. Mitä hyötyä implisiittisestä osittamisesta on? Milloin hyötyä voidaan odottaa? Perustele.

3. Ositettu otanta (*Stratified sampling*)

- a) Selosta alkiotasoisien ositetun otannan tavoite ja työvaiheet. Mitä lisäinformaatiota tarvitaan perusjoukon alkioista ositettua otantaa varten?
- b) Millaisissa tilanteissa voidaan odottaa, että ositettu otanta on totaalini estimoinnissa tehokkaampaa kuin yksinkertainen satunnaisotanta palauttamatta (SRSWOR)? Perustele vastauksesi.
- c) Selosta suhteellisen kiintiöinnin (*proportional allocation*) ja tasakiintiöinnin (*equal allocation*) periaatteet ja toteuttaminen alkiotasoisessa ositetussa otannassa.

4. PPS-otanta (*Sampling with Probabilities Proportional to Size*)

- a) Selosta systemaattisen PPS-otannan poimintaproseduuri. Mitä lisäinformaatiota tarvitaan perusjoukon alkioista systemaattista PPS-otantaa varten?
- b) Muotoile ehdot, joiden voimassa ollessa PPS-otanta on tehokkaampaa kuin yksinkertainen satunnaisotanta palauttamatta (SRSWOR).

5. Malliavusteinen estimointi (*Model-assisted estimation*)

- a) Selosta regressioestimoinnin tavoite, toimintaperiaate ja työvaiheet totaalini estimoinnin yhteydessä.
- b) Tulosuuttujan y SRSWOR-otoksesta estimoitu kokonaismäärä on $\hat{t} = 26\,440$ ja apumuuttujan z kokonaismäärän estimaatti on $\hat{t}_z = 164\,952$. Otoksesta on lisäksi saatu y :n ja z :n regressiokertoimeksi $\hat{b} = 0.152$. Perusjoukosta tiedetään apumuuttujan z kokonaismäärä; se on $T_z = 91\,753$. Konstruoi tulosuuttujan y totaalini T regressioestimaattori \hat{t}_{reg} ja laske annetuilla tiedoilla regressioestimaatti (= 15 312).

Otantamenetelmät (78143)

Loppuentti 22.1.2009

AINEOPINNOT (6 op)

Valitse seuraavista neljä tehtävää.

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

- a) Kehikkoperusjoukon alipeitto ja ylipeitto.
- b) Analyysipaino.
- c) Äärellisyyskorjaus (*Finite population correction*).

2. Systemaattinen otanta (*Systematic sampling*)

- a) Selosta Bernoulli-tyyppisen systemaattisen otannan työvaiheet, kun perusjoukossa on N alkioita ja tavoitteena on n alkion otos.
- b) Mitä erityispiirteitä Bernoulli-otantaan sisältyy ”tavanomaiseen” systemaattiseen otantaan verrattuna?
- c) Selosta implisiittisen osittamisen menetelmä systemaattisessa otannassa. Mitä hyötyä implisiittisestä osittamisesta on? Milloin hyötyä voidaan odottaa? Perustele.

3. Ositettu otanta (*Stratified sampling*)

- a) Selosta alkiotasoisien ositetun otannan työvaiheet, kun estimoitavana on muuttujan y perusjoukon kokonaismäärä (totaali).
- b) Selosta tasakiintiöinnin (*equal allocation*) periaate ja toteuttaminen alkiotasoisessa ositetussa otannassa, kun estimoitavana on muuttujan y perusjoukon kokonaismäärä.
- c) Selosta optimaalisen (Neyman) kiintiöinnin periaate ja toteuttaminen alkiotasoisessa ositetussa otannassa, kun estimoitavana on muuttujan y perusjoukon kokonaismäärä.

4. PPS-otanta (*Sampling with Probabilities Proportional to Size*)

- a) Selosta systemaattisen PPS-otannan työvaiheet, kun estimoitavana on muuttujan y perusjoukon kokonaismäärä (totaali).
- b) Mitä lisäinformaatiota tarvitaan perusjoukon alkioista kohdan a) systemaattista PPS-otantaa varten?
- c) Muotoile ehdot, joiden voimassa ollessa PPS-otanta on tehokkaampaa kuin yksinkertainen satunnaisotanta palauttamatta (SRSWOR).

5. Malliavusteinen estimointi (*Model-assisted estimation*)

- a) Selosta regressioestimoinnin tavoite, toimintaperiaate ja työvaiheet perusjoukon kokonaismäärän (totaalin) estimoinnin yhteydessä.
- b) Olkoon totaalin regressioestimaattori muotoa $\hat{t}_{reg} = \hat{b}_0 N + \hat{b}_1 T_{z_1} + \hat{b}_2 T_{z_2}$. Tulomuuttuja on UE91 ja apumuuttujat ovat HOU85 ja URB85. Sijoita liitteenä olevista SURVEYMEANS- ja SURVEYREG-tulostuksista saatavat tunnusluvut sopivasti regressioestimaattorin kaavaan ja laske regressioestimaatti.

LIITE

The SURVEYMEANS Procedure

Data Summary

Number of Observations	8
Sum of Weights	32

Statistics

Variable	Sum	Std Dev
UE91	26440	13282
HOU85	164952	87299
URB85	12.000000	5.070926

The SURVEYREG Procedure

Regression Analysis for Dependent Variable UE91

Data Summary

Number of Observations	8
Sum of Weights	32.00000
Weighted Mean of UE91	826.25000
Weighted Sum of UE91	26440.0

Estimated Regression Coefficients

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	29.7768913	19.7517828	1.51	0.1754
HOU85	0.1495578	0.0023199	64.47	<.0001
URB85	68.1072704	62.7319985	1.09	0.3136

Coefficients of Estimate "UE91 Total"

Effect	Row 1
Intercept	32
HOU85	91753
URB85	7

Analysis of Estimable Functions

Parameter	Estimate	Standard Error	t Value	Pr > t
UE91 Total	15151.9849	568.987386	26.63	<.0001

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

Otantamenetelmät (78143)

Lopputentti 3.3.2009

AINEOPINNOT (6 op)

Valitse seuraavista neljä tehtävää.

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

- a) Kehikkoperusjoukko (*Frame population*).
- b) Asetelmapaino (*Design weight*).
- c) Asetelmakerroin (*Design effect, deff*).

2. Systemaattinen otanta (*Systematic sampling*)

- a) Selosta systemaattisen otannan työvaiheet, kun perusjoukossa on N alkioita ja otoskoko on n alkioita.
- b) Selosta implisiittisen osittamisen tavoite ja toteutus systemaattisessa otannassa. Milloin implisiittisestä osittamisesta voidaan odottaa olevan hyötyä? Millaista hyötyä? Perustele.

3. Ositettu otanta (*Stratified sampling*)

- a) Selosta alkiotasoisien ositetun otannan tavoite ja työvaiheet.
- b) Selosta suhteellisen kiintiöinnin (*proportional allocation*) periaate ja toteuttaminen alkiotasoisessa ositetussa otannassa.
- c) Selosta tasakiintiöinnin (*equal allocation*) periaate ja toteuttaminen alkiotasoisessa ositetussa otannassa.

4. PPS-otanta (*Sampling with Probabilities Proportional to Size*)

- a) Selosta palauttaen-tyyppisen, kumulatiivisen summan menetelmään perustuvan PPS-otannan työvaiheet.
- b) Mitä lisäinformaatiota tarvitaan perusjoukon alkioista kohdan a) PPS-otantaa varten?
- c) Muotoile ehdot, joiden voimassa ollessa PPS-otanta on tehokkaampaa kuin yksinkertainen satunnaisotanta palauttamatta (SRSWOR).

5. Malliavusteinen estimointi (*Model-assisted estimation*)

- a) Selosta regressioestimoinnin tavoite, toimintaperiaate ja työvaiheet totaalin estimoinnin yhteydessä.
- b) Regressioestimaattorin \hat{t}_{reg} varianssiestimaattori voidaan SRSWOR-tilanteessa kirjoittaa muotoon

$$\hat{v}_{SRS}(\hat{t}_{reg}) = N^2 \left(1 - \frac{n}{N}\right) \left(\frac{1}{n}\right) s_y^2 (1 - \hat{\rho}_{yz}^2).$$

Päättele kaavan perusteella, milloin regressioestimoinnin voidaan odottaa tehostavan estimointia (siis tuottavan pienemmän varianssiestimaatin) yksinkertaiseen satunnaisotantaan (SRSWOR) verrattuna. Perustele vastauksesi.

Otantamenetelmät (78143)

Lopputentti 2.4.2009

AINEOPINNOT (6 op)

Valitse seuraavista neljä tehtävää.

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

- a) Imputointi (*Imputation*).
- b) Tehokas otoskoko (*Efficient sample size*).
- c) Asetelmakerroin (*Design effect, deff*).

2. Systemaattinen otanta (*Systematic sampling*)

- a) Selosta Bernoulli-tyyppisen systemaattisen otannan periaate ja työvaiheet, kun perusjoukossa on N alkia ja tavoitteena on n alkion otos.
- b) Mitä erityispiirteitä Bernoulli-otantaan sisältyy ”tavanomaiseen” systemaattiseen otantaan verrattuna?
- c) Selosta implisiittisen osittamisen tavoite ja toteutus systemaattisessa otannassa. Milloin implisiittisestä osittamisesta voidaan odottaa olevan hyötyä? Millaista hyötyä? Perustele.

3. Ositettu otanta (*Stratified sampling*)

- a) Selosta alkiotasaisen ositetun otannan periaate ja työvaiheet.
- b) Selosta optimaalisen (Neyman) kiintiöinnin (*Optimal allocation*) periaate ja toteuttaminen alkiotasoisessa ositetussa otannassa.
- c) Selosta tasakiintiöinnin (*equal allocation*) periaate ja toteuttaminen alkiotasoisessa ositetussa otannassa.

4. PPS-otanta (*Sampling with Probabilities Proportional to Size*)

- a) Selosta systemaattisen PPS-otannan periaate ja työvaiheet.
- b) Mitä lisäinformaatiota tarvitaan perusjoukon alkioista kohdan a) PPS-otantaa varten?
- c) Muotoile ehdot, joiden voimassa ollessa PPS-otanta on tehokkaampaa kuin yksinkertainen satunnaisotanta palauttamatta (SRSWOR).

5. Malliavusteinen estimointi (*Model-assisted estimation*)

- a) Selosta regressioestimoinnin periaate ja työvaiheet perusjoukon kokonaismäärän (totaalin) estimoinnin yhteydessä.
- b) Olkoon totaalin regressioestimaattori muotoa $\hat{t}_{reg} = \hat{b}_0 N + \hat{b}_1 T_{z_1} + \hat{b}_2 T_{z_2}$. Perusjoukossa on 32 alkia ja otoskoko on 8 alkia. Tulosuuttuja on UE91 ja apumuuttujat ovat HOU85 ja URB85. Sijoita liitteenä olevista SURVEYMEANS- ja SURVEYREG-tulostuksista saatavat tunnusluvut sopivasti regressioestimaattorin kaavaan ja laske regressioestimaatti. (Vihje: oikea estimaatti löytyy myös tulostusotteesta.)

Helsingin yliopisto
Matematiikan ja tilastotieteen laitos

LIITE

The SURVEYMEANS Procedure

Data Summary

Number of Observations 8
Sum of Weights 32

Statistics

Variable	Sum	Std Dev
UE91	26440	13282
HOU85	164952	87299
URB85	12.000000	5.070926

The SURVEYREG Procedure

Regression Analysis for Dependent Variable UE91

Data Summary

Number of Observations 8
Sum of Weights 32.00000
Weighted Mean of UE91 826.25000
Weighted Sum of UE91 26440.0

Estimated Regression Coefficients

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	29.7768913	19.7517828	1.51	0.1754
HOU85	0.1495578	0.0023199	64.47	<.0001
URB85	68.1072704	62.7319985	1.09	0.3136

Coefficients of Estimate "UE91 Total"

Effect	Row 1
Intercept	32
HOU85	91753
URB85	7

Analysis of Estimable Functions

Parameter	Estimate	Standard Error	t Value	Pr > t
UE91 Total	15151.9849	568.987386	26.63	<.0001

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