

Surveyometriikka 2013

Esimerkki Puolan ESS:n otannasta kuudennella kierroksella eli 2013

Aluksi on otteita hyväksytystä otanta-asetelmasta

Target Population, Population Coverage

Persons aged 15 years or over resident in private households in Poland.

Remark: A large part of the Polish population is permanently staying abroad for short- or long-term job migrations (80 percent of these for more than 12 months). According to estimations made by Central Statistical Office ([1], p. 3; [2], p. 23), the number of emigrants staying abroad for more than 3 months has been growing steadily: from 1 million in 2004 (2.6% of population), through 1.95 million in 2006 (5.1%), up to 2.21 million in 2008 (5.7%). After 2008, the figure has slightly dropped, to 1,99 million in 2010 (5.2%), but generally these figures may be systematically underestimated. This trend is reflected in the results of the ESS study. In 2006, 3.7% of the sample were found to be living abroad and two years later the percentage went up to 5.7%. Then, in 2010, it decreased to 5.1%. However, given that according to the official statistics, the number of emigrants has been rather stable since the beginning of 2008, we assumed the percentage of the population staying abroad to be equal to 5.5%.

Surveyometriikka 2013

Esimerkki Puolan ESS:n otannasta kuudennella kierroksella eli 2013

Sampling Frame

Computer-based National Register of Citizens (PESEL).

Sampling Design

A total sample is divided into two parts:

(1) a simple random sample (srs) for towns over 50,000 inhabitants (37.84% of population 15+);

(2) a clustered sample for towns below 50,000 and villages (62.16% of population 15+). Towns over 50,000. This part of the sample covers all 86 towns over 50,000 inhabitants. The corresponding population size is 12,275,805 individuals aged 15 or over. Individuals in each town are selected using the simple random sampling scheme, so this part of the sample produces no design effect. However, towns are divided into strata according to response rates. The estimated response rates are: 60.0% in Warsaw (the only town over 1 million inhabitants); 60.0% in towns sized from 100,000 to 999,999 inhabitants; and 65.0% in towns sized from 50,000 to 99,999 inhabitants. (These percentages were estimated using the results of previous ESS rounds in Poland.) The values of response rates have been applied during calculation of gross sample size in each town, which is proportional to the town's population 15+ and inversely related to the assumed response rate. Additionally, 7.1% was added for non-eligibility.

Surveymetodiikka 2013

Esimerkki Puolan ESS:n otannasta kuudennella kierroksella eli 2013

Towns below 50,000 and villages. This part of the sample is selected using a two-stage sampling design. During the first stage a required number of towns and villages (PSUs) is selected with probabilities proportional to the number of inhabitants aged 15+ and inversely proportional to the expected response rates. The PSUs are selected with replacement, so a PSU can be selected more than once. Response rates are assumed to be 70.0% for towns 10,000-49,999, 75% for towns below 10,000 and 80.0% for villages. When the selection of towns and villages is completed, a constant number of 4 persons (a cluster) is selected within each PSU, what constitutes the second stage of sampling. The individuals are selected following the simple random sample schema, so the second stage does not produce additional design effect. Ineligibility correction (7.1%) results in selection of additional clusters (PSUs) instead of selecting additional individuals in each cluster.

Design Effects

$$DEFF_c = 1 + (\bar{b} - 1)\rho = 1.1516$$

$$DEFF_p = 1.0146$$

$$DEFF = DEFF_c \cdot DEFF_p = (1.1516 \cdot 1.0146) = 1.1684$$

Surveyometriikka 2013

Esimerkki Puolan ESS:n otannasta kuudennella kierroksella eli 2013

Remarks

$DEFF_c = 1.0000$ in towns over 50,000 (not clustered part of a sample)

$DEFF_c = 1.1516$ (total; including weighting by sample shares in population categories).

In the clustered part of the sample only, \bar{b} stands for the actual cluster size, which is equal to 3.2 for rural, 3 for towns with less than 10,000 inhabitants, and 2.8 for towns with population between 10,000 and 49,999 inhabitants. The estimated value of ρ (intra-cluster correlation) is 0.13. This is the average value of ρ calculated separately for selected items in the ESS4 (183 items) and ESS5 (90 items).

Target Response Rate & Rate of Ineligibles

$rr = 69.74\%$

$ri = 7.1\%$

Remarks

An estimated total response rate (rr) weighted total of response rates calculated independently for categories of towns of different sizes and for villages (see above for details).

Estimated rate of ineligibles = 7.1%

5.5 % (emigrants)+

0.8 % (situations in which selected subjects deceased between sampling time and a day of interviewer's visit) +

0.3 % (people living in institutions) +

0.5 % (selected address turns out not to exist; see above for details)

Surveyometriikka 2013

Esimerkki Puolan ESS:n otannasta kuudennella kierroksella eli 2013

Sample Size

$$\begin{aligned}n_{eff} &= 1500 \\n_{net} &= 1500 \cdot DEFF = 1500 \cdot 1.1684 = 1752 \\n_{gross} &= \frac{n_{net}}{rr(1 - ri)} = \frac{1752}{0.6974 \cdot (1 - 0.071)} = 2705\end{aligned}$$

Remarks

Round 1 (2002): gross sample size = 2978 (cluster size = 12; ineligibility rate = 5%; $\rho = 0.02$)

Round 2 (2004): gross sample size = 2399 (cluster size = 4; ineligibility rate = 2%; $\rho = 0.05$)

Round 3 (2006): gross sample size = 2574 (cluster size = 4; ineligibility rate = 5%; $\rho = 0.05$)

Round 4 (2008): gross sample size = 2428 (cluster size = 4; ineligibility rate = 5%; $\rho = 0.05$)

Round 5 (2010): gross sample size = 2661 (cluster size = 4; ineligibility rate = 7%; $\rho = 0.12$)

Surveymetodiikka 2013

Esimerkki Puolan ESS:n otannasta kuudennella kierroksella eli 2013

Suunnitelma oli tehty vuonna 2012. Kenttätyön jälkeen vuoden 2012 lopulla pääosin saatiin siis data kerättyä ja myös tiedettiin ketkä tosiasiallisesti vastasivat ja kuinka moni brutto-otoksesta oli ylipeittoa (ineligible). Siis nyt voitiin muodostaa otantatiedosto. Puolan tiedosto on varsin hyvä kuten seuraavassa näytän (monen maan tiedosto ei kelpaa ja joudumme keksimään miten saada se kunnolliseksi). Aluksi seuraavalla sivulla on tiedoston muutamista kohdista valokuvia. Selitän termejä myös.

Tunnus Sisältymistodennäköisyydet Osite

Tulos Apumuuttujat

1=vastannut

$$\pi_k = \frac{nx_k}{\sum_U x_k} \frac{m}{x_k}$$

Tuo ekakaava on oikeana se mikä on kirjan pdf:ssä jostain teknisestä syystä sotkuin

Responden identification number	Country	Selection probability at first stage (village and town below 50,000; town over 50,000; PROB1=1)	Selection probability at second stage (respondent in PSU)	NO AP	N A	First stage unit	NO AP	NO AP	Explicit stratum (villages and towns below 50,000; STRATEX1 is combination of NUTS2 and urbanization category; every town over 50,000 is separate value of STRATEX1)	N AP	N A	N AP	N AP	N A	Summary filed outcome	NO APP	NO AP	Population aged 15+ in STRATEX1	Gender of selecte individu	Age of selected individual (in 2012)	Sampling register: urbanisatio category	Current living place: urbanisation category	Current living place: NUTS-1 (=REGIO)	Current living place: NUTS-2	Sampling register: NUTS-2
1	2242	PL	0.00433572	0.01666667	.	18907	.	.	10200	1	.	.	719603.2	2	72	1	1	5	51	51
2	3854	PL	0.00433572	0.01666667	.	18907	.	.	10200	1	.	.	719603.2	1	50	1	1	5	51	51
3	4435	PL	0.00433572	0.01666667	.	18907	.	.	10200	1	.	.	719603.2	1	62	1	1	5	51	51
4	5326	PL	0.00433572	0.01666667	.	18907	.	.	10200	3	.	.	719603.2	2	37	1	1	5	51	51
5	1523	PL	0.00440798	0.01639344	.	19312	.	.	10200	1	.	.	719603.2	1	67	1	1	5	51	51
6	4446	PL	0.00440798	0.01639344	.	19312	.	.	10200	1	.	.	719603.2	1	35	1	1	5	51	51
7	7414	PL	0.00440798	0.01639344	.	19312	.	.	10200	1	.	.	719603.2	2	39	1	1	5	51	51
8	8411	PL	0.00440798	0.01639344	.	19312	.	.	10200	1	.	.	719603.2	2	23	1	1	5	51	51
9	1959	PL	0.0047693	0.01515152	.	19315	.	.	10200	3	.	.	719603.2	1	40	1	1	5	51	51
10	2559	PL	0.0047693	0.01515152	.	19315	.	.	10200	1	.	.	719603.2	2	74	1	1	5	51	51
11	3921	PL	0.0047693	0.01515152	.	19315	.	.	10200	2	.	.	719603.2	1	47	1	1	5	51	51
12	7135	PL	0.0047693	0.01515152	.	19315	.	.	10200	1	.	.	719603.2	2	63	1	1	5	51	51
13	2327	PL	0.00607001	0.01190476	.	19316	.	.	10200	1	.	.	719603.2	2	81	1	1	5	51	51
14	5499	PL	0.00607001	0.01190476	.	19316	.	.	10200	1	.	.	719603.2	1	27	1	1	5	51	51
15	6495	PL	0.00607001	0.01190476	.	19316	.	.	10200	1	.	.	719603.2	2	89	1	1	5	51	51
16	7864	PL	0.00607001	0.01190476	.	19316	.	.	10200	1	.	.	719603.2	1	32	1	1	5	51	51
17	559	PL	0.00337825	0.02139037	.	19318	.	.	10200	2	.	.	719603.2	2	48	1	1	5	51	51
18	1195	PL	0.00337825	0.02139037	.	19318	.	.	10200	1	.	.	719603.2	2	16	1	1	5	51	51
19	2168	PL	0.00337825	0.02139037	.	19318	.	.	10200	1	.	.	719603.2	1	49	1	1	5	51	51
20	9735	PL	0.00337825	0.02139037	.	19318	.	.	10200	1	.	.	719603.2	1	25	1	1	5	51	51
21	511	PL	0.0186978	0.00386473	.	19319	.	.	10200	1	.	.	719603.2	1	22	1	8	4	41	51
22	1033	PL	0.0186978	0.00386473	.	19319	.	.	10200	1	.	.	719603.2	2	40	1	1	5	51	51
23	2091	PL	0.0186978	0.00386473	.	19319	.	.	10200	3	.	.	719603.2	1	29	1	1	5	51	51
24	2503	PL	0.0186978	0.00386473	.	19319	.	.	10200	1	.	.	719603.2	1	63	1	1	5	51	51
25	6361	PL	0.0186978	0.00386473	.	19319	.	.	10200	2	.	.	719603.2	1	36	1	1	5	51	51
26	7168	PL	0.0186978	0.00386473	.	19319	.	.	10200	1	.	.	719603.2	2	18	1	1	5	51	51
27	8470	PL	0.0186978	0.00386473	.	19319	.	.	10200	1	.	.	719603.2	2	32	1	5	5	51	51
28	9725	PL	0.0186978	0.00386473	.	19319	.	.	10200	3	.	.	719603.2	1	30	1	1	5	51	51
29	1575	PL	0.00688296	0.01049869	.	19320	.	.	10200	2	.	.	719603.2	2	15	1	1	5	51	51
30	3942	PL	0.00688296	0.01049869	.	19320	.	.	10200	1	.	.	719603.2	1	24	1	1	5	51	51
31	3979	PL	0.00688296	0.01049869	.	19320	.	.	10200	1	.	.	719603.2	2	25	1	1	5	51	51
32	9415	PL	0.00688296	0.01049869	.	19320	.	.	10200	1	.	.	719603.2	1	56	1	1	5	51	51
33	1152	PL	0.00619647	0.01166181	.	19321	.	.	10200	1	.	.	719603.2	1	34	1	1	5	51	51
34	3947	PL	0.00619647	0.01166181	.	19321	.	.	10200	2	.	.	719603.2	2	56	1	1	5	51	51
35	6305	PL	0.00619647	0.01166181	.	19321	.	.	10200	1	.	.	719603.2	2	33	1	1	5	51	51
36	8148	PL	0.00619647	0.01166181	.	19321	.	.	10200	1	.	.	719603.2	1	57	1	1	5	51	51
37	1818	PL	0.00592549	0.01219512	.	19322	.	.	10200	2	.	.	719603.2	1	61	1	1	5	51	51
38	3965	PL	0.00592549	0.01219512	.	19322	.	.	10200	1	.	.	719603.2	1	31	1	1	5	51	51
39	4229	PL	0.00592549	0.01219512	.	19322	.	.	10200	1	.	.	719603.2	2	53	1	1	5	51	51
40	5022	PL	0.00592549	0.01219512	.	19322	.	.	10200	3	.	.	719603.2	2	62	1	1	5	51	51
41	5181	PL	0.00592549	0.01219512	.	19322	.	.	10200	1	.	.	719603.2	2	70	1	1	5	51	51
42	5420	PL	0.00592549	0.01219512	.	19322	.	.	10200	1	.	.	719603.2	2	75	1	1	5	51	51
43	6435	PL	0.00592549	0.01219512	.	19322	.	.	10200	1	.	.	719603.2	2	38	1	1	5	51	51
44	9961	PL	0.00592549	0.01219512	.	19322	.	.	10200	1	.	.	719603.2	1	37	1	1	5	51	51
45	523	PL	0.00953859	0.00757576	.	19324	.	.	10200	3	.	.	719603.2	2	36	1	1	5	51	51
46	3135	PL	0.00953859	0.00757576	.	19324	.	.	10200	1	.	.	719603.2	1	46	1	1	5	51	51
47	5132	PL	0.00953859	0.00757576	.	19324	.	.	10200	1	.	.	719603.2	1	55	1	1	5	51	51
48	6817	PL	0.00953859	0.00757576	.	19324	.	.	10200	1	.	.	719603.2	2	49	1	1	5	51	51

Seloste
n = ryp-
päiden
määrä

m = 15+
vuoti-
aiden
määrä
ryppäas-
sä

Ryväs k=
Pienalue
x = sen
15+ -
väki-
määrä.

Havainnosta
1541
Alkaa yksin-
kertainen
satunnais-
otanta;
Sitä ennen
kaksiasteinen
ryväotanta:

Tätä kutsu-
taan
Two-domain
Sampling

Halvempi
otanta
käynti-
haastatte-
lussa

	Respondent identifica number	Countr	Selection probability at first stage (village and town below 50,000; town over 50,000: PROB1=1)	Selection probability at second stage (respondent in PSU)	NO AP	N A	First stage unit	NO AP	NO AP	Explicit stratum (villages and towns below 50,000: STRATEX1 is combination of NUTS2 and urbanization category; every town over 50,000 is separate value of STRATEX1)	N AP	N A	N AP	N AP	N A	Summary filed outcome	NO APP	NO AP	Population aged 15+ in STRATEX1	Gender of selecte individu	Age of selected individual (in 2012)	Sampling register: urbanisatio category	Current living place: urbanisation category	Current living place: NUTS-1 (=REGIO)	Current living place: NUTS-2	Sampling register: NUTS-2
1531	3328	PL	0.5338	0.00014	.	.	48104	.	.	43200	1	.	.	216972.2	2	63	4	4	4	42	42
1532	4424	PL	0.5338	0.00014	.	.	48104	.	.	43200	1	.	.	216972.2	2	41	4	4	4	42	42
1533	5303	PL	0.5338	0.00014	.	.	48104	.	.	43200	1	.	.	216972.2	2	81	4	4	4	42	42
1534	6628	PL	0.5338	0.00014	.	.	48104	.	.	43200	1	.	.	216972.2	1	43	4	4	4	42	42
1535	6781	PL	0.5338	0.00014	.	.	48104	.	.	43200	2	.	.	216972.2	1	48	4	4	4	42	42
1536	6961	PL	0.5338	0.00014	.	.	48104	.	.	43200	1	.	.	216972.2	2	30	4	4	4	42	42
1537	7179	PL	0.5338	0.00014	.	.	48104	.	.	43200	1	.	.	216972.2	2	16	4	4	4	42	42
1538	7712	PL	0.5338	0.00014	.	.	48104	.	.	43200	2	.	.	216972.2	2	55	4	4	4	42	42
1539	9148	PL	0.5338	0.00014	.	.	48104	.	.	43200	3	.	.	216972.2	2	35	4	4	4	42	42
1540	9575	PL	0.5338	0.00014	.	.	48104	.	.	43200	1	.	.	216972.2	1	51	4	4	4	42	42
1541	117	PL	1	0.00008	.	.	50119	.	.	50119	1	.	.	48212	1	48	5	5	1	12	12
1542	2210	PL	1	0.00008	.	.	50119	.	.	50119	1	.	.	48212	2	32	5	5	1	12	12
1543	3300	PL	1	0.00008	.	.	50119	.	.	50119	1	.	.	48212	1	60	5	5	1	12	12
1544	4240	PL	1	0.00008	.	.	50119	.	.	50119	1	.	.	48212	2	40	5	5	1	12	12
1545	203	PL	1	0.00009	.	.	50120	.	.	50120	2	.	.	43728	2	24	5	5	1	12	12
1546	3412	PL	1	0.00009	.	.	50120	.	.	50120	2	.	.	43728	2	56	5	5	1	12	12
1547	4738	PL	1	0.00009	.	.	50120	.	.	50120	2	.	.	43728	2	20	5	5	1	12	12
1548	9829	PL	1	0.00009	.	.	50120	.	.	50120	1	.	.	43728	1	57	5	5	1	12	12
1549	435	PL	1	0.00009	.	.	50502	.	.	50502	2	.	.	58255	2	76	5	5	3	34	34
1550	736	PL	1	0.00009	.	.	50502	.	.	50502	1	.	.	58255	1	35	5	5	3	34	34
1551	833	PL	1	0.00009	.	.	50502	.	.	50502	1	.	.	58255	2	65	5	5	3	34	34
1552	2231	PL	1	0.00009	.	.	50502	.	.	50502	1	.	.	58255	2	56	5	5	3	34	34
1553	5183	PL	1	0.00009	.	.	50502	.	.	50502	2	.	.	58255	2	49	5	5	3	34	34
1554	5159	PL	1	0.00008	.	.	50503	.	.	50503	2	.	.	48485	2	27	5	5	6	62	62
1555	5539	PL	1	0.00008	.	.	50503	.	.	50503	1	.	.	48485	2	53	5	5	6	62	62
1556	7074	PL	1	0.00008	.	.	50503	.	.	50503	3	.	.	48485	1	41	5	5	6	62	62
1557	7796	PL	1	0.00008	.	.	50503	.	.	50503	1	.	.	48485	1	71	5	5	6	62	62
1558	88	PL	1	0.00009	.	.	50902	.	.	50902	1	.	.	65773.9	2	72	5	5	6	61	61
1559	641	PL	1	0.00009	.	.	50902	.	.	50902	2	.	.	65773.9	2	36	5	5	6	61	61
1560	5311	PL	1	0.00009	.	.	50902	.	.	50902	1	.	.	65773.9	2	60	5	5	6	61	61
1561	5839	PL	1	0.00009	.	.	50902	.	.	50902	2	.	.	65773.9	1	36	5	5	6	61	61
1562	8826	PL	1	0.00009	.	.	50902	.	.	50902	2	.	.	65773.9	1	34	5	5	6	61	61
1563	9728	PL	1	0.00009	.	.	50902	.	.	50902	1	.	.	65773.9	2	40	5	5	6	61	61
1564	5519	PL	1	0.00009	.	.	50903	.	.	50903	3	.	.	63387	1	25	5	5	4	41	41
1565	6286	PL	1	0.00009	.	.	50903	.	.	50903	2	.	.	63387	2	43	5	5	4	41	41
1566	6337	PL	1	0.00009	.	.	50903	.	.	50903	2	.	.	63387	1	50	5	5	4	41	41
1567	6482	PL	1	0.00009	.	.	50903	.	.	50903	1	.	.	63387	2	72	5	5	4	41	41
1568	7696	PL	1	0.00009	.	.	50903	.	.	50903	1	.	.	63387	1	62	5	5	4	41	41
1569	9075	PL	1	0.00009	.	.	50903	.	.	50903	1	.	.	63387	2	20	5	5	4	41	41
1570	342	PL	1	0.00008	.	.	51903	.	.	51903	3	.	.	50419	2	39	5	5	6	63	63
1571	3358	PL	1	0.00008	.	.	51903	.	.	51903	1	.	.	50419	2	24	5	5	6	63	63
1572	3980	PL	1	0.00008	.	.	51903	.	.	51903	2	.	.	50419	1	56	5	5	6	63	63
1573	6620	PL	1	0.00008	.	.	51903	.	.	51903	1	.	.	50419	1	32	5	5	6	63	63
1574	4428	PL	1	0.00009	.	.	52502	.	.	52502	2	.	.	68442	1	21	5	5	4	41	41
1575	4782	PL	1	0.00009	.	.	52502	.	.	52502	1	.	.	68442	1	58	5	5	4	41	41
1576	5359	PL	1	0.00009	.	.	52502	.	.	52502	1	.	.	68442	1	66	5	5	4	41	41
1577	5595	PL	1	0.00009	.	.	52502	.	.	52502	2	.	.	68442	2	45	5	5	4	41	41
1578	8749	PL	1	0.00009	.	.	52502	.	.	52502	2	.	.	68442	1	22	5	5	4	41	41

Otantatiedoston loppuosa josta näet brutto-otoksen koon.

Respondent identifica number	Countr	Selection probability at first stage (village and town below 50,000; town over 50,000: PROB1=1)	Selection probability at second stage (respondent in PSU)	NO AP	N A	First stage unit	NO AP	NO AP	Explicit stratum (villages and towns below 50,000: STRATEX1 is combination of NUTS2 and urbanization category; every town over 50,000 is separate value of STRATEX1)	N AP	N A	N AP	NO AP	N A	Summary filed outcome	NO APP	NO AP	Population aged 15+ in STRATEX1	Gender of selecte individu	Age of selected individual (in 2012)	Sampling register: urbanisatio category	Current living place: urbanisation category	Current living place: NUTS-1 (=REGIO)	Current living place: NUTS-2	Sampling register: NUTS-2
2661	341	PL	1	0.00009		90111			90111						1			42620.9	2	37	9	9	1	12	1
2662	472	PL	1	0.00009		90111			90111						1			42620.9	2	30	9	9	1	12	1
2663	1350	PL	1	0.00009		90111			90111						1			42620.9	1	68	9	9	1	12	1
2664	9531	PL	1	0.00009		90111			90111						2			42620.9	2	20	9	9	1	12	1
2665	396	PL	1	0.00009		90112			90112						1			129550	2	33	9	9	1	12	1
2666	620	PL	1	0.00009		90112			90112						1			129550	2	79	9	9	1	12	1
2667	1336	PL	1	0.00009		90112			90112						1			129550	2	58	9	9	1	12	1
2668	2773	PL	1	0.00009		90112			90112						1			129550	1	50	9	9	1	12	1
2669	3257	PL	1	0.00009		90112			90112						2			129550	2	42	9	9	1	12	1
2670	4286	PL	1	0.00009		90112			90112						1			129550	2	18	9	9	1	12	1
2671	4879	PL	1	0.00009		90112			90112						2			129550	1	34	9	9	1	12	1
2672	5825	PL	1	0.00009		90112			90112						2			129550	1	28	9	9	1	12	1
2673	6814	PL	1	0.00009		90112			90112						1			129550	1	48	9	9	1	12	1
2674	7764	PL	1	0.00009		90112			90112						2			129550	1	39	9	9	1	12	1
2675	9090	PL	1	0.00009		90112			90112						1			129550	1	56	9	9	1	12	1
2676	9833	PL	1	0.00009		90112			90112						1			129550	2	55	9	9	1	12	1
2677	1136	PL	1	0.0001		90113			90113						1			59613	2	77	9	9	1	12	1
2678	2032	PL	1	0.0001		90113			90113						1			59613	2	70	9	9	1	12	1
2679	2118	PL	1	0.0001		90113			90113						1			59613	1	57	9	9	1	12	1
2680	3052	PL	1	0.0001		90113			90113						1			59613	1	29	9	9	1	12	1
2681	4269	PL	1	0.0001		90113			90113						2			59613	2	46	9	9	1	12	1
2682	8170	PL	1	0.0001		90113			90113						2			59613	1	65	9	9	1	12	1
2683	3217	PL	1	0.00011		90114			90114						2			18947	1	39	9	9	1	12	1
2684	8193	PL	1	0.00011		90114			90114						1			18947	2	61	9	9	1	12	1
2685	1703	PL	1	0.00012		90115			90115						1			16720.1	1	40	9	9	1	12	1
2686	5389	PL	1	0.00012		90115			90115						2			16720.1	2	28	9	9	1	12	1
2687	1126	PL	1	0.00012		90116			90116						3			34639.9	1	29	9	9	1	12	1
2688	2880	PL	1	0.00012		90116			90116						2			34639.9	1	30	9	9	1	12	1
2689	4629	PL	1	0.00012		90116			90116						1			34639.9	2	74	9	9	1	12	1
2690	9994	PL	1	0.00012		90116			90116						1			34639.9	1	42	9	1	1	12	1
2691	139	PL	1	0.0001		90117			90117						1			122955.9	1	43	9	9	1	12	1
2692	1440	PL	1	0.0001		90117			90117						1			122955.9	2	58	9	9	1	12	1
2693	1636	PL	1	0.0001		90117			90117						2			122955.9	1	27	9	9	1	12	1
2694	3771	PL	1	0.0001		90117			90117						1			122955.9	2	34	9	9	1	12	1
2695	4038	PL	1	0.0001		90117			90117						1			122955.9	1	17	9	9	1	12	1
2696	4108	PL	1	0.0001		90117			90117						1			122955.9	2	16	9	9	1	12	1
2697	4992	PL	1	0.0001		90117			90117						1			122955.9	2	36	9	9	1	12	1
2698	5491	PL	1	0.0001		90117			90117						2			122955.9	1	41	9	9	1	12	1
2699	7590	PL	1	0.0001		90117			90117						2			122955.9	2	52	9	9	1	12	1
2700	7785	PL	1	0.0001		90117			90117						2			122955.9	2	22	9	9	1	12	1
2701	8684	PL	1	0.0001		90117			90117						1			122955.9	1	31	9	9	1	12	1
2702	9060	PL	1	0.0001		90117			90117						1			122955.9	2	22	9	9	1	12	1
2703	416	PL	1	0.00009		90118			90118						1			42421.8	2	15	9	9	1	12	1
2704	4190	PL	1	0.00009		90118			90118						1			42421.8	1	33	9	9	1	12	1
2705	4327	PL	1	0.00009		90118			90118						1			42421.8	2	44	9	9	1	12	1
2706	7145	PL	1	0.00009		90118			90118						1			42421.8	1	23	9	9	1	12	1

Tuosta datasta saadaan tehtyä helposti peruspaino = $w2_resp$

Eli otospaino vastaajille $resp=1$.

Tämä on SAS-koodia mutta sitä ei tule harjoitukseen. Perusymmärrys riittää.

Loppuun koodattu keskiarvotulostus on seuraavalla sivulla.

```
data poland; set z.sddf6pl;  
w1=1/prob1;  
w2=1/(prob1*prob2);
```

On siis käänteisluku kahden sisällyttämistodennäköisyyden tulosta

```
if outcome=1 then resp=1; else if outcome=2 then resp=0;
```

Tästä saadaan vastaajat eli niiden indikaattori.

```
proc sort ; by stratex1 ; run;
```

```
proc summarY nway data=poland; class stratex1; var w1 w2 ; output out=sum_w2  
sum(w1 w2)=sumw1_1 sum_w2_2 ; run;
```

```
proc sort; by stratex1; run;
```

```
proc summarY nway data=poland; where resp=1; class stratex1; var w2 w1; output  
out=sum_w2_resp sum(w1 w2)=sumw1_1_resp sum_w2_2_resp ; run;
```

```
proc sort; by stratex1;
```

```
data poland2; merge poland sum_w2 sum_w2_resp; by stratex1;
```

```
if resp=1 then w2_resp=w2*(sum_w2_2/sum_w2_2_resp);
```

Näin meillä on lopulta peruspaino = $w2_resp$.

```
proc means n mean cv min max sum; var w1 w2 w2_resp; run;
```

Tässä on siis kolme painoa joista käyttökelpoisin on alin. Tutki sen tunnuslukuja. Summahan tarkoittaa estimoitua tavoiteperusjoukon kokoa eli 15+ -vuotiaiden kotona asuvien määrää. Tämä paino muunnetaan ESS-datassa joka ilmestyy lokakuun lopulla 2013 sellaiseksi että keskiarvo = 1. Kutsutaan suomeksi analyysipainoksi, ESS:ssä Design weight (DWEIGHT). Se on alempana.

Variable	N	Mean	Coeff of Variation	Minimum	Maximum	Sum
w1	2706	47.3494	164.243516	1.000000	346.478402	128127.60
w2	2706	11978.4	14.0513881	7577.71	15409.82	32413744.5
w2_resp	1898	17077.8	22.6927088	8832.71	54096.00	32413744.5

Analysis Variable : w_analyysi					
N	Mean	Coeff of Variation	Minimum	Maximum	Sum
1898	1.0000000	22.6927088	0.5172032	3.1676137	1898.00

Katso miten otanta-asetelman suunnitelma onnistui käytännössä? Laske vastausprosentti, kun kerron että ylipeitto (number of in-eligibles) = 171. Laske myös Ylipeittoprosentti (In-eligibility rate).