Chapter 6

Weighting adjustments due to unit missingness

Weighting and reweighting can be considered to cover the following **7 actions**:

(i) Sampling design before the fieldwork

(ii) Weights for the gross-sample (*n*) using (i), the outcome being 'design weights'

(iii) Creation of the sampling design data file before and after the fieldwork

(iv) Computation of 'Basic weights' for the net sample or for the respondents (*r*), assuming *MARS (Missing at random under sampling design)*

(v) Re-weighting assuming MAR(C): specification, estimation, outputs

(vi) Estimation: point-estimates, variance estimation = sampling variance plus variance due to missingness.

(vii) Critical look at the results including benchmarking these against recent results (how plausible they are?)

Re-weighting methods

We here do not try to explain all possible re-weighting methods since they are too many. Often it is however difficult to recognise what a certain method is since so many different terms are used. This is not any exception. Our terms are somewhat new but they are in our opinion clear and logical.

We concentrate on the two core methodology families:

1. **Response propensity weighting**
2. **Calibration weighting**

Post-stratification

Post-stratification is a basic calibration method that is useful to apply if there exists such population level data (macro auxiliary data) that are not yet exploited in the sampling design. This is often the case.). The target is

- to reduce the bias due to frame error if the post-strata statistics (margins) are more correct than initial ones; if an updated frame exist it helps
- to calibrate estimates to more detailed level than initially, thus to post-strata level
- to reduce the bias due to unit nonresponse but it is not automatically ensured.
- to reduce sampling error that occurs if post-strata are more homogenous than initial strata.

All these targets cannot be to be definitely achieved but some at least should be.

Optional example of post-stratification when found differences in response rates within pre-strata as explained

	Initial stratification = Pre-stratification						
	Region 1		Region 2		Region R		
Post-strata within pre-strata	Little educated	More educated	Males	Females	Little educated males	Little educated females	More educated males and females

other regional pre-strata may need other post-stratification. This flexible strategy also allows to use some initial pre-strata if any benefit is not believed to get by post-strata. These requirements thus are found in the nonresponse analysis (Chapter 6). The form of the post-stratified weights is similar to that of the stratified weights

$$w_k = \frac{N_{hg}}{r_{hg}}$$

The target population figures of cells 'N_{hg}' are possibly available as more up-to-date than initially and these should be used in this case (using The updated frame population). The number of the respondents r_{hg} is calculated from the sampling data file. It is important to avoid too small post-strata so that the denominator is not too small, in particular. However, any problem is not the case when $N_{hg} = r_{hg} = 1$, since the weight is equal to one.

Table 7.2: The first 11 post-strata including the re-weights

Strata	Lower limit of Age group	Target population size	Respon-dents	Post-stratified weights
1	15	987995	42	23523.69
2	30	1299525	72	18048.96
3	45	1308426	60	21807.1
4	60	1059202	41	25834.2
5	75	551853	13	42450.24
6	15	258129	11	23466.27
7	30	329337	21	15682.71
8	45	409446	13	31495.85
9	60	267030	19	14054.21
10	75	186921	4	46730.25
11	15	97911	5	19582.2

Table 7.3. Basic and post-stratified sampling weights and their analysis weights from the same basis as Table 7.2

.

Weight	Number of respondents	Mean	Coefficient of Variation	Minimum	Maximum	Sum
Basic	605	16030	42.2	7970	22837	9698424
Basic Analysis	605	1	42.2	0.50	1.43	605
Post-stratified	605	16030	51.7	6328	46730	9698424
Post-stratified analysis	605	1	51.7	0.39	2.92	605

Table 7.4 Parameter estimates of the logistic regression model

Example from the test data, the complete table in the Word version.

Parameter		Estimate	Standard Error	P-value
Intercept		1.31	0.2966	<.0001
Gender	Male	-0.3824	0.1292	0.0031
	Female	0	0	.
Education	Degree 2	-2.7349	0.2483	<.0001
	Degree 3	-1.2829	0.2072	<.0001
	Degree 4	-0.4259	0.2814	0.1301
	Degree 5	0.8743	0.3071	0.0044
	Degree 6	-0.1133	0.2623	0.6657
	Degree 7	0	0	.
Citizen of the country	Yes	0.2044	0.3011	0.4972
	No	0	0	.
Members of household	1	0.5642	0.2985	0.0587
	2	0.7966	0.2803	0.0045
	3	0.0772	0.2745	0.7785
	4	0.1825	0.2749	0.5068
	5	0	0	.

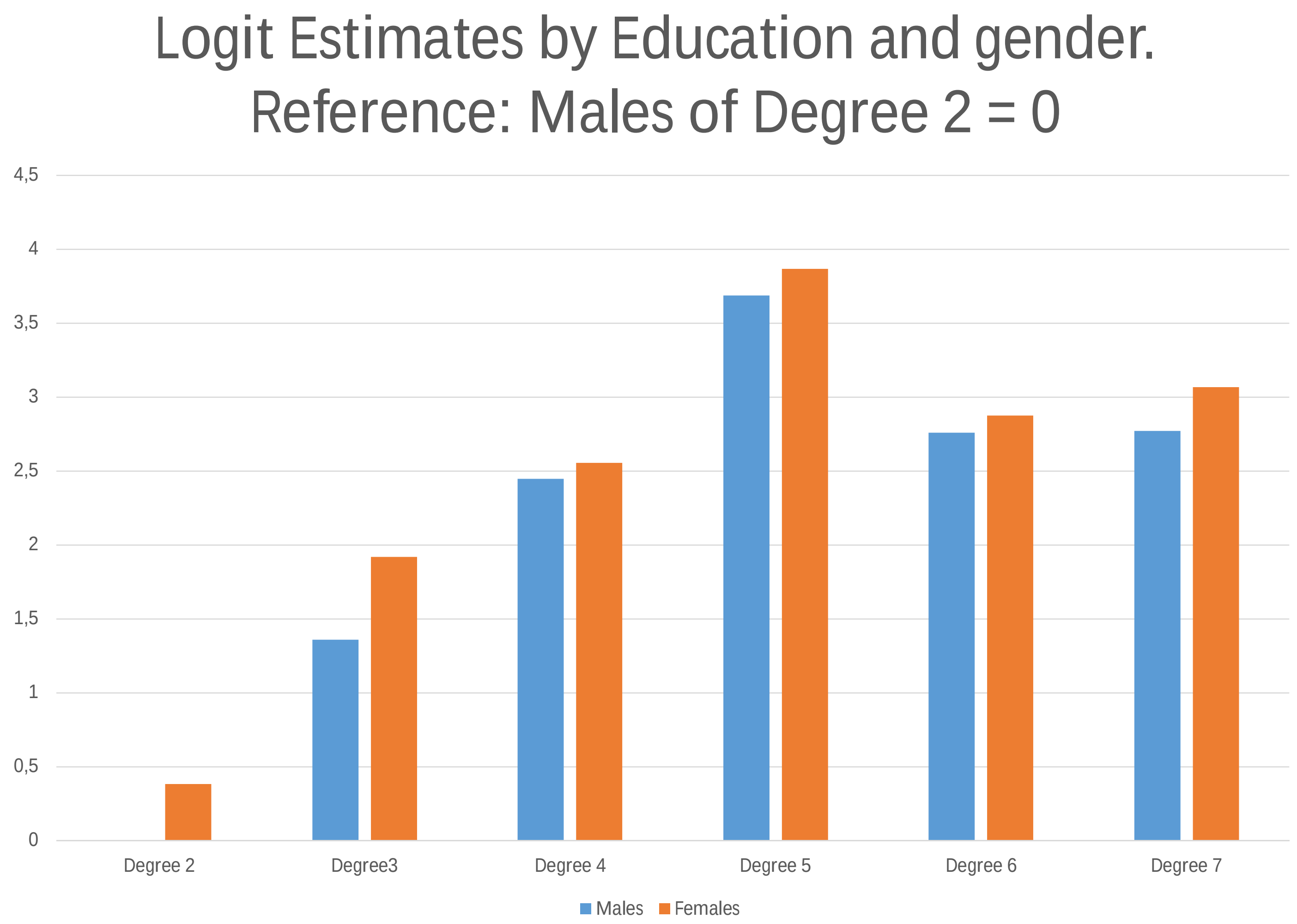
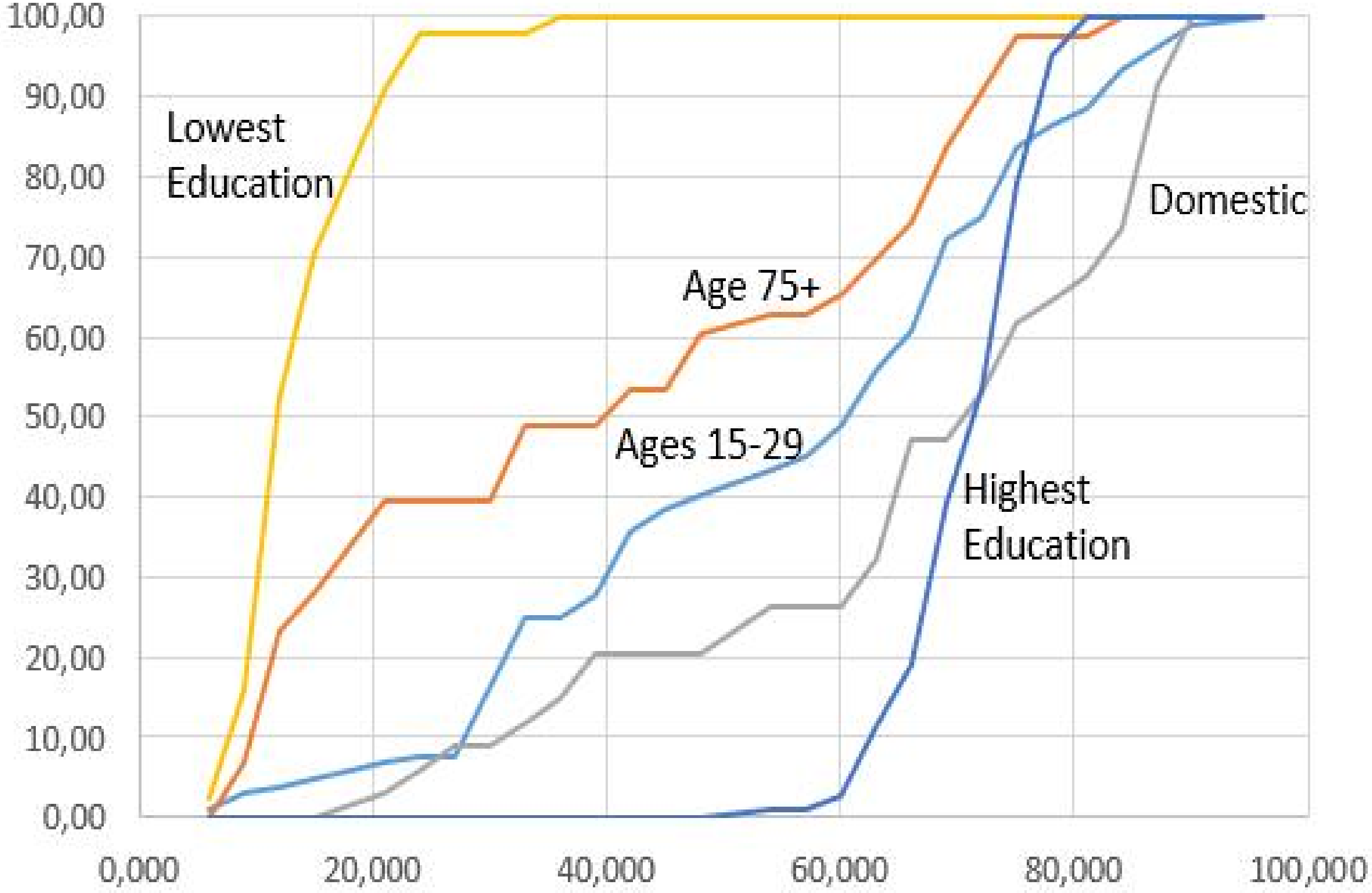

Logit Estimates by Education and gender.
Reference: Males of Degree 2 = 0
4,5
4
3,5
3
2,5
2
1,5
1
0,5
0
Degree 2
Degree3
Degree 4
Degree 5
Degree 6
Degree 7
Males
Females

Figure 7.2 Cumulative frequencies of response propensities for some population responded domains



Steps of response propensity weighting when the response model is estimated

This is for the stratified simple random sampling:

(i) We assume that the response mechanism within each stratum is ignorable, and hence the initial (basic) weights thus have been calculated. These are available only for the respondents k, and symbolized by w_k.

(ii) Next we take those initial weights and divide these by the estimated response probabilities of each respondent obtained from the probit or logit model, and symbolized by p_k.

(iii) Before going forward, it is good to check that the probabilities p_k are realistic, that is, they are not too small, for instance. All probabilities are below 1, naturally.

(iv) Since the sum of the weights (ii) does not match to the known population statistics by strata h, they should be calibrated or scaled so that the sums are equal to the sums of the initial weights in each stratum. This is made by multiplying the weights (ii) by the ratio in each stratum h:

$$q_h = \frac{\sum_h w_k}{\sum_h w_k / p_k}$$

Table 7.3 The weights and their characteristics from the SDDF of the stratified simple random sampling

Sampling Weight	Respondents r	Mean	Minimum	Maximum	Coeff of Variation	Sum
Ordinary						
Basic	605	16030	7970	22838	42.2	9698424
Post-stratified	605	16030	6328	46730	51.7	9698424
Adjusted Basic	605	16030	3163	137329	78.3	9698424
Adjusted post-stratified	605	16030	2273	108195	78.0	9698424
Analysis/relative						
Basic	605	1.0	0.5	1.4	42.2	605
Post-stratified	605	1.0	0.4	2.9	51.7	605
Adjusted Basic	605	1.0	0.2	8.6	78.3	605
Adjusted post-stratified	605	1.0	0.1	6.7	78.0	605

able 7.4: The two analysis weights of the European Social Survey by country

Country	Sampling weight	Respondents	Minimum	Maximum	Coeff of Variation (CV)
Bulgaria	Basic	2260	0..20	3..09	41
	Post-stratified	2260	0..16	4..00	59
Cyprus	Basic	1116	0..40	2..85	45
	Post-stratified	1116	0..22	4..01	60
Czech	Basic	2009	0..04	4..00	52
	Post-stratified	2009	0..00	4..03	63
Germany	Basic	2958	0..54	1..24	33
	Post-stratified	2958	0..02	4..00	57
Spain	Basic	1889	0..70	1..66	11
	Post-stratified	1889	0..50	4..00	31
France	Basic	1968	0..21	4..00	53
	Post-stratified	1968	0..16	4..01	63
United Kingdom	Basic	2286	0..51	4..00	52
	Post-stratified	2286	0..29	4..01	57
Hungary	Basic	2014	0..87	1..17	5
	Post-stratified	2014	0..57	1..70	21

Table 7.5 Characteristics of the weights of the two Finnish surveys that used response propensity weighting

Survey	Response rate	Respondents	Minimum	Maximum	Coeff of Variation
Finnish Security Survey					
Face-to-face sub-survey	50	366	0.10	4.01	48
Web sub-survey	25	971	0.05	17.60	58
Phone sub-survey	62	1866	0.08	4.79	37
Southern Finland Grid-based survey	36	9618	0.17	10.10	68

Scheme 7.1 Summary of all the sampling weights of this book.