LIFE SATISFACTION - General Target Variable Report (GVR)

1. General Information

We provide two harmonized variables on respondent's self-declared level of satisfaction with life: life satisfaction assessment on an 11-point scale T_LIFESAT_11 and on a distribution-preserving scale T_LIFESAT_DISTRIB.

T_LIFESAT_11 is a harmonized measure on an 11-point scale with target values ranging from 0 (lowest degree) to 10 (highest degree).

T_LIFESAT_DISTRIB is a harmonized measure of the relative position of a respondent in the distribution of self-declared life satisfaction in a national sample on the range from 0 (the lowest percentile point in a distribution) to 100 (the highest percentile point in a distribution). The scores of T_LIFESAT_DISTRIB target variable are percentiles within the national sample that indicate what share of respondents reports the same or poorer life satisfaction than the individual.

Both target variables on life satisfaction are accompanied by the control variables, which capture special features of the source response scales: source scale length (C_LIFESAT_LENGTH), source scale direction (C_LIFESAT_ASCEND) and source scale polarity (C_LIFESAT_POLARITY) (see Table 1.1, and Section 3.3).

The target variable report for Life Satisfaction is accompanied by the following Excel documents:

- The Detailed Variable Report (DVR): T_HAPPY_LIFESAT_DVR_SDR2.xlsx. DVR Excel files in SDR2 systemize all information about source variables that were used for harmonization into a given target variable of the SDR2 database;
- The Crosswalk Tables (CWT): T_LIFESAT_CWT_SDR2.xlsx. CWT Excel files in SDR2 contain details about mapping of source values to target values.

Table 1. LIFE SATISFACTION: Description of the target, source, and control variables

	Variable description	**Variable name**	**Variable values***
Target variables	Life satisfaction (11-point scale)	T_LIFESAT_11	0 = lowest degree 10 = highest degree
	Life satisfaction (distribution-preserving scale)	T_LIFESAT_DISTRIB	0 = Lowest percentile point in distribution 100 = Highest percentile point in distribution

Source variables			See: T_HAPPY_LIFESAT_DVR_SDR2.xlsx and T_LIFESAT_CWT_SDR2.xlsx
Control variables	Source life satisfaction scale length	C_LIFESAT_LENGTH	3 = 3-point scale 4 = 4-point scale 5 = 5-point scale 7 = 7-point scale 10 = 10-point scale 11 = 11-point scale
	Source life satisfaction scale direction	C_LIFESAT_ASCEND	0 = descending 1 = ascending
	Source life satisfaction polarity	C_LIFESAT_POLARITY	1 = unipolar (satisfied) 2 = bipolar (satisfied/dissatisfied) 3 = bipolar (agree/disagree)

[a] Missing values are assigned according to the SDR2 missing codes schema, provided in the Appendix.

2. Survey Projects

Source variables that we used for T_LIFESAT appear in 14 international survey projects: AMB, ASES, CB, EB, EQLS, ESS, EVS, ISJP, LB, LITS, NBB, NEB, PA1, WVS, 101 waves and 2050 national surveys. The data cover 130 countries and years from 1973 to 2017.

3. General Rules and Procedures

3.1. Source data description

The majority of the source variables containing information on LIFE SATISFACTION ask about life satisfaction in general (e.g. ESS/1: *All things considered, how satisfied are you with your life as a whole nowadays? Please answer using this card, where 0 means extremely dissatisfied and 10 means extremely satisfied*, EB/40: O*n the whole, are you very satisfied, fairly satisfied, not very satisfied or not at all satisfied with the life you lead?*). Many questions use time phrases emphasizing that respondents are asked about current time, such as "*these days*", "*now*" or "*nowadays*", e.g. ASES: *All things considered, how satisfied are you with your life as a whole these days?* The majority of questions used to create source variables can be divided into two groups: a) asking respondents "how satisfied (would you say you are)", b) asking respondents to declare whether he/she is "very satisfied, fairly satisfied, not very satisfied, or not at all satisfied". Exceptions from these divisions are described in Special Cases section below.

The response scale variation in source variables is presented in Table 3.1.1.

Table 3.1.1. Life satisfaction: Types of scales

Length of scale	Direction of scale	
	Descending	**Ascending**
3	NEB/1	
4	AMB/2004-2016; EB/0.2, 8, 9, 11, 13, 15, 18, 21, 24, 25, 27, 29, 31, 31A, 34.1, 35, 40, 42, 44.2bis, 47.1, 49, 52.1, 54.1, 55.1, 56.1, 57.1, 60.1, 62.2, 63.4, 66.1, 68.1, 73.4, 77.3, 79.3, 81.4, 83.3, 85.2, 87.3; LB/1997, 2000, 2001, 2003-2011, 2013, 2015, 2016; NBB/3	
5	ASES	LITS/1, 3; WVS/4-5
7		ISJP/1-2
10		CB/2010-2015; EB/74.1, 76.2; EQLS/1-3; EVS/1-4; LITS/2; WVS/1-6
11		EB/3, 5, 19; ESS/1-8; PA1

3.2 Rules of transformation of the source variable into target variables

To construct the 11-point scale and distribution-preserving target variables, we first create preparatory scales. This involves recording the values of the source scales using the consecutive numbers *k*, where *k* ranges from 1 to n. The value 1 of the preparatory scale corresponds to the lowest life satisfaction, and higher scores correspond to higher life satisfaction (ascending direction). Each preparatory scale is of the same length as the source scale it was derived from.

3.2.1 Constructing the 11-point target scale

To construct the 11-point target scale, we use the preparatory scales and assign scores to them in the interval from 0 to 10, according to the following linear transformation:

$$l(k) = \frac{10}{n*2} + (k - 1) * \frac{10}{n}$$

where: $l(k)$ is a target score corresponding to the preparatory score k, and n is the number of k-values.

This process involves "stretching" preparatory (and thus, source) scales that have fewer than 10 points and keeping 11-point scales as such. Figure 3.2.1 and Table 3.2.1 depict transformations using this type of rescaling.

Figure 3.2.1. Transformation of source values into the target 11-point scale with 0 to 10 range

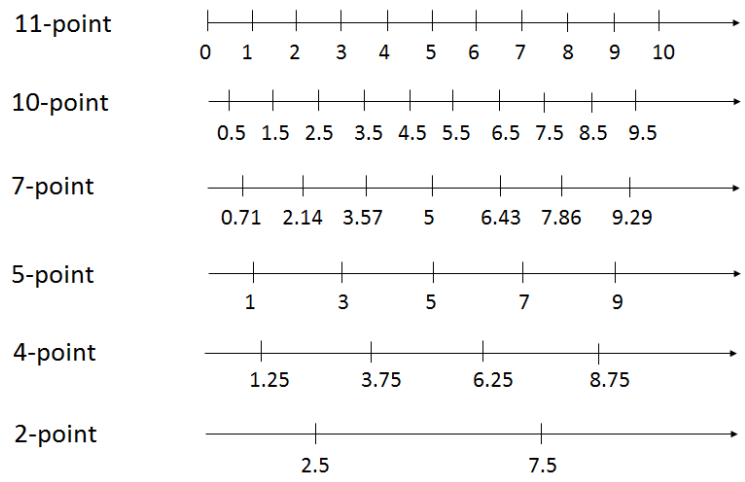

Table 3.2.1. Creating the 11-point scale (from 0 to 10), with median and mean values 5, and minimized inter-scale differences in the variability

Source scale length	Recodes	Median Mean	Average of absolute deviations	Variance	Standard deviation
11-point	0,1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0	5.0	2.72	10.00	3.16
10-point	0.5, 1.5, 2.5, 3.5, 4.5, 5.5, 6.5, 7.5, 8.5, 9.5	5.0	2.50	8.25	2.87
7-point	0.7, 2.1, 3,6, 5.0. 6,4, 7.9, 9,3	5.0	2.46	8.25	2.87
5-point	1.0, 3.0, 5.0, 7.0, 9.0	5.0	2.40	8.00	2.83
4-point	1.25, 3.75, 6.25, 8.75	5.0	2.50	7.81	2.80
3-point	1.7, 5.0, 8.3	5.0	2.20	7.26	2.69
2-point	2.5, 7.5	5.0	2.50	6.25	2.50

3.2.2 Constructing the distribution-preserving target scale

To construct the distribution-preserving target scale, we take into account respondents' position in the distribution of reported life satisfaction values in a given national sample. For an *n*-point preparatory scale, for values *k* that range from 1 to *n,* where X_i is the percent distribution of the variable in sample *s*, *k* is recoded to:

$$k = \sum_{i=1}^{k-1} X_i + \frac{X_k}{2}$$

The distributional score for the answer option *k* is the sum of percentiles of all previous answer options up to *k-1* plus half of the percentile of the answer option *k*.

For a given sample, each scale point of the distribution target scale corresponds to the midpoint of the cumulative distribution of scores *k* (see Table 3.2.2). Put differently, the scores of the distributional target scale are percentiles that indicate what share of respondents within a national sample reports the same or lower value than the individual. The target variable is computed using unweighted samples.

Table 3.2 illustrates how we transform **preparatory** variables (which recode **source** variables' values in ascending direction) with 5 response options into the distribution-based target variable.

Table 3.2. Example of the distribution-based transformation of 5-point preparatory variables into T_LIFESAT_DISTRIB.

Preparatory variable values, based on source values k	**Percentage distribution** X_k	**Cumulative percentage distribution** $\sum_{i=1}^{k} X_i$	**Interval** $\sum_{i=1}^{k-1} X_i$	**Interval lower bound plus interval midpoint** $\sum_{i=1}^{k-1} X_i + \frac{X_k}{2}$	**Target value (rounded to integer)**
1 = lowest satisfaction	10.68	10.68	0	= 10.68/2 = 5.34	5
2	32.75	43.44	10.68	= (10.68 + 32.75)/2 = 27.05	27
3	32.11	75.55	43.44	= (43.44 + 32.11)/2 = 59.49	59
4	21.69	97.23	75.55	= (75.55 + 21.69)/2 = 86.39	86
5 = highest satisfaction	2.77	100	97.23	= (97.23 + 2.77)/2 = 98.61	99

Missing values and different situations that warrant to be treated as missing data are coded according to the SDR2 missing codes schema, provided in Table A.1 in the Appendix.

3.3. Methodological variables that accompany LIFE SATISFACTION

The main variation in source questions about life satisfaction stems from differences in response scales. We provide three harmonization control variables that store specific features of the source variables (see Table 1.1):

1. C_LIFESAT_LENGTH is a nominal variable that identifies the length (i.e. number of answer options) of any given source scale used to construct T_LIFESAT_11 and T_LIFESAT_DISTRIB. It takes on the following values: 3, 4, 5, 7, 10, 11.

2. C_LIFESAT_ASCEND is dichotomous. It takes the value 1 for source scales whose values are in ascending order (i.e. response options are ordered from least to most satisfied); it takes the value 0 when source scale values are in descending order (i.e. response options are ordered from most to least satisfied).
3. C_LIFESAT_POLARITY is coded 1 if the source scale is unipolar i.e. the concept is measured along one dimension (from not at all satisfied with life to very satisfied), 2 if the scale is bipolar (e.g. from completely dissatisfied to completely satisfied), and 3 in the case of agree - disagree scale of answers.

4. Special Cases

- In LITS/2 the questionnaire contains two variables on life satisfaction:

  a) `q722` "All things considered, how satisfied or dissatisfied are you with your life as a whole these days? Please answer on a scale from 1 to 10, where 1 means completely dissatisfied and 10 means completely satisfied."

  b) `q301e` "To what extent do you agree with the following statements? All things considered, I am satisfied with my life now: Strongly disagree = 1, Disagree = 2, Neither disagree nor agree = 3, Agree = 4, Strongly agree = 5".

  Both questions appear in the questionnaire and have realized values for all countries. We select `q722` as most source questions in SDR2 ask respondents to rate their life satisfaction on the dissatisfied/satisfied scale (rather than agree/disagree).

- In the three waves in EB (3, 5, and 19) there are two source variables on life satisfaction:

  a) "We have talked about various parts of your life. All things considered, how satisfied or dissatisfied are you with your life as a whole these days?" followed by 11-point scale from 1 = "Completely dissatisfied" to 11 = "Very satisfied". (This question has the same structure as in those asked in ESS (all waves) and PA1.)

  b) "Now let's talk about your life in general. On the whole, are you very satisfied, fairly satisfied, not very satisfied or not at all satisfied with the life you lead?" followed by 4-point scale of answers (1=very satisfied, 2 = fairly satisfied, 3 = not very satisfied, 4 = not at all satisfied). (Such questions were asked in 38 waves of EB. 4-point scale of answers is present also in AMB, LB and NBB questions concerning life satisfaction but the question is phrased slightly differently.)

  From these datasets, we selected the 11-point scale in order to retain as much detailed information about respondents' answers as possible.

Appendix: Codes for missing values in SDR2

In the SDR database v.2 we identify different situations that warrant to be treated as missing data. Table A.1 lists all SDR2 missing value codes:

Table A.1. Codes for missing values in SDR2

SDR tag [a]	**SPSS (STATA) codes**	**Label**
		Standardized source codes for missing values
DK	-1 (.a)	Don't know
NA	-2 (.b)	No answer
REF	-3 (.c)	Refusal
DU	-4 (.d)	Don't understand the question
DNR	-5 (.e)	Any combination of DK, NA, REF, DU
INAP	-6 (.f)	Inapplicable
NEC	-7 (.g)	Not elsewhere classified
		SDR created codes for missing values
UNFIT	-8 (.h)	Source value does not fit to target
ERR	-9 (.i)	Errors in source data and undocumented source values
COMBI	-10 (.j)	Different missing codes on multiple sources taken for a target
CINAP	-11 (.k)	For control variables only: inapplicable
INSUF	-12 (.l)	For survey: Insufficiently defined response categories
QNA	-20 (.t)	For survey: Question not available

[a] Abbreviations for the labels corresponding to the SDR2 codes for missing values. These tags are used in the Crosswalk Table (CWT) files (Excel) that accompany documentation of SDR2 target variables.

In exceptional situations when codes for missing data listed in Table A.1 cannot be used, we apply a system missing <null> value.