**TRUST IN PEOPLE - General Target Variable Report (GVR)**

## 1. General Information

We provide three harmonized measures of respondents' trust in people: T_TRPEOPLE, T_TRPEOPLE_11 and T_TRPEOPLE_DISTRIB. Section 3.2 outlines the harmonization rules for each target variable.

T_TRPEOPLE is a dichotomous variable, where 1 indicates that respondents express trust in people, and 0 means otherwise, i.e. they are ambivalent, or distrust people. In the case of odd-numbered source scales, the midpoint is coded as 0.

T_TRPEOPLE_11 measures respondents' reported trust on an 11-point scale. Target values range from 0 (highest degree/intensity of trust) to 10 (lowest degree/intensity of trust).

T_TRPEOPLE_DISTRIB measures the relative position of a respondent in the distribution of trust in people in a given source survey (national sample). The scores of this target variable are percentiles within the national sample that indicate what share of respondents reports the same or lower trust than the individual.

The three target variables are accompanied by harmonization control variables that capture special features of the source variables. Two of them refer to the response scales: length (C_TRPEOPLE_LENGTH) and direction (C_TRPEOPLE_ASCEND). One control variable captures special features of the source variables with regard to wording (C_TRPEOPLE_NONSTANDARD) (see Table 1.1 and Section 3.3).

The target variable report for TRUST IN PEOPLE is accompanied by the following Excel documents:

- The Detailed Variable Report (DVR) T_TRUSTS_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;
- A Crosswalk Table (CWT): T_TRPEOPLE_CWT_SDR2.xlsx. CWT Excel files in SDR2 contain details about mapping of source values to target values.

**Table 1.1.** TRUST IN PEOPLE: Description of the target, source, and control variables

| | **Variable description** | **Variable name** | **Variable values** [a] |
|---|---|---|---|
| **Target variables** | Trust in people (binary) | T_TRPEOPLE | 0 = false<br>1 = true |
| | Trust in people (11-point scale) | T_TRPEOPLE_11 | 0 = lowest degree<br>10 = highest degree |

| | | | |
|---|---|---|---|
| | Trust in people (distribution-preserving scale) | T_TRPEOPLE_DISTRIB | 0 = lowest percentile point in distribution<br>100 = highest percentile point in distribution |
| **Source variables** | | | See: T_TRUSTS_DVR_SDR2.xlsx and T_TRPEOPLE_CWT_SDR2.xlsx |
| **Control variables** | The length of the rating scale (i.e. number of answer options) that the source variable uses | C_TRPEOPLE_LENGTH | 2 = 2-point scale<br>4 = 4-point scale<br>5 = 5-point scale<br>7 = 7-point scale<br>10 = 10-point scale<br>11 = 11-point scale |
| | Source values: scale direction | C_TRPEOPLE_ASCEND | 0 = descending<br>1 = ascending |
| | Nonstandard wording of source question on 'trust in people' | C_TRPEOPLE_NONSTANDARD | 0 = standard question wording<br>1 = nonstandard question wording |

[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_TRPEOPLE appear in 16 international survey projects: ABS, AFB, AMB, ARB, CB, EB, EQLS, ESS, EVS, ISSP, LB, LITS, NBB, NEB, VPCPCE, WVS, 76 waves and 1591 national surveys. The data cover 143 countries and years from 1981 to 2017.

## 3. General Rules and Procedures

### 3.1. Source data description

The source question wording and, especially, the type of response scales and coding of answers for items on trust in people vary across survey projects. Table 3.1 illustrates variation in properties of response scales.

**Table 3.1. TRUST IN PEOPLE: Types of source scales**

| **Length of scale** | **Direction of scale** | |
|---|---|---|
| | **Descending** | **Ascending** |

| | | |
|---|---|---|
| 2 | ABS/1-4; AFB/1; AMB/2008; ARB/1-4; EB/25, 62.2; EVS/1-4; LB/1996-2016; NBB/2-3; WVS/1-6 | AFB/3, 5 |
| 4 | ISSP/1998, 2004, 2007, 2008, 2014 | |
| 5 | VPCPCE | ISSP/2010; LITS/1-3 |
| 7 | | NBB/5-6; NEB/5-7 |
| 10 | | CB/2010, 2011, 2013, 2015; EB/72.1, 74.1; EQLS/1-3 |
| 11 | | ESS/1-8 |

### 3.2. Rules of transformation of the source variable into target variables

To construct T_TRPEOPLE, we recode source values that express respondents' trust in people into the target value 1, and values that express respondents' ambivalence (e.g. neither trust nor distrust) or distrust into the target value 0. In the case of odd-numbered source scales, the midpoint is coded as 0. We harmonize into T_TRPEOPLE source items with 2-, 4-, 5-, 7-, 10- and 11-point scales.

To construct the 11-point scale and distribution-preserving target variables, use only 4-, 5-, 7-, 10- and 11-point scales. First, we 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 trust, and higher scores correspond to higher trust (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

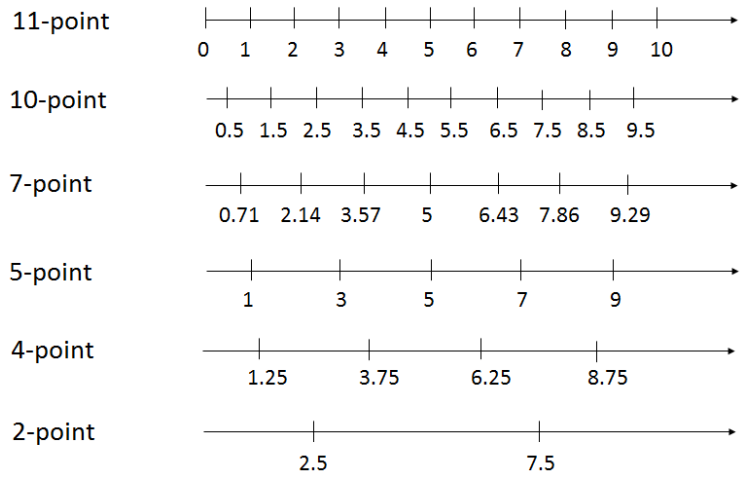


**Table 3.2.1. Creating the 11-point scale (from 0 to 10), with median and mean values 5, and minimized inter-scale differences in the variability**

| 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 trust 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_TRPEOPLE_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 trust | 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 trust | 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 TRUST IN PEOPLE

The main variation in source questions about trust in people 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_TRPEOPLE_LENGTH is a nominal variable that identifies the length (i.e. number of answer options) of source scales used to construct either of the three target variables. It can take the values: 2, 4, 5, 7, 10, 11.

   Note that we do not harmonize 2-point scales into T_TRPEOPLE_11 and T_TRPEOPLE_DISTRIB. For these two targets, C_TRPEOPLE_LENGTH takes the values 4, 5, 7, 10, 11.

2. C_TRPEOPLE_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 trust); it takes the value 0 when source scale values are in descending order (i.e. response options are ordered from most to least trust).
3. C_TRPEOPLE_NONSTANDARD is coded 1 if the wording of the source question departs from the formulation treated as standard in SDR, namely, "*Generally speaking, would you say that most people can be trusted or that you must be very careful in dealing with people?"* When the source question is formulated in this standard (i.e. most frequent type of wording in the source data) way, we assign the control the value 0.

**4. Special cases**

- ABS waves 1 and 2 feature a discrepancy between the scale direction presented in source questionnaires on the one hand, and codebooks and data dictionaries on the other. Specifically, for wave 1 the questionnaire provides the following scale:

  *1 = A great deal of trust; 2 = Quite a lot of trust; 3 = Not very much trust; 4 = None at all;*

  <u>while the codebook and data dictionary provide:</u>

  *0 = Not sure; 1= None at all; 2 = Not very much trust; 3 = Quite a lot of trust; 4 = A great deal of trust; 98 = Don't know; 99 = No Answer*.

  For wave 2, the scale according to the questionnaire is:

  *1 = A great deal of trust; 2 = Quite a lot of trust; 3 = Not very much trust; 4 = None at all; 7 =DU; 8 = CC; 9 = DA*

  <u>while the scale from codebook and data dictionary reads:</u>

  *1 = None at all; 2 = Not Very Much Trust; 3 = Quite a Lot of Trust; 4 = A Great Deal of Trust; 7 = Do not understand the question; 8 = Can't choose; 9 = Decline to answer*).

  We use the scales as provided in the codebook and data dictionary, whose values are in ascending order, to construct the target variable. Yet, for the control variable C_TRPEOPLE_ASCEND it is important what scale direction the respondent was presented with; hence, we assign C_TRPEOPLE_ASCEND the code 0 (descending), based on the direction of the scale as available in the questionnaire.

- Some variables used in SDR1.0 were removed from the SDR2.0 dataset, as it was decided they were not well enough fitting our target concepts. That concerns e.g. variables in ISSP/2006 (question wording: *To what extent do you agree or disagree with the following statements.(...) There are only a few people I can trust completely*.), CB/2009 (question wording: *For each of the following statements, please tell me whether it describes your feelings, more or less describes your feelings, or does not describe your feelings. (...) There are many people I can trust completely*.)

There are also instances of nonstandard question wording, which we code C_TRPEOPLE_NONSTANDARD = 1 (see T_TRUSTS_DVR_SDR2.xlsx for details). These are:

- NBB/1-6 and NEB/6-7: *To what extent do you trust each of the following institutions to look after your interests? Please indicate on a scale with 1 for no trust at all and 7 great trust: Most people you know.*
- NEB/5: *There are many different institutions in this country, for example, government, courts, police, civil servants. Please show me on this 7-point scale, where 1 represents no trust and 7 great trust, how great is your personal trust in each of these institutions: q Most people you meet*
- VPCPCE: *I'd like to ask how much you feel you can trust some people and other things. Please could you choose the answer that best represents your opinion. Most ordinary people that you meet in everyday life*

We treat the following source question formulations as 'standard', despite some differences in wording that appear in the English language version of the survey documentation (see T_TRUSTS_DVR_SDR2.xlsx for details on source question wording by survey datafile). The cases below are coded C_TRPEOPLE_NONSTANDARD = 0.

- *Generally speaking, would you say that most people can be trusted, or that **you can't be too careful** in dealing with people?*
- *Generally speaking, would you say that "Most people can be trusted" or "that **you must be very careful** in dealing with people"?*
- *Generally speaking, would you say that "Most people can be trusted" or "you can"t be too careful i**n dealing with them**"?*
- *Generally speaking, would you say that you can trust most people, or that you can **never** be too careful **when dealing with others**?*
- *Generally speaking, would you say that most people **in /country/** can be trusted, or that you can't be too careful in dealing with people?*
- ARB/1 and EQLS: *"Generally speaking, would you say that most people can be trusted?"* [Note: unbalanced question]
- ARB/2-4: *Generally speaking, do you think most people are trustworthy or not?* [Note: unbalanced question]
- ISSP/2014: *Generally speaking, would you say that people can be trusted or that you can't be too careful in dealing with people? (fully labeled 4-point scale)*. Same survey wave (question wording in Venezuela only): *Generally speaking, would you say that people can always, almost always, almost never or never be trusted?* Values: *1. People*

*can almost always be trusted (VE: always) 2. People can usually be trusted (VE: almost always) 3. You usually can't be too careful in dealing with people (VE: almost never) 4. You almost always can't be too careful in dealing with people (VE: never)* [Note: all four-point length scales in our source data are fully labeled; these are five ISSP survey waves from 1998, 2004, 2007, 2014]

**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.