

Health Equity Assessment Toolkit Plus

Upload Database Edition

TECHNICAL NOTES

© Copyright World Health Organization, 2016–2018.

Disclaimer

Your use of these materials is subject to the Terms of Use and Software Licence agreement – see Readme file, pop-up window or License tab under About in HEAT Plus – and by using these materials you affirm that you have read, and will comply with, the terms of those documents.

Suggested Citation

Health Equity Assessment Toolkit Plus (HEAT Plus): Software for exploring and comparing health inequalities in countries. Upload database edition. Version 2.0. Geneva, World Health Organization, 2018.

Contents

1. Introduction	1
2. Disaggregated data	1
2.1 Health indicators.....	1
2.2 Dimensions of inequality	2
3. Summary measures	2
3.1 Absolute measures.....	3
3.1.1 Absolute concentration index	3
3.1.2 Between-group standard deviation	4
3.1.3 Between-group variance	4
3.1.4 Difference	5
3.1.5 Population attributable risk	6
3.1.6 Slope index of inequality	6
3.1.7 Unweighted mean difference from best performing subgroup	7
3.1.8 Unweighted mean difference from mean	7
3.1.9 Weighted mean difference from best performing subgroup	8
3.1.10 Weighted mean difference from mean	8
3.2 Relative measures.....	9
3.2.1 Coefficient of variation	9
3.2.2 Mean log deviation	9
3.2.3 Population attributable fraction	10
3.2.4 Ratio	10
3.2.5 Relative concentration index	11
3.2.6 Relative index of inequality	12
3.2.7 Theil index.....	12
3.2.8 Unweighted index of disparity	13
3.2.9 Weighted index of disparity.....	13

Supplementary tables

Supplementary table 1 Summary measures of inequality: formulas, characteristics and interpretation.....	15
---	----

1. Introduction

The Health Equity Assessment Toolkit Plus (HEAT Plus) enables the assessment of within-country inequalities, i.e. inequalities that exist between population subgroups within a country, based on disaggregated data and summary measures of inequality. Disaggregated data show the level of health by population subgroup of a given dimension of inequality. Summary measures build on disaggregated data and present the degree of inequality across multiple population subgroups in a single numerical figure. These technical notes provide information about disaggregated data in general (section 2) and detailed information about the summary measures calculated in HEAT Plus (section 3).

2. Disaggregated data

Assessing within-country health inequalities requires the use of health data that are disaggregated according to relevant dimensions of inequality (i.e. demographic, socio-economic, or geographical factors). Disaggregated data break down overall averages, revealing differences in health indicators between different population subgroups. This is essential for the development of equity-oriented interventions, especially those on the path to universal health coverage, serving as the basis for the design and re-orientation of rights-based, gender-responsive and equitable health systems.

Two types of data are required calculating disaggregated data: data about “health indicators” that describe an individual’s experience of health and data about “dimensions of inequality” that allow populations to be organized into subgroups according to their demographic, geographic and/or socioeconomic characteristics.

The following two sections provide more information about the different types and characteristics of health indicators and dimensions of inequality. Further information about disaggregated data can be found in the *Handbook on health inequality monitoring: with a special focus on low- and middle-income countries*.¹ For instructions on how to calculate disaggregated data using household survey data in R, SAS, SPSS and Stata, please refer to http://www.who.int/gho/health_equity/statistical_codes/en/. For instructions on how to prepare disaggregated data for upload to HEAT Plus, please refer to the template and to the user manual.

2.1 Health indicators

There are different types of health indicators, that may be reported at different scales. Differentiating between the different types and scales is important as these characteristics have implications for the calculation of summary measures (see section 3).

Health indicators can be divided into favourable and adverse health indicators. **Favourable health indicators** measure desirable health events that are promoted through public health action. They include health intervention indicators, such as antenatal care coverage, and desirable health outcome indicators, such as life expectancy. For these indicators, the ultimate goal is to achieve a maximum level, either in health intervention coverage or health outcome (for example, complete coverage of antenatal care or the highest possible life expectancy). **Adverse health indicators**, on the other hand, measure undesirable events, that are to be reduced or eliminated through public health action. They include undesirable health outcome indicators, such as stunting prevalence in children aged less

¹ World Health Organization (2013). Handbook on health inequality monitoring: with a special focus on low- and middle-income countries. Geneva: World Health Organization. Available from: www.who.int/gho/health_equity/handbook/en/

than five years or under-five mortality rate. Here, the ultimate goal is to achieve a minimum level in health outcome (for example, a stunting prevalence or mortality rate of zero).

Furthermore, health indicators can be reported at different **indicator scales**. For example, while total fertility rate is usually reported as the number of children *per woman* (indicator scale = 1), coverage of skilled birth attendance is reported as a *percentage* (indicator scale = 100) and neonatal mortality rate is reported as the number of deaths *per 1000 live births* (indicator scale = 1000).

In HEAT Plus, you must provide information about the indicator type (favourable vs. adverse) and the indicator scale for each indicator by filling in the variables "favourable_indicator" and "indicator_scale" in the template. Please refer to the user manual or the template for instructions on how to correctly fill in these variables.

2.2 Dimensions of inequality

Similarly to health indicators, there are different types of dimensions of inequality with different characteristics. It is important to take these characteristics into account as they have implications for the calculation of summary measures, too (see section 3).

At the most basic level, dimensions of inequality can be divided into **binary dimensions**, i.e. dimensions that compare the situation in two population subgroups (e.g. males and females), versus dimensions that look at the situation in more than two population subgroups (e.g. economic status quintiles).

In the case of dimensions with more than two population subgroups it is possible to differentiate between dimensions with ordered subgroups and non-ordered subgroups. **Ordered dimensions** have subgroups with an inherent positioning that can be ranked. For example, education has an inherent ordering of subgroups in the sense that those with less education unequivocally have *less* of something compared to those with more education. **Non-ordered dimensions**, by contrast, have subgroups that are not based on criteria that can be logically ranked. Subnational regions are an example of non-ordered groupings.

For ordered dimensions, subgroups can be ranked from the most-disadvantaged to the most-advantaged subgroup. The **subgroup order** defines the rank of each subgroup. For example, if education is categorized in three subgroups (no education, primary school, and secondary school or higher), then subgroups may be ranked from no education (subgroup order = 1) to secondary school or higher (subgroup order = 3).

For non-ordered dimensions, while it is not possible to rank subgroups, it is possible to identify a **reference subgroup**, that serves as a benchmark. For example, for subnational regions, the region with the capital city may be selected as the reference subgroup in order to compare the situation in all other regions with the situation in the capital city.

HEAT Plus automatically recognizes the number of subgroups for each dimension, but you must provide information about the dimension type (ordered or non-ordered), the subgroup order and the reference subgroup for each dimension by filling in the variables "ordered_dimension", "subgroup_order" and "reference_subgroup" in the template. Please refer to the user manual or the template for instructions on how to correctly fill in these variables.

3. Summary measures

Summary measures are calculated based on disaggregated data, combining estimates of a given health indicator for two or more subgroups into a single numerical figure. Supplementary table 1 lists

the 19 summary measures currently calculated in HEAT Plus along with their basic characteristics, formulas and interpretation.

As indicated in supplementary table 1, summary measures of inequality can be divided into absolute measures and relative measures. For a given health indicator, **absolute inequality measures** indicate the magnitude of difference in health between subgroups. They retain the same unit as the health indicator.² **Relative inequality measures**, on the other hand, show proportional differences in health among subgroups and have no unit.

Furthermore, summary measures may be weighted or unweighted. **Weighted measures** take into account the population size of each subgroup, while **unweighted measures** do not. Importantly, simple measures are always unweighted and complex measures may be weighted or unweighted.

Simple measures make pairwise comparisons between two subgroups, such as the most and least wealthy. They can be calculated for all health indicators and dimensions of inequality. The characteristics of the indicator and dimension determine which two subgroups are compared to assess inequality. Contrary to simple measures, **complex measures** make use of data from all subgroups to assess inequality. They can be calculated for all health indicators, but they are only calculated for dimensions with more than two subgroups.³

Complex measures can further be divided into ordered complex measures and non-ordered complex measures of inequality. **Ordered measures** can only be calculated for ordered dimensions. Here, the calculation is also influenced by the type of indicator (favourable vs. adverse). **Non-ordered measures** are only calculated for non-ordered dimensions.⁴

Lastly, each summary measure makes use of a **reference group** in their calculation. The reference group serves as a benchmark from which differences among subgroups are measured. This can be the setting average or a subgroup estimate, depending on the measure. In HEAT Plus, you can select a reference subgroup for binary dimensions and non-ordered dimensions, which impacts on the calculation of some summary measures.

The following sections give further information about the definition, calculation and interpretation of each summary measure of inequality. Further information about summary measures of inequality can be found in the *Handbook on health inequality monitoring: with a special focus on low- and middle-income countries*.⁵

3.1 Absolute measures

3.1.1 Absolute concentration index

Definition

The absolute concentration index (ACI) is a complex, weighted measure of inequality that shows the gradient across multiple subgroups with a natural ordering, on an absolute scale. It indicates the extent to which an indicator is concentrated among the disadvantaged or the advantaged.

Calculation

The calculation of ACI is based on a ranking of the whole population from the most-disadvantaged subgroup (at rank zero or 0) to the most-advantaged subgroup (at rank 1), which is inferred from the

² One exception to this is the between-group variance (BGV), which takes the squared unit of the indicator.

³ Exceptions to this are the population attributable risk (PAR) and the population attributable fraction (PAF), which can be calculated for all dimensions of inequality.

⁴Non-ordered complex measures could also be calculated for ordered dimensions, however, in practice, they are not used for such dimensions and are therefore only reported for non-ordered dimensions.

⁵ World Health Organization (2013). Handbook on health inequality monitoring: with a special focus on low- and middle-income countries. Geneva: World Health Organization. Available from: www.who.int/gho/health_equity/handbook/en/

ranking and size of the subgroups. The relative rank of each subgroup is calculated as: $X_j = \sum_j p_j - 0.5p_j$. Based on this ranking, ACI can be calculated as:

$$(1) \quad ACI = \sum_j p_j (2X_j - 1)y_j,$$

where y_j indicates the estimate for subgroup j , p_j the population share of subgroup j and X_j the relative rank of subgroup j .

ACI is calculated for ordered dimensions. It is missing if at least one subgroup estimate or subgroup population share is missing.

Interpretation

If there is no inequality, ACI takes the value zero. Positive values indicate a concentration of the indicator among the advantaged, while negative values indicate a concentration of the indicator among the disadvantaged. The larger the absolute value of ACI, the higher the level of inequality.

3.1.2 Between-group standard deviation

Definition

The between-group standard deviation (BGSD) is a complex, weighted measure of inequality.

Calculation

BGSD is calculated as the square root of the weighted sum of squared differences between the subgroup estimates y_j and the setting average μ . Squared differences are weighted by each subgroup's population share p_j :

$$(2) \quad BGSD = \sqrt{\sum_j p_j (y_j - \mu)^2}.$$

BGSD is calculated for non-ordered dimensions. It is missing if at least one subgroup estimate or subgroup population share is missing.

Interpretation

BGSD takes only positive values with larger values indicating higher levels of inequality. BGSD is zero if there is no inequality. BGSD is more sensitive to outlier estimates as it gives more weight to the estimates that are further from the setting average.

3.1.3 Between-group variance

Definition

The between-group variance (BGV) is a complex, weighted measure of inequality.

Calculation

BGV is calculated as the weighted sum of squared differences between the subgroup estimates y_j and the setting average μ . Squared differences are weighted by each subgroup's population share p_j :

$$(3) \quad BGV = \sum_j p_j (y_j - \mu)^2,$$

where μ is calculated as the weighted average of subgroup estimates y_j (weighted by each subgroup's population share p_j): $\mu = \sum_j p_j * y_j$.

BGV is calculated for non-ordered dimensions. It is missing if at least one subgroup estimate or subgroup population size is missing.

Interpretation

BGV takes only positive values with larger values indicating higher levels of inequality. BGV is zero if there is no inequality. BGV is more sensitive to outlier estimates as it gives more weight to the estimates that are further from the setting average.

3.1.4 Difference

Definition

The difference (D) is a simple, unweighted measure of inequality that shows the absolute inequality between two subgroups.

Calculation

D is calculated as the difference between two subgroups. For ordered dimensions, the most-advantaged and most-disadvantaged subgroups are compared. For non-ordered dimensions and binary dimensions, either a selected reference subgroup is compared with another subgroup or, if no reference subgroup has been selected, the two subgroups with the highest and lowest estimates are used:

$$(4) \quad D = y_{high} - y_{low}.$$

Note that the selection of y_{high} and y_{low} depends on the characteristics of the dimension of inequality and the type of indicator, for which D is calculated. For ordered dimensions, y_{high} refers to the most-advantaged subgroup and y_{low} to the most-disadvantaged subgroup in the case of favourable indicators, and vice versa in the case of adverse indicators. For non-ordered and binary dimensions, for which a reference subgroup has been selected, y_{high} refers to the selected reference subgroup and y_{low} to the subgroup that maximizes the absolute difference in the case of favourable indicators, and vice versa in the case of adverse indicators. For non-ordered and binary dimensions without a selected reference subgroup, the lowest estimate is subtracted from the highest estimate, regardless of the indicator type.

D is calculated for all dimensions of inequality. In the case of ordered dimensions, D is missing if the estimates for the most-advantaged and/or most-disadvantaged subgroup are missing. In the case of non-ordered dimensions and binary dimensions, D is missing if at least one subgroup estimate is missing.

Interpretation

If there is no inequality, D takes the value zero. Greater absolute values indicate higher levels of inequality. For favourable indicators, positive values indicate a higher concentration of the indicator among the advantaged subgroups and negative values indicate a higher concentration among the disadvantaged subgroups. For adverse indicators, positive values indicate a higher concentration of the indicator among the disadvantaged and negative values indicate a higher concentration among the advantaged.

Other calculations

For non-ordered dimensions with more than 30 subgroups, additional difference measures are calculated, including

- Difference between percentile 80 and percentile 20
- Difference between the mean estimates in quintile 5 and quintile 1.

In addition, for non-ordered dimensions with more than 60 subgroups, the following difference measures are calculated:

- Difference between percentile 90 and percentile 10
- Difference between the mean estimates in decile 10 and decile 1.

Finally, for non-ordered dimensions with more than 100 subgroups, the following difference measures are also calculated:

- Difference between percentile 95 and percentile 5
- Difference between the mean estimates in the top 5% and the bottom 5%.

These measures are displayed in the Explore Inequality – Disaggregated data (detailed bar graphs) tab, provided a non-ordered dimension with at least 30 subgroups has been selected.

3.1.5 Population attributable risk

Definition

The population attributable risk (PAR) is a complex, weighted measure of inequality that shows the potential for improvement in the setting average of an indicator that could be achieved if all subgroups had the same level as a reference subgroup.

Calculation

PAR is calculated as the difference between the estimate for the reference subgroup y_{ref} and the setting average μ :

$$(5) \quad PAR = y_{ref} - \mu,$$

where μ is calculated as the weighted average of subgroup estimates y_j (weighted by each subgroup's population share p_j): $\mu = \sum_j p_j * y_j$.

Note that the selection of the reference subgroup y_{ref} depends on the characteristics of the dimension of inequality and the type of indicator, for which PAR is calculated. For ordered dimensions, y_{ref} refers to the most-advantaged subgroup, regardless of the indicator type. For non-ordered dimensions and binary dimensions, y_{ref} either refers to the selected reference subgroup or, if no reference subgroup has been selected, y_{ref} refers to the subgroup with the highest estimate in the case of favourable indicators and to the subgroup with the lowest estimate in the case of adverse indicators.

PAR is calculated for all dimensions. In the case of ordered dimensions, PAR is missing if the estimate for the reference subgroup or the population size for at least one subgroup is missing. In the case of non-ordered dimensions and binary dimensions, PAR is missing if at least one subgroup estimate or subgroup population size is missing.

Interpretation

PAR takes positive values for favourable indicators and negative values for adverse indicators. The larger the absolute value of PAR, the higher the level of inequality. PAR is zero if no further improvement can be achieved, i.e. if all subgroups have reached the same level as the reference subgroup.

3.1.6 Slope index of inequality

Definition

The slope index of inequality (SII) is a complex, weighted measure of inequality that indicates the absolute difference in estimated values of an indicator between the most-advantaged and most-disadvantaged (or vice versa for adverse indicators), while taking into consideration all the other subgroups – using an appropriate regression model.

Calculation

To calculate SII, a weighted sample of the whole population is ranked from the most-disadvantaged subgroup (at rank zero or 0) to the most-advantaged subgroup (at rank 1). This ranking is weighted, accounting for the proportional distribution of the population within each subgroup. The population of each subgroup is then considered in terms of its range in the cumulative population distribution, and the midpoint of this range. According to the definition currently used in HEAT Plus, the indicator of interest is then regressed against this midpoint value using a generalized linear model with logit link, and the predicted values of the indicator are calculated for the two extremes (rank 1 and rank 0).

For favourable indicators, the difference between the estimated values at rank 1 (v_1) and rank 0 (v_0) (covering the entire distribution) generates the SII value:

$$(6) \quad (a) \quad SII = v_1 - v_0.$$

For adverse indicators, the calculation is reversed and the SII value is calculated as the difference between the estimated values at rank 0 (v_0) and rank 1 (v_1) (covering the entire distribution):

$$(6) \quad (b) \quad SII = v_0 - v_1.$$

SII is calculated for ordered dimensions. It is missing if at least one subgroup estimate or subgroup population size is missing.

Interpretation

If there is no inequality, SII takes the value zero. Greater absolute values indicate higher levels of inequality. For favourable indicators, positive values indicate a higher concentration of the indicator among the advantaged subgroups and negative values indicate higher a higher concentration among the disadvantaged subgroups. For adverse indicators, positive values indicate a higher concentration of the indicator among the disadvantaged and negative values indicate a higher concentration among the advantaged.

3.1.7 Unweighted mean difference from best performing subgroup

Definition

The unweighted mean difference from best performing subgroup (MDBU) is a complex, unweighted measure of inequality that shows the difference between each subgroup and a reference subgroup, on average.

Calculation

MDBU is calculated as the sum of absolute differences between the subgroup estimates y_j and the estimate for the reference group y_{ref} :

$$(7) \quad MDBU = \sum_j |y_j - y_{ref}|.$$

y_{ref} either refers to the selected reference subgroup or, if no reference subgroup has been selected, y_{ref} refers to the subgroup with the highest estimate in the case of favourable indicators and to the subgroup with the lowest estimate in the case of adverse indicators.

MDBU is calculated for non-ordered dimensions. It is missing if at least one subgroup estimate is missing. Note that the 95% confidence intervals calculated for MDBU are simulation-based estimates.

Interpretation

MDBU takes only positive values with larger values indicating higher levels of inequality. MDBU is zero if there is no inequality.

3.1.8 Unweighted mean difference from mean

Definition

The unweighted mean difference from mean (MDMU) is a complex, unweighted measure of inequality that shows the difference between each subgroup and the setting average, on average.

Calculation

MDMU is calculated as the sum of absolute differences between the subgroup estimates y_j and the setting average μ :

$$(8) \quad MDMU = \sum_j |y_j - \mu|,$$

where μ is calculated as the weighted average of subgroup estimates y_j (weighted by each subgroup's population share p_j): $\mu = \sum_j p_j * y_j$.

MDMU is calculated for non-ordered dimensions. It is missing if at least one subgroup estimate or subgroup population share is missing. Note that the 95% confidence intervals calculated for MDMU are simulation-based estimates.

Interpretation

MDMU takes only positive values with larger values indicating higher levels of inequality. MDMU is zero if there is no inequality.

3.1.9 Weighted mean difference from best performing subgroup

Definition

The weighted mean difference from best performing subgroup (MDBW) is a complex, weighted measure of inequality that shows the difference between each subgroup and a reference subgroup, on average.

Calculation

MDBW is calculated as the weighted sum of absolute differences between the subgroup estimates y_j and the estimate for the reference subgroup y_{ref} . Absolute differences are weighted by each subgroup's population share p_j :

$$(9) \quad MDBW = \sum_j p_j |y_j - y_{ref}|.$$

y_{ref} either refers to the selected reference subgroup or, if no reference subgroup has been selected, y_{ref} refers to the subgroup with the highest estimate in the case of favourable indicators and to the subgroup with the lowest estimate in the case of adverse indicators.

MDBW is calculated for non-ordered dimensions. It is missing if at least one subgroup estimate is missing. Note that the 95% confidence intervals calculated for MDBW are simulation-based estimates.

Interpretation

MDBW takes only positive values with larger values indicating higher levels of inequality. MDBW is zero if there is no inequality.

3.1.10 Weighted mean difference from mean

Definition

The weighted mean difference from mean (MDMW) is a complex, weighted measure of inequality that shows the difference between each subgroup and the setting average, on average.

Calculation

MDMW is calculated as the weighted sum of absolute differences between the subgroup estimates y_j and the setting average μ . Absolute differences are weighted by each subgroup's population share p_j :

$$(10) \quad MDMW = \sum_j p_j |y_j - \mu|,$$

where μ is calculated as the weighted average of subgroup estimates y_j (weighted by each subgroup's population share p_j): $\mu = \sum_j p_j * y_j$.

MDMW is calculated for non-ordered dimensions. It is missing if at least one subgroup estimate or subgroup population size is missing. Note that the 95% confidence intervals calculated for MDMW are simulation-based estimates.

Interpretation

MDMW takes only positive values with larger values indicating higher levels of inequality. MDMW is zero if there is no inequality.

3.2 Relative measures

3.2.1 Coefficient of variation

Definition

The coefficient of variation (COV) is a complex, weighted measure of inequality.

Calculation

COV is calculated by dividing the between-group standard deviation (BGSD) by the setting average μ and multiplying the fraction by 100:

$$(11) \quad COV = \frac{BGSD}{\mu} * 100,$$

where μ is calculated as the weighted average of subgroup estimates y_j (weighted by each subgroup's population share p_j): $\mu = \sum_j p_j * y_j$.

COV is calculated for non-ordered dimensions. It is missing if at least one subgroup estimate or subgroup population share is missing.

Interpretation

COV takes only positive values with larger values indicating higher levels of inequality. COV is zero if there is no inequality.

3.2.2 Mean log deviation

Definition

The mean log deviation (MLD) is a complex, weighted measure of inequality.

Calculation

MLD is calculated as the sum of products between the negative natural logarithm of the share of the indicator of each subgroup ($-\ln\left(\frac{y_j}{\mu}\right)$) and the population share of each subgroup (p_j). MLD may be more easily interpreted when multiplied by 1000:

$$(12) \quad MLD = \sum_j p_j \left(-\ln\left(\frac{y_j}{\mu}\right)\right) * 1000,$$

where y_j indicates the estimate for subgroup j , p_j the population share of subgroup j and μ the setting average, which is calculated as the weighted average of subgroup estimates y_j (weighted by each subgroup's population share p_j): $\mu = \sum_j p_j * y_j$.

MLD is calculated for non-ordered dimensions. It is missing if at least one subgroup estimate or subgroup population size is missing.

Interpretation

If there is no inequality, MLD takes the value zero. Greater absolute values indicate higher levels of inequality. MLD is more sensitive to differences further from the setting average (by the use of the logarithm).

3.2.3 Population attributable fraction

Definition

The population attributable fraction (PAF) is a complex, weighted measure of inequality that shows the potential for improvement in the setting average of an indicator, in relative terms, that could be achieved if all subgroups had the same level as a reference subgroup.

Calculation

PAF is calculated by dividing the population attributable risk (PAR) by the setting average μ and multiplying the fraction by 100:

$$(13) \quad PAF = \frac{PAR}{\mu} * 100,$$

where μ is calculated as the weighted average of subgroup estimates y_j (weighted by each subgroup's population share p_j): $\mu = \sum_j p_j * y_j$.

PAF is calculated for all dimensions. For ordered dimensions, PAF is missing if the estimate for the reference subgroup or the population size for at least one subgroup is missing. For non-ordered and binary dimensions, PAF is missing if at least one subgroup estimate or subgroup population size is missing.

Interpretation

PAF takes positive values for favourable indicators and negative values for adverse indicators. The larger the absolute value of PAF, the larger the degree of inequality. PAF is zero if no further improvement can be achieved, i.e. if all subgroups have reached the same level as the reference subgroup.

3.2.4 Ratio

Definition

The ratio (R) is a simple, unweighted measure of inequality that shows the relative inequality between two subgroups.

Calculation

R is calculated as the ratio of two subgroups. For ordered dimensions, the most-advantaged and most-disadvantaged subgroups are compared. For non-ordered and binary dimensions,, either a selected reference subgroup is compared with another subgroup or, if no reference subgroup has been selected, the two subgroups with the highest and lowest estimates are used:

$$(14) \quad R = y_{high}/y_{low}.$$

Note that the selection of y_{high} and y_{low} depends on the characteristics of the dimension of inequality and the type of indicator, for which R is calculated. For ordered dimensions, y_{high} refers to the most-advantaged subgroup and y_{low} to the most-disadvantaged subgroup in the case of favourable indicators, and vice versa in the case of adverse indicators. For non-ordered dimensions and binary

dimensions, for which a reference subgroup has been selected, y_{high} refers to the selected reference subgroup and y_{low} to the subgroup that maximizes the ratio in the case of favourable indicators, and vice versa in the case of adverse indicators. For non-ordered dimensions and binary dimensions without a selected reference subgroup, the highest estimate is divided by the lowest estimate, regardless of the indicator type.

R is calculated for all dimensions of inequality. In the case of ordered dimensions, R is missing if the estimates for the most-advantaged and/or most-disadvantaged subgroup are missing. In the case of non-ordered dimensions and binary dimensions,, R is missing if at least one subgroup estimate is missing.

Interpretation

If there is no inequality, R takes the value one. It takes only positive values (larger or smaller than 1). The further the value of R from 1, the higher the level of inequality.

Other calculations

For non-ordered dimensions with more than 30 subgroups, additional ratio measures are calculated, including

- Ratio of percentile 80 to percentile 20
- Ratio of the mean estimates in quintile 5 to quintile 1.

In addition, for non-ordered dimensions with more than 60 subgroups, the following ratio measures are calculated:

- Ratio of percentile 90 to percentile 10
- Ratio of the mean estimates in decile 10 to decile 1.

Finally, for non-ordered dimensions with more than 100 subgroups, the following ratio measures are also calculated:

- Ratio of percentile 95 to percentile 5
- Ratio of the mean estimates in the top 5% to the bottom 5%.

These measures are displayed in the Explore Inequality – Disaggregated data (detailed bar graphs) tab, provided a non-ordered dimension with at least 30 subgroups has been selected.

3.2.5 Relative concentration index

Definition

The relative concentration index (RCI) is a complex, weighted measure of inequality that shows the gradient across multiple subgroups with natural ordering, on a relative scale. It indicates the extent to which an indicator is concentrated among the disadvantaged or the advantaged.

Calculation

RCI is calculated by dividing the absolute concentration index (ACI) by the setting average μ . This fraction may be more easily interpreted when multiplied by 100:

$$(15) \quad RCI = \frac{ACI}{\mu} * 100,$$

where μ is calculated as the weighted average of subgroup estimates y_j (weighted by each subgroup's population share p_j): $\mu = \sum_j p_j * y_j$.

RCI is calculated for ordered dimensions. It is missing if at least one subgroup estimate or subgroup population size is missing.

Interpretation

RCI is bounded between -1 and +1 (or -100 and +100 if multiplied by 100) and takes the value zero if there is no inequality. Positive values indicate a concentration of the indicator among the advantaged, while negative values indicate a concentration of the indicator among the disadvantaged. The greater the absolute value of RCI, the higher the level of inequality.

3.2.6 Relative index of inequality

Definition

The relative index of inequality (RII) is a complex, weighted measure of inequality that indicates the ratio of estimated values of an indicator of the most-advantaged to the most-disadvantaged (or vice versa for adverse indicators), while taking into consideration all the other subgroups – using an appropriate regression model.

Calculation

To calculate RII, a weighted sample of the whole population is ranked from the most-disadvantaged subgroup (at rank zero or 0) to the most-advantaged subgroup (at rank 1). This ranking is weighted, accounting for the proportional distribution of the population within each subgroup. The population of each subgroup is then considered in terms of its range in the cumulative population distribution, and the midpoint of this range. According to the definition currently used in HEAT Plus, the indicator of interest is then regressed against this midpoint value using a generalized linear model with logit link, and the predicted values of the indicator are calculated for the two extremes (rank 1 and rank 0).

For favourable indicators, the ratio of the estimated values at rank 1 (v_1) to rank 0 (v_0) (covering the entire distribution) generates the RII value:

$$(16) \quad (a) \quad RII = v_1/v_0.$$

For adverse indicators, the calculation is reversed and the RII value is calculated as the ratio of the estimated values at rank 0 (v_0) to rank 1 (v_1) (covering the entire distribution):

$$(16) \quad (b) \quad RII = v_0/v_1.$$

RII is calculated for ordered dimensions. It is missing if at least one subgroup estimate or subgroup population size is missing.

Interpretation

If there is no inequality, RII takes the value one. RII takes only positive values, with values larger than one indicating a concentration of the indicator among the advantaged and values smaller than one indicating a concentration of the indicator among the disadvantaged. The further the value of RII from one, the higher the level of inequality.

3.2.7 Theil index

Definition

The theil index (TI) is a complex, weighted measure of inequality.

Calculation

TI is calculated as the sum of products of the natural logarithm of the share of the indicator of each subgroup ($\ln \frac{y_j}{\mu}$), the share of the indicator of each subgroup ($\frac{y_j}{\mu}$) and the population share of each subgroup (p_j). TI may be more easily interpreted when multiplied by 1000:

$$(17) \quad TI = \sum_j p_j \frac{y_j}{\mu} \ln \frac{y_j}{\mu} * 1000,$$

where y_j indicates the estimate for subgroup j , p_j the population share of subgroup j and μ the setting average, which is calculated as the weighted average of subgroup estimates y_j (weighted by each subgroup's population share p_j): $\mu = \sum_j p_j * y_j$.

TI is calculated for non-ordered dimensions. It is missing if at least one subgroup estimate or subgroup population size is missing.

Interpretation

If there is no inequality, TI takes the value zero. Greater absolute values indicate higher levels of inequality. TI is more sensitive to differences further from the setting average (by the use of the logarithm).

3.2.8 Unweighted index of disparity

Definition

The unweighted index of disparity (IDISU) is a complex, unweighted measure of inequality that shows the proportional difference between each subgroup and the setting average, on average.

Calculation

IDISU is calculated as the sum of absolute differences between the subgroup estimates y_j and the setting average μ , divided by the setting average μ and the number of subgroups n :

$$(18) \quad IDISU = \frac{1}{n} * \frac{\sum_j |y_j - \mu|}{\mu} * 100,$$

where μ is calculated as the weighted average of subgroup estimates y_j (weighted by each subgroup's population share p_j): $\mu = \sum_j p_j * y_j$.

IDISU is calculated for non-ordered dimensions. It is missing if at least one subgroup estimate or subgroup population size is missing. Note that the 95% confidence intervals calculated for IDISU are simulation-based estimates.

Interpretation

IDISU takes only positive values with larger values indicating higher levels of inequality. IDISU is zero if there is no inequality.

3.2.9 Weighted index of disparity

Definition

The weighted index of disparity (IDISW) is a complex, weighted measure of inequality that shows the proportional difference between each subgroup and the setting average, on average.

Calculation

IDISW is calculated as the weighted sum of absolute differences between the subgroup estimates y_j and the setting average μ , divided by the setting average μ . Absolute differences are weighted by each subgroup's population share p_j :

$$(19) \quad IDISW = \frac{\sum_j p_j |y_j - \mu|}{\mu} * 100,$$

where μ is calculated as the weighted average of subgroup estimates y_j (weighted by each subgroup's population share p_j): $\mu = \sum_j p_j * y_j$.

IDISW is calculated for non-ordered dimensions. It is missing if at least one subgroup estimate or subgroup population size is missing. Note that the 95% confidence intervals calculated for IDISW are simulation-based estimates.

Interpretation

IDISW takes only positive values with larger values indicating higher levels of inequality. IDISW is zero if there is no inequality.

Supplementary tables

Supplementary table 1 Summary measures of inequality: formulas, characteristics and interpretation

Summary measure (abbreviation)	Formula	Simple or complex	Weighted or unweighted	Ordered or non-ordered (complex only)	Unit	Value of no inequality	Interpretation
Absolute measures							
Absolute concentration index (ACI)	$ACI = \sum_j p_j (2X_j - 1)y_j$	Complex	Weighted	Ordered	Unit of indicator	Zero	Positive (negative) values indicate a concentration of the indicator among the advantaged (disadvantaged). The larger the absolute value of ACI, the higher the level of inequality.
Between-group standard deviation (BGSD)	$BGSD = \sqrt{\sum_j p_j (y_j - \mu)^2}$	Complex	Unweighted	Non-ordered	Unit of indicator	Zero	BGSD takes only positive values with larger values indicating higher levels of inequality.
Between-group variance (BGV)	$BGV = \sum_j p_j (y_j - \mu)^2$	Complex	Weighted	Non-ordered	Squared unit of indicator	Zero	BGV takes only positive values with larger values indicating higher levels of inequality.
Difference (D)	$D = y_{high} - y_{low}$	Simple	Unweighted	-	Unit of indicator	Zero	The larger the absolute value of D, the higher the level of inequality.
Population attributable risk (PAR)	$PAR = y_{ref} - \mu$	Complex	Weighted	Non-ordered	Unit of indicator	Zero	PAR takes only positive values for favourable indicators and only negative values for adverse indicators. The larger the absolute value, the higher the level of inequality.
Slope index of inequality (SII)	$SII = v_1 - v_0$ for favourable indicators; $SII = v_0 - v_1$ for adverse indicators	Complex	Weighted	Ordered	Unit of indicator	Zero	For favourable (adverse) indicators, positive values indicate a concentration among the advantaged (disadvantaged) and negative values indicate a concentration among the disadvantaged (advantaged). The larger the absolute value of SII, the higher the level of inequality.
Unweighted mean difference from best performing subgroup (MDBU)	$MDBU = \sum_j y_j - y_{ref} $	Complex	Unweighted	Non-ordered	Unit of indicator	Zero	MDBU takes only positive values with larger values indicating higher levels of inequality.
Unweighted mean difference from mean (MDMU)	$MDMU = \sum_j y_j - \mu $	Complex	Unweighted	Non-ordered	Unit of indicator	Zero	MDMU takes only positive values with larger values indicating higher levels of inequality.
Weighted mean difference from best performing subgroup (MDBW)	$MDBW = \sum_j p_j y_j - y_{ref} $	Complex	Weighted	Non-ordered	Unit of indicator	Zero	MDBW takes only positive values with larger values indicating higher levels of inequality.
Weighted mean difference from mean (MDMW)	$MDMW = \sum_j p_j y_j - \mu $	Complex	Weighted	Non-ordered	Unit of indicator	Zero	MDMW takes only positive values with larger values indicating higher levels of inequality.
Relative measures							

Coefficient of variation (COV)	$COV = \frac{BGSD}{\mu} * 100$	Complex	Unweighted	Non-ordered	Unit of indicator	Zero	COV takes only positive values with larger values indicating higher levels of inequality.
Mean log deviation (MLD)	$MLD = \sum_j p_j (-\ln(\frac{y_j}{\mu})) * 1000$	Complex	Weighted	Non-ordered	No unit	Zero	The larger the absolute value of MLD, the higher the level of inequality.
Population attributable fraction (PAF)	$PAF = \frac{PAR}{\mu} * 100$	Complex	Weighted	Non-ordered	No unit	Zero	PAF takes only positive values for favourable indicators and only negative values for adverse indicators. The larger the absolute value of PAF, the larger the degree of inequality.
Ratio (R)	$R = y_{high}/y_{low}$	Simple	Unweighted	-	No unit	One	R takes only positive values. The further the value of R from 1, the higher the level of inequality.
Relative concentration index (RCI)	$RCI = \frac{ACI}{\mu} * 100$	Complex	Weighted	Ordered	No unit	Zero	RCI is bounded between -1 and +1 (or -100 and +100 if multiplied by 100). Positive (negative) values indicate a concentration of the indicator among the advantaged (disadvantaged). The larger the absolute value of RCI, the larger the degree of inequality.
Relative index of inequality (RII)	$RII = v_1/v_0$ for favourable indicators; $RII = v_0/v_1$ for adverse indicators	Complex	Weighted	Ordered	No unit	One	RII takes only positive values. The further the value of RII from 1, the higher the level of inequality.
Theil index (TI)	$TI = \sum_j p_j \frac{y_j}{\mu} \ln \frac{y_j}{\mu} * 1000$	Complex	Weighted	Non-ordered	No unit	Zero	The larger the absolute value of TI, the greater the level of inequality.
Unweighted index of disparity (IDISU)	$IDISU = \frac{1}{n} * \frac{\sum_j y_j - \mu }{\mu} * 100$	Complex	Unweighted	Non-ordered	No unit	Zero	IDISU takes only positive values with larger values indicating higher levels of inequality.
Weighted index of disparity (IDISW)	$IDISW = \frac{\sum_j p_j y_j - \mu }{\mu} * 100$	Complex	Weighted	Non-ordered	No unit	Zero	IDISW takes only positive values with larger values indicating higher levels of inequality.

y_j = Estimate for subgroup j.

y_{high} = Estimate for subgroup high. For ordered dimensions, subgroup high refers to the most-advantaged subgroup in the case of favourable indicators and to the most-disadvantaged subgroup in the case of adverse indicators. For non-ordered dimensions with a selected reference subgroup, subgroup high refers to the reference subgroup in the case of favourable indicators and to the subgroup that maximizes the absolute difference in the case of adverse indicators. For non-ordered dimensions without a selected reference subgroup, subgroup high refers to the subgroup with the highest estimate.

y_{low} = Estimate for subgroup low. For ordered dimensions, subgroup low refers to the most-disadvantaged subgroup in the case of favourable indicators and to the most-advantaged subgroup in the case of adverse indicators. For non-ordered dimensions and binary dimensions, for which a reference subgroup has been selected, subgroup low refers to the subgroup that maximizes the absolute difference in the case of favourable indicators and to the reference subgroup in the case of adverse indicators. For non-ordered dimensions and binary dimensions without a selected reference subgroup, subgroup high refers to the subgroup with the lowest estimate.

y_{ref} = Estimate for reference subgroup. For ordered dimensions, the reference group refers to the most-advantaged subgroup. For non-ordered dimensions and binary dimensions, the reference group either refers to the selected reference subgroup or, if no reference subgroup has been selected, to the subgroup with the highest estimate in the case of favourable indicators and to the subgroup with the lowest estimate in the case of adverse indicators.

p_j = Population share for subgroup j.

$X_j = \sum_j p_j - 0.5p_j$ = Relative rank of subgroup j.

$\mu = \sum_j p_j * y_j$ = Setting average.

v_0 = Predicted value of the hypothetical person at the bottom of the social-group distribution (rank 0).

v_1 = Predicted value of the hypothetical person at the top of the social-group distribution (rank 1).

n = Number of subgroups.