Computing IPD scores

DVRPC's Indicators of Potential Disadvantage (IPD) identify populations of interest under Title VI of the Civil Rights Act and the Executive Order on Environmental Justice (#12898) using American Community Survey (ACS) five-year estimates from the U.S. Census Bureau. IPD analysis assists both DVRPC and outside organizations in equity work by identifying populations of interest, including youth, older adults, female, racial minority, ethnic minority, foreign-born, limited English proficiency, disabled, and low-income populations at the census tract level in DVRPC's nine-county region.

This document provides an overview of the calculations required to conduct IPD analysis and is meant to provide enough information to replicate prior years of IPD analysis without getting bogged down on programmatic implementation. It is meant to be the less technical companion of `script_reference.pdf`. See the home of the Github repository at `ipd_2017` and a detailed discussion of the way IPD analysis is implemented in `R` at `script_reference.pdf`.

Data for download and analysis

For each indicator in IPD analysis, download the count, universe, and percentage (if available) from ACS five-year estimates, as well as their associated margins of error (MOEs). The table below gives an overview of the nine indicators used in IPD analysis, their primary ACS data source, and their justification for inclusion as a population of interest.

As an example, for the Older Adults indicator, the count is the number of adults aged 65 or older, the universe is the total population, and the percentage is the percentage of older adults at the census tract level for the DVRPC region.

The file `variables.csv` provides a three-year summary of ACS fields used in IPD analysis. If you're confused about what fields to download, this serves as a useful starting point.

Indicator in IPD analysis	ACS data table for indicator in IPD analysis	Protected class indicator represents	Authorizing source or guiding document
Youth	B09001: Population Under 18 Years by Age	Age	FHWA's Title VI Program and Related Authorities: 23 CFR 200
Older Adults	S0101: Age and Sex	Age	FHWA's Title VI Program and Related Authorities: 23 CFR 200
Female	S0101: Age and Sex	Sex	FHWA's Title VI Program and Related Authorities: 23 CFR 200
Racial Minority	B02001: Race	Race and Minority	Executive Order 12898, Title VI of the Civil Rights Act of 1964, FHWA's Title VI Program and Related Authorities: 23 CFR, **and** Title VI Requirements and Guidelines for FTA Recipients
Ethnic Minority	B03002: Hispanic or Latino Origin by Race	Minority and National Origin	Executive Order 12898, Title VI of the Civil Rights Act of 1964, FHWA's Title VI Program and Related Authorities: 23 CFR, **and** Title VI Requirements and Guidelines for FTA Recipients
Foreign-Born	B05012: Nativity in the United States	National Origin	Title VI of the Civil Rights Act of 1964. FHWA's Title VI Program and Related Authorities: 23 CFR, **and** Title VI Requirements and Guidelines for FTA Recipients
Limited English Proficiency	S1601: Language Spoken at Home	Limited English Proficiency and National Origin	Title VI of the Civil Rights Act of 1964. FHWA's Title VI Program and Related Authorities: 23 CFR 200, **and** Title VI Requirements and Guidelines for FTA Recipients
Disabled	S1810: Disability Characteristics	Disability	FHWA's Title VI Program and Related Authorities: 23 CFR
Low-Income	S1701: Poverty Status in the Past 12 Months	Low-Income	Executive Order 12898 **and** FHWA's Title VI Program and Related Authorities: 23 CFR 200

Compute fields

First, remove "No Data" tracts. These outliers skew standard deviation computations and therefore affect final IPD scores. All but one "No Data" tract have populations of zero; Tract 34021002400 is a prison. GEOIDs of "No Data" tracts in the DVRPC region are:

- 34021002400
- 42045980000
- 42017980000
- 42101980800
- 42101980300
- 42101980500
- 42101980400
- 42101980900
- 42101980700
- 42101980600
- 42101005000

Required fields

The fields required to compute IPD scores are: 1) count estimates, 2) universe estimates, and 3) percentage estimates. IPD scores can be calculated as 4) individual IPD scores, 5) IPD classifications, or 6) IPD composite scores. All are discussed in this section.

1) Count estimates

Count estimates can often be downloaded directly from ACS five-year estimates. The racial minority indicator is the only exception to this rule, where the count estimate is the result of aggregating several subfields from ACS table B02001.

2) Universe estimates

Universe estimates can be downloaded directly from ACS five-year estimates.

3) Percentage estimates

Percentage estimates can sometimes be found in ACS five-year estimates, but other indicators require that the percentage be computed by dividing the count estimate by the universe estimate. Ensure that all percentage estimates range from 0 to 100 and are rounded to the tenths place.
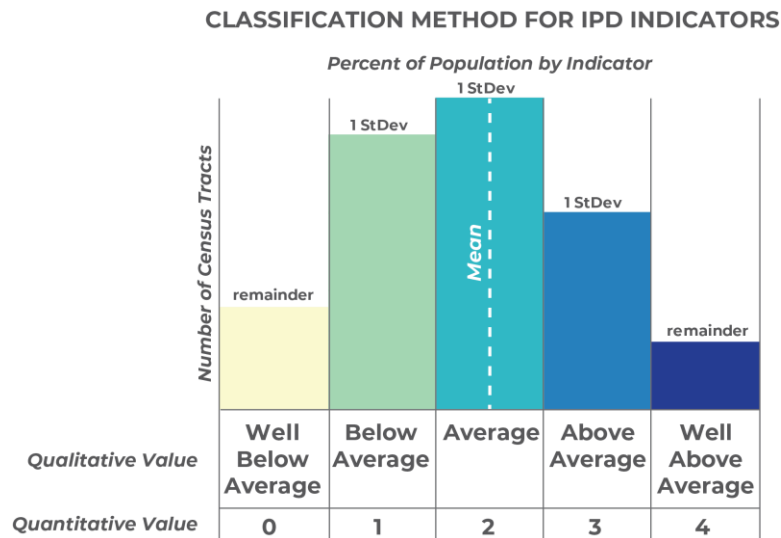
4) IPD score

IPD scores are calculated for each individual indicator and added together to create a composite IPD score. For a given indicator, percentage estimates are split into five bins, where the bin breaks are generally equal to $-1.5 \cdot stdev$, $-0.5 \cdot stdev$, $0.5 \cdot stdev$, and $1.5 \cdot stdev$. There is one exception: if $-1.5 \cdot stdev$ is a negative value, then it is manually reassigned to equal 0.1. This ensures that at least some census tracts fall in the bottom bin regardless of the spread of the indicator. Note that break values range from 0 to 100, that they are unrounded, and that they must be applied to percentage values rounded to the tenths place.

Scores are assigned using the following rubric, where p is the percentage value of a given census tract:

IPD score	Standard deviations
0	$p < -1.5 \cdot stdev$
1	$-1.5 \cdot stdev \leq p < -0.5 \cdot stdev$
2	$-0.5 \cdot stdev \leq p < 0.5 \cdot stdev$
3	$0.5 \cdot stdev \leq p < 1.5 \cdot stdev$
4	$p \geq 1.5 \cdot stdev$

For a given indicator, an IPD score of 2 represents a census tract whose percentage estimate is close to the regional average.

CLASSIFICATION METHOD FOR IPD INDICATORS

Percent of Population by Indicator

5) IPD classification

The IPD classification is a textual explanation of the IPD score.

IPD score	Classification
0	Well Below Average
1	Below Average
2	Average
3	Above Average
4	Well Above Average

6) Composite IPD score

The composite IPD score is computed by adding up all the IPD scores for individual indicators. In theory, the composite IPD score can range from 0 to 36, since each indicator's IPD score can be as high as 4. In practice, the mean composite IPD score in 2017 is 17.63, and the highest observed composite score is 31.

Optional fields

Other fields included in IPD score readouts but not necessary to compute scores are: 1) count MOEs, 2) universe MOEs, 3) percentage MOEs, and 4) percentiles.

1) Count MOEs

Count MOEs can often be downloaded directly from ACS five-year estimates. The racial minority indicator is the only exception to this rule. Because the racial minority count estimate is the result of aggregating several subfields, its count MOE is calculated using variance replicate tables.

Where VR is the sum of the variance replicate subfields and E is the sum of the estimate subfields for Table B02001; and $(MOE)C$ is the count MOE:

$$(MOE)C = 1.645 \sqrt{\frac{4}{80} \sum_{i=1}^{80} (VR_i - E_i)^2}$$

For additional guidance on computing margins of error using variance replicate tables, see Documentation for the 2013-2017 Variance Replicate Estimates Tables.

2) Universe MOEs

Universe MOEs can be downloaded directly from ACS five-year estimates.

3) Percentage MOEs

There's no such thing as a 100% certain tract-level estimate from survey data. If any percentage MOE equals 0, manually reassign its value to 0.1.

Use percentage MOEs when they are available from ACS estimates. Ensure that all percentages range from 0 to 100 and are rounded to the tenths place. When they are not available for direct download, compute MOEs using the following approximation formula.

Where A is the count estimate; $MOE(A)$ is the count MOE; B is the universe estimate; $MOE(B)$ is the universe MOE; P is the percentage estimate ranging from 0 to 100; and $MOE(P)$ is the MOE of the percentage, also ranging from 0 to 100:

$$(MOE)P = \frac{\sqrt{MOE(A)^2 - (P^2 \cdot MOE(B)^2)}}{B}$$

If the value under the square root is negative, then $MOE(P)$ is:

$$(MOE)P = \frac{\sqrt{MOE(A)^2 + (P^2 \cdot MOE(B)^2)}}{B}$$

For additional guidance on computing margins of error for user-derived percentages, see Understanding and Using American Community Survey Data: What All Data Users Need to Know.

4) Percentiles

For a given indicator, compute the empirical cumulative distribution of the percentage estimates. Ensure that values range from 0 to 1 and are rounded to the hundredths place. There are many ways to implement the empirical cumulative distribution including by hand in Excel and the `ecdf` function in R and MATLAB.