Reall Household Income and Affordability Calculator

Methodology Paper: Philippines 2018

This document outlines the methodology used to process the 2018 Philippines Family Income and Expenditure Survey (FIES) data for use in Reall's Household Income and Affordability Calculator. The calculator is available at www.reall.net/calculator.

Background and Data Source

The **Family Income and Expenditure Survey (FIES)**[1] is a nationwide triennial survey conducted by the **Philippine Statistics Authority (PSA)**[2] to collect data on household income, expenditure, and related socio-economic indicators. It is one of the country's key instruments for estimating poverty and income distribution. The FIES adopts a stratified, multi-stage sampling design based on the 2020 Census, ensuring representation at national, regional, and provincial levels. The survey captures both urban and rural households and is conducted in two rounds within a year to minimize recall bias.

This study uses **Volume 2** data from the **2018** FIES rounds, which provide household-level summaries on total income, total expenditure (food and non-food), household size, urban-rural status, and imputed rent values. Each household is uniquely represented, and the inclusion of final survey weights allows for nationally representative estimates

In contrast to many national surveys where individual- and household-level data are stored in separate files (necessitating a merge using a household identifier such as hhid), the Philippines Family Income and Expenditure Survey (FIES) used in this study provides a consolidated household-level dataset in Volume 2. Each row in this dataset represents a unique household and includes aggregated variables such as total monthly expenditure, household size, number of employed members, and rental expenses.

Because the dataset is already structured at the household level, there is no need to use hhid as a merging key or grouping variable during the processing stage. All computations, such as quantile summaries, median household sizes, and sample sizes, are performed directly on the household-level data without any additional joins.

[1] http://psada.psa.gov.ph/catalog/FIES/about
[2] https://psada.psa.gov.ph/home

Dataset Creation

Input files used:

The following data files and mappings were used:

- **2018_FIES_vol2.csv** – main dataset with household expenditure, income, rent, and demographic info
- **region_code_mapping_Philippines.xlsx** – mapping for W_REGN (region code) to region names
- **province_code_mapping_Philippines.xlsx** – mapping for W_PROV (province code) to province names

Variables extracted and transformations performed:

- HH_exp : Monthly total household expenditure (TTOTEX)
- HH_size : Household size (FSIZE)
- no_of_earners Number of employed members (EMPLOYED_PAY)
- monthly_rent : Monthly rent expenditure (RENT)
- urban_rural : Set based on URB (1 = Urban, 2 = Rural)
- Percent_Income_Spent_on_Housing : Calculated as (RENT / TTOTEX) * 100
- norm_weight : Normalized survey weights (RFACT / 100)
- Region and Province names were merged using lookup tables and cleaned to ensure consistent naming for Location and City

  Note: We manually corrected the region names for two special regions:

  - Provinces like *Batangas, Cavite, Quezon* etc. were grouped under **Region IV-A - CALABARZON**
  - Provinces like *Palawan, Marinduque, Romblon* etc. were grouped under **Region IV-B – MIMAROPA**
  - Some columns in the dataset, such as the region code, province code, household size, and survey weight, were changed to number format to make them easier to work with. This helped ensure the data could be grouped and analyzed correctly.

```
# Clean and process
ph_2018_clean <- ph_2018 %>%
  mutate(
    W_REGN = as.numeric(W_REGN),
    W_PROV = as.numeric(W_PROV),
    HH_exp = as.numeric(TTOTEX),
    HH_size = as.numeric(FSIZE),
    no_of_earners = as.numeric(EMPLOYED_PAY),
    monthly_rent = as.numeric(RENT),
    urban_rural = ifelse(URB == 1, "Urban", "Rural"),
    Percent_Income_Spent_on_Housing = (monthly_rent / HH_exp) * 100,
    norm_weight = round(as.numeric(RFACT) / 100)
  ) %>%
  left_join(region_lookup_2018, by = "W_REGN") %>%
  left_join(province_lookup_2018, by = "W_PROV") %>%
  mutate(
    Region = case_when(
      ProvinceName %in% c("Batangas", "Cavite", "Laguna", "Quezon", "Rizal") ~ "Region IV-A - CALABARZON",
      ProvinceName %in% c("Marinduque", "Occidental Mindoro", "Oriental Mindoro", "Palawan", "Romblon") ~ "Region IV-B - MIMAROPA",
      TRUE ~ RegionName
    ),
    Location = Region,
    City = ProvinceName
  ) %>%
  drop_na(HH_exp, HH_size, monthly_rent)
```

Expanding Dataset by Weight

Each row was replicated by its norm_weight value to simulate the weighted population. To reflect actual population distributions, we **expanded** the dataset based on the survey weights (RFACT).

Each household record was duplicated based on its weight so that the final dataset better approximates real-world population figures.

This expansion is important to:

- Avoid bias from under-sampled or over-sampled groups
- Accurately calculate percentiles and median statistics

```
all_data_ph_2018 <- ph_2018_clean %>%
  uncount(weights = norm_weight)
```

Calculating Sample Sizes

We computed the number of valid survey households (i.e., sample size) for each **region-province-urban/rural** group. Two sets of sample sizes were calculated:

1. **Urban/Rural disaggregated**
2. **Combined (All)**

```
# Sample size
sample_size_ph_2018 <- ph_2018_clean %>%
  group_by(Location, City, urban_rural, Year = 2018) %>%
  summarise(sample_size = n(), .groups = "drop") %>%
  mutate(Country = "Philippines")


# Urban + Rural Combined
sample_size_ph_u_r_2018 <- ph_2018_clean %>%
  group_by(Location, City, urban_rural = "All", Year = 2018) %>%
  summarise(sample_size = n(), .groups = "drop") %>%
  mutate(Country = "Philippines")
```

Calculating Percentiles

Using a defined quantile probability range (1% to 99%), percentile values were computed for:

- HH_exp (household expenditure)
- HH_size (household size)
- no_of_earners
- Percent_Income_Spent_on_Housing

These were grouped by:

- Location (region)
- City (province)
- urban_rural (Urban / Rural / All)

Each quantile represents a snapshot of affordability for that income segment in a given location and classification.

These were calculated for both urban/rural disaggregated and combined cases.

```
ph_quantiles_2018 <- all_data_ph_2018 %>%
  group_by(Location, City, urban_rural) %>%
  group_modify(~ calculate_quantiles(.x, quantile_probs))


ph_quantiles_u_r_2018 <- all_data_ph_2018 %>%
  group_by(Location, City, urban_rural = "All") %>%
  group_modify(~ calculate_quantiles(.x, quantile_probs))
```

Final Dataset Preparation

- **Quantile and sample size datasets were merged**: The two datasets (ph_quantiles_2018, ph_quantiles_u_r_2018) were merged with sample_size_ph_2018 and sample_size_ph_u_r_2018 respectively using left_join() on Country, Location, City, urban_rural, and Year. This ensured that each quantile summary included the correct sample size.
- **Columns were aligned using a common schema**: Variable names and structure were standardized to ensure compatibility across country files.
- **Combine relevant sample size datasets**: All disaggregated and combined sample size datasets across countries (including sample_size_ph_2018 and sample_size_ph_u_r_2018) were appended using bind_rows() into a unified dataset called combined_sample_size.

- **Join combined sample size to summary dataset**: The combined_sample_size dataset was merged with summary_dataset using left_join() on Country, urban_rural, Location, City, and Year.

```
# Combine all

ph_final_combined_2018 <- bind_rows(

  ph_quantiles_2018 %>%

    left_join(sample_size_ph_2018, by = c("Country", "Location", "City", "urban_rural", "Year")),

  ph_quantiles_u_r_2018 %>%

    left_join(sample_size_ph_u_r_2018, by = c("Country", "Location", "City", "urban_rural",
"Year"))) %>%

  select(Country, Location, Year, Quantile, urban_rural, sample_size,

        HH_exp, no_of_earners, HH_size, Percent_Income_Spent_on_Housing, City)

# Align columns

ph_final_combined_2018 <- align_columns(ph_final_combined_2018, common_columns)

# Add to final dataset

summary_dataset <- bind_rows(summary_dataset, ph_final_combined_2018)
```

```
combined_sample_size <- bind_rows(

  sample_size_india,  sample_size_india_u_r,  sample_size_uganda,  sample_size_uganda_u_r,

  sample_size_nigeria,  sample_size_nigeria_u_r,   sample_size_pakistan,
sample_size_pakistan_u_r,  sample_size_kenya,   sample_size_kenya_u_r,

  sample_size_kenya_2020,  sample_size_kenya_2020_u_r,  sample_size_india_2022,

  sample_size_india_u_r_2022,  sample_size_ph_2023,  sample_size_ph_u_r_2023,
sample_size_ph_2018,  sample_size_ph_u_r_2018,  sample_size_rw,  sample_size_rw_u_r,
sample_size_mw, sample_size_mw_u_r, sample_size_ci, sample_size_ci_u_r,sample_size_bf,

 sample_size_bf_u_r, sample_size_kenya_2021,  sample_size_all_kenya_2021

)

summary_dataset <- summary_dataset %>%

  left_join(combined_sample_size, by = c("Country","urban_rural","Location","City","Year"))
```

In addition to the quantile-level summaries, we also generated **aggregated metrics at both state and national levels** to support broader comparisons within and across countries:

- **State-Level Aggregation:**
  The dataset was grouped by Country, Location (region), Year, Quantile, and urban_rural. Within each group, we calculated average household expenditure, household size, number of earners, and percentage of income spent on housing. The sample sizes were summed to reflect total representation. Kenya was excluded from this step as it is handled separately due to unique formatting.
- **National-Level Aggregation:**
  All regions within each country were collapsed into a national summary by grouping on Country, Year, Quantile, and urban_rural. This allowed for country-level benchmarking across time and income segments.
- **Combined Dataset Creation:**
  The original quantile-level data, state-level aggregates, and national-level aggregates were combined using bind_rows() into a single unified dataset.

```
state_aggregated <- summary_dataset %>%
  group_by(Country, Location, Year, Quantile,urban_rural) %>%
    summarise(
    sample_size = sum(sample_size),
    HH_exp = mean(HH_exp, na.rm = TRUE),
    no_of_earners = mean(no_of_earners, na.rm = TRUE),
    HH_size = mean(HH_size, na.rm = TRUE),
    Percent_Income_Spent_on_Housing = mean(Percent_Income_Spent_on_Housing, na.rm =
TRUE)
  ) %>%
 ungroup()
%>%
mutate(City = "All")
```

```
# ---------------------------------------
# National-Level Aggregation
# ---------------------------------------
# Aggregate data for the national level by combining all states within each country
national_aggregated <- summary_dataset %>%
  group_by(Country, Year, Quantile,urban_rural) %>%
  summarise(
    sample_size = sum(sample_size),
    HH_exp = mean(HH_exp, na.rm = TRUE),
    no_of_earners = mean(no_of_earners, na.rm = TRUE),
    HH_size = mean(HH_size, na.rm = TRUE)
    #Percent_Income_Spent_on_Housing = mean(Percent_Income_Spent_on_Housing, na.rm = TRUE)
  ) %>%
  ungroup() %>%
  mutate(Location = "All",
         City = "All")
# Combine State and National Data
# Combine both the state-level and national-level data
final_dataset <- bind_rows(state_aggregated, national_aggregated,summary_dataset)
```

Calculating Income Growth

All household income data for the Philippines was projected using **GDP per capita growth rates**.

For the historical period **2016–2024**, actual annual GDP per capita growth rates (sourced from the World Bank)[3] were applied directly to the dataset. For the forward-looking period **2025–2040**, a **median annual growth rate** of 4.667% was used.

Year	GDP per capita growth rate (%)
2000	1.92828
2001	0.80081
2002	1.63441
2003	3.024477
2004	4.518295
2005	3.009221
2006	3.559722
2007	4.687505
2008	2.516571
2009	-0.36296
2010	5.158694
2011	1.837605
2012	4.840765
2013	4.762666
2014	4.615141
2015	4.787251
2016	5.721215
2017	5.56223
2018	5.034291
2019	4.835775
2020	-10.5489
2021	4.761639
2022	6.765954
2023	4.667697
2024	4.823012
2025	4.667697

[3] https://data.worldbank.org/indicator/NY.GDP.PCAP.KD.ZG