ISI, Kolkata

Economic Statistics: Assignment 2

Insights into Household Consumption Patterns: A Comprehensive Analysis of Budget-Share Engel Curves

Advisor
Prof. Nachiketa Chattopadhyay

Name
Manas Sharma: BS2137

Contents

Introduction

This document investigates the distribution of age groups and the dependency ratio, using the 2001 Census data for the state of Gujarat, consisting of 26 districts at the time. The study focuses on Urban, Rural, and Total areas across diverse districts, aiming to uncover insights into the demographic composition and dependency patterns within the state.

1 Methodology

- To visualize age distributions across Urban, Rural, and Total populations, histograms are employed as visual aids, provided alongside this report. All tables are duly attached.
- The dependency ratio table is included within the document for easy reference.
- The tabulated distributions can be easily understood by referring to the attached codes.

2 Data Processing

- Data is sourced from the Census of India website, specifically concentrating on marital status by age and sex.
- The analysis zeroes in on age-related aspects. Given the unevenly sized age groups, particularly the 0-9 category, it is divided into two equal-sized classes for the creation of histograms.
- To accommodate the limited data frequency, the population bin for 80+ is assumed to be 80-85.

3 R-Code

The ensuing section presents the R-Code utilized for generating both the histograms and the dependency ratio table.

Listing 1: R Code

```
library(dplyr)
library(haven)
library(ggplot2)
library(rstatix)
library(caret)
library(ineq)

split_row_half <- function(data, row_index) {
  original_row <- data[row_index, ]
  new_row1 <- c("0-4", original_row$Persons / 2)
  new_row2 <- c("4-9", original_row$Persons / 2)
  data <- rbind(new_row1, new_row2, data[-row_index, ])
  data$Persons <- as.numeric(data$Persons)
  return(data)
}

census_dat_init <- read.csv("C:/Users/Manas/Downloads/Census_Data_2001.csv")
census_dat <- census_dat_init[,1:9]
dist_dat <- split(census_dat, census_dat$Distt.)
```

```
dependency_ratio_table <- data.frame()


for (name in names(dist_dat)) {

  current_dat <- dist_dat[[name]]
  urban_data <- current_dat[current_dat$Popn_Type == "Urban", ][2:17,]
  rural_data <- current_dat[current_dat$Popn_Type == "Rural", ][2:17,]
  total_data <- current_dat[current_dat$Popn_Type == "Total", ][2:17,]

  coltotal <- c("Age.group", "Persons")

  work_urban <- urban_data[,coltotal]
  work_rural <- rural_data[,coltotal]
  work_total <- total_data[,coltotal]

  work_rural1 <-  split_row_half(work_rural, 1)
  work_urban1 <-  split_row_half(work_urban, 1)
  work_total1 <-  split_row_half(work_total, 1)


  work_total1$Age.group <- factor(work_total1$Age.group,
                                  levels = work_total1$Age.group,
                                  ordered = TRUE)
  work_total1$Persons <- work_total1$Persons/sum(work_total1$Persons)

  work_urban1$Age.group <- factor(work_urban1$Age.group,
                                  levels = work_urban1$Age.group,
                                  ordered = TRUE)
  work_urban1$Persons <- work_urban1$Persons/sum(work_urban1$Persons)

  work_rural1$Age.group <- factor(work_rural1$Age.group,
                                  levels = work_rural1$Age.group,
                                  ordered = TRUE)
  work_rural1$Persons <- work_rural1$Persons/sum(work_rural1$Persons)

  width <- rep(0.05,17)
  main_rural <- paste("Rural-Age-Distribution:", total_data$Area.Name[1])
  main_tot <- paste("Total-Age-Distribution:", total_data$Area.Name[1])
  main_urban <- paste("Urban-Age-Distribution:", total_data$Area.Name[1])



  barplot(work_rural1$Persons, names.arg = work_rural1$Age.group,
          col = "skyblue",
          xlab = "Class Intervals",
          ylab = "Frequency", border = "black",
          density = 20, width = width, main = main_rural)

  barplot(work_urban1$Persons, names.arg = work_urban1$Age.group,
          col = "pink",
          xlab = "Class Intervals",
```

```
          ylab = "Frequency", border = "black",
          density = 20, width = width, main = main_urban)

  barplot(work_total1$Persons, names.arg = work_total1$Age.group,
          col = "green",
          xlab = "Class Intervals",
          ylab = "Frequency", border = "black",
          density = 20, width = width, main = main_tot)

  non_dependents1 <- work_total[3:12,]
  total_dependency_ratio <-
    (sum(work_total$Persons) -
       sum(non_dependents1$Persons))/sum(non_dependents1$Persons)

  non_dependents2 <- work_urban[3:12,]
  urban_dependency_ratio <-
    (sum(work_urban$Persons) -
       sum(non_dependents2$Persons))/sum(non_dependents2$Persons)

  non_dependents3 <- work_rural[3:12,]
  rural_dependency_ratio <-
    (sum(work_rural$Persons) -
       sum(non_dependents3$Persons))/sum(non_dependents3$Persons)

  new_entry <- data.frame(District = total_data$Area.Name[1],
                          Urban_Dependency_Ratio = urban_dependency_ratio,
                          Rural_Dependency_Ratio = rural_dependency_ratio,
                          Total_Dependency_Ratio = total_dependency_ratio,
                          Dependent_Popn = (sum(work_total$Persons) -
                                              sum(non_dependents1$Persons)),
                          Non_Dependents = sum(non_dependents1$Persons))

  dependency_ratio_table <- rbind.data.frame(dependency_ratio_table, new_entry)

}

Total_DR <-
  sum(dependency_ratio_table$Dependent_Popn)/sum(dependency_ratio_table$Non_Depen

plots.dir.path <- list.files(tempdir(), pattern="rs-graphics",
                             full.names = TRUE);
plots.png.paths <- list.files(plots.dir.path, pattern=".png",
                              full.names = TRUE)

file.copy(from=plots.png.paths, to="C:/Users/Manas/Downloads/Eco_Plots")

plots.png.detials <- file.info(plots.png.paths)
plots.png.detials <- plots.png.detials[order(plots.png.detials$mtime),]
sorted.png.names <- gsub(plots.dir.path, "C:/Users/Manas/Downloads/Eco_Plots",
                         row.names(plots.png.detials), fixed=TRUE)
numbered.png.names <- paste0("C:/Users/Manas/Downloads/Eco_Plots/",
                             1:length(sorted.png.names), ".png")

# Rename all the .png files as: 1.png, 2.png, 3.png, and so on.
```

```
file.rename(from=sorted.png.names, to=numbered.png.names)
```

4 Insights

Several noteworthy insights emerge:

- Gujarat boasts a considerable young population, accompanied by a substantial dependent segment in the younger age group. The older dependents don't have significant numbers
- The dependency ratio is lower in urban regions compared to rural areas, with some areas reaching below 0.45 in urban regions
- The overall dependency ratio stands at around 0.51, reflecting a positive scenario where there are approximately two working-age adults for every dependent. This is comparatively good even in the international scenario.

4.1 A look at Tapi - A district in Gujarat

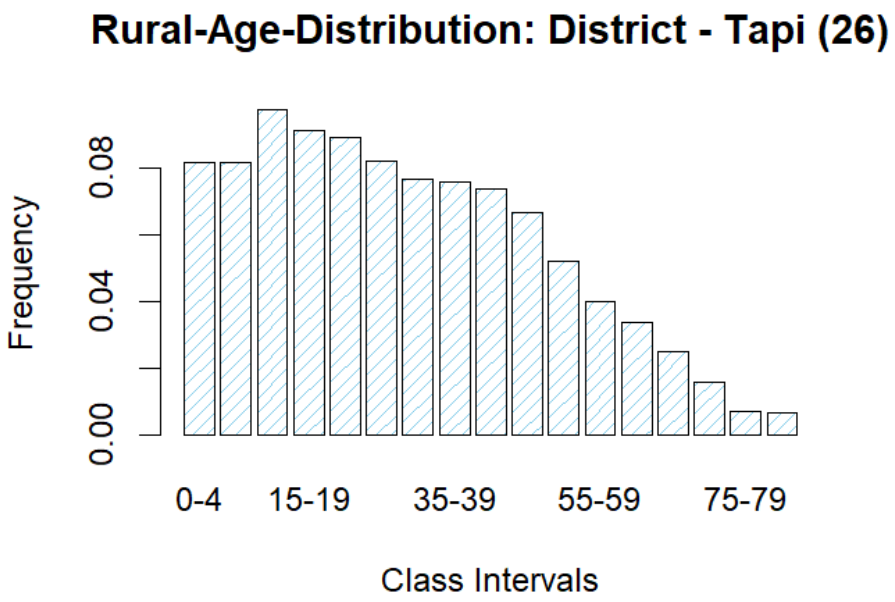

Figure 1: Tapi-Rural

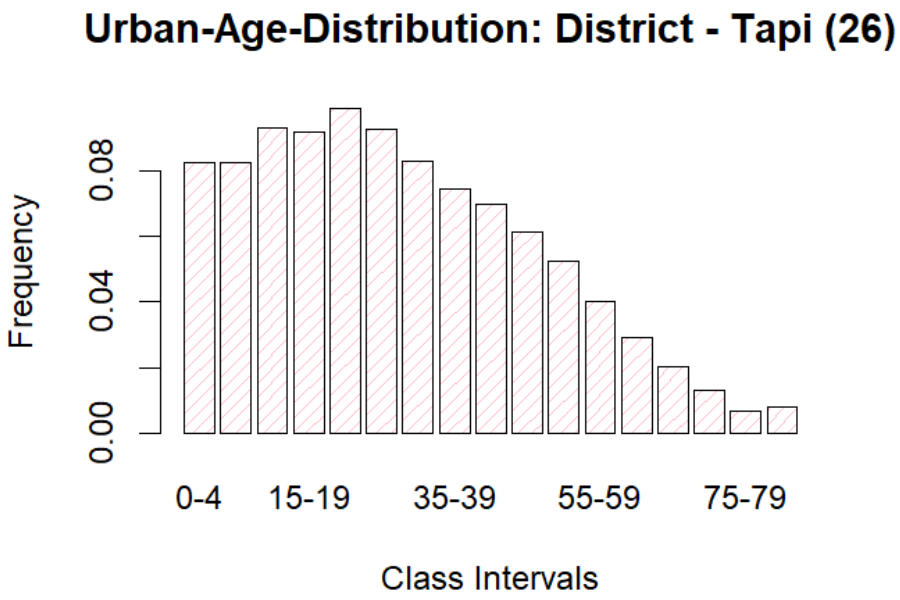

Figure 2: Tapi-Urban

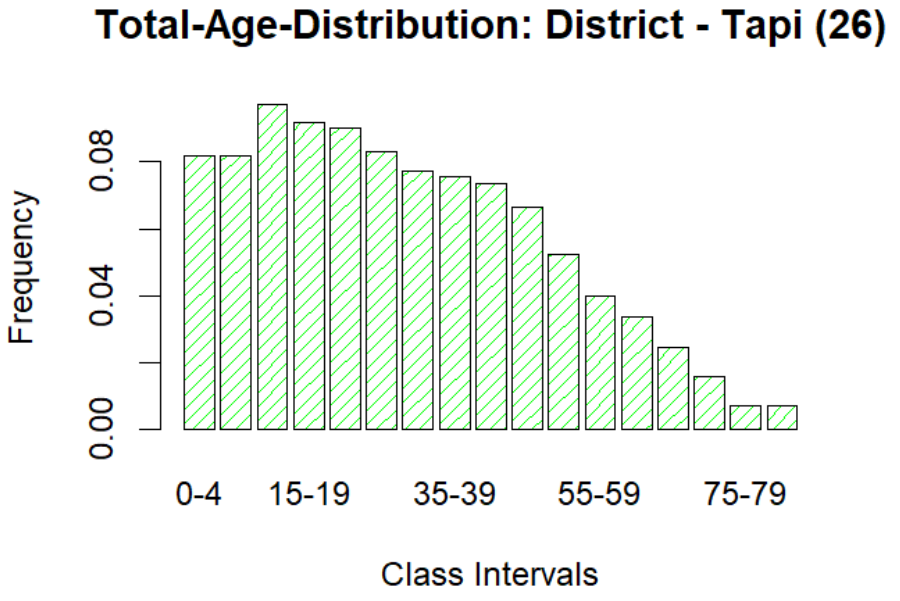

Figure 3: Tapi-Total

5 Dependency Table

District	Urban-DR	Rural-DR	Total-DR
District - Kachchh (01)	0.490419717	0.636219675	0.582422261
District - Banas Kantha (02)	0.522661672	0.687079559	0.663196324
District - Patan (03)	0.471487034	0.607274403	0.576808088
District - Mahesana (04)	0.432339639	0.516290126	0.494187158
District - Sabar Kantha (05)	0.486561111	0.58506499	0.569502451
District - Gandhinagar (06)	0.426221502	0.507818754	0.47145681
District - Ahmadabad (07)	0.433003608	0.581599481	0.454825505
District - Surendranagar (08)	0.478538764	0.627858353	0.582647258
District - Rajkot (09)	0.438743403	0.529523239	0.475376136
District - Jamnagar (10)	0.473260356	0.559791622	0.519684806
District - Porbandar (11)	0.459633601	0.541688969	0.500577836
District - Junagadh (12)	0.472523411	0.534531105	0.513486121
District - Amreli (13)	0.483433358	0.554043884	0.535376439
District - Bhavnagar (14)	0.483926642	0.63446168	0.569132907
District - Anand (15)	0.448731104	0.508237977	0.489688429
District - Kheda (16)	0.468906094	0.53537427	0.519746461
District - Panch Mahals (17)	0.500494036	0.630242142	0.610732975
District - Dohad (18)	0.56396333	0.815804101	0.789827958
District - Vadodara (19)	0.410610562	0.5547099	0.479800587
District - Narmada (20)	0.445742241	0.579708449	0.564485842
District - Bharuch (21)	0.434348246	0.503656517	0.479500239
District - The Dangs (22)	0.572051003	0.752882577	0.731444979
District - Navsari (23)	0.409598624	0.440266628	0.430695158
District - Valsad (24)	0.427285468	0.539432442	0.495603825
District - Surat (25)	0.413278377	0.464190854	0.423303954
District - Tapi (26)	0.440681988	0.464007471	0.461677598

Table 1: Dependency Ratios