PLSC 502: "Statistical Methods for Political Research"

Exercise Two
September 16, 2022

Introduction

The point of this exercise is to get you accustomed to thinking about, calculating, and discussing summary statistics. The data are census-based, and consist of the 3,141 counties (or county-like entities; e.g., parishes in Louisiana) in the United States. Most of the data come from 2004 or 2005, and most of the 38 variables are relatively self-explanatory (e.g., *Population*, *Land Area*, etc.). A few of the stranger / less familiar variables include:

- A variable indicating whether (=1) or not (=0) that county was designated a "warm-humid" county by the IECC. (The definition means more-or-less exactly what it sounds like...).
- A 21-category variable denoting the *land surface form typography* code of the county, as designated by the U.S. Geological survey.
- A *Natural Amenities Scale* "constructed by combining six measures of climate, typography, and water area that reflect environmental qualities most people prefer."
- A *Natural Amenities Ranking* variable that "breaks up" the *Natural Amenities Scale* into a seven-point ordinal variable.
- Three dichotomous variables indicating whether (=1) or not (=0) the county was designated a *Health Professional Shortage Area* (i.e., whether or not they have a shortage of doctors) as of 2007. This distinction is separated into the areas of *primary care*, *mental health care*, and *dental care*.

A full description of the variables is available in the Appendix to this exercise.

Exercise

1. Choose *one* demographic variable, *one* economic or political variable, *one* climatic variable, *one* of the three HPSA variables, and *either* `Typography` or `AmenityRank`. For each:
   (a) Discuss the central tendency and variation ("spread") of the variable, using what we learned in class.
   (b) If appropriate, discuss whether and how skewed the variable is, and whether its distribution is leptokurtic, platykurtic, or somewhere in between.
2. Reconsider each of the five variables you picked above, but this time, examine *only* data for counties in the state of Pennsylvania. Compare each of the variables distributions in PA with that for the U.S. as a whole, and discuss whether and how they are different.
3. Finally, reexamine the summary statistics for your first four variables above, distinguishing between counties that were and were not HPSAs (using the HPSA variable you selected for parts one and two above). Discuss if, and how, HPSA-designated counties differ in their characteristics on those four traits from those not so designated.

Use plots, statistics, or combinations thereof to complete the exercise. Submit your answers **in PDF format**. For each answer, provide both the figure(s) and a short textual description of your "answer." In addition to your answers, please include a copy of all computer code used to generate your figures. This can be in any form – a separate `.R` or `.do` file, an appendix in the PDF, or as a `.Rmd` or similar format containing both content and code. This homework exercise is due by 11:59 p.m. ET on Friday, September 23, 2022; submit your materials in electronic format – via e-mail attachment – to Tuba (`tzs5636@psu.edu`) *and* to me (`zorn@psu.edu`). This exercise is worth 50 possible points.

Variable Descriptions

Variable Name	Description
State	State FIPS Code
County	County FIPS Code
FIPS	State and county FIPS Code
StateName	State name
CountyName	County name
LandArea	Land area (square miles)
Pop05	Total resident population, 7/1/05
PopGr0005	Percent population change, 7/1/00 to 7/1/05
HU05	Housing units, 7/1/05
Pop85plus05	Resident population aged 85 years and over, 7/1/05
SexRatio05	Sex Ratio, 7/1/05
MedianAge05	Median age of total resident population, 7/1/05
White05	White alone resident population, 7/1/05
Black05	Black alone resident population, 7/1/05
AIAN05	American Indian and Alaska native alone resident population, 7/1/05
Asian05	Asian alone resident population, 7/1/05
NHOPI05	Native Hawaiian and other Pacific islander alone resident population, 7/1/05
LF05	Labor force, annual average estimate, 2005
UnempRate05	Unemployment rate, annual average estimate, 2005
PersIncome05	Personal income ($1,000s), 2005
PCPersIncome05	Per capita personal income (dollars), 2005
JanTemp	Mean temperature for January, 1941-1970
JulTemp	Mean temperature for July, 1941-1970
JulHumid	Mean relative humidity for July, 1941-1970
Typography	Land surface form typography
AmenityScale	ERS Natural Amenity Scale
AmenityRank	ERS Natural Amenity Rank
WarmHumid	2004 IECC (supplement to 2003 IECC) warm-humid counties
PropTax02	Local government property tax revenue ($1,000s) , 2001-2002 fiscal year
PropTaxPerCap02	Local government property tax revenue per capita, 2001-2002 fiscal year
CrimeRate04	Index crime rate (per 100,000 persons), 2004
TotalVotes04	2004 presidential election: Total votes
BushVotes04	2004 presidential election: Votes for Bush
KerryVotes04	2004 presidential election: Votes for Kerry
OtherVotes04	2004 presidential election: Votes for other candidates
PHPSACty07	Whole county designated as a Single County Primary Care HPSA, 6/21/07
MHPSACty07	Whole county designated as a Single County Mental Health Care HPSA, 6/21/07
DHPSACty07	Whole county designated as a Single County Dental Care HPSA, 6/21/07