CUSP-GX 8153 Complex Urban Systems: HW3 Questions: Lectures 9–12

Qingyang Zhao, qz2567

May 3, 2024

1 Problem 1

Report the equation of the scaling law and the coefficient of determination, R2. The Scaling Figure 1 below:

The equation(After converting to the normal format from the log format):

$$Obesity_Prevalence = 47.919220809685044 \times Population^{-0.02492204271590146}$$

the coefficient: 47.919220809685044

the R2: 0.08400818414751543

The analysis result doesn't match my expectations.

In the case of obesity prevalence, a $\beta < 0$ could indicate that as the urban population grows, the obesity prevalence decreases and shows negative growth. In my mind, as the population of an area increases, the rate of increase in obesity prevalence slows down but still increase in general. Because big cities have access to more high-calorie fast food and a higher proportion of the population eats fast food. At the same time, work and life in big cities are more likely to lack exercise, so the obesity rate will increase with the size of the city. But people in big cities also have better sports and health and related facilities, and most people pay more attention to body management. So my guess is that the result is sub-linear. This could be due to the greater influence of factors such as better access to healthcare, increased awareness about healthy lifestyle choices, or the presence of more recreational facilities in larger urban areas.

The SAMIs distribution is the Figure 2.

County with Smallest SAMI: Boulder.

Boulder County is a county located in the U.S. state of Colorado of the United States. As of the 2020 census, the population was 330,758. The most populous municipality in the county and the

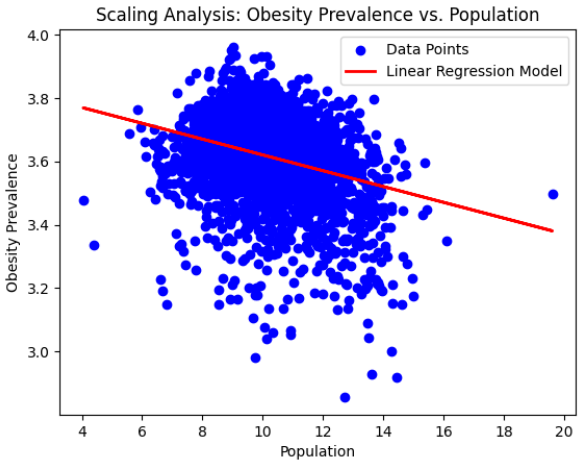

Figure 1: Scaling Analysis

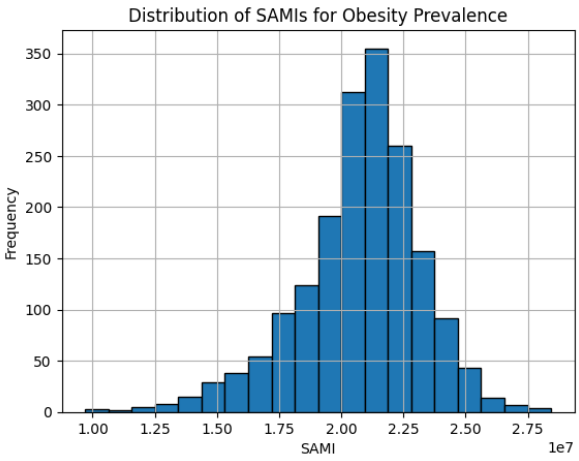

Figure 2: SAMIs distribution

county seat is Boulder. Boulder County comprises the Boulder, Colorado Metropolitan Statistical Area, which is included in the Denver–Aurora, Colorado Combined Statistical Area.

County with Largest SAMI: Coahoma

Coahoma County is a county located in the U.S. state of Mississippi. As of the 2020 census, the population was 21,390. Its county seat is Clarksdale. The Clarksdale, MS Micropolitan Statistical Area includes all of Coahoma County. It is located in the Mississippi Delta region of Mississippi. In 2023, the Clarksdale, Mississippi Micropolitan area was added to form the new Memphis-Clarksdale-Forrest City Combined Statistical Area. The Memphis-Clarksdale-Forrest City Combined Statistical Area has a population of roughly 1.4 million.

2 Problem2

Report the equation of the scaling law and the coefficient of determination, R2. The Scaling of diabetes Figure 3 below:

The equation(After converting to the normal format from the log format):

$$Diabetes_Prevalence = 11.75144622950784 \times Population^{-0.012450545727809348}$$

the coefficient: 411.75144622950784

the R2: 0.008519472886346802

The analysis result matches my expectations.

In the case of diabetes prevalence, a $\beta < 0$,like the β in the obesity scaling, could indicate that as the urban population grows, the diabetes prevalence decreases and shows negative growth. In my mind, obesity and diabetes are closely related. And as the population of an area increases, the rate of increase in diabetes prevalence decrease because bigger cities have access to better healthcare and related facilities, and increased awareness about healthy lifestyle choices.

The SAMIs distribution of diabetes is the Figure 4.

Please report the R2 and comment on your results, possibly using references from the public health field.

The correlation coefficient (R) of 0.691 and the associated p-value of 2.95e-257 indicate a strong positive correlation between the two variables being analyzed, Obesity and Diabetes.

Obesity and diabetes are closely related, and there is a strong correlation between the two. Here are some key relationships between obesity and diabetes:

1.Increased Risk of Type 2 Diabetes: Obesity is one of the leading risk factors for type 2 diabetes. Excess body fat, especially abdominal fat, makes cells more resistant to the effects of insulin, leading to high blood sugar levels characteristic of type 2 diabetes.

2.Insulin Resistance: Obesity-induced insulin resistance is a primary mechanism linking obesity to type 2 diabetes. In obese individuals, adipose tissue releases inflammatory molecules that interfere with insulin signaling, impairing glucose uptake by cells and contributing to insulin resistance.

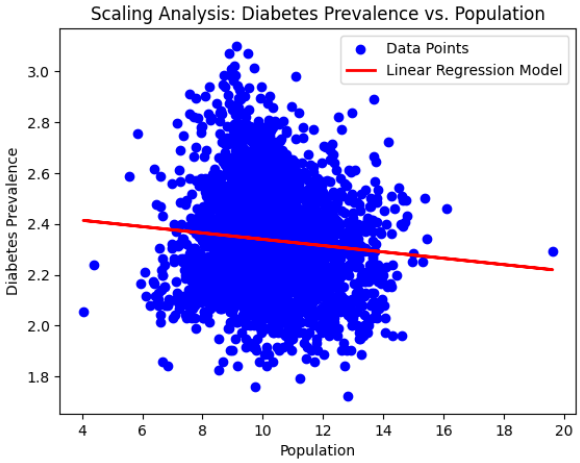

Figure 3: Scaling Analysis: Diabetes Prevalence vs. Population

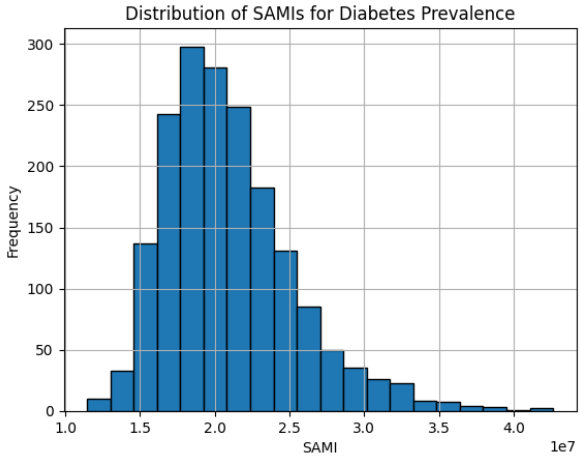

Figure 4: Distribution of SAMIs for Diabetes Prevalence

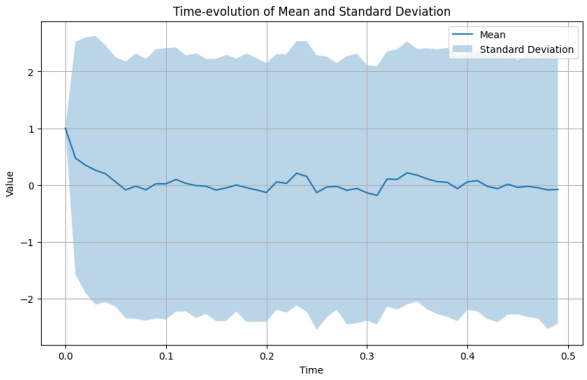

Figure 5: Time Series of Mean and Standard Deviation

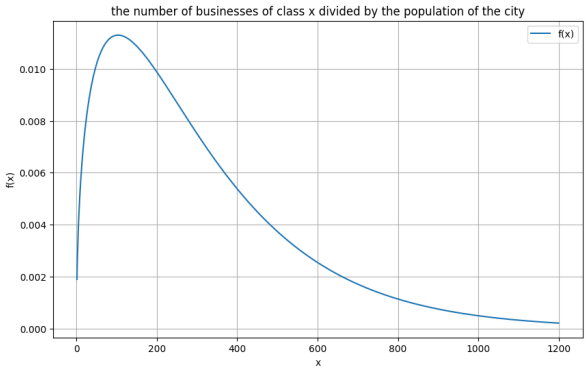

Figure 6: Enter Caption

3 Problem3

The plot is the Figure 5. And the time-evolution of the mean with 500 independent realizations show me that the x tends to 0 as t becomes larger, and the time-evolution of the standard deviation is relatively stable, it's about 2.3

4 Problem 4

Compute the Shannon entropy and Herfindahl-Hirschman index and compare to their maximum and minimum values, respectively.

Shannon Entropy: 2.1137427585458974
Maximum Shannon Entropy: 2.321928094887362
Minimum Shannon Entropy: 0
Herfindahl-Hirschman Index: 0.25124444444444444
Maximum Herfindahl-Hirschman Index: 1
Minimum Herfindahl-Hirschman Index: 0.2

5 Problem 5

Please plot the function for x = 1, 2, . . . 1200.

See Figure 6.
For any positive integer value x, f(x) is not a probability mass function
Shannon Entropy: 9.601245924783095

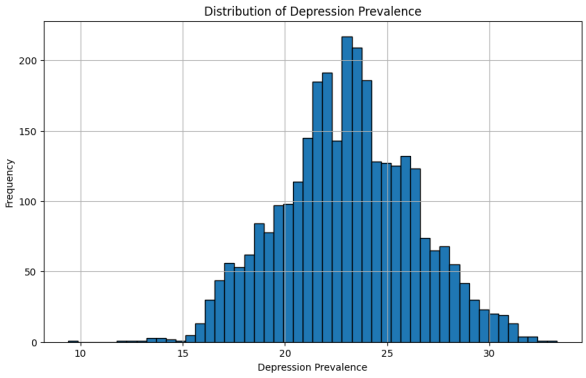

Figure 7: Distribution of Depression Prevalence

6 Problem 6

Plot the distribution using 50 bins and compute the Theil index of inequality

The distribution is Figure 7.

Theil Index of Inequality: 0.14586786275050512

7 Problem 7

Compute Gini's index for average unemployment rate in 2021 and median household income. Please comment on your results.

Gini index for average unemployment rate in 2021: 0.11951356727378615 Gini index for median house hold income: 0.11103833713212885

The Gini Index for average unemployment rate in 2021 is about 0.12 and The Gini Index for median house hold income is about 0.11, they both small. It illustrates that unemployment rates and median household incomes are relatively even across counties.

8 Problem 8

Moran's index for average unemployment rate in 2021: 2.554987468075973e-05 Moran's index for median household income: 6.837135612357777e-05

Both of them are super small and tend to 0 with positive, they suggest that the two viarables' spatial clusterings and autocorrelations are relatively weak. And so that we can think the overall spatial structures of them are relatively random or evenly distributed.

9 Problem 9

s_{ij} values:

$$\begin{bmatrix} 0 & 1076.92307692 & 10000 & 1030.92783505 & 1200 \\ 1076.92307692 & 0 & 7000 & 8750 & 10500 \\ 10000 & 7000 & 0 & 8620.68965517 & 15000 \\ 1030.92783505 & 8750 & 8620.68965517 & 0 & 8333.33333333 \\ 1200 & 10500 & 15000 & 8333.33333333 & 0 \end{bmatrix}$$

10 Problem 10

1) Compute the elements of the probability transition matrix p_{ij} defining the likelihood of an emission/absotption event between city pairs.

Probability Transition Matrix (p_{ij}):

$$\begin{bmatrix} 0 & 0.03940887 & 0.24617997 & 0.03856105 & 0.03425309 \\ 0.08092389 & 0 & 0.17232598 & 0.32728693 & 0.29971456 \\ 0.75143613 & 0.25615764 & 0 & 0.32245018 & 0.42816365 \\ 0.07746764 & 0.32019704 & 0.21222411 & 0 & 0.2378687 \\ 0.09017234 & 0.38423645 & 0.36926995 & 0.31170184 & 0 \end{bmatrix}$$

2) Compute the probabilities that people will not need to commute outside of their city for work for all the five cities.

Probabilities of not commuting outside of each city:

$$[0.82079851, 0.55987432, 0.1208962, 0.57612125, 0.42230971]$$

3) Assuming that the flux T_{ij} is equal to p_{ij} scaled by the population of origin times a factor of 1.1, how much is the fraction of the population of city 4 that commutes outside for work?

Fraction of the population of city 4 that commutes outside for work: 0.4502930943982753