

MAT1372, Classwork2, Fall2025

ID: _____

Name: _____

1.2 Data Basics (Conti.)

6. By using Scatterplots, one can study the relationship between two numerical variables.

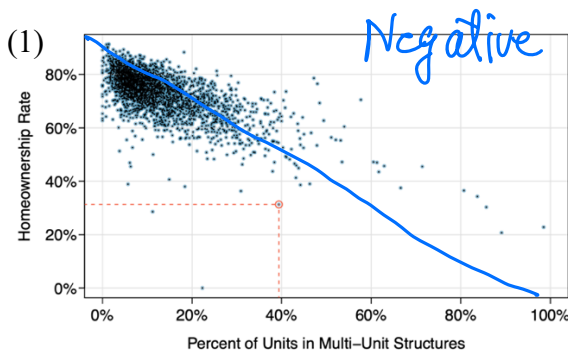


Figure 1.8: A scatterplot of homeownership versus the percent of units that are in multi-unit structures for US counties.

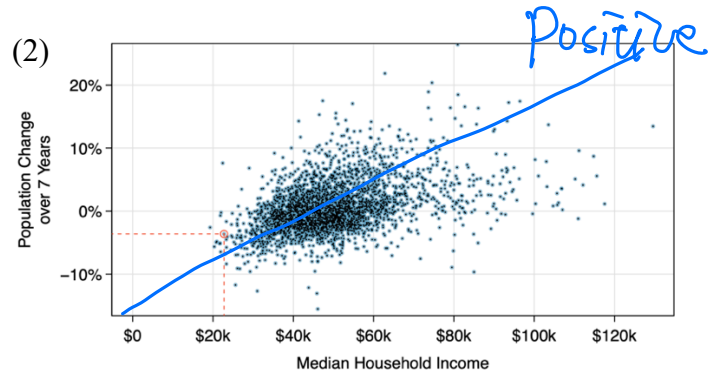


Figure 1.9: A scatterplot showing pop. change against median hh. income.

The scatterplot (1) and (2) suggest:

- (1) Counties with a higher rate of multi-unit tend to have lower homeownership rates.
- (2) Counties with a higher median household income tend to have higher than average increase in county population.

7. Associated (or dependent) variable: two variables show some connection with one another.

Positive association: a variable with **higher** value tends to have the other variable with **higher** value.

Negative association: a variable with **higher** value tends to have the other variable with **lower** value.

Independent variables: two variables are not associated.

8. In 6 (1), homeownership rates and multi-unit rates are a negative association or negatively associated.

In 6 (2), median household income and increase in county population are a positive association

9. Explanatory variable and Response variable:

When we suspect one variable might causally affect another, we label the first variable the explanatory variable and the second the response variable. For example, if we can ask:

“If there is an increase in the median household income in a county, does this drives an increase in its pop.?”

In this question, the explanatory variable is median household income, and the response variable is the population change.

10. Association and Causation:

Association \neq Causation

1.3 Sampling principles and strategies

1. Populations and Samples.

In a research study, the population is the entire group of individuals, **cases**, or units that share a specific characteristic and about which the study aims to draw conclusions. Because it is often **impractical** or **impossible** to study the entire population, researchers typically select a representative subset called a sample from the population to collect data from and then generalize the findings back to the entire population.

2. Find the target population and a sample in each research question:

(1) *What is the average mercury content in swordfish in the Atlantic Ocean?*

Population: All swordfish in the Atlantic Ocean; Sample: 125 swordfishes in Atlantic Ocean

(2) *Over the last 5 years, what is the average time to complete a degree for City Tech undergrads?*

Population: All the City Tech undergrads who graduated in the last five years.; Sample: _____.

(3) *Does a new drug reduce the number of deaths in patients with severe heart disease?*

Population: All people with severe heart disease.; Sample: _____.

3. Sampling from a population:

Biased sample: a sample is collected in such a way that some cases of the population have a lower or higher sampling rate than others.

Random sample: a sample is called **simple random sample** which is equivalent to using a raffle to select cases.

4. Types of Biased Sample:

Non-response bias: This occurs when a specific group of people refuse to participate in a study or drop out.

Convenience Sampling bias: This involves selecting individuals who are easiest to reach rather than using a random selection process.

5. Examples of Biased sample:

(1) *A health survey only got a 30% response rate.*

Non-response bias causes underestimation the prevalence of a disease

(2) *Surveying shoppers at a single grocery store on a Tuesday morning.*

Convenience Sampling bias that missed people from other store at different times.