Chapter 3 review

Fred Azizi

2023-09-12

Reminders

Reminder: Chp 2 Homework is due this weekend

Please scan this QR code if you want this slides and worksheet on your phone:

You can also use the link in the Blackboard

Or go to fredazizi.github.io/Teaching

Quick review

- We want to describe interval/ratio data, what is the appropriate way to visualize?
- We use Grouped Frequency Distribution when size of data (n) is very large and/or the range of the values is large
- The steps necessary to define the classes:
  - How to choose the number of non-overlapping classes?
  - How to determine the width of each class?
  - Where to start the first class?

Quick review (2)

- How to choose number of classes? **Sturges's formula**. (Round up to an integer if necessary)

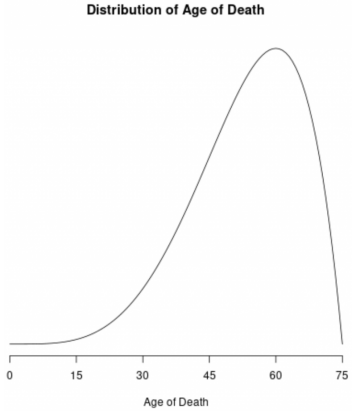
$$\text{Number of classes} = 1 + 3.3 \log_{10} n$$

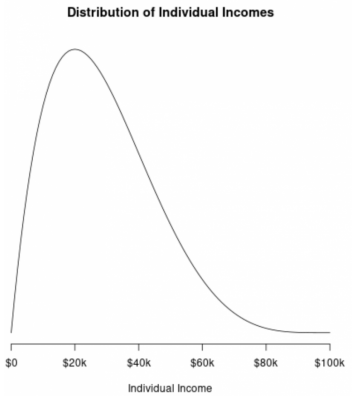
- Class width $= \dfrac{\text{Largest Observation} - \text{Smallest Observation}}{\text{Number of Classes}}$.
- Select a lower limit for the first class. If the measurements have k places of decimals, you should deduct a number that has $k + 1$ decimals, from the **minimum measurement**.

Quick review (3)

Graphical representations of distribution $\Rightarrow$ Histogram

- Show the frequency distribution for quantitative data over a set of class intervals (Similar to bar chart but works over class intervals).
- Constructed by rectangles whose bases are the intervals and whose heights are the frequencies (or relative frequencies or percent frequencies). **No gaps** between bars.
- We can potentially identify symmetry, distribution skewness, etc.





Quick check (4)

Graphical representations for Cumulative Frequency:

- Ogive
  - We plot the class end points on the horizontal axis and the cumulative frequencies on the vertical axis. Start from 0 go up to the amount of cumulative frequency toward the end of class.
  - End point is always 1, 100 or total frequency depending on cumulative frequency we are using.

Quick review (5)

- Time series data are values that correspond to specific measurements taken over a range of time periods.
- Cross-section data are values collected from a number of subjects during a single time period

Figure 1.5 - A Time Series Graph of U.S. Unemployment Rates, 2008–2012

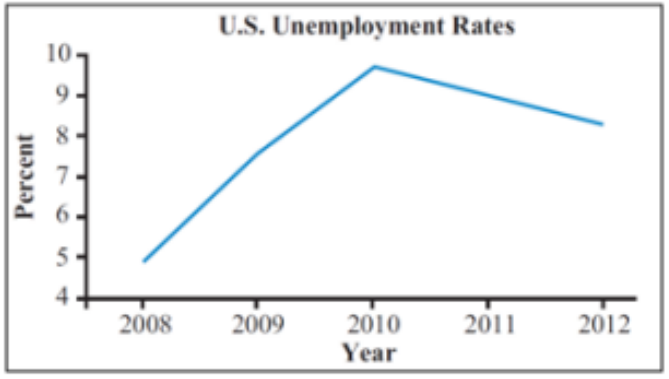


Figure 1.6 - A Cross-Sectional Graph of 2012 Unemployment Rates

Quick review (6)

- Scatter plot: A graph that displays pairs of values as points on a two-dimensional grid.
- The independent/explanatory variable is placed on the horizontal axis, or x-axis. The dependent/response variable is placed on the vertical axis, or y-axis.