Day 9 of #100daysofmathandstats: Exploration of binary and categorical data

By Harsh Kathiriya

Outline

- Mode
- Expected value
- Bar charts
- Pie charts

Mode

- The mode is the value — or values in case of a tie — that **appears most often** in the data.
- It is especially useful when you want to see which value for **particular features is repeating**.
- Mode is sometimes useful in **imputing data for NULLs** in categorical columns of the dataset.

Some real examples of box and whisker plot

- The mode of the cause of delay at **Dallas/Fort Worth airport** is “Inbound.”
- In most parts of the United States, the **mode for religious preference** would be Christian.
- The mode is a simple **summary statistic for categorical** data, and it is generally not used for numeric data.

Expected value

- When the categories can be associated with a numeric value, this gives an average value based on a category's probability of occurrence.
- A special type of categorical data is data in which the categories represent or can be mapped to discrete values on the same scale.
- One of the most important term for categorical data.

Real example of expected value

- A marketer for a new cloud technology, for example, offers two levels of service, one priced at $300/month and another at $50/month.
- The marketer offers free webinars to generate leads, and the firm figures that 5% of the attendees will sign up for the $300 service, 15% for the $50 service, and 80% will not sign up for anything.
- This data can be summed up, for financial purposes, in a single “expected value,” which is a form of weighted mean in which the weights are probabilities.

Real example of expected value (Contd…)

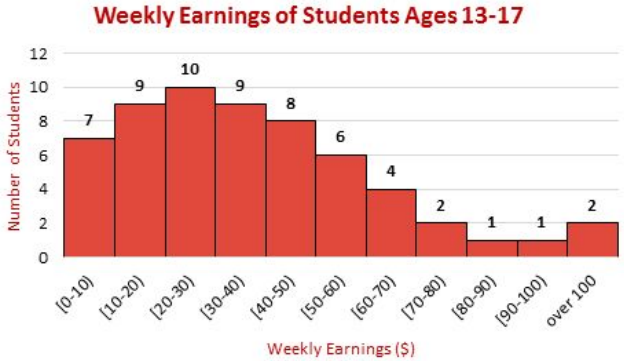
- The expected value is calculated as follows:
- 1. Multiply each outcome by its probability of occurring.
- 2. Sum these values. In the cloud service example,
- The expected value of a webinar attendee is thus $22.50 per month, calculated as follows:
- **EV = (0.05) * (300) + (0.8) * (0) + (0.15) * (50) = 22.5**
- Expected value is a fundamental concept in business valuation and capital budgeting — for example, the expected value of five years of profits from a new acquisition, or the expected cost savings from new patient management software at a clinic.

Bar charts

- The **frequency or proportion for each category** plotted as bars.
- A bar chart is a type of chart which shows the **values of different categories** of data as rectangular bars with different lengths.
- If you have **comparative data** that you would like to represent through a chart then a bar chart would be the best option.

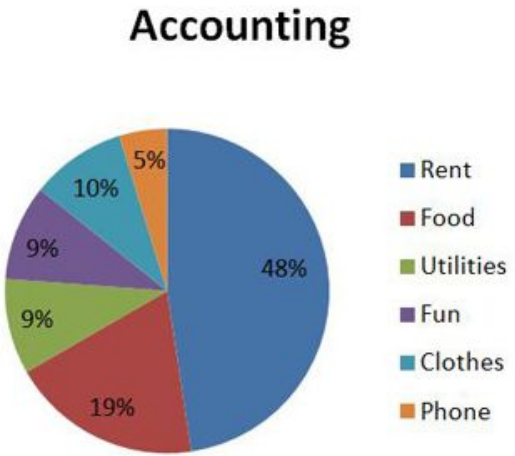
Real example of bar chart

- For example, you can make a bar chart about what the average time and activities on our smartphones.
- On the left side you list all of the activities people do on their phones and than you have a graph representing how much time they spend on those activities. Bar chart are important to track the progress of something.



Pie charts

- The **frequency or proportion** for each category plotted as **wedges in a pie.**
- It can be used as an alternative for bar chart to make it visually better and simple.



Thank you

Github Link: https://github.com/harsh9898/100daysofstatandmath

Don’t forget to post your queries or feedbacks on the post.

Share or like for the benefit of others.