

INTRODUCTION TO PROBABILITY MODELS

Lecture 36

Qi Wang, Department of Statistics

Nov 15, 2018

GRAPHS: QUALITATIVE DATA

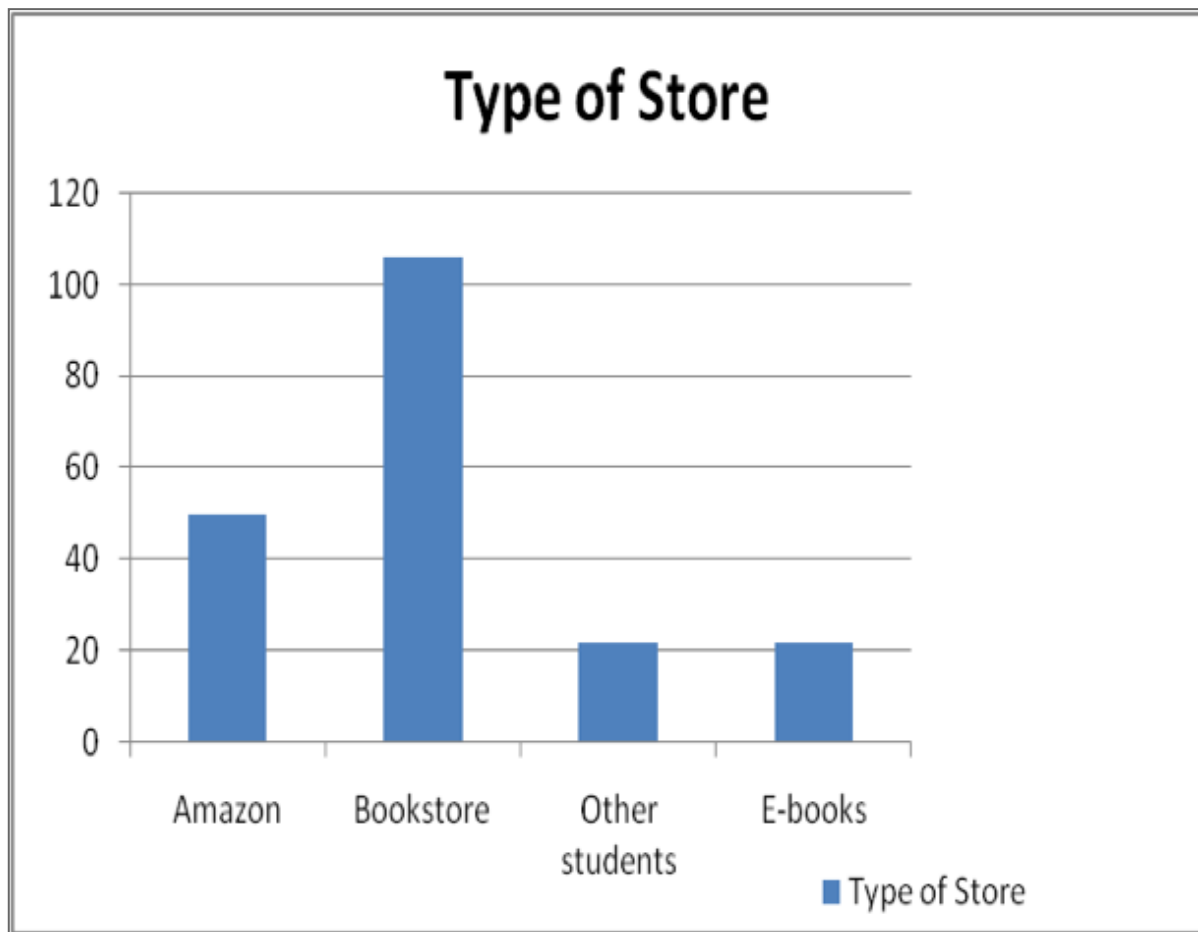
EXAMPLE 1

- **Frequency distribution** is the frequency or count of the occurrences within a particular category
- **Relative frequency distribution** is the proportion/percentage of the occurrences within a particular category

Place of Purchase	# of Students	% of Students
University/Follett's Bookstore	106	53%
Amazon.com or Half.com	50	25%
Previous students	22	11%
Purchase e-book	22	11%

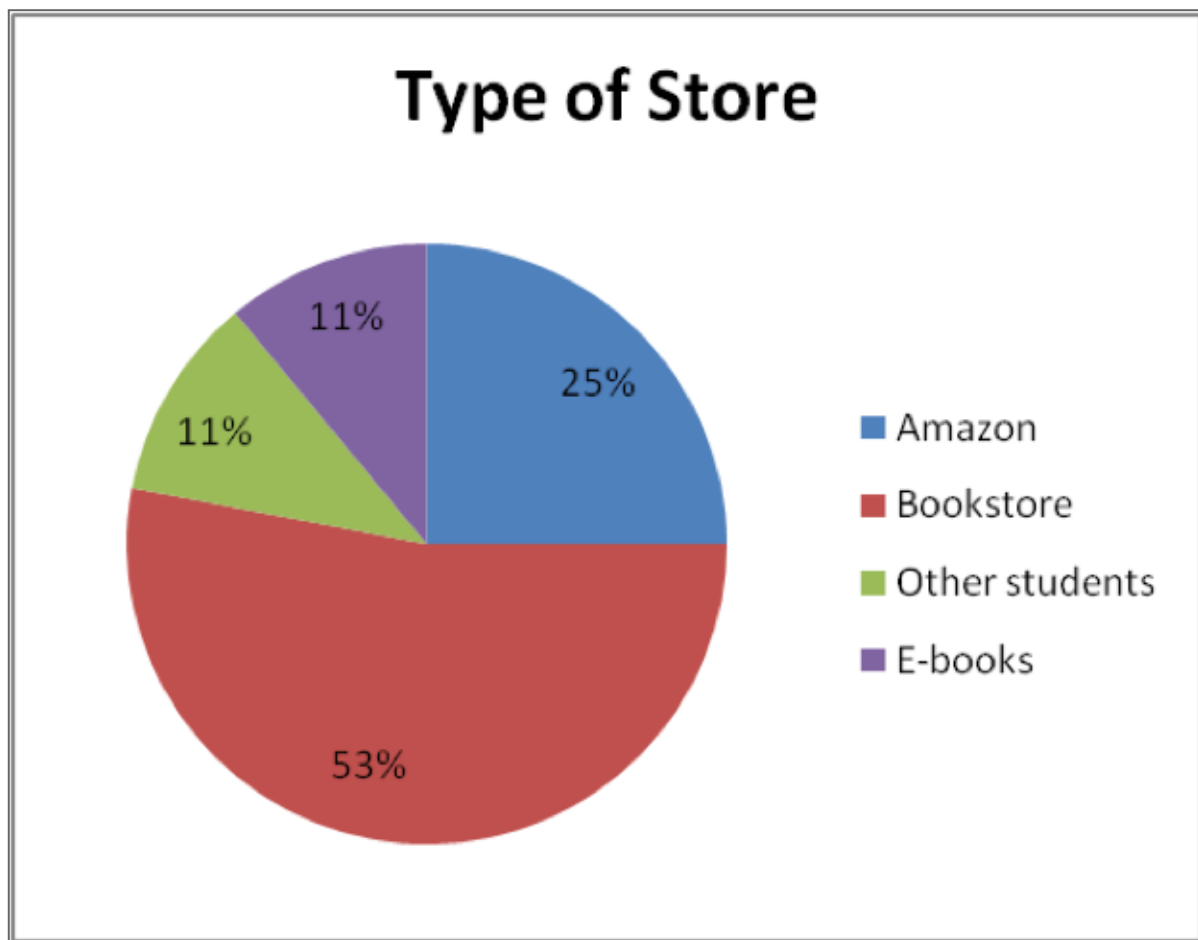
BAR GRAPH

Typically with bar graphs, the y-axis represents the frequency (# of observations) in the categories. Elements can give multiple answers or no answer. Pay attention to the axes.



PIE CHART

Pie Chart is circular graph in which the relative size of the “slice” represents the percents or proportion of responses in that category. Percents must add to 100%. Each element/subject is allowed to give only one answer.



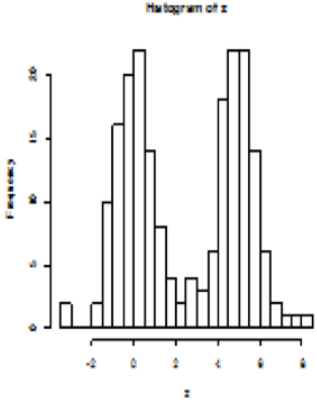
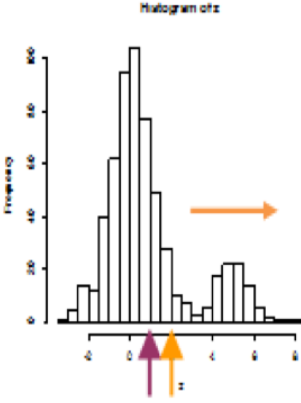
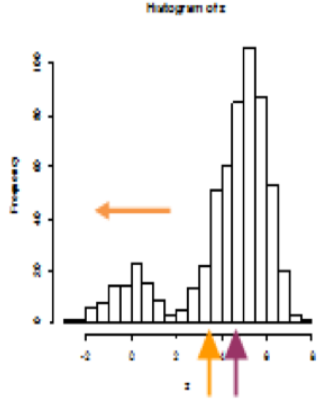
GRAPHS: QUANTITATIVE DATA

HISTOGRAM

Histogram is used for large amount of data, displays a count (frequency distribution) or percent or relative frequency distribution (probabilities)

QUANTITATIVE DATA ONLY!!!! It is constructed by placing the class intervals on the horizontal axis and the frequencies, relative frequencies, or percent frequencies on the vertical axis. When making a histogram, you need to pick an adequate number of classes (or, equivalently, an appropriate width of the interval for each class).

SYMMETRY AND SKEWNESS

Symmetric	Skewed Right	Skewed Left
<p>Histogram of z</p> 	<p>Histogram of z</p> 	<p>Histogram of z</p> 
Mean \approx Median	Mean $>$ Median	Mean $<$ Median

WHEN TO USE MEAN AND WHEN TO USE MEDIAN

Suppose there are 7 people eating in a diner. We ask “what is their income/ salary? Then Bill Gates walks into the diner and we add in his salary.

which measure is more representative of the group after Bill Gates arrives?

- When data is symmetric --want to use mean and standard deviation
- When data is skewed or have outlier(s) --want to use 5 number summary (or median and IQR)

Since the median is relatively unchanged by a few very large or very small measurements in the data set, we say that the median is resistant. The mean is non-resistant.

DOT PLOT

Dot Plot is a graphical device that summarizes data by the number of dots above each data value on the horizontal axis.

LINE GRAPH

Line Graph is a graphical device that summarizes time series data. Time/date is usually on the horizontal axis and the value of the measured variable on the vertical axis. Look for upward or downward trends. Look for seasonal patterns

SCATTERPLOT

COMPARING SOME OF THE GRAPHS

HISTOGRAM VS. BAR GRAPH

HISTOGRAM VS. BAR GRAPH

HISTOGRAM VS. STEMPLOT

HISTOGRAM VS. STEM PLOT

EXAMPLE 2

What type of graph would be appropriate in order to answer the following questions?

- What percentage of a household budget was spent on housing, utilities, food, insurance, entertainment, transportation and other?
- What values represent the middle 50% of the data for miles per gallon (mpg) of a sample of 2008 Toyota Camrys?
- How did the daily closing price of one share of Intel Corporation (INTC) stock fluctuate in the First quarter of 2012?
- Are gas prices and daily high temperature related
- A poll of 300 Purdue University students was taken. How many students in the poll purchased food from each of the following restaurants in the past month: Jimmy Johns, Chipotle, Noodles, Buffalo Wild Wings and Panera?