Author: Adapted from previous MASH handout by Jenny Freeman

Displaying data using charts

Date amended: 9/8/2018

WHAT GRAPHS DO I USE TO INTERPRET MY DATA?

Type	When to use?	Use for data features
Box plot	Summarises data with median, upper quartile, lower quartile, maximum and minimum: • Highlights outliers • Quick view of the centre, spread and interquartile range	For large datasets where stem & leaf or dot plots are not suitable For one or more continuous variables
Histogram	Shows distribution of the number of observations in equal intervals: • Shows the shape of distribution • Helps assess symmetry/skewness, modality(number of peaks)	For a continuous variable
Stem & leaf plot	Retains the original data to at least two significant digits, and puts the data in order. It separates the two digits by a vertical line; the left side contains the *stems* and the right side contains the *leaves* • Highlights outliers • Shows frequencies within each stem	For a single discrete variable, or a continuous variable of integers For small/moderate datasets
Dot plot	Describes the data on a single axis: • Highlights clusters, gaps and outliers • Identifies the mode • Shows the shape of distribution	For a single discrete variable, or a continuous variable For small datasets
Scatter plot	Plots one continuous variable against another: • Shows the correlation between them, i.e. strength of the relationship, positive/negative etc • Shape of the relationship, linear/curved? • Shows the outliers	For two continuous variables
Line chart	Uses lines to join the observed points. • It is very effective when exploring a time-dependent continuous variable	For one or more continuous variables, e.g. compares the changes in groups over the time
Pie chart	A circular chart (pie-shaped); it is split into segments to show percentages or the relative contributions of the categories of data: • Compares the proportions with the whole pie	For categorical data: percentages/ proportions with the same basis
Bar chart	A chart with rectangular bars with heights proportional to the frequencies that they represent: • Shape of the distribution • Can be used to compare the frequencies for two or more groups	For one and more categorical variables
Other plots:		
QQ plot	A graphical method for comparing two probability distributions by plotting their quantiles against each other: • Widely used to check the normality assumption	For one/two continuous variables
Kaplan Meier	A special graphical method used in medical research to measure the fraction of patients living for a certain amount of time after treatment	For survival time variables

Author: Adapted from previous
MASH handout by Jenny Freeman

Displaying data using charts

Date amended: 9/8/2018