

1.1.5 Descriptive measures for numerical variables

Probability density function: likelihood that a continuous random variable is equal to some value

Central tendency: mean (average), median (middle observation in sorted data), mode (most common value)

minimum and maximum

percentiles: value below which given percentage of data falls

Dispersion: range, variance (average squared deviation from mean), mean absolute deviation (MAD, average absolute deviation)

skew: negatively or positively

kurtosis: flatness of tails

1.1.6 Descriptive measures for categorical variables

Probability mass function: probability that a discrete random variable is exactly equal to some value

Frequency distribution

Central tendency: mode (most common value), median (middle observation in sorted ordinal values)

Graphical summaries: bar chart

1.1.7 Outliers and missing values

Unusual values that lie outside the norm.

Should be either filtered out or treated separately.

Generally a rare case and contributes to noise in the data.