Applied Statistics

Domain	Exploring Data	
Cluster	**Select appropriate graphical and numerical methods to explore data.**	
Standard(s)	M.ASHS.7	Use appropriate measures of center and spread to describe a distribution. Instructional Note: Emphasize that only some data are well described by a normal distribution.

Content Examples

» Describing the shape, center and spread of a distribution:
   https://www.youtube.com/watch?v=KTkizHDaFWw

» Measures of center and spread to describe skewness:
   https://www.youtube.com/watch?v=DNabd1us_Y4

» Using Chebyshev's Theorem:
   https://www.youtube.com/watch?v=uMgK000XFhA

» Using the Empirical Rule:
   https://www.youtube.com/watch?v=T7-eeg6rhjY

Relevant Content

Vocabulary

» Normal distribution: continuous, symmetric, bell-shaped curve showing a particular distribution of probability over the values of a random variable

Formulas

Pearson's Index of Skewness: $P = \dfrac{3(\bar{x} - median)}{s}$, $P < 0 \Rightarrow$ the data are skewed left, $P = 0 \Rightarrow$ the data are symmetric, $P > 0 \Rightarrow$ the data are skewed right

Skewness Facts

» Measures of location
   • Mean < median < mode, the distribution is negative skewed
   • Mean = median = mode, the distribution is symmetric
   • Mean > median > mode, the distribution is positive skewed

» Measures of dispersion
   • $Q_3 - Q_2 < Q_2 - Q_1$, the distribution is negative skewed
   • $Q_3 - Q_2 = Q_2 - Q_1$, the distribution is symmetric
   • $Q_3 - Q_2 > Q_2 - Q_1$, the distribution is positive skewed

Normal Distribution Facts

- » The distribution curve is bell-shaped.
- » The mean, median, and mode are equal and are located midway on the distribution.
- » The distribution is unimodal.
- » The curve is symmetric about the mean.
- » The curve is continuous.
- » The curve approaches the x axis as $x \rightarrow -\infty$ and as $x \rightarrow \infty$.
- » The area under the curve is equal to 1.00, or 100%.
- » The area under the curve that lies within 1 standard deviation of the mean is approximately 0.68, or 68%.
- » The area under the curve that lies within 2 standard deviations of the mean is approximately 0.95, or 95%.
- » The area under the curve that lies within 3 standard deviations of the mean is approximately 0.997, or 99.7%.

Theorems

Chebyshev's Theorem: The proportion of values from a data set within k standard deviations of the mean is at least $1 - \frac{1}{k^2}$, where $k > 1$. This applies to any distribution regardless of its shape.

The Empirical (Normal) Rule

- » Approximately 68% of the data values will fall within 1 standard deviation of the mean.
- » Approximately 95% of the data values will fall within 2 standard deviations of the mean.
- » Approximately 99.7% of the data values will fall within 3 standard deviations of the mean.

Assessment Links or Tasks

- » Should Statistics be Shapely?
  https://www.cpalms.org/Public/PreviewResourceLesson/Preview/127994
- » Normal distribution tasks:
  https://study.com/academy/lesson/normal-distribution-activities-games.html