About Distributions

UNIVERSITY of GUELPH

TABLE OF CONTENTS

About Distributions

What is DISTRIBUTION?

- A distribution is a type of graph with a smooth curve to illustrate the relative changes occurring from one score to the next.

Introduction

- When a population consists of interval- or ratio- data, distribution is drawn with a smooth curve, as opposed to the jagged-shape seen in histograms
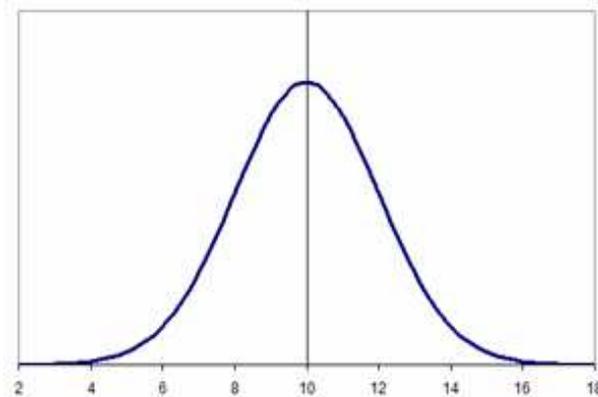

Three Characteristics to Describe Distributions

1)   Shape

- The shape of a specific distribution is defined by an equation that describes the relationship between X and Y.

  There are two main shapes for a distribution:

A)   *Normal*

- The distribution is symmetrical
- The greatest frequencies of scores are located in the middle, while the relatively smaller frequencies of scores exist at opposing extremes
- An easy way to visualize the normal distribution is to think of a bell. It is often referred to as bell-shaped.
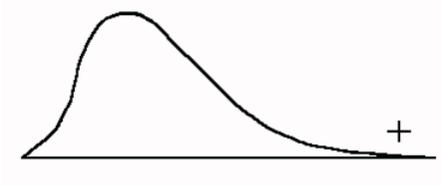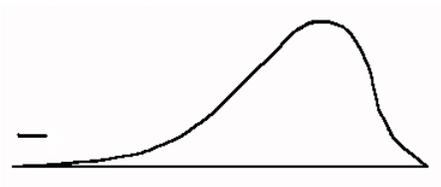
\
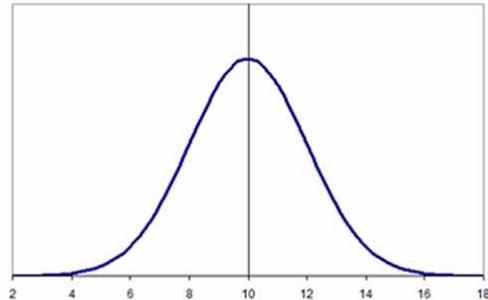
B)   *Skewed*

- The distribution is asymmetrical.
- The greatest frequencies of scores are piled towards one end of the distribution, while the smaller frequencies of scores are tapered towards the other end of the distribution (this end is called the tail of the distribution).

There are two types of skewed distribution:

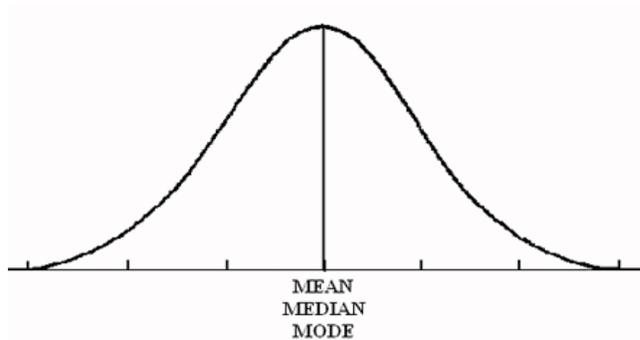1) Positively skewed

2) Negatively skewed

2) Central Tendency

- Measures of central tendency (i.e., mean, median, and mode) describe the center of a distribution.

  For a normal distribution, the mean, median, and mode will be in the center of the distribution.

  If you draw a line down the middle of a normal distribution, there is a mirror image on both sides of the line

  MEAN
  MEDIAN
  MODE

  For a skewed distribution, the median and mode will be closer to the greatest frequency of scores. The mean will be pulled towards the tail end of the distribution by the extreme scores. This is because the mean uses every score in the distribution.

  In this type of distribution, the median is the most accurate measure of central tendency.

  MODE   MEAN
  MEDIAN

3
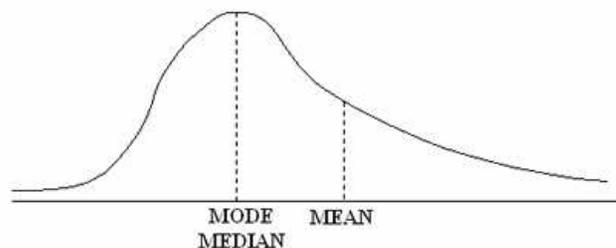
3)   Variability

● This refers to how spread out the scores are over a wide range.

For a normal distribution, the greatest frequencies of scores are in the center of distribution, and the remaining scores taper towards the extreme ends of the distribution.

For skewed distributions, the greatest frequencies of scores are located at one end of the distribution.

● In a normal distribution, the scores are spread out over a greater range; in a skewed distribution, the scores are more clustered together. Therefore a normal distribution has greater variability than a skewed distribution

Normal Distribution

Characteristics of Normal Distribution

● A normal (bell-shaped) distribution has several important characteristics:
- symmetrical
- unimodal
- extends to positive or negative infinity
- area under the curve = 1

● The normal distribution can be defined by two parameters:
1) The mean
2) The standard deviation

The empirical rule

- The empirical rule is an estimate of the spread of data, given the mean and standard deviation, in a normal distribution.

- The empirical rule states that:
  - 68% of the scores will fall within 1 standard deviation of the mean.
  - 95% of the scores will fall within 2 standard deviations of the mean.
  - Almost all (99.7%) of the score will fall within 3 standard deviations of the mean.

Importance of Normal Distributions

- Normal distributions are considered as the most important type of distribution for 3 main reasons:

  1) Many dependent variables are commonly assumed to be normally distributed in the population

  2) Most statistical techniques have an assumption that the population is normally distributed. This assumption allows for making inferences about that population.

  3) A sampling distribution of the mean approximates a normal distribution under a wide variety of conditions

Central Limit Theorem

- A data set that approximates a normal distribution will become more normal as the sample size increases. This relates to the *central limit theorem.*

- The central limit theorem states that, regardless of the distribution of the population, the distributions of the means of random samples will approach a normal distribution for a large sample size.

- In many statistical tests, there is the assumption for the population from which the sample is drawn to be normally distributed. This assumption can be violated if the sample size is large enough (usually sample size is greater than 30)

Skewed Distributions

- Skewness describes the asymmetry of a distribution around its mean.

- A skewed distributions can be:

    A) Positively skewed

    The tail of the distribution is on the right side.

    A positively skewed distribution may describe the testing scores of a difficult exam: a high frequency of students scored poor marks, and a low frequency of students scored high marks. In this case, the value of the mean would be greater than the values for the median and the mode.

    B) Negatively skewed

    The tail of the distribution is on the left side.

    A negatively skewed distribution may describe testing scores on an easy exam: a high frequency of students scored high marks, and a low frequency of students scored low marks. In this case the value of the mean would be smaller than the values for median and mode.

- Positively skewed distributions are more common than negatively skewed distributions.

Glossary

Central Limit Theorem: given a population distribution, the sampling distribution of the mean approaches a normal distribution as the sample size increases, despite the shape of the original population distribution.

Empirical rule: for data sets having a normal, bell-shaped distribution, the following properties apply: about 68% of all values fall within 1 standard deviation of the mean; about 95% of all values fall within 2 standard deviation of the mean; and about 99.7% of all values fall within 3 standard deviation of the mean.

Mean: the arithmetic average of a set of scores

Measures of central tendency: describe the average or mid-most score in a distribution.

Median: the middle score in a distribution

Mode: the most frequently-occurring score in a distribution.

Negatively skewed distribution:

A type of skewed distribution where the greatest frequencies of scores are located in the left-hand side of the distribution and the fewest scores are located in the right-hand side of the distribution.

Normal distribution: a symmetrical, bell-shaped curve that has the greatest frequency of scores in the middle of the distribution and relatively

fewer scores towards the ends of the distribution.

Positively skewed distribution:

a type of skewed distribution where the greatest frequencies of scores are located in the right-hand side of the distribution and the fewest scores located in the left-hand side of the distribution.

Skewed distribution:

an asymmetrical curve that has the greatest frequency of scores towards one end of the distribution and fewer scores that taper towards the other end.

Variability:

describes the range (or spread) of the scores in a distribution.

References

Gravetter, F. J. & Wallnau, L. B. (2004). Statistics for the
   behavioural sciences. (6th ed.) Belmont: Wadsworth/Thomson
   Learning.