

Math 1332
Chapter 12

Section 1

Data: Numerical information

Statistics: a method of collecting, organizing, analyzing, and interpreting data, as well as drawing conclusions based on the data.

Descriptive Statistics: Concerned with collecting, organizing, summarizing and presenting the data.

Inferential Statistics: concerning the generalizations and the drawing conclusions from the data collected.

Population: all the objects whose properties are in question

Sample: subset or subgroup of the population

Random Sample: every element in the population has an equal chance to be selected.

Representative Sample: a sample that exhibits characteristics that the population has.

Data item: piece of data

Frequency Distribution: a table that holds the data values and the number of times the data value occurs, the number of times is referred to as the frequency.

Classes: groups to arrange data, has a lower and upper bounds.

Grouped frequency distribution: a frequency distribution where data is collected into classes.

Class width: difference between two consecutive lower class limits.

Visual displays of data

Histogram

Frequency Polygon

Stem-and-leaf plot

Section 2

Measure of Central Tendency: measurement that is the “typical” or “average” value of the data. It is a measurement that occurs in the center of the distribution. Mean, Median, Mode, and Midrange.

Mean: $\frac{\sum x}{n}$ Sum of all the data values divided by number of data values.

Mean of Frequency Distribution: $\frac{\sum xf}{n}$ Sum of the product of each data value and its frequency divided by the number of data values.

Median: value in middle of data

1. Arrange data in order, smallest to largest
2. If n is odd, the data value in the middle is the mean
3. If n is even, the median is the mean of the two data values in the middle

The median is located at $\frac{n+1}{2}$ position.

Mode: the data value that appears the most, if there is two modes that data set is bimodal.

Midrange: $\frac{\text{lowest value} + \text{highest value}}{2}$

Section 3

Measures of dispersion: measurements used to describe the spread of data items in a data set. Range and Standard Deviation.

Range: difference between the highest value and the lowest value

Standard Deviation: measurement that shows how each data item differs from the mean. A kind of average the deviations from the mean.

1. Find the mean
2. Find the difference of each data item and the mean
3. Square the value found in part 2.
4. Sum all the values in 3.
5. Divide the result by $(n - 1)$.
6. Take the square root of part 5. The result is the standard deviation.

Section 4

Symmetric: a histogram that as you move out from the middle, the heights of the bars on the left side are the same as the heights on right side.

Normal Distribution: Bell shaped data that is symmetric about a vertical line in the center of the data. The mean, median, and the mode are all equal.

68 – 95 – 99.7 Rule

1. Approximately 68% of the data falls with 1 standard deviation of the mean.
2. Approximately 95% of the data falls with 2 standard deviations of the mean.
3. Approximately 99.7% of the data falls with 3 standard deviations of the mean.

Z-score: describes how many standard deviations a data item in a normal distribution lies above or below the mean.

$$z - score = \frac{\text{data item} - \text{mean}}{\text{standard deviation}}$$

Percentile: if n% falls below an item it is in the nth percentile of the distribution. To find the percentile of the z score use Table 12.10, Remember that it reads the percentile that is less than the item in question. To calculate percentile between, subtract the lesser percentile from the greater percentile.

Margin of error: $\frac{1}{\sqrt{n}}$

Skewed Distributions: Large number of data items are located to one side with a “tail” at the opposite end. The side the tail is out is the direct of the skew.

Skewed to the right is when all the data is piled on the left side and the tail is to the right.

Skewed to the left is when all the data is piled on the right side and the tail is to the left.

Section 5

Scatter plot: collection of data points, one point per value.

Correlated: data having a clear relationship

Positively correlated: data variables increase or decrease together.

Negatively Correlated: one variable decreases while the other increases.

Correlation: measurement of the relationship between two variables, and the strength and direction of the relationship if one exists.

Correlation Coefficient: r : variable used to describe the correlation. Always between -1 and 1.

Perfect positive correlation: $r = 1$

Perfect negative correlation: $r = -1$

Computing Correlation Coefficient:
$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \sqrt{n(\sum y^2) - (\sum y)^2}}$$

Regression line: line that best fits the data values.

Equation of Regression Line: $y = mx + b$

$$m = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \qquad b = \frac{\sum y - m(\sum x)}{n}$$

Significance Level: probability that the variable in question is correlated. We use 1% and 5% most often. If the absolute value of r is greater than the corresponding value in Table 12.12, (the intersection of n and significance level) then there is a correlation.