Measures of Central Tendency

LECTURE 3

Objectives

- Define the terms.
- Calculate mean, mode, and median.
- Understand simple statistical notation.

Central tendency

- A large proportion of the values for most variables are found near the middle of the range of observed values.
  - Define statistical range
  - Define mathematical range
- The concentration is typically refereed to as the average or the central tendency.

Central tendency

- Can be useful for characterizing a population

Figure 1. Number of eggs laid by red-eared slider per clutch

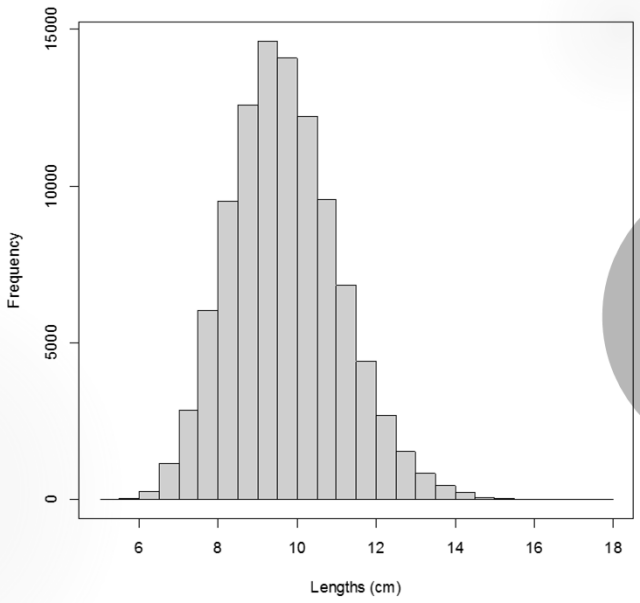


Figure 1. Lengths of pine needles (cm) from Larimer county, CO.

Arithmetic mean

- The commonly used measure of central tendency is the arithmetic mean.
- Frequently referred to as the mean.
- Most commonly referred to (incorrectly) as the average.

Statistical notation

- X_i -- the i^{th} value of the variable in a data array.

1 3 7 13 21 31

$X_1=1$

$X_3=7$

$X_6=31$

Statistical notation

- N = The total number of values in the population.
- μ = Mean of the population (lower case mu)

Statistical notation

- Σ = Summation

$$\sum_{i=1}^{N} X_i$$

- Summation of all X_i values from X_1 through X_N

Statistical notation

1 3 7 13 21 31

$$\sum_{i=1}^{N} X_i = 1+3+7+13+21+31 = 76$$

$$\sum_{i=1}^{N} X_i = 76$$

Arithmetic mean

- Population mean (μ)

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$$

Arithmetic mean

1 3 7 13 21 31

$$\sum_{i=1}^{N} X_i = 76 \qquad N=6$$

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{76}{6} = 12.667$$

Arithmetic mean

- Sample mean

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

n = sample size

Median

- The median (M) is the middle measurement in an ordered set of data.
- In a symmetrical distribution, the sample median is unbiased and is an effective estimate of the population mean.
  - However, most distributions are not perfectly symmetrical and then median becomes a poor measure of the population mean.
  - The median is rarely used.

Median

- Order the data in ascending order.

$$\mathrm{M} = X_{\frac{(n+1)}{2}}$$

  - If the sample size (n) is odd then the subscript in the equation will be an integer and will indicate which datum is the median
  - If the sample size (n) is even then the subscript will be midway between two integers. The data on each side of midpoint are then used and the median is the midpoint between the values of the two data.

Median

1 3 7 13 21 31 43

$$\mathrm{M} = X_{\frac{(n+1)}{2}} = X_{\frac{7+1}{2}} = X_4 = 13$$

Median

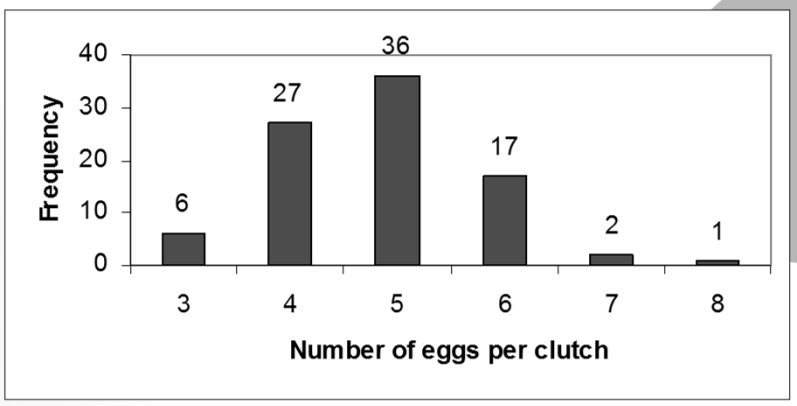
1 3 7 13 21 31

$$\mathrm{M} = X_{\frac{(n+1)}{2}} = X_{\frac{6+1}{2}} = X_{3.5} = \frac{X_3 + X_4}{2} = \frac{7+13}{2} = 10$$

Mode

- The mode is the most frequently occurring value in a data set or a measure of relatively great concentration.
- The mode is easily determined from a frequency table, histogram, or polygon.
- If two adjacent values have equal number of occurrences then the mode is the midpoint between the two values.

Mode



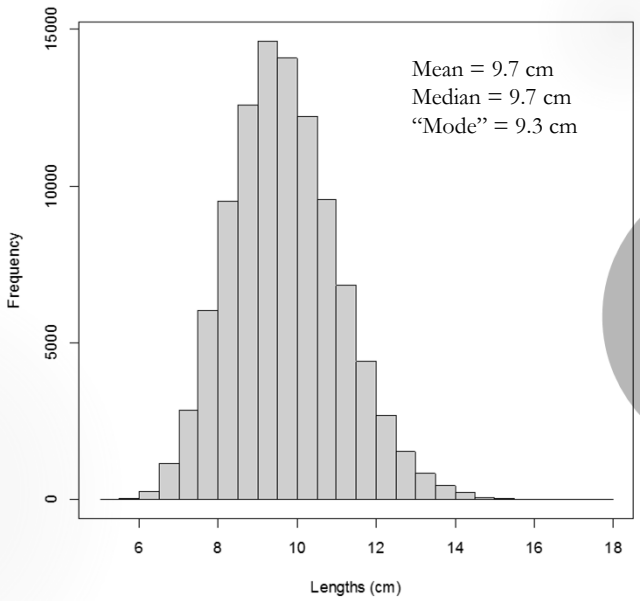


Figure 1. Lengths of pine needles (cm) from Larimer county, CO.

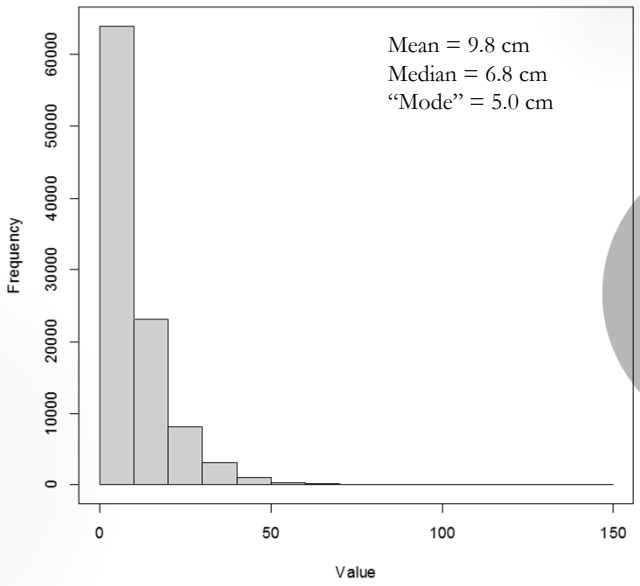


Figure 2. Fake data showing an exponential distribution.