Key

Math III Statistics Review

S.ID.4, S.IC.1, S.IC.3, S.IC.4, SIC.5

Sampling

Population – a group that has something in common (ex/ freshmen at high school, employees at Burger King). Populations are often people but they can refer to things (ex/ bolts produced at a particular factory on Friday night)

Often, researchers want to know things about a population but it is not practical or perhaps possible to observe every member of a population, so a sample is taken from the population.

Sample - a smaller group of a population selected to represent the population.

In order for a sample to be a good representation of the population, the sample needs to be random. A random sample means every member of the population has an equal chance of being selected.

A **parameter** is a characteristic of a population. A **sample statistic** is a characteristic of a sample. For example, if you are interested in finding the average height of freshmen girls at a particular high school, and you find the average height to be 5' 5" tall. The parameter of the population is freshmen girls and the statistic is the average height of 5' 5".

Different types of sampling:

Simple Random Sample – choosing n elements of a population that each have an equal chance of being selected.

Systematic Random Sampling – elements of the population are put into a list and every nth element is chosen. For example, using every 10^{th} student in an alphabetical list of students enrolled at a high school.

Stratified Random Sample – When sub-populations vary considerably, members of the population are grouped into subgroups before sampling and samples are taken from each subgroup.

Cluster/Multistage Sample – In a cluster sampling, a total population of interest is divided into 'clusters', (ex/ a total population grouped into geographic regions) and from each cluster of interest, individual subjects are selected randomly. Multistage – divide the first stage of clusters further into second stage clusters (ex/ first 'cluster' of a total population by geographic region, and next dividing each regional cluster by neighborhood).

Convenience Sample – selecting sample elements from volunteers or other convenient methods

Experiments and observational studies

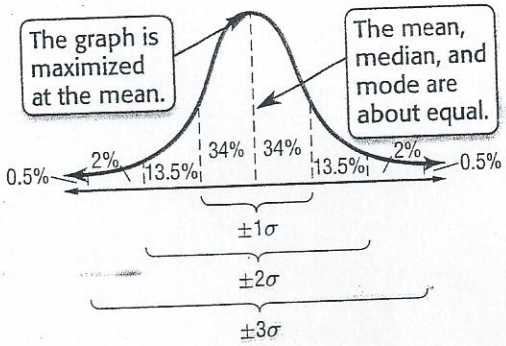
Sample surveys are used to collect data from human subjects to describe the population of interest.

Experiments and observational studies are used to establish a cause and effect relationship. In an experiment, a treatment is imposed on the experimental units. The randomization occurs when experimental units are assigned to treatments in order to ensure that the treatment groups are equivalent. In an observational study, the treatment is not imposed but the relationship between the variables of interest is observed (ex/ smoking and birth defects). There is no random assignment of treatments (ex/ when looking at the relationship between smoking and cancer, we do not "assign" people to be smokers, they choose to be one or not. We then observe the rate of cancer for smokers and non-smokers and compare.)

Empirical Rule

A normal distribution with mean μ and standard deviation σ has the following properties.

- About 68% of the values are within 1σ of the mean.
- About 95% of the values are within 2σ of the mean.
- About 99% of the values are within 3σ of the mean.

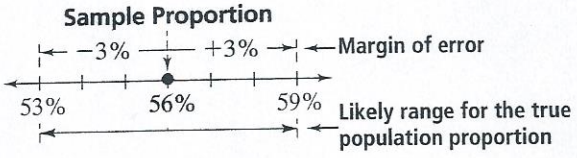


Z-Scores

The formula to find the z-score is $z - score = \frac{Observed\ Value - Mean}{Standard\ Deviation}$. The z-score is how many standard deviations the observation is from the mean.

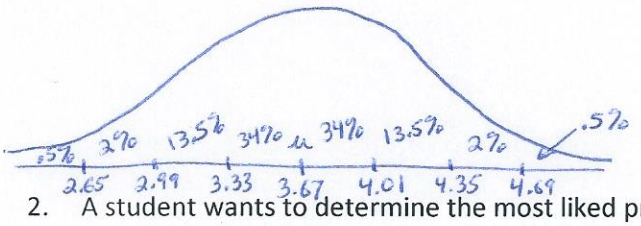
Margin of error

When a random sample of size n is taken from a large population, the sample proportion has a margin of error of approximately $\pm \frac{1}{\sqrt{n}}$



1. A town has 685 households. The number of people per household is normally distributed with a mean, μ, of 3.67 and a standard deviation, σ, of 0.34. ***Approximately*** how many households have between 2.99 and 4.01 people?

   A. 493 households B. 520 households (C.) 558 households D. 575 households



2.99 to 4.01 is 81.5%

685(.815) = 558.275

2. A student wants to determine the most liked professor at her college. Which type of study would be the most ***practical*** to obtain this information?

   A. a simulation B. an experiment (C.) a survey D. an observation

3. A principal wants to survey 150 students to determine which electives to offer during the next school year. There are 1,800 students in the school. Which procedure could the principal use to select a sample using a systematic random sample?

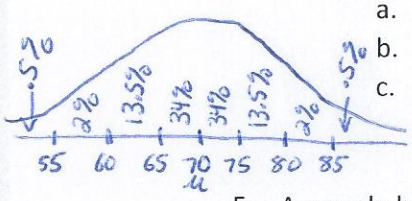
   (A.) Obtain a list of all students. Start with the eighth student, and select every twelfth student until 150 students have been selected.

   B. Select the first 150 students who enter the school.

   C. Choose the fifth student to come to the cafeteria, and then select every third student who comes into the cafeteria until 150 students have been selected.

   D. Place students' names on slips of paper and select 150 slips.

4. 2000 freshmen at State University took a math test. The scores were distributed normally with a mean of 70 and a standard deviation of 5.



   a. What percent of the scores were between 60 and 75? 81.5%

   b. What percent of the scores were between 75 and 80? 13.5%

   c. Approximately how many students scored between 65 and 70?

65 to 70 = 34% 2000(.34) = 680

5. A sample has a mean of 85 and a standard deviation of 2. Find the z-score of an observation of 82 and explain where it is located on the normal distribution.

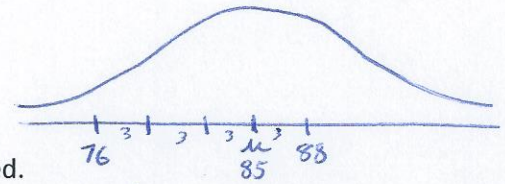
$$\text{z-Score} = \frac{\text{observation} - \text{mean}}{\text{std. dev}}$$

$$= \frac{82 - 85}{2} = 1.5$$

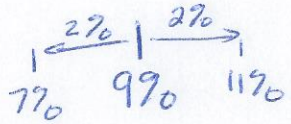
82 is located 1.5 standard deviations below the mean

6. A reporter wants to know the percentage of voters in the state who support building a new highway. What is the reporter's population?
   A. The number of people who live in the state
   B. The people who were interviewed in the state
   C. All voters over 25 years old in the state
   D. All eligible voters in the state

7. In a set of test scores that are normally distributed, a test score of 76 is 3 standard deviations below the mean. A score of 88 is 1 standard deviation above the mean. What is the mean of the data?
   A. 79 B. 82 C. 84 D. 85



8. A survey of 2580 students found that 9% are left-handed.
   - Find the margin of error for the sample $\pm \frac{1}{\sqrt{n}} = \pm \frac{1}{\sqrt{2580}} = \pm 0.0197 \approx 2\%$
     $\pm 2\%$
   - Use the margin of error to find an interval that is likely to contain the true population proportion.



The proportion of students who are left-handed is likely to be from 7% to 11%