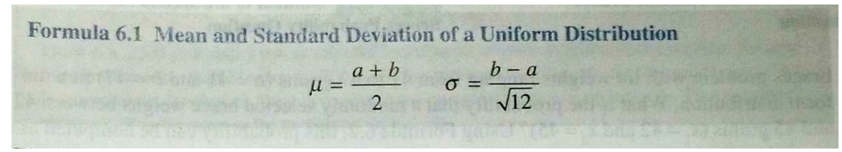
Week 4 - Review

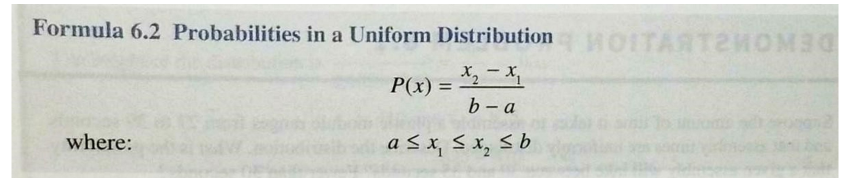
Sunday, February 04, 2018 6:20 AM

Different Types of Sampling Plans

- Reasons for sampling
  - The sample can save money and time
  - The sample ban broaden the scope of the study (resource dependent)
  - If access to the population is impossible, the sample is the only options.
- Reasons for taking a census
  - Eliminate the possibility that by chance a randomly selected sample may not be representative of the population
  - Safety of the consumer
  - Benchmark data for future studies
- Random Versus Nonrandom Sampling
  - **Random sampling** *(aka probability sampling)* means that every unit of the population has the same probability of being selected into the sample.
    - Techniques for random sampling
      - ***Simple Random Sampling***
        - This is the most basic random sampling technique.
        - Each unit of the set is number from 1 to N (where N is the size of the population).
        - A random number generator is used to select n items into the sample.
      - ***Stratified Random Sampling***
        - The population is divided into non-overlapping sub-populations called strata.
        - The main reason for using this technique is that it has the potential to reduce sampling error, which occurs when the sample does not represent the population.
        - Strata selection is usually based on available information which could be from previous censuses or surveys.
        - ***Proportionate stratified random sampling –*** this occurs when the percentage of the sample taken from each stratum is proportionate to the percentage that each stratum is within the whole population.
        - ***Disproportionate stratified random sampling –*** this occurs whenever the proportions of the strata in the sample are different from the proportions of the strata in the populations.
      - ***Systematic Sampling***
        - Every kth item is selected to produce a sample of size n from a population of size N. K is sometimes called sampling cycle
          - k = N/n
            - *n = sample size, N = population size, k = size of the interval for selection*
      - ***Cluster (or area) Sampling***
        - This involves dividing the population into non-overlapping areas, or

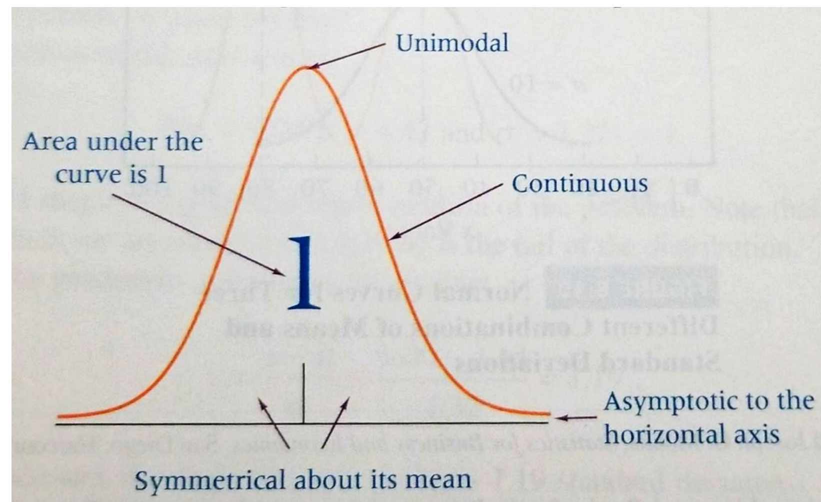

Formula 6.1 Mean and Standard Deviation of a Uniform Distribution

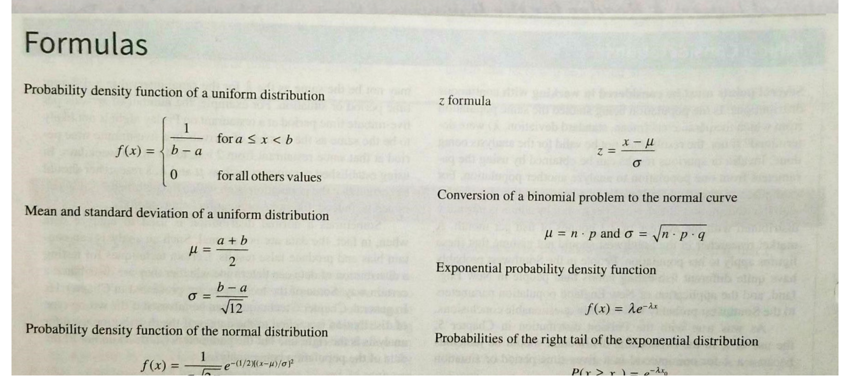
$$\mu = \frac{a+b}{2} \qquad \sigma = \frac{b-a}{\sqrt{12}}$$


Formula 6.2 Probabilities in a Uniform Distribution

$$P(x) = \frac{x_2 - x_1}{b-a}$$

where: $a \le x_1 \le x_2 \le b$




Formulas

Probability density function of a uniform distribution

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x < b \\ 0 & \text{for all others values} \end{cases}$$

Mean and standard deviation of a uniform distribution

$$\mu = \frac{a+b}{2}$$

$$\sigma = \frac{b-a}{\sqrt{12}}$$

Probability density function of the normal distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(1/2)[(x-\mu)/\sigma]^2}$$

z formula

$$z = \frac{x-\mu}{\sigma}$$

Conversion of a binomial problem to the normal curve

$$\mu = n \cdot p \text{ and } \sigma = \sqrt{n \cdot p \cdot q}$$

Exponential probability density function

$$f(x) = \lambda e^{-\lambda x}$$

Probabilities of the right tail of the exponential distribution

$$P(x \ge x_0) = e^{-\lambda x_0}$$

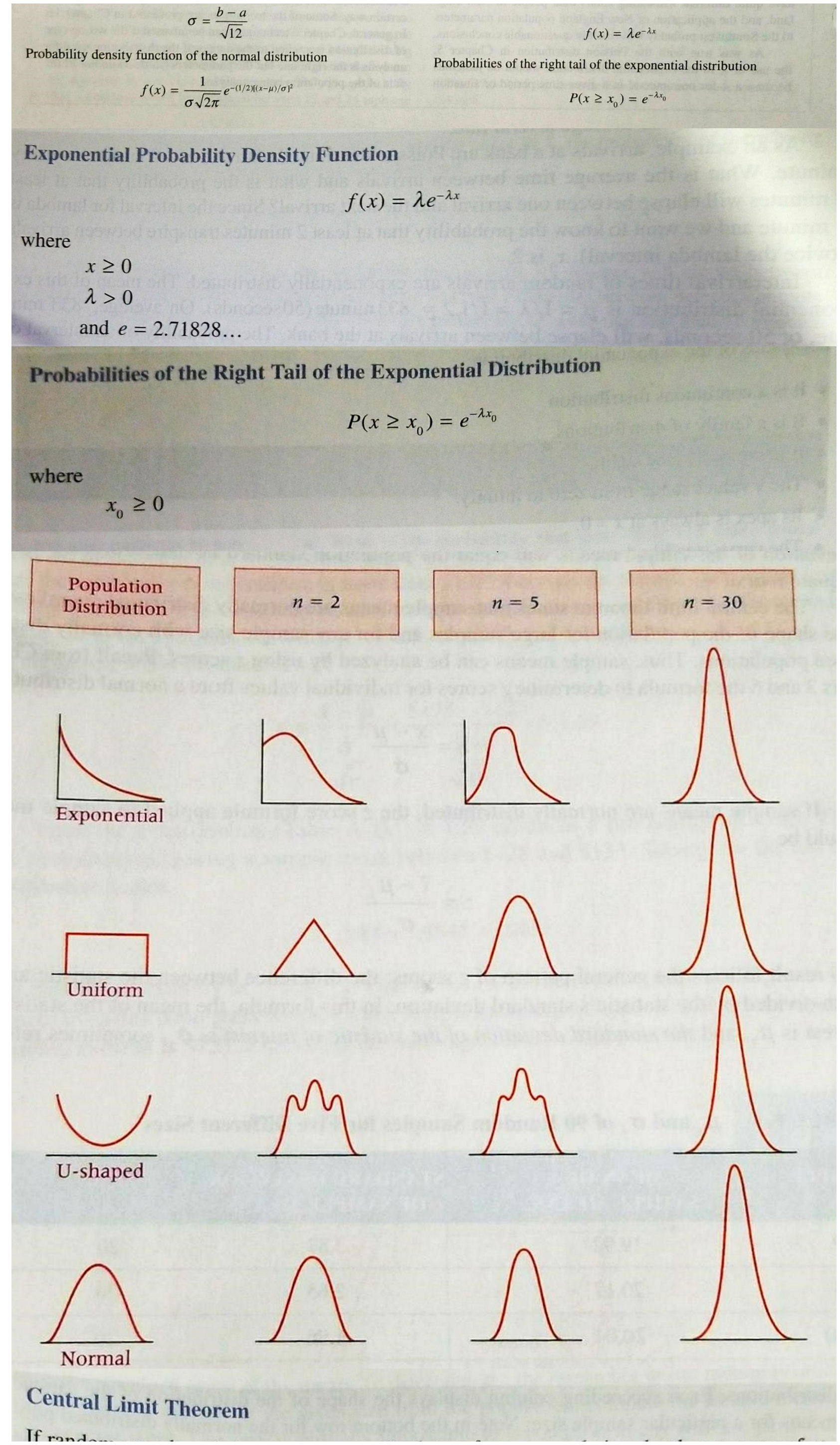
selection

- ☐ ***Cluster (or area) Sampling***
  - ◆ This involves dividing the population into non-overlapping areas, or clusters.
  - ◆ This differs from stratified random sampling since their strata are homogeneous within. Whereas cluster sampling contains a wide variety of elements.
  - ◆ ***Two-stage sampling-*** used whenever clusters are too large and second clusters are taken from each original cluster.

- ○ **Nonrandom sampling** *(aka non-probability sampling)* means that not every unit of the population has the same probability of being selected into the sample. *Nonrandom sampling methods are not appropriate techniques for gathering data to be analyzed by most of the statistical methods presented and have limitations.*
  - ▪ Techniques for nonrandom sampling
    - ☐ ***Convenience Sampling***
      - ◆ Elements for the sample are selected for the convenience of the researcher. These chosen elements are readily available, nearby, or willing to participate.
    - ☐ ***Judgment Sampling***
      - ◆ Elements selected for the sample are chosen by the judgment of the researcher.
      - ◆ Problems with judgment sampling
        - ◇ Researchers tend to make errors of judgment in one direction which can lead to biases.
        - ◇ It is also unlikely that researches will include extreme elements.
        - ◇ There is no objective method for determining if one researcher's judgment is better than another's.
    - ☐ ***Quota Sampling***
      - ◆ This appears to be similar to stratified random sampling however instead of randomly sampling each stratus, the researcher uses a nonrandom sampling method to gather data from a stratum until the desired quote of samples is filled.
    - ☐ ***Snowball Sampling***
      - ◆ Survey subjects are selected based on referral from other survey respondents.
      - ◆ The researcher identifies a person who fits the profile of the subjected wanted for the study. The researcher then asks this person for the names and locations of other who would also file the profile of the subjects wanted for the study.
  - ▪ **Types of Error**
    - ☐ ***Sampling error –*** this occurs when the sample is not representative of the population
    - ☐ ***Non-sampling error –*** all errors other than sampling errors

- ○ **Sampling from a Finite Population**
  - ▪ Finite correction factor is a statistical adjustment can be made to the z formula for sample means.

- ○ **Sampling Distribution of p-hat**


$\sigma = \frac{b-a}{\sqrt{12}}$

Probability density function of the normal distribution

$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(1/2)[(x-\mu)/\sigma]^2}$

$f(x) = \lambda e^{-\lambda x}$

Probabilities of the right tail of the exponential distribution

$P(x \geq x_0) = e^{-\lambda x_0}$

Exponential Probability Density Function

$$f(x) = \lambda e^{-\lambda x}$$

where

$x \geq 0$

$\lambda > 0$

and $e = 2.71828\ldots$

Probabilities of the Right Tail of the Exponential Distribution

$$P(x \geq x_0) = e^{-\lambda x_0}$$

where

$x_0 \geq 0$

Central Limit Theorem

- Sampling from a Finite Population
    - Finite correction factor is a statistical adjustment can be made to the z formula for sample means.
- **Sampling Distribution of p-hat**
    - ***Sample proportion*** is computed by dividing the frequency with which a given characteristic occurs in a sample by the number of items in the sample.
    - *Standard error of the proportion*
- **Sampling Distribution of the Mean Under Normality**
    - Under random sampling, the average value of the sample mean, over millions of studies can be shown to equal the population mean. Chances are that the sample mean will not be equal to the population mean but on average, the sample mean provides a correct estimate of the population mean.
    - The variance of the sample mean can be shown to the average squared difference between the sample mean and the population mean.
- ***Estimator***
    - It is some expression, based on the observations made, intended to estimate some feature of the population under study.
    - Example: the sample mean is the estimator of the population mean and its observed value is called an *estimate*.
    - It is unbiased if is average value over millions of studies is equal to the quantity it is trying to estimate.

<u>Central Limit Theorem</u>

- Under random sampling, as the sample size gets larger, the sampling distribution of the sample mean approaches a normal distribution with mean and variance.
- If the sample size is sufficiently large, we can assume that the sample mean has a normal distribution.
    - Sufficiently large: a common claim is that n = 40 generally suffices
- The central limit theorem creates the potential for applying the normal distribution to many problems when the sample size is sufficiently large.
- ***Approximating the binomial distribution***
    - The central limit theorem says that if the sample size is sufficiently large, p-hat will have a normal distribution with mean *p* (the true probability of success) and variance *p(1-p)/n*.

<u>Standard Normal Distribution</u>

- Properties
    - It is a continuous distribution
    - It is a symmetrical distribution about its mean
    - It is asymptotic to the horizontal axis
    - It is unimodal
    - It is a family of curves
    - Area under the curve is 1
- Find z scores
    - Z score – is the number of standard deviations that a value, x, is above or below the mean.
    - Z distribution – is a normal distribution with a mean of 0 and a standard deviation of 1.
    - *z = (x – mean) / standard deviation*
- Normal Distribution to Approximate Binomial Distributions
    - Correction for Continuity – is made during conversion of a discrete distribution into a

Normal

Central Limit Theorem

If random samples of size n are repeatedly drawn from a population that has a mean of μ and a standard deviation of σ, the sample means, $\bar{x}$, are approximately normally distributed for sufficiently large sample sizes ($n \geq 30$) regardless of the shape of the population distribution. If the population is normally distributed, the sample means are normally distributed for any size sample.

From mathematical expectation,* it can be shown that the mean of the sample means is the population mean

$$\mu_{\bar{x}} = \mu$$

and the standard deviation of the sample means (called the standard error of the mean) is the standard deviation of the population divided by the square root of the sample size.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

z Formula for Sample Means of a Finite Population (7.3)

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}}$$

Sample Proportion (7.4)

$$\hat{p} = \frac{x}{n}$$

where

x = number of items in a sample that have the characteristic
n = number of items in the sample

z Formula for Sample Proportions for $n \cdot p > 5$ and $n \cdot q > 5$ (7.5)

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}}$$

where

$\hat{p}$ = sample proportion
n = sample size
p = population proportion
$q = 1 - p$

Formulas

Determining the value of k

$$k = \frac{N}{n}$$

z formula for sample means

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

z formula for sample means when there is a finite population

$$\bar{x} - \mu$$

Sample proportion

$$\hat{p} = \frac{x}{n}$$

z formula for sample proportions

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}}$$

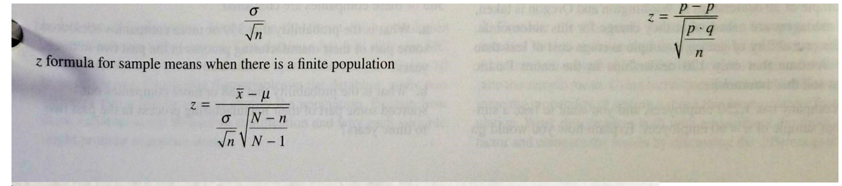
- Z distribution – is a normal distribution with a mean of 0 and a standard deviation of 1.
    - *z = (x – mean) / standard deviation*
- Normal Distribution to Approximate Binomial Distributions
    - Correction for Continuity – is made during conversion of a discrete distribution into a continuous distribution
- Exponential Distributions
    - Properties
        - It is a continuous distribution
        - It is a family of distributions
        - It is skewed to the right
        - The x values range from zero to infinity
        - Its apex is always at x = 0
        - The curve steadily decreases as x gets larger
- Probability Density Function for a Continuous Variables
    - Features of continuous distributions
        - Created from measured variables: volume, distance, temperature, weight, dollars
        - Areas under a continuous distribution curve is 1
        - Probability are found by determining the area

<u>Probability Density Function</u>

- In probability theory, a probability density function (PDF), or density of a continuous random variable, is a function, whose value at any given sample in the sample space (the set of possible values taken by the random variable) can be interpreted as providing a relative likelihood that the value of the random variable would equal that sample.

<u>Expected Value (or Mean) and Variance</u>

- The ***expected value (mean)*** of X, where X is a discrete random variable, is a weighted average of the possible values that X can take. Each value being weighted according to the probability of that event occurring.
- The ***variance*** of a discrete random variable X measures the spread, or variability of the distribution.

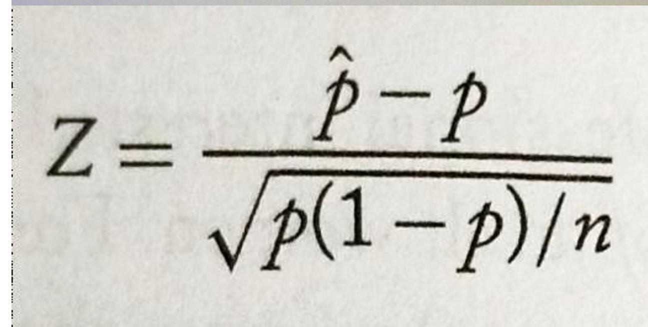


$$\frac{\sigma}{\sqrt{n}}$$

z formula for sample means when there is a finite population

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}}$$

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot q}{n}}}$$



$$Z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$$