**Important Concepts**

Biased sample

* A sample that is not completely random because it favors observations with a certain characteristic

Central Limit Theorem (CLT)

* The mean for a sample taken from any data has an approximate normal distribution regardless of the distribution of the original data if the sample size is large enough (n30).
* Applies to nominal, ordinal, and interval data.

Distribution

* A table, graph, or function that describes all possible outcomes of a random variable.

Margin of error (MOE)

* A measure of chance variance
* For percentage statistics, MOE ≈

Parameter

* A number that summarizes a variable for an entire population.

Population proportion (*p*)

* The proportion of observations in the population that have a certain characteristic.

Random variable

* A characteristic whose value for any given sample cannot be predicted.

Sample proportion ()

* The proportion of observations in the sample that have a certain characteristic.

Statistic

* A number that summarizes the data collected for a variable from a sample of a population.

Standard Error (SE)

* The difference between a population parameter and a sample result.
* Similar to standard deviation of a population.
* Measures variability in sample results.
* Applies to sample means and sample percentages
* Standard error decreases as sample size increases.
* Standard error increases as population standard deviation increases (i.e., diversity increases).

Variable

* A characteristic that varies from element to element within a population.

**Symbols**

|  |  |  |
| --- | --- | --- |
| **Denotation** | **Pronunciation** | **Meaning** |
| μ | Mu | Population mean |
| σ | Sigma | Population standard deviation |
| σ2 | Sigma squared | Population variance |
| ρ | Rho | Population correlation coefficient |
| df | “dee-ehf” | Degrees of freedom |
| n | “en” | Sample size or number of trials |
| p | “pee” | Sample proportion |
| MOE | “em-oh-ee” | Margin of error |
| r | “ahr” | Sample correlation coefficient |
| r2 | “ahr” squared | Sample correlation coefficient |
| S | “es” | Sample standard deviation |
| S2 | “es” squared | Sample variance |
| SSY | “es-es-wahy” | Sum of the squares of deviations from the mean for variable Y |
| SSE | “es-es-ee” | Sum of the squares due to error (residuals) |

**Summary Table**

| Objective | Constraints and Assumptions | Concept or Tool | Equations and Algorithms |
| --- | --- | --- | --- |
| Analyze individual values within a population:  Find the probability of a value. | * The outcome is a discrete random variable. * There are a fixed number of trials (n). * Each trial has only two possible outcomes. * The probability of success (p) is the same for each trial. * The outcome of one trial does NOT influence the outcome of any other trial. | Binomial distribution  Use the normal distribution to approximate the binomial distribution if both of the following are true:   * (n)(p) 10 * (n)(1-p) 10 | 1. Locate the binomial table associated with the value of n. 2. Find the column that represents p. 3. Find the row that represents the number of successes. 4. The intersection of the row and column provides the probability of x successes, p(x). |
| Analyze individual values within a population:  Find the probably for an outcome greater than some specific value.  Find the probably for an outcome less than some specific values.  Find the probably for an outcome between two specific values. | * The outcome is a discrete random variable. * There are a fixed number of trials (n). * Each trial has only two possible outcomes. * The probability of success (p) is the same for each trial. * The outcome of one trial does NOT influence the outcome of any other trial. | Binomial distribution  Use the normal distribution to approximate the binomial distribution if both of the following are true:   * (n)(p) 10 * (n)(1-p) 10 | Add the probabilities of the individual values.  For n=3  p(1≤x≤3)=p(1)+p(2)+p(3)  p(x>1)=p(2)+p(3) |
| Analyze individual values within a population:  Find the mean, variance, and standard deviation for a specific outcome. | * The outcome is a discrete random variable. * There are a fixed number of trials (n). * Each trial has only two possible outcomes. * The probability of success (p) is the same for each trial. * The outcome of one trial does NOT influence the outcome of any other trial. | Binomial distribution  Use the normal distribution to approximate the binomial distribution if both of the following are true:   * (n)(p) 10 * (n)(1-p) 10 | Mean of X is μ=np  Variance of X is  σ2=np(1-p)  Standard deviation of X is σ = |
| Analyze individual values within a population:  Find the probability of a value.  Compare values from different distributions. | * The probability of a random outcome has a normal distribution. * The outcome is a continuous random variable. * The mean and standard deviation are known. | Normal distribution  Standard normal distribution (Z) | Normal distribution depends on its mean and standard deviation.  Any normal distribution can be converted to the Z-distribution.  Z-value represents the number of standard deviations above or below the mean.  z =  x = μ + zσ  Probabilities on Z-table represent the area under the curve to the left of the value. |
| Analyze individual values within a population:  Find the value of X given the percentage or probability of X   * Find the pth percentile for X {Find a where p(X<a)=p} * Find the (1-p)th percentile for X {Find b where p(X>b)=p which is rewritten as p(X<b)=1-p. | * The probability of a random outcome has a normal distribution. * The outcome is a continuous random variable. | Normal distribution  Standard normal distribution (Z) | Normal distribution depends on its mean and standard deviation.  Any normal distribution can be converted to the Z-distribution.  Z-value represents the number of standard deviations above or below the mean.  z =  x = μ + zσ  Probabilities on Z-table represent the area under the curve to the left of the value. |
| Analyze individual values within a population:  Approximate the binomial distribution when the number of trials (n) is large. | * (*n*)(*p*) 10 * (*n*)(1-*p*) 10 | Normal distribution  Standard normal distribution (Z) | μ = np  σ = |
| Analyze a population parameter:  Find the probability of a population parameter given a specific sample size | * The population is fairly close to normal distribution. * Data set is small * The standard deviation of the population is unknown | t-distribution  t-table  Columns=prob  Rows=degr of free | t-distribution has a mean of zero like the z-distribution.  t-distribution with sample size of 30 approximately equal to z-distribution  t-distribution depends on sample size.  Each t-distribution is characterized by its degrees of freedom.  t-table classified by degrees of freedom and probabilities.  t-table values represent greater-than probabilities (right-side tail of the t-distribution). |
| Analyze a population parameter:  Find the percentile of a population parameter given a specific sample size (i.e., the number on a distribution whose less-than probability [left-side tail] is the given percentage) | * The population is fairly close to normal distribution. * Data set is small * The standard deviation of the population is unknown | t-distribution  t-table  Columns=prob  Rows=degr of free  Bottom=confid inter | 1. Calculate degrees of freedom. 2. Determine the greater-than probability of interest (right-side tail). |
| Analyze a population parameter:  Find the confidence interval for population parameter given a confidence interval | * The population is fairly close to normal distribution. * Data set is small * The standard deviation of the population is unknown | t-distribution  t-table  Columns=prob  Rows=degr of free  Bottom=confid inter  n30, use Z-distribution  n<30, use t-distribution | statisticmargin of error  margin of error defined by the critical value  critical value = std errors (see MOE below)  To find critical value:   1. Find confidence interval on bottom row of table 2. Move up to the degrees of freedom for sample size |
| Analyze a population parameter when the distribution is non-normal or unknown | * The sample size is large (n30) * (*n*)(*p*) 10 * (*n*)(1-*p*) 10 | Central Limit Theorem  Approximate with the Normal distribution and the Z-distribution  Confidence interval  Standard error (SE) | Mean of the sampling distribution equals the population proportion:  = *p*  Standard error equals:  = |
| Analyze a population parameter when the distribution is non-normal or unknown:  Find the probability for a sample proportion given a population proportion | * The sample size is large (n30) * (*n*)(*p*) 10 * (*n*)(1-*p*) 10 | Central Limit Theorem  Approximate with the Normal distribution and the Z-distribution  Confidence interval  Standard error (SE) | 1. Confirm constraints 2. Convert sample proportion to z-value 3. Find probability of z |
| Define the confidence interval for the difference of two population means at a given confidence level | * The standard deviation for both populations is known | Confidence interval  Z-distribution | 1 - 2 ± (z\*)( ) |
| Define the confidence interval for the difference of two population means at a given confidence level | * The standard deviation for both populations is NOT known   OR   * The sample size is small (n<30) | Confidence interval  t-distribution | (1 - 2 ) ± ()() |
| Estimate the confidence interval for the difference of two proportions at a given confidence level | * The sample size is large (n30) * (*n*)(*p*) 10 * (*n*)(1-*p*) 10 | Central Limit Theorem  Confidence interval  Z-distribution | (1 - 2) ± (z\*)() |
| Calculate the margin of error (MOE) for a sample proportion, where the sample proportion is the decimal version of the sample percentage | * The sample size is large (n30) * (*n*)(*p*) 10 * (*n*)(1-*p*) 10 | Central Limit Theorem  Approximate with the Normal distribution and the Z-distribution  Confidence interval  Standard error (SE)  z\*-value, which is the standard normal value for a given confidence level. | SE =  MOE = (z\*)(SE) ≈  MOE = (z\*) ( )   1. Specify the sample size, n 2. Specify the sample proportion, 3. Calculate the standard error, SE 4. Determine the  z\*-value for the desired confidence level 5. Multiply the z\*-value and the SE   To determine z\*-value:   1. Determine the desired confidence level, CL 2. Determine the  less-than probability as CL + 3. Look up the less-than probability in the  Z-table |
| Calculate the margin of error (MOE) for a sample mean where the population standard deviation is known | * The original population has a normal distribution   OR   * The sample size is large (n30) * (*n*)(*p*) 10 * (*n*)(1-*p*) 10   AND   * Population standard deviation is known | Central Limit Theorem  Approximate with the Normal distribution and the Z-distribution  Confidence interval  Standard error (SE)  z\*-value, which is the standard normal value for a given confidence level. | SE =  MOE = (z\*)(SE)  MOE = (z\*) ()   1. Specify the sample size, n 2. Determine the population standard deviation, σ 3. Calculate the standard error, SE 4. Determine the  z\*-value for the desired confidence level 5. Multiply the z\*-value and the SE   To determine z\*-value:   1. Determine the desired confidence level, CL 2. Determine the  less-than probability as CL + 3. Look up the less-than probability in the  Z-table |
| Calculate the margin of error (MOE) for a sample mean where the population standard deviation is unknown | * The original population has a normal distribution   AND Either   * The sample size is small (n<30)   OR   * Population standard deviation is unknown | Central Limit Theorem  Approximate with the Normal distribution and the Z-distribution  Confidence interval  Standard error (SE)  z\*-value, which is the standard normal value for a given confidence level. | SE =  MOE = ()(SE)  MOE = () ()   1. Specify the sample size, n 2. Determine the population standard deviation, σ 3. Calculate the standard error, SE 4. Determine the  t\*-value for the desired confidence level 5. Multiply the t\*-value and the SE   To determine t\*-value:   1. Determine the desired confidence level, CL 2. Determine the degrees of freedom 3. Find the CL column in the t-table 4. Move down to the row for the degrees of freedom |
| Calculate the minimum sample size to obtain a desired margin of error (MOE) for a population mean μ | * The population has a normal distribution | Central Limit Theorem  Approximate with the Normal distribution and the Z-distribution  Confidence interval  Standard error (SE)  z\*-value, which is the standard normal value for a given confidence level. | n ≥  Round up n |
| Test whether two nominal or ordinal (i.e., categorical) variables are associated in a population | * Observations are independent * Variables are categorical * Variables are non-negative   For 2 x 2 table   * All expected frequencies are greater than 5   For large data   * No more than 20% of expected frequencies are less than 5 * All expected frequencies are greater than 1 | Chi-square test  Chi-square distribution  Expected frequencies, which are the frequencies we expect if the null hypothesis is true | *X* 2-distribution =  Degrees of freedom df = (i – 1)(j – 1)  Test statistic *X* 2 =  Oi , j ≡ observed value  Ei , j ≡ expected value  i denotes row of contingency table  j denotes column of contingency table  The probability of finding a given association in the sample data with a given number of degrees of freedom if the null hypothesis is true (i.e., there is no association in the population) |
| Test hypothesis about the equality of two population variances.  Test the validity of a multiple regression equation. | * Continuous random variables * Variables are non-negative * Samples are independent * Population approximately normally distributed | F-test  F-distribution  Numerator degrees of freedom, n1-1  Denominator degrees of freedom, n2-1 | F =  F = /  SSE ≡ residual sum of squares  m ≡ number of restrictions  k ≡ no. of independent var  F = , larger variance always in the numerator  To perform F-test:   1. Decide on the confidence level 2. Determine the variance for each sample 3. Divide the larger variance by the smaller variance 4. Determine the degrees of freedom for each sample 5. Select the appropriate table for the confidence level 6. Look up the F-value in the appropriate table 7. If table F-value is less than calculated F-value then reject the null hypothesis |