**Business Statistics – Confidence Intervals**

Point Estimate – Sample Mean

Interval Estimate – Confidence Interval

|  |  |
| --- | --- |
| **Confidence Interval Summary** | |
| Let X~N( m , s2 ).  The Random Variable, X, follows a Normal distribution  with a Population Mean, m ,  and Population Variance, s 2 .  The Sample Mean,`X = SX/n, estimates the Population Mean, m , using a sample of size n.  The Sample Variance, S2 = S(X –`X)2/(n–1) , estimates the Population Variance, s 2 . | |
| The Population Mean, m, can be estimated using a Point Estimate and an Interval Estimate:  Point Estimate. Sample Mean:`X = SX/n  Interval Estimate. Confidence Interval: `X ± Z(1–a/2) ( s / √n ) | |
| The Population Mean, m , and Population Variance, s2 , are Parameters.  Parameters are Constants.  The Sample Mean,`X, and Sample Variance, S2, are Point Estimates.  Point Estimates are Random Variables with a Population Mean and Population Variance.  Thus, `X~N( m , s2/n ).  Where the Population Mean of the Sample Mean,`X, is m .  and the Population Variance of the Sample Mean,`X, is s2/n . | |
| **Confidence Interval Definitions** | |
| Random Variable Distribution | X~N( m , s2 ) |
| Parameter | m |
| Point Estimate of Parameter | `X = SX/n |
| Sample Variance of the Point Estimate | 2X = s2/n |
| Point Estimate Distribution | `X ~N( m , s2 / n ) |
| 1–a = Level of Confidence | 1–a |
| Z(1–a/2) = Random Variate | Z(1–a/2) = norm.s.inv(1 – a/2 ) |
| Sample Standard Deviation of the Point Estimate | X = (s/√n) |
| Confidence Interval | `X ± Z(1–a/2) ( s / √n ) |
| **Confidence Interval Statements**. | |
| A 100(1–a)% ***confidence interval*** for ***parameter***  Example: A 90% confidence interval for m is `X ± Z(0.95) ( s / √n ) | |
| The ***confidence interval*** contains the ***parameter*** with a ***probability*** of 1–a | |
| The ***parameter*** lies in the ***confidence interval*** with a ***confidence*** of 1–a | |
|  | |

|  |  |  |
| --- | --- | --- |
|  | **Example** |  |

|  |
| --- |
| 1. A local department store wants to investigate the mean age of adult customers in one of their marketing areas to help manage its advertising. |

|  |  |
| --- | --- |
| To estimate the population mean age,  a random sample of 40 customers yielded  a sample mean of 36.0 years and  a sample standard deviation of 6.842 years. | n=40  X=36.0  S=6.842 |

However, from previous reports on the demographics of this target market, the store assumes the age follows a normal distribution with a standard deviation of 5 years. [ X~N(  ,  =52 ]

|  |  |  |  |
| --- | --- | --- | --- |
| Normal Distribution. [ X~N(  ,  =52 ] | | | |
|  |  |  |  |
|  |  |  |  |
|  X  X~N(  ,  =52  | | | |

Thus, the information available is [ n=40 ,X=36.0 , s = 5 , S=6.842 ].

To estimate the parameter m, use the point estimate,X=36.0 years. How good is this estimate?

What statements can be made about the estimate of the parameter m?

1. Statement about the Estimate of the parameter m.

|  |  |  |
| --- | --- | --- |
| One way is to use  `X and s | X ±  | In the expression [X ± ],  is the standard deviation of X, notX. The standard deviation  represents the variability of an individual person randomly selected and not a sample mean of individuals randomly selected. |
| 36.0 years ± 5 years |
| (Sample Mean) ± (Standard Deviation) |

2. Statement about the Estimate of the parameter m.

|  |  |  |
| --- | --- | --- |
| One way is to use  `X and s/√n | X ± s/√n | In the expression [X ± s/√n ],  s/√n is the standard deviation of the  sample mean,X=SX/n. The interval around the sample mean is using the standard deviation of the sample mean which represents the point estimate with the variability of the point estimate. |
| 36.0 years ± 5/√40 years |
| (Sample Mean) ±  (Standard Deviation of`X) |

Now, consider the interval, `X ± Z(1–a/2) ( s / √n ), where the interval is a multiple of the standard deviation of the sample mean, X

3. Statement about the Estimate of the parameter m. A confidence interval is defined as

|  |
| --- |
| **X ± Z(1–) ( X )**  **X ± Z(1–) ( s/√n )** |

|  |  |  |
| --- | --- | --- |
|  | **Confidence Intervals** |  |

A ‘Confidence Interval’ can be described as an interval estimate of a parameter. Consider the estimate of the population mean of the normal distribution. The sample mean, X, is the point estimate of the population mean, , which is a population parameter. An estimate is a random variable and a parameter is a constant. The confidence interval will be the interval estimate of the population mean  .

From X~N( /n ) consider the probability, P[ Z)

As an example, let =0.10, the probability expression becomes, P[ Z

|  |  |  |  |
| --- | --- | --- | --- |
| =0.05 |  |  | =0.05 |
|  | 🡨 - - - - (1–a)=0.90 - - - - 🡪 | |  |
| Z  =  +Z Z | | | |
| P[ Z  P[ ZX – sqrt(n))   P[ Zsqrt(n)) X – sqrt(n))  | | | |

|  |
| --- |
| Now we may proceed in two ways: |
| 1. Isolate X in the middle,  P[ + Zsqrt(n)) X sqrt(n))   The interval defines the 5th and 95th percentiles of X,  5th Percentile of X: X0.05 =  + Zsqrt(n))  95th Percentile of X: X0.95 = sqrt(n)) |
| 2. Isolate  in the middle,  P[X – Zsqrt(n)) X –sqrt(n))   The interval around the population mean, , is  X ± Z9sqrt(n))  and is called a “90% Confidence Interval for ”  A “Confidence Interval” is an “Interval Estimate” of .  And the sample mean,X=X/n, is a “Point Estimate” of . |

There are two ways to correctly make statements using a confidence interval.

|  |
| --- |
| A “Confidence Interval” is an estimate which is a random variable. Probability is used for an estimate.  A “Population Mean m” is a parameter which is a constant. Confidence is used for a parameter. |
| The **confidence interval** will contain the parameter with a **probability** of 0.90. |
| The **parameter** will lie in the confidence interval with a **confidence** of 90%. |

|  |  |  |
| --- | --- | --- |
|  | **Confidence Interval Terminology** |  |

|  |
| --- |
| **A 100(1–)% Confidence Interval for  is**  **X ± Z(1–) ( s/√n )** |
| In general,  A Confidence Interval is an interval estimate.  1 –  = Statistical Confidence of Estimate (Reliability)  Error Bound of Estimate (Precision or Accuracy)  Point Estimate ± (Random Variate) \* sqrt{ Var(Point Estimate) }  . . . |

|  |  |  |
| --- | --- | --- |
|  | **Confidence Interval Example** |  |

|  |
| --- |
| A 100(1–)% Confidence Interval for the mean, m.  **X ± Z(1–) ( s/√n )** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| =0.05 | |  |  | =0.05 | |
|  | | 🡨 - - - (1– a)=0.90 - - - 🡪 | |  | |
| Z  =  +Z Z | | | | | |
| Given the information, [ n=40 ,`X=36.0 , s = 5 , S=6.842 ].  Find a 90% Confidence Interval for the mean, m. | | | | | |
| **X ± Z(1–) ( s/√n )** | | | | | |
| `X=36.0 | 1–a=0.9, a/2=0.05, 1–a/2=0.95  Z(1–a/2) = Z0.95 = norm.s.inv(0.95) ≈ 1.645 | | | | ( s/√n ) = ( 5/√40 ) ≈ 0.79 |
| 36 ± (1.645)\*(0.79)  36 ± (1.3)  Point Estimate ± (Error Bound)  ( 34.7 , 37.3 )  The Confidence Interval is “36 ± (1.3)” or “( 34.7 , 37.3 )”  The Point Estimate is 36. The Error Bound is 1.3. The Level of Confidence is 90% | | | | | |
| The ***interval*** **( 34.7 , 37.3 )** contains the parameter  with a ***probability*** of 0.9.  The ***parameter***  lies the interval **( 34.7 , 37.3 )** with a ***confidence*** of 90%. | | | | | |

|  |  |  |
| --- | --- | --- |
|  | **Confidence Intervals using Sample Variance** |  |

|  |  |
| --- | --- |
| For **known s2**, use Normal distribution.  A 100(1–)% Confidence Interval for  is  **`X ± Z(1–a/2) ( s/√n )** | For **unknown s2**, use S2 and t-distribution.  A 100(1–)% Confidence Interval for  is  **X ± t(1–a/2),(n–1) (S/√n )**  **X ± t(1–a/2),(n–1) ( SX )** |
| If population variance, s2 is unknown, then use sample variance, S2 = S(X –`X)2/(n–1)  Since sample variance is being used, use t-distribution instead of the Normal distribution.  The sample variance of the random variable, X, is S2 = S(X –`X)2/(n–1).  The sample variance of the sample mean,`X, is S2X = S2/n . | |

**Example**.

|  |  |  |
| --- | --- | --- |
| Given the information, [ n=40 ,`X=36.0 , S=6.842 ]. (Notice s2 is unknown)  Find a 90% Confidence Interval for the mean, m. | | |
| **X ± t(1–a/2),(n–1) (S/√n )** | | |
| `X=36.0 | 1–a=0.9, a/2=0.05, 1–a/2=0.95  n=40, n–1=39  t(1–a/2),(n–1)= t.inv(0.95,39) ≈ 1.685 | S = 6.842  S/√n = 6.842/√40 ≈ 1.082 |
| 36 ± (1.685)\*(1.082)  36 ± (1.82)  Point Estimate ± (Error Bound)  ( 34.18 , 37.82 ) | | |
| The ***interval*** **( 34.18 , 37.82 )** contains the parameter  with a ***probability*** of 0.9.  The ***parameter***  lies the interval **( 34.18 , 37.82 )** with a ***confidence*** of 90%. | | |

|  |  |  |
| --- | --- | --- |
|  | **Numerical Examples: Confidence Intervals** |  |

|  |
| --- |
| **A 100(1–)% Confidence Interval for  is**  **X ± Z(1–) ( s/√n )** |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Sample**  **Mean** | **Level of**  **Confidence** | **Z**  **Variate** | **Standard**  **Deviation** | **Sample**  **Size** | **Error**  **Bound** | **Lower**  **Limit** | **Upper**  **Limit** |
| **Example** | **`X** | **1 – a** | **Z(1–a/2)** | **s** | **n** | **Z\*(s/√n)** |  |  |
| **1** | 36 | 0.9 | 1.644854 | 5 | 40 | 1.300371 | 34.70 | 37.30 |
| **2** | 30 | 0.9 | 1.644854 | 4 | 43 | 1.003351 | 29.00 | 31.00 |
| **3** | 39 | 0.9 | 1.644854 | 5 | 40 | 1.300371 | 37.70 | 40.30 |
| **4** | 28 | 0.95 | 1.959964 | 3 | 24 | 1.200228 | 26.80 | 29.20 |
| **5** | 45 | 0.95 | 1.959964 | 6 | 19 | 2.697879 | 42.30 | 47.70 |
| **6** | 62 | 0.8 | 1.281552 | 7 | 28 | 1.695333 | 60.30 | 63.70 |
| **7** | 71 | 0.8 | 1.281552 | 4 | 41 | 0.800579 | 70.20 | 71.80 |

|  |
| --- |
| **A 100(1–)% Confidence Interval for  is**  **X ± t(1–a/2),(n–1) (S/√n )** |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Sample**  **Mean** | **Level of**  **Confidence** | **t**  **Variate** | **Standard**  **Deviation** | **Sample**  **Size** | **Error**  **Bound** | **Lower**  **Limit** | **Upper**  **Limit** |
| **Example** | **`X** | **1 – a** | **t(1–a/2),(n-1)** | **S** | **n** | **t\*(S/√n)** |  |  |
| **1** | 36 | 0.9 | 1.684875 | 5 | 40 | 1.332011 | 34.67 | 37.33 |
| **2** | 30 | 0.9 | 1.681952 | 4 | 43 | 1.025981 | 28.97 | 31.03 |
| **3** | 39 | 0.9 | 1.684875 | 5 | 40 | 1.332011 | 37.67 | 40.33 |
| **4** | 28 | 0.95 | 2.068658 | 3 | 24 | 1.266789 | 26.73 | 29.27 |
| **5** | 45 | 0.95 | 2.100922 | 6 | 19 | 2.891907 | 42.11 | 47.89 |
| **6** | 62 | 0.8 | 1.313703 | 7 | 28 | 1.737866 | 60.26 | 63.74 |
| **7** | 71 | 0.8 | 1.303077 | 4 | 41 | 0.814026 | 70.19 | 71.81 |

|  |  |  |
| --- | --- | --- |
|  | **Confidence Intervals for Population Variance** |  |

Population Variance, s2 = S( X – m )2 / n

Sample Variance, S2 = S( X –`X )2 / (n – 1)

If X~N( m , s2 ),

then the expression, (n – 1)S2/s2 follows a chi-squared distribution, c2(n–1) , with (n–1) degrees of freedom.

Thus, the probability may be expressed,

P[ c2(1 – a/2) ≤ (n – 1)S2/s2 ≤ c2( a/2) ] = (1 – a )

Isolate the parameter, s2 , in the middle yields,

P[ (n – 1)S2/ c2( a/2) ≤ s2 ≤ (n – 1)S2/ c2(1 – a/2) ] = (1 – a )

Which describes a confidence interval.

|  |
| --- |
| **A 100(1–)% Confidence Interval for s2 is**  ( (n – 1)S2/ c2( a/2) , (n – 1)S2/ c2(1 – a/2) ) |

Given the values from a random sample [ X=36.0, S=6.842, n=40 ]

The 90% confidence interval for the variance becomes

|  |
| --- |
| Find a 100(1–)% Confidence Interval for the variance, s2  ( (n – 1)S2/ c2( a/2) , (n – 1)S2/ c2(1 – a/2) ) |
| **S2 = 6.8422**  **1 – a = 0.9**  **a/2 = 0.05**  **n = 40**  **degrees of freedom = n – 1 = 39**  **c2(0.05) = chisq.inv(0.05,39) ≈ 25.695**  **c2(0.95) = chisq.inv.rt(0.05,39) ≈ 54.572** |
| **39\*6.8422/54.572 , 39\*6.8422/25.695**  **( 33.45 , 71.05 )** |
| The ***interval*** **( 33.45 , 71.05 )** contains the parameter s2 with a ***probability*** of 0.9.  The ***parameter*** s2 lies the interval **( 33.45 , 71.05 )** with a ***confidence*** of 90%. |

|  |  |  |
| --- | --- | --- |
|  | **Central Limit Theorem** |  |

Let the random variable, X, follow any probability distribution.

|  |
| --- |
| **Central Limit Theorem** |
| For sufficiently large n, the random variable, X , will approximately follow a Normal Probability Distribution. |

This also applies to any linear function of X such as X = X/n .

Specifically, if X has mean  and variance 2,

then X will have mean of  and a variance of 2/n.

And by CLT,X ~N(  , 2/n ).

|  |
| --- |
| **Transformation to Standard Normal Distribution** |
| Now for the random variable, X, with mean  and variance 2,  consider the transformation, Z = ( X –  ) / sqrt(2 ). [Note “sqrt” represents “square root”.]  Or, for the random variable, X, with mean , and variance 2/n ,  Z = ( X –  ) / ( sqrt(2/n ) = ( X –  ) / ( /sqrt(n) )  The mean of Z is 0 and the variance of Z is 1.  The random variable, Z ~ N( 0 , 1 ), is called the Standard Normal Distribution. |

|  |
| --- |
| **Procedure** |
| Any Random Variable  🡪 Random Sample of sufficient sample size  🡪 by Central Limit Theorem Sum is Normal  🡪 Transformation to Standard Normal  🡪 Use Standard Normal Tables for calculations |

|  |
| --- |
| For a sufficiently large sample size, n, and a mean and variance exists,  the following holds by the CLT. |
| If X~B(n1,p) and E[X]= and Var[X]=2, then X~N(,2/n) whereX=X/n.  If X~Pssn() and E[X]= and Var[X]=2, then X~N(,2/n) whereX=X/n.  If X~P(X) and E[X]= and Var[X]=2, then X~N(,2/n) whereX=X/n.  etc. for other probability distributions. |