|  |
| --- |
| **UNIVERSIDADE FEDERAL DE LAVRAS**  PRÓ-REITORIA DE PÓS-GRADUAÇÃO  INSTITUTO DE CIÊNCIAS NATURAIS  DEPARTAMENTO DE BIOLOGIA  PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E  MELHORAMENTO DE PLANTAS |



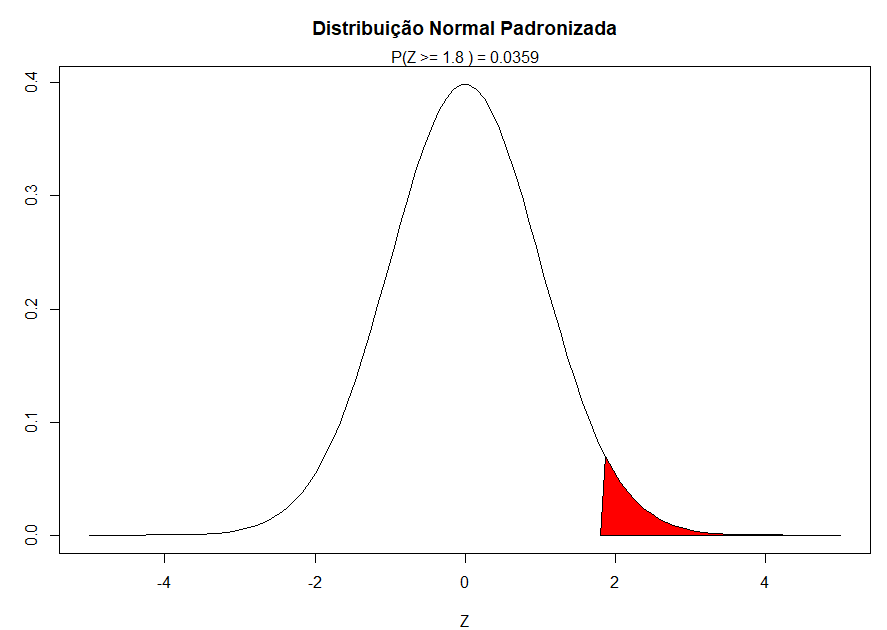
**Ricardo Antonio Ruiz Cardozo**

**PGM522 – ANÁLISE DE EXPERIMENTOS EM GENÉTICA E MELHORAMENTO DE PLANTAS**

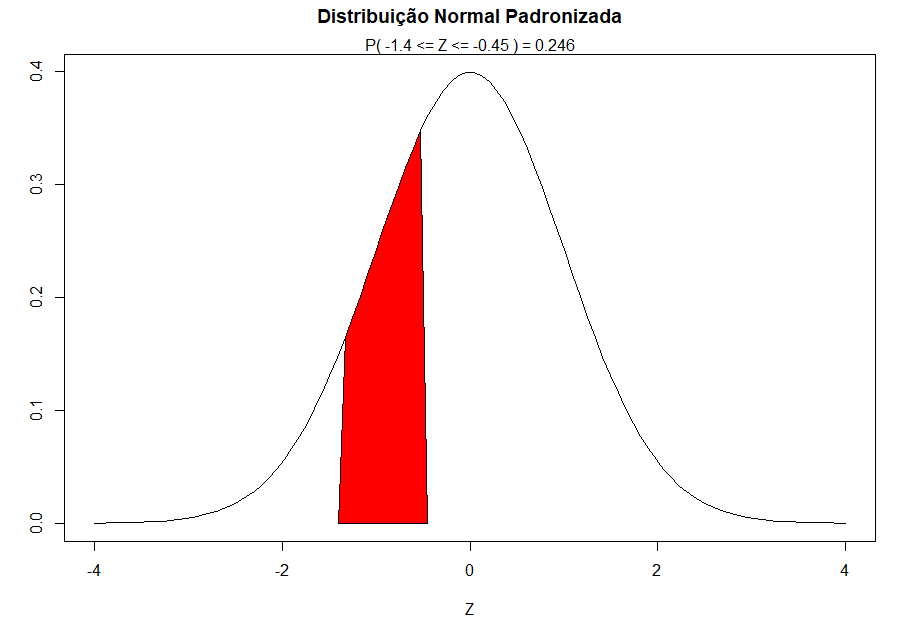
# 1ª LISTA DE EXERCÍCIOS

## Distribuição Normal e Estimação de parâmetros

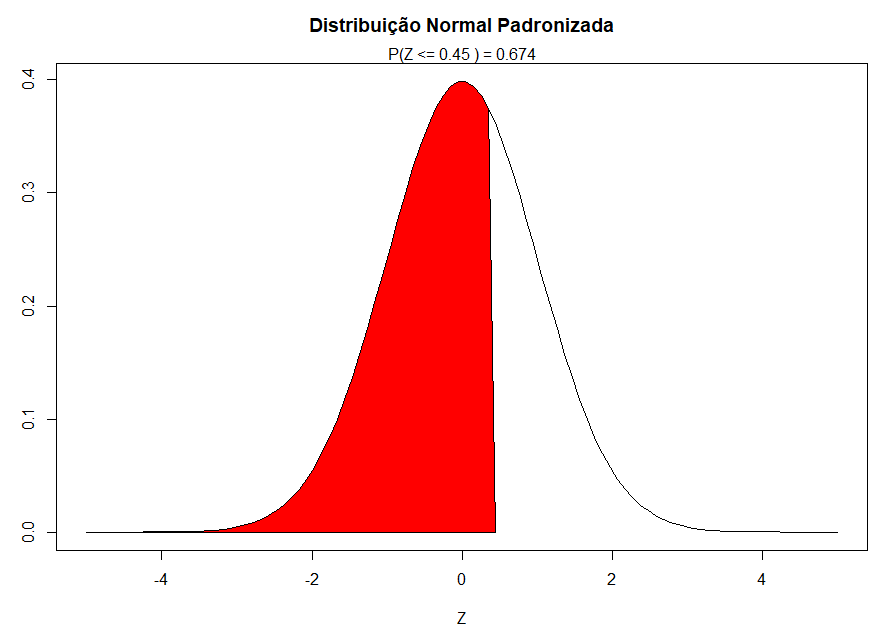
1. Seja Z uma variável normal padronizada [Z ~ N(0,1)]. Determine as seguintes probabilidades e as represente em gráficos:
   1. P(Z ≥ 1,8) = **0.0359**



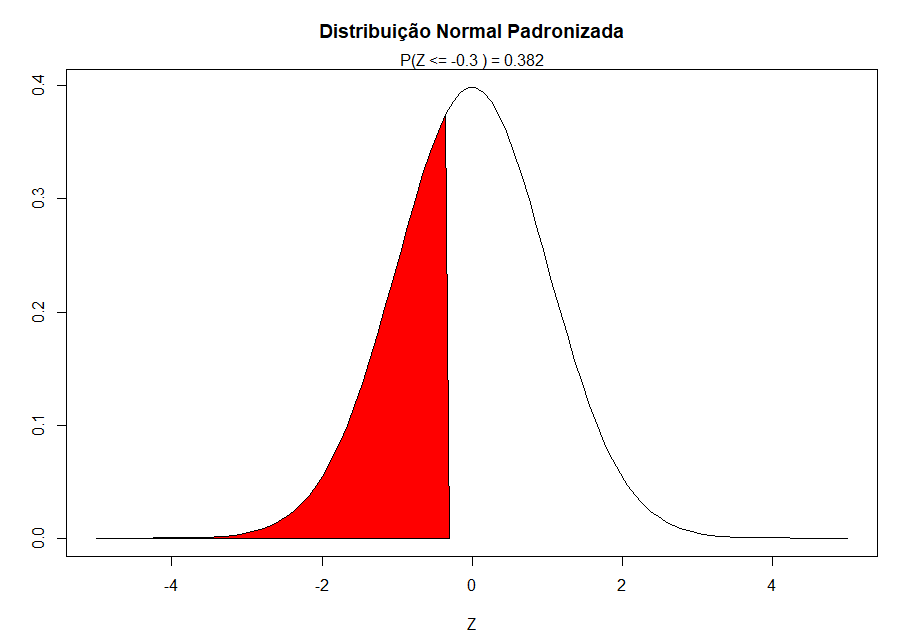
* 1. P(-1,4 ≤ Z ≤ -0,45) = **0.246**



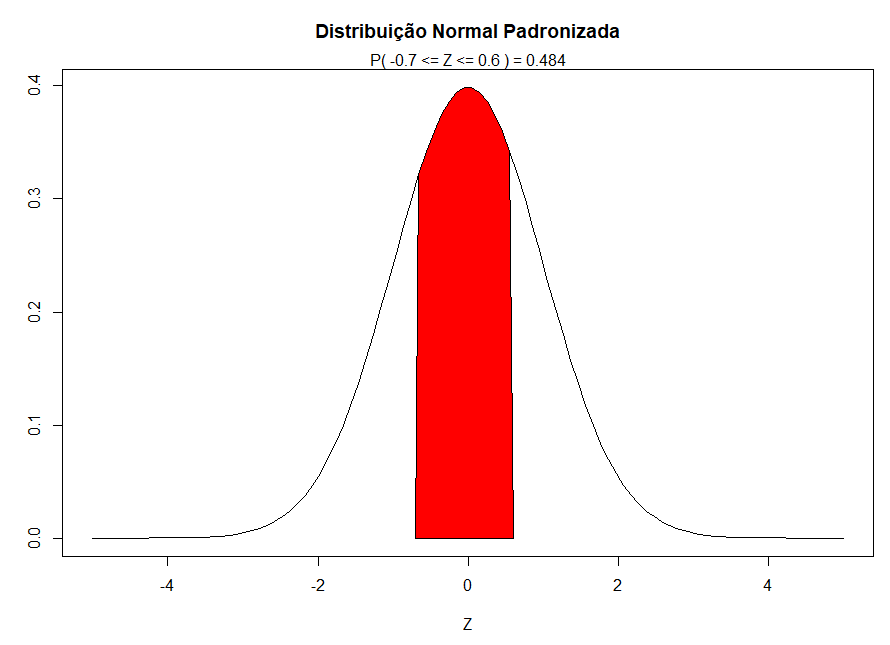
* 1. P(Z ≤ 0,45) = **0.674**



* 1. P(Z ≤ -0,3) = **0.382**



* 1. P(-0,7 ≤ Z ≤ 0,6) = **0.484**



1. Encontre os valores de z da distribuição N(0,1), tais que:
   1. P(Z > z) = 0,9798 🡪 ***z = -2.049636***
   2. P(Z < z) = 0,063 🡪 ***z = -1.530068***
   3. P(1< Z < z) = 0,10 🡪 ***z = 1.566163***
   4. P(-1,5 < Z < z)= 0,30 🡪 ***z = -0.3403215***
2. Foi realizado um estudo sobre a altura de plantas de trigo em Latossolo Vermelho do Cerrado com adubação fosfatada. Observou-se que este caráter se distribui normalmente com média 1,70 m e variância de 400 cm2. Calcule a probabilidade da altura (X) de uma planta sorteada desse estudo esteja de acordo com os seguintes eventos:
   1. X > 1,90 m; 🡪 **Probabilidade (Prob.) = 0.1586553; o 15.87% de probabilidade que uma planta seja maior a 1.90 m**
   2. 1,90 < X < 2,10 m; 🡪 **Prob. = 0.1359051; o 13.59% de probabilidade que uma planta esteja entre 1.90 e 2.10 m**
   3. X < 1,50 m; 🡪 **Prob. =** **0.1586553; o 15.87% de probabilidade que uma planta seja menor que 1.50 m**
   4. A partir de que altura (m) se encontra 30% das plantas mais altas?

* P (Z > z) = 0.3 🡪 **A partir da altura X = 1.80488 m se encontram o 30% das plantas mais altas.**
  1. A partir de que altura (m) se encontra 40% das plantas mais baixas?
* P (Z > z) = 0.3 🡪 **A partir da altura X = 1.649331 m se encontram o 40% das plantas mais baixas.**
  1. Suponha que as plantas de trigo apresentem a seguinte classificação:

Classe E – 10% mais baixas 🡪 P (Z <= z) = 0.1; a altura do 10% das plantas mais baixas é até 1.44369 m

Classe D – 30% seguintes à classe E 🡪 P (Z <= z) = 0.4; a classe D com o 40% das plantas, são plantas com altura até 1.649331 m.

Classe C – 15% seguintes à classe D 🡪 P (Z <= z) = 0.55; a classe C com o 55% das plantas, são plantas com altura até 1.725132 m

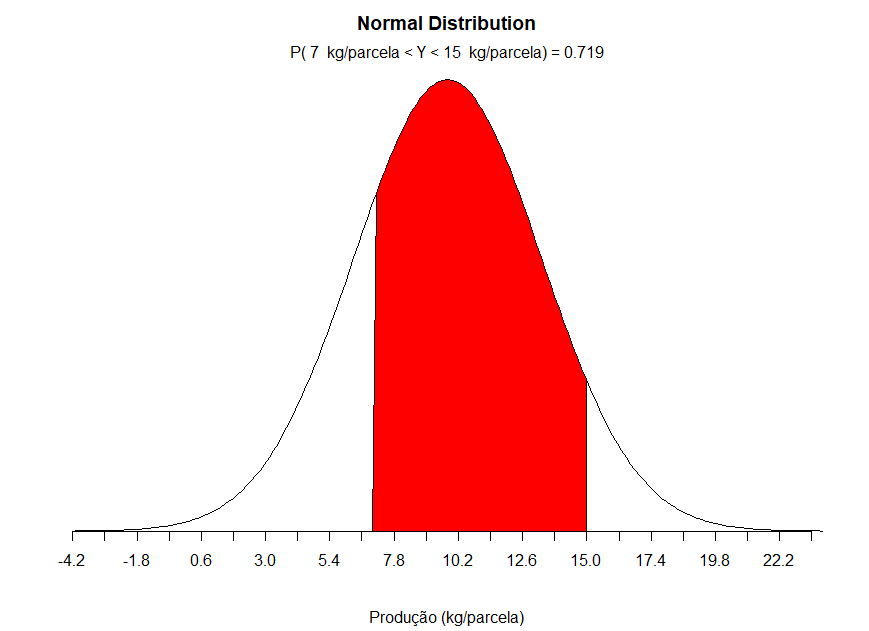
Classe B – 25% seguintes à classe C 🡪 P (Z <=z) = 0.8; a classe B com o 80% das plantas, são plantas com altura até 1.868324 m

Classe A – 20% mais altas 🡪 P (Z >z) = 0.2; a classe A com o 20% das plantas restantes, são plantas com altura maior de 1.868324 m

Com base nesta classificação, determine os limites de altura das plantas de cada classe.

1. Em um estudo quanto à produção de grãos (kg/parcela) de uma amostra de linhagens de feijão, observou-se que a produtividade é uma variável normalmente distribuída com média μ=9,8 (kg/parcela) e desvio padrão σ=3,5 (kg/parcela). Num programa de melhoramento, entre outras características, uma cultivar deve satisfazer a condição 7,0 < x < 15,0 kg/parcela. Nessas condições, tendo-se 169 linhagens de feijão, pergunta-se:
   1. Qual a proporção de linhagens que deverá ser aceita?

O 71,9% das linhagens deverão ser aceitas



* 1. Qual o número esperado de linhagens que continuará participando do programa de melhoramento?
  2. Qual o número esperado de linhagens cuja produção por parcela é superior a 14,0 kg?

1. Suponha que as medidas dos grãos de pólen de *Euterpe oleracea* (açaizeiro) em vista equatorial, em µm, seja uma variável normalmente distribuída (Dados adaptados da dissertação de Oliveira, 2011). Assume-se que o comprimento do colpo tenha média 96,60 µm e desvio padrão de 12,00 µm, e que a largura do colpo tenha média 1,23 µm e desvio padrão de 0,30 µm. Qual a probabilidade de sortear um grão de pólen com:
   1. comprimento do colpo maior que 97,20 µm e largura do colpo menor que 1,19 µm?

Há uma probabilidade do 21,46% de ter grãos com comprimento maior que 97,20 µm e largura do colpo menor que 1,19 µm.

* 1. comprimento do colpo menor que 95,60 µm ou largura do colpo maior que 1,26 µm?

Há uma probabilidade do 92,70% de ter grãos com comprimento menor que 95,60 µm ou largura do colpo maior que 1,26 µm.

1. Os dados a seguir referem-se à produção de grãos, em g/planta, obtidos numa amostra de 20 plantas de feijão da geração F2 do cruzamento das cultivares Flor de Maio e Carioca.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1,38 | 4,14 | 6,23 | 12,13 | 17,12 |
| 3,65 | 4,54 | 6,79 | 12,56 | 19,68 |
| 3,72 | 5,64 | 8,21 | 13,19 | 21,26 |
| 3,87 | 5,67 | 9,79 | 15,60 | 24,57 |

Calcule:

1. = **199.74 g/planta**
2. = **2852.213 g2/planta**
3. Média amostral: = **9.987 g/planta**
4. Soma dos desvios em relação à média: = **0 g/planta**
5. Calcule a soma de quadrados de desvios em relação à média: = **857.4096 (g/planta)2**
6. Variância amostral: = **45.12682 (g/planta)2**
7. Desvio padrão amostral: **= 6.71765** **g/planta**
8. Somar aos dados da amostra a constante k (k=10) e calcular novamente a média amostral, a variância amostral, o desvio padrão amostral e o coeficiente de variação para o novo conjunto de dados. Compare os resultados obtidos com os da amostra original.

|  |  |  |
| --- | --- | --- |
| **Estimador** | **Original** | **Somando a constante k = 10** |
| Média Amostral (g/planta) | 9,987 | 19,987 |
| Variância Amostral (g2/planta) | 45,127 | 45,127 |
| Desvio Padrão (g/planta) | 6,718 | 6,718 |
| Coeficiente de variação (%) | 67, 263 | 33,610 |

Gráfico, Gráfico de líneas

Descripción generada automáticamente Gráfico, Gráfico de líneas, Histograma

Descripción generada automáticamente

Pode se observar que ao somar a constante k o único valor que mudou foi a média, sendo maior, fazendo que o coeficiente de variação seja menor que os dados originais

1. Multiplicar os dados da amostra por uma constante k (considerar k=10) e calcular novamente a média amostral, a variância amostral, o desvio padrão amostral e o coeficiente de variação para o novo conjunto de dados. Compare os resultados obtidos com os da amostra original.

|  |  |  |
| --- | --- | --- |
| **Estimador** | **Original** | **Somando a constante k = 10** |
| Média Amostral (g/planta) | 9,987 | 99,87 |
| Variância Amostral (g2/planta) | 45,127 | 4512,682 |
| Desvio Padrão (g/planta) | 6,718 | 67,176 |
| Coeficiente de variação (%) | 67, 263 | 67,263 |

Gráfico, Gráfico de líneas

Descripción generada automáticamenteGráfico, Gráfico de líneas, Histograma

Descripción generada automáticamente

Pode se observar que ao multiplicar a constante k os valores da média, variância e desvio padrão mudaram, mas mantem o coeficiente de variação

1. Qual é a diferença entre parâmetro, estimador e estimativa?

* **Parâmetro** é uma característica da população, e podem ser utilizados vários

estimadores para medir essa característica.

* **Estimador,** é a metodologia ou uma estatística utilizada para estimar determinado parâmetro, como por exemplo, média, mediana, moda.
* **Estimativa** é referente a uma aproximação do parâmetro de uma população

através de dados coletados de amostras dessa população. Entre outras palavras é um valor obtido a partir de um estimador, que pode se aproximar ao parâmetro.

1. A partir dos dados a seguir da circunferência à altura do peito (cm) de 30 árvores de candeia de uma população, determine:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 20,00 | 9,20 | 12,00 | 14,40 | 12,50 |
| 5,80 | 10,80 | 15,70 | 21,50 | 13,20 |
| 12,80 | 14,80 | 15,80 | 22,20 | 11,20 |
| 8,70 | 12,00 | 20,60 | 17,50 | 8,80 |
| 19,00 | 15,40 | 18,50 | 13,30 | 13,80 |
| 19,30 | 17,60 | 23,20 | 8,20 | 11,40 |

1. Calcule o intervalo de confiança da média a 95% de probabilidade. Intérprete.

|  |  |
| --- | --- |
| **Limite inferior (2.5%)** | **Limite Superior (97.5%)** |
| 12.9421 cm | 16,3379 cm |

Os valores do intervalo de confiança da média indicam que o 95% dos valores da média amostral podem se encontrar em esse intervalo, sabendo que a média foi de 14,64 cm

1. Calcule o intervalo de confiança da variância a 95% de probabilidade. Intérprete.

|  |  |
| --- | --- |
| **Limite inferior (2.5%)** | **Limite Superior (97.5%)** |
| 13.11378 cm2 | 37.36457 cm2 |

É esperado que em 95% dos casos de uma **n** amostras a variância esteja dentro desse intervalo, sabendo que a variância foi de 20,68 cm2

1. Qual é o significado do nível de confiança?

O nível de confiança significa que se sortearmos uma quantidade de valores aleatórios de dados, de acordo com uma média e o desvio padrão amostral, é esperado que os valores fiquem dentro de um intervalo de confiança estipulado pelo rigor do pesquisador ou pessoa. A exemplo, a um nível de confiança de 95%, se sorteamos 100 valores aleatórios é esperado que 95 destes valores fiquem dentro do intervalo de confiança. Pelo tanto pode determinar a probabilidade de aceitar uma hipótese corretamente.

1. Represente os dados num histograma.

Gráfico, Histograma

Descripción generada automáticamente

1. Verifique a normalidade dos dados pelo método do “Q – Q plot”. Calcule o coeficiente de correlação associado conforme Johnson e Wichern (1998) e faça a inferência a 5% de probabilidade.

Gráfico, Gráfico de líneas

Descripción generada automáticamente

O coeficiente de correlação foi de **0,9917**, pelo tanto, os dados têm distribuição normal.

De acordo com o teste Q-Q plot sob um nível de significância de 5%, os dados estão em normalidade, pois a correlação calculada 0,9917 é maior que a correlação tabelada 0,964.

1. Verifique a normalidade dos dados pelo teste de Shapiro-Wilk a 5% de probabilidade.

|  |  |
| --- | --- |
| **Shapiro-Wilk normality test** | |
| W = 0,97786 | *p-value* = 0,7664 |

Sob o nível de 5% de significância no teste de Shapiro-Wilk, os dados possuem se encontram com uma distribuição normal, pois o p-value é maior que o α = 5%.

1. Qual importância de se verificar a normalidade dos dados?

* É importante para verificar se a distribuição de probabilidade de um conjunto de dados possui uma distribuição normal. Sendo essa um pressuposto para poder utilizar diferentes testes estatísticos de maneira adequadas.

1. Em um estudo quanto à produção, em t/ha, de variedades de batata, tomaram-se os seguintes valores:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 9,2 | 15,4 | 23,1 | 27,0 | 18,0 | 24,6 | 24,2 | 20,0 | 9,2 | 12,3 |
| 21,1 | 12,7 | 18,0 | 29,9 | 21,1 | 11,0 | 10,1 | 26,4 | 25,7 | 17,1 |
| 22,6 | 20,0 | 13,4 | 11,9 | 24,2 | 26,4 | 18,2 | 24,0 | 25,1 | 28,0 |

Usando o programa R, pede-se:

* 1. Verifique a normalidade dos dados pelo procedimento Q-Q plot e pelo teste de Shapiro-Wilk a 5% de probabilidade.

Gráfico, Gráfico de líneas

Descripción generada automáticamente

De acordo com o teste Q-Q plot sob um nível de significância de 5%, os dados

estão em normalidade, pois a correlação calculada 0,975 é maior que a

correlação tabelada 0,964.

|  |  |
| --- | --- |
| **Shapiro-Wilk normality test** | |
| W = 0,94012 | *p-value* = 0,09166 |

Sob o nível de 5% de significância no teste de Shapiro-Wilk, os dados estão com a distribuição normal, pois o p-value é maior que o α = 5%.

* 1. Determine o intervalo para a média da produção com um grau de confiança de 99%. Interprete o resultado.
* Média=19,6633 t/ha
* Variância = 38,417575 (t/ha)2
* Erro padrão (SE mean) = 1,131630 t/ha
* t = 17.376, df (graus de liberdade) = 29, p-value < 2.2e-16

|  |  |
| --- | --- |
| **Limite inferior** | **Limite Superior** |
| 16.5441 t/ha | 22.7825 t/ha |

Ao nível de 99% de confiança estima-se que a média da produção de batata está entre o intervalo de 16,54 e 22,78. A média foi de 19,66 t/ha encontrando-se no intervalo de confiança de 99%

* 1. Determine o intervalo de confiança de 99% para a variância da produção. Interprete o resultado.

|  |  |
| --- | --- |
| **Limite inferior (0.5%)** | **Limite Superior (99.5%)** |
| 21,2878 | 84,9095 |

Nesses resultados, ao nível de 99% de confiança estima-se que a variância

populacional da produção de batata está entre o intervalo de 21,29 e 84,91. A variância é de 38,42 encontrando-se nesse intervalo

* 1. Caso tivesse escolhido um nível de significância de 5%, qual seria a implicação na amplitude do intervalo confiança dos parâmetros.

Intervalo de confiança de 95% para a média

|  |  |
| --- | --- |
| **Limite inferior** | **Limite Superior** |
| 17.348891 | 21.977776 |

Intervalo de confiança de 95% para a variância

|  |  |
| --- | --- |
| **Limite inferior (2.5%)** | **Limite Superior (97.5%)** |
| 24,36688 | 69,42760 |

Ao nível de 95% de confiança os intervalos se estreitam mais em relação aos intervalos obtidos sob o nível de 99% de confiança.