Distribuição Amostral

Intervalo de Confiança

Reinaldo Madarazo - 2015

Distribuição de Amostragens da Média Aritmética

- Corresponde à distribuição das médias aritméticas de todas as amostras possíveis, caso selecione todas as amostras possíveis de um determinado tamanho.
- A média aritmética é isenta de viés, uma vez que a média aritmética de todas as possíveis médias aritméticas da amostra (de tamanho n) é igual a média aritmética da população, μ.

Erro-Padrão da Média Aritmética

- É o valor do desvio-padrão de todas as possíveis médias aritméticas de amostras de tamanho *n*.
- O erro-padrão pode ser obtido pela seguinte expressão:

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

Onde σ é o desvio-padrão da população.

Exemplo

- A Oxford Cereals abastece milhares de caixas de cereais durante um turno de 8 horas. O Gerente de Operações é também responsável por monitorar a quantidade de cereal colocada em cada caixa. Para ser coerente com conteúdo especificado na embalagem, as caixas devem conter uma média aritmética de 368 gramas de cereal. Em razão da velocidade do processo, o peso do cereal varia de caixa para caixa, fazendo com que algumas caixas fiquem mal abastecidas enquanto outras ficam hiperabastecidas. Se o processo não estiver funcionando de modo apropriado, o peso médio das caixas pode se desviar demasiadamente do peso especificado no rótulo, 368 gramas, e se tornar inaceitável.
- Como a pesagem de cada caixa individual consome muito tempo, é dispendiosa e ineficiente, deve-se extrair uma amostra de caixas e para cada amostra selecionada, pesar as caixas e calcular a média aritmética para a amostra. Essa média aritmética das amostras se aproxima ou não da média aritmética da população, 368 gramas? Com base na análise, deve-se manter, alterar ou interromper o processo de embalagem?

Exemplo

- Se for selecionada uma amostra de 25 caixas sem reposição dos milhares de caixas abastecidas durante um determinado turno, a amostra contém menos de 5% da população. Considerando que o desvio-padrão do processo de abastecimento de cereais seja de 15 gramas, calcule o erro-padrão da média aritmética:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{25}} = \frac{15}{5} = 3$$

- A variação nas médias aritméticas de amostras com tamanho n=25 é bem menor do que a variação nas caixas individuais de cereais (3 para 15).

Amostragem de Populações Distribuídas nos Moldes da Distribuição Normal

- A amostra age como uma representação em miniatura da população, de modo que, se os valores na população forem distribuídos nos moldes da distribuição normal, os valores na amostra devem ser distribuídos aproximadamente nos moldes da distribuição normal.
- Utilizando o exemplo dado, se a média aritmética da população for 368 gramas, a média aritmética da amostra possui uma boa chance de estar próxima de 368 gramas.

Probabilidades

- Pode-se determinar a probabilidade de que uma amostra de 25 caixas venha a ter uma média aritmética inferior a 365 gramas (a média da população é de 368 gramas)?
- Na distribuição normal, é possível determinar probabilidades encontrando a área abaixo de qualquer valor de X fazendo a conversão para valores Z padronizados:

$$Z = \frac{X - \mu}{\sigma}$$

Probabilidades

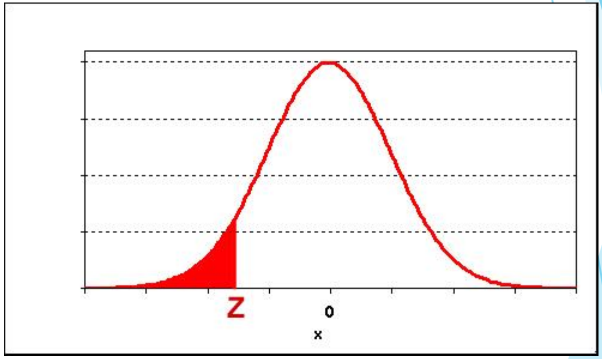
- Para determinar o valor de Z em uma distribuição de amostragens da Média aritmética deve-se substituir $X \; por \; \bar{X}, \mu \; por \; \mu_{\bar{x}} \; e \; \sigma \; por \; \sigma_{\bar{x}}$:

$$Z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Exemplo

- Qual a probabilidade da média aritmética de uma amostra de 25 caixas, seja menor do que 365 gramas?

$$Z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{365 - 368}{\frac{15}{\sqrt{25}}} = \frac{-3}{3} = -1{,}00$$

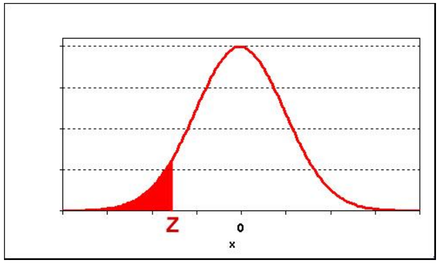


- Usando a tabela da distribuição normal, podemos encontrar a área correspondente a Z=-1,00, que vale 0,1587.
- Assim, 15,87% de todas as amostras possíveis de 25 caixas apresentam uma média aritmética de amostra inferior a 365 gramas.

Outro exemplo

- Qual a percentagem de caixas individuais que terão menos de 365 gramas de cereal?
- Este caso é diferente do anterior. Antes, determinou-se a probabilidade para uma distribuição amostral, agora a probabilidade será obtida pela população:

$$Z = \frac{X - \mu}{\sigma} = \frac{365 - 368}{15} = \frac{-3}{15} = -0{,}20$$



- Pela tabela, a área correspondente a Z=-0,20 vale 0,4207. Portanto, espera-se que 42,07% das caixas individuais contenham menos do que 365 gramas.

Amostras x Individuais

- Os resultados anteriores são explicados pelo fato de que cada amostra consiste em 25 valores diferentes, alguns pequenos e outros grandes. O processo de obtenção da média aritmética dilui a importância de qualquer valor individual, particularmente quando o tamanho da amostra é grande. Por conseguinte, a chance de que a média aritmética de 25 caixas esteja bem distante da média aritmética da população e menor que a chance de uma única caixa venha a estar distante de média aritmética.

Exercício

- De que maneira o erro-padrão da média aritmética é afetado pelo crescimento do tamanho da amostra de 25 para 100 caixas?

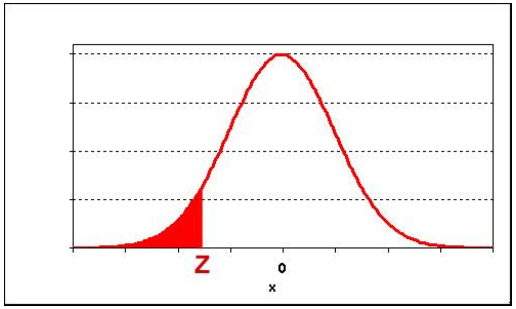
$$n = 100\ caixas.\, Então\ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{100}} = \frac{15}{10} = 1{,}5$$

- O crescimento de quatro vezes no tamanho da amostra, de 25 para 100, reduz o erro padrão da média aritmética pela metade, de 3 gramas para 1,5 gramas. Isso demonstra que adotar uma amostra maior resulta em menor variabilidade nas médias aritméticas das amostras.

Exercício

- Se for selecionada uma amostra de 100 caixas, qual é a probabilidade de que a média aritmética da amostra seja inferior a 365 gramas?

$$Z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{365 - 368}{\frac{15}{\sqrt{100}}} = \frac{-3}{1,5} = -2.00$$



- Com base na tabela da distribuição normal, a área correspondente a menos de Z=-2,00 vale 0,0228. Por conseguinte, 2,28% das amostras de 100 caixas apresentam médias aritméticas inferiores a 365 gramas, comparada a 15,87% para amostras de 25 caixas.

Encontrando a média

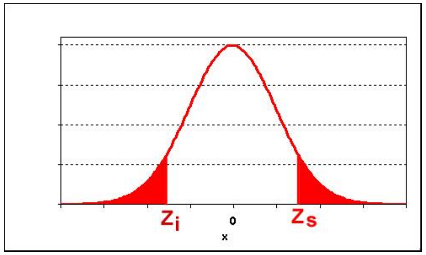
- Algumas vezes precisamos encontrar o intervalo que contém uma proporção fixa de médias aritméticas de amostras. Precisa-se determinar uma distância acima e abaixo de média aritmética da população que contenha uma área específica da curva normal.
- Para isso utilizamos:

$$\bar{X} = \mu + Z\frac{\sigma}{\sqrt{n}}$$

Exemplo:

- No exemplo do abastecimento de cereais, encontre um intervalo simetricamente distribuído em torno da média aritmética da população que inclua 95% das médias aritméticas de amostras com base em amostras de 25 caixas.

- Se 95% das médias aritméticas de amostras estão contidas no intervalo, então 5% estão fora, sendo 2,5% na cauda inferior e 2,5% na cauda superior. O valor de Z na tabela correspondente a uma área de 0,0250 na cauda inferior é de -1,96 e o valor de Z correspondente a uma área de 0,0250 na cauda superior é de +1,96.



$$\overline{X_1} = 368 + (-1{,}96)\frac{15}{\sqrt{25}} = 368 - 5{,}88 = 362{,}12$$

$$\overline{X_2} = 368 + (+1{,}96)\frac{15}{\sqrt{25}} = 368 + 5{,}88 = 373{,}88$$

- Por conseguinte, 95% de todas as médias aritméticas de amostras de 25 caixas estão entre 362,12 e 373,88 gramas.

Teorema do Limite Central

- O Teorema do Limite Central afirma que, *à medida que o tamanho da amostra (ou seja, o número de valores em cada amostra)* ***fica grande o suficiente****, a distribuição de amostragem da média aritmética passa a ser distribuída aproximadamente nos moldes da distribuição normal. Isso é verdadeiro independentemente do formato da distribuição dos valores individuais na população.*
- O que "grande o suficiente"? Para a maior parte das distribuições de população, independentemente do formato, a distribuição de amostragens da média aritmética será distribuída aproximadamente nos moldes da distribuição normal se forem selecionadas amostras com tamanho de pelo menos igual a 30.

Intervalo de Confiança

- Nos exemplos vistos anteriormente, usávamos o valor da média aritmética da população (μ) para estimar o intervalo simetricamente distribuído em torno de μ que continham a média mínima e máxima com uma dada porcentagem de ocorrência, ou seja, no exemplo anterior, 95% das caixas da amostra tinham média de peso compreendidas entre 362,12g e 373,88g.
- Mas em geral, o valor da média aritmética da população (μ) é desconhecido.
- Dessa forma, podemos utilizar o valor da média aritmética de uma amostra para estimar o valor da média aritmética da população. Dizemos que ***$\bar{X}$ é um estimador de μ***.
- Assim, se conhecermos o valor de $\bar{X}$, é também possível estimar um intervalo simétrico em torno de $\bar{X}$ onde podemos ter o valor da média aritmética da população μ.

Intervalo de Confiança

- Para determinar esse intervalo, basta substituir o valor de μ pelo valor de $\bar{X}$ na equação:

$$\bar{X} = \mu + Z\frac{\sigma}{\sqrt{n}}$$

- Obtendo:

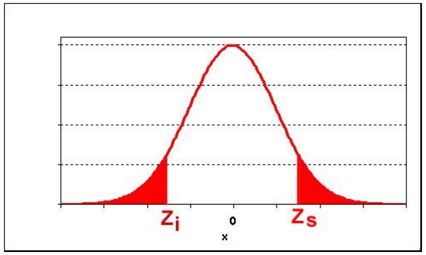
$$\mu = \bar{X} + Z\frac{\sigma}{\sqrt{n}}$$

Estimativa do intervalo de confiança

- Exemplo: suponha que uma amostra de n=25 caixas tenha uma média aritmética igual a 362,3g. Determine o intervalo simetricamente distribuído em torno de $\bar{X}$ com probabilidade de ocorrência de 95% onde pode estar o valor da média aritmética da população (μ).
- Como queremos 95% de probabilidade, cada cauda deve conter 2,5%:

Pela tabela, os valores de Z são:

Z_{inf}=-1,96 e Z_{sup}=+1,96



Assim, o intervalo pedido será:

$$\mu_i = \bar{X} + Z_{inf}\frac{\sigma}{\sqrt{n}} = 362{,}3 - 1{,}96\frac{15}{\sqrt{25}} = 356{,}42$$

$$\mu_s = \bar{X} + Z_{sup}\frac{\sigma}{\sqrt{n}} = 362{,}3 + 1{,}96\frac{15}{\sqrt{25}} = 368{,}18$$

Assim, estima-se que a média aritmética da população esteja dentro do seguinte intervalo com 95% de confiança:

$$356{,}42 \leq \mu \leq 368{,}18$$

Intervalo de Confiança para a média aritmética

$$\bar{X} \pm Z\frac{\sigma}{\sqrt{n}}$$

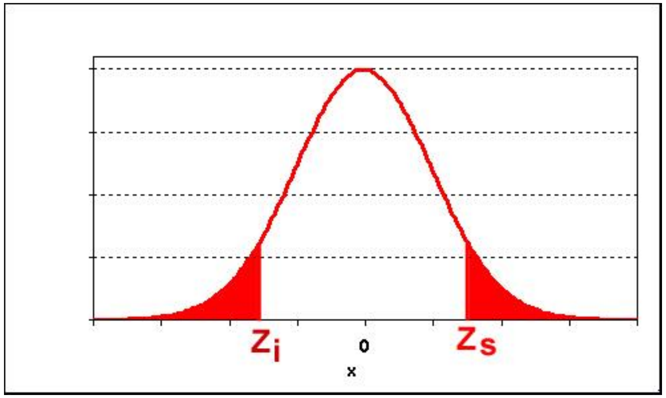
Ou

$$\bar{X} - Z\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z\frac{\sigma}{\sqrt{n}}$$

O valor de Z necessário para construir o intervalo de confiança é chamado de **valor crítico** para a distribuição.

Nível de confiança

- É representado por (1-α)x100%, onde α é a proporção nas caudas da distribuição que se encontra fora do intervalor de confiança.
- A proporção na cauda superior da distribuição vale $\alpha/2$ e a proporção na cauda inferior da distribuição vale também $\alpha/2$.
- Por exemplo, se queremos um nível de confiança de 95%, α=0,05 e a proporção na cauda superior vale 0,025 e na cauda inferior vale 0,025.



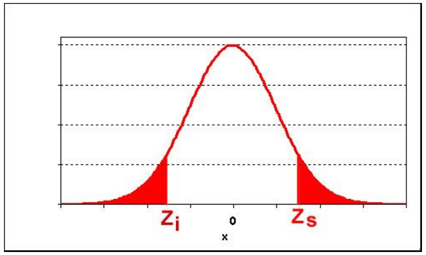
Exemplo

- Utilizando o exemplo das caixas de cereais, encontre o intervalo de confiança para um nível de confiança de 99%:

Para um nível de confiança de 99%, α=0,01 e a proporção em cada cauda deve ser de 0,005.

Pela tabela, o valor de Z crítico é:

Z=2,58



Dessa forma, o intervalo de confiança será:

$$\bar{X} - Z\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z\frac{\sigma}{\sqrt{n}} \rightarrow 362{,}3 - 2{,}58\frac{15}{\sqrt{25}} \leq \mu \leq 362{,}3 + 2{,}58\frac{15}{\sqrt{25}}$$

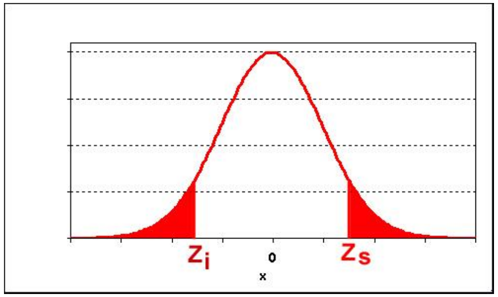
$$354{,}56 \leq \mu \leq 370{,}04$$

Exercício

- Um fabricante de papel possui um processo de produção que opera continuamente ao longo de todo um turno de produção. Espera-se que o papel possua uma média aritmética de comprimento de 11 polegadas, e o desvio-padrão do comprimento é igual a 0,02 polegadas. Em intervalos periódicos, amostras são selecionadas para determinar se a média aritmética do comprimento do papel é ainda igual a 11 polegadas, ou se ocorreu algo de errado no processo de produção, de modo a alterar o comprimento do papel produzido. É selecionada uma amostra de 100 folhas, e a média aritmética do comprimento do papel é igual a 10,998 polegadas. Determine um estimativa para o intervalo de confiança de 95% para a média aritmética da população correspondente ao comprimento do papel.

Resolução

- Para um nível de confiança de 95%, o valor de Z crítico (pela tabela) é 1,96.
- Assim, como a média é 10,998, o desvio-padrão 0,02 e o tamanho da amostra é 100, temos:



$$\bar{X} - Z\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z\frac{\sigma}{\sqrt{n}} \rightarrow$$

$$10{,}998 - 1{,}96\frac{0{,}02}{\sqrt{100}} \leq \mu \leq 10{,}998 + 1{,}96\frac{0{,}02}{\sqrt{100}} \rightarrow$$

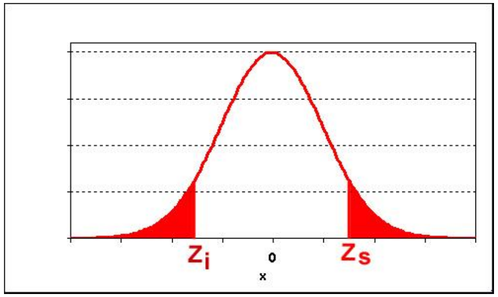
$$10{,}99408 \leq \mu \leq 11{,}00192$$

Exercício

- Construa a estimativa para o intervalo de confiança de 99% para a média aritmética do comprimento do papel.

Resolução

- Para um nível de confiança de 99%, o valor de Z crítico (pela tabela) é 2,58.
- Assim, como a média é 10,998, o desvio-padrão 0,02 e o tamanho da amostra é 100, temos:



$$\bar{X} - Z\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z\frac{\sigma}{\sqrt{n}} \rightarrow$$

$$10{,}998 - 2{,}58\frac{0{,}02}{\sqrt{100}} \leq \mu \leq 10{,}998 + 2{,}58\frac{0{,}02}{\sqrt{100}} \rightarrow$$

$$10{,}99284 \leq \mu \leq 11{,}00316$$

- Nos dois exercícios, uma vez que o intervalo de confiança inclui o valor 11, não existe nenhuma razão para acreditar que existe algum problema na produção.

Desvio-padrão desconhecido

- É também possível se determinar o intervalo de confiança da média aritmética quando não se conhece o valor do desvio-padrão.
- Para isso utiliza-se uma outra distribuição para se determinar os valores críticos.
- Essa distribuição é chamada de Distribuição t de Student e não será tratada neste curso.

O que é Teste de Hipóteses?

- Geralmente queremos utilizar a média aritmética de uma amostra para validar uma afirmativa sobre a média aritmética da população.
- O método inferencial para resolver esse problema é chamado de Teste de Hipóteses.
- A hipótese que que o parâmetro da população é igual ao parâmetro especificado (da amostra) é chamado de ***Hipótese nula (H₀)***:

$$H_0: \mu = \bar{X}$$

- A ***Hipótese alternativa*** H_1 corresponde ao oposto da Hipótese nula:

$$H_1: \mu \neq \bar{X}$$

Pontos chaves:

- A hipótese nula representa o status quo, ou aquilo que se acredita no momento, em uma determinada situação.
- A hipótese alternativa é o oposto da hipótese nula, e representa uma declaração a ser investigada,, ou a uma inferência específica que se gostaria de provar.
- Se a hipótese nula for rejeitada, tem-se comprovação estatística de que a hipótese alternativa está correta.
- Se a hipótese nula não for rejeitada, não se comprova a hipótese alternativa. O fato de não se comprovar a hipótese alternativa, entretanto, não significa que tenha comprovado a hipótese nula.
- A hipótese nula sempre se refere a um parâmetro da população, e não a uma estatística da amostra.
- A hipótese nula sempre contém um sinal de igualdade.
- A hipótese alternativa nunca contém um sinal de igualdade.

Exercício

- Você é o gerente de uma lanchonete e deseja determinar se o tempo de espera para atender a um pedido se modificou, no mês anterior, em relação ao valor anterior de 4,5 minutos, correspondente à média aritmética da população. Declare a hipótese nula e a hipótese alternativa.

Resolução

- A hipótese nula é de que a média aritmética da população não se modificou em relação ao seu valor anterior de 4,5 minutos:
- H_0: $\mu = 4,5$
- A hipótese alternativa corresponde ao oposto da hipótese nula. Uma vez que a hipótese nula é que a média da população é igual a 4,5 minutos, a hipótese alternativa é de que a média da população não é igual a 4,5 minutos:
- H_1: $\mu \neq 4,5$

Metodologias

- Teste Z para a média com desvio-padrão conhecido.
- Testes Unicaudais.
- Teste t para a média com desvio-padrão desconhecido.
- Teste Z para a proporção.
- Teste Z para a diferença entre duas médias.
- Teste t em pares.
- Teste Z para a diferença entre duas proporções.
- Teste F para a diferença entre duas variâncias.
- Análise de Variância de fator único (ANOVA fator único).
- Procedimento de Tukey-Kramer.
- Modelo fatorial.
- Testes Qui-Quadrados.
- Testes não paramétricos.

Bibliografia

- Levine, Stephan, Krehbiel, Berenson. Estatística Teoria e Aplicações. Quinta Edição. LTC.