Distribuição Normal

Introdução

- É a distribuição de probabilidade mais importante na estatística
- Abrange um grande número de fenômenos
- Oferece base para inferência estatística clássica devido à sua afinidade com o teorema do limite central

Introdução

- Possui gráfico simétrico, em formato de sino

- As medidas de tendência central: média, moda e mediana; são todas idênticas (simetria)

O Modelo Matemático

- A função de densidade da probabilidade da distribuição normal é:

$$f(x) = \frac{1}{2\pi\sigma} e^{-(0,5)[(X-\mu)/\sigma]^2}$$

- Felizmente, não precisamos usar esta exaustiva fórmula, uma vez que podemos trabalhar com padronização de dados, usando apenas uma tabela

Padronizando a Distribuição Normal

- Utilizando a fórmula de transformação, qualquer variável aleatória normal X é convertida em uma variável normal padronizada Z

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

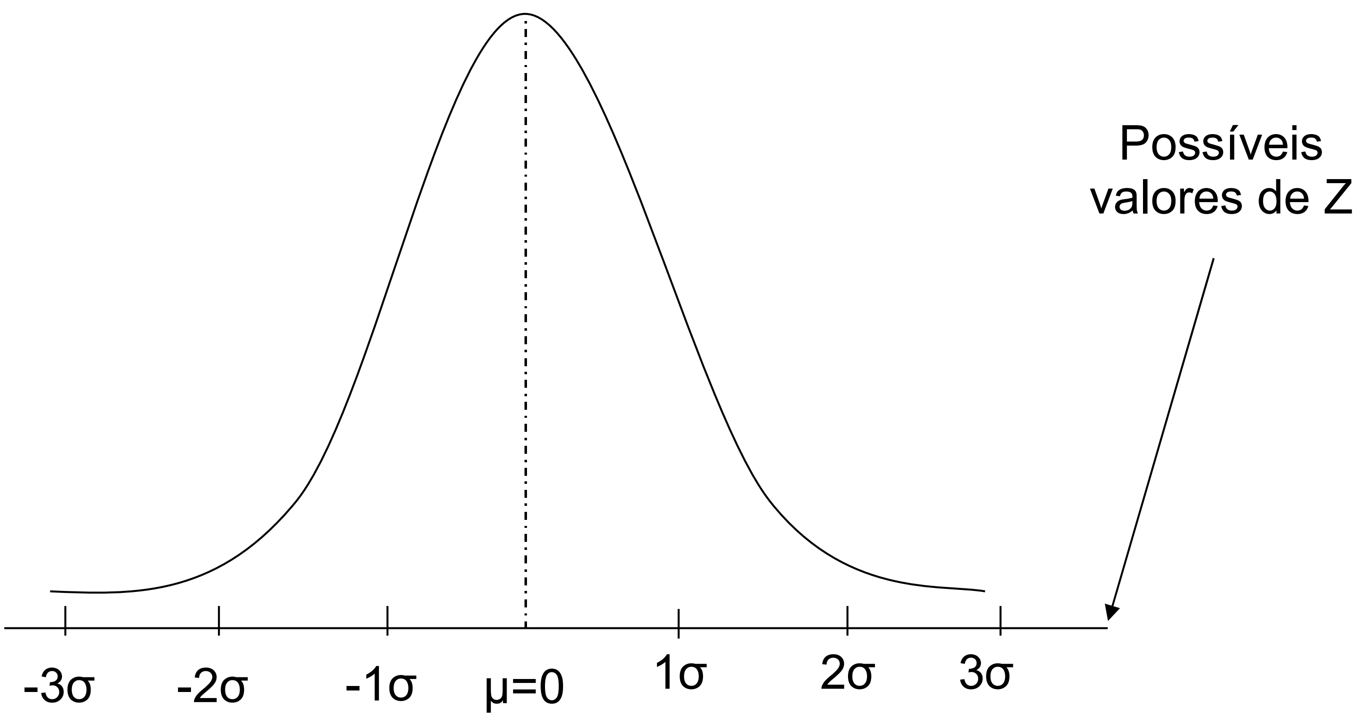
onde:

σ é o desvio padrão

μ é a média aritmética

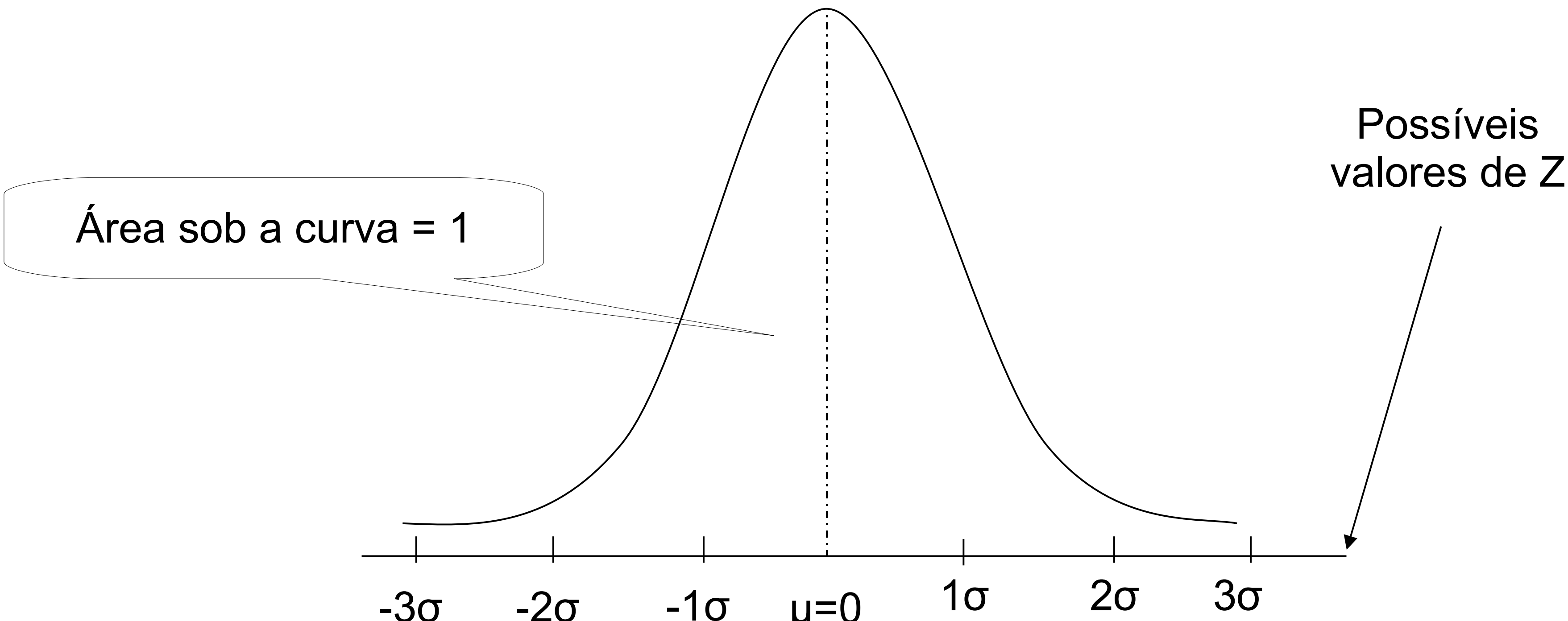
Análise Gráfica

- Na distribuição normal padronizada, a variável Z possui média 0 e desvio padrão 1
- Z é variável contínua que representa o número de desvios a contar da média



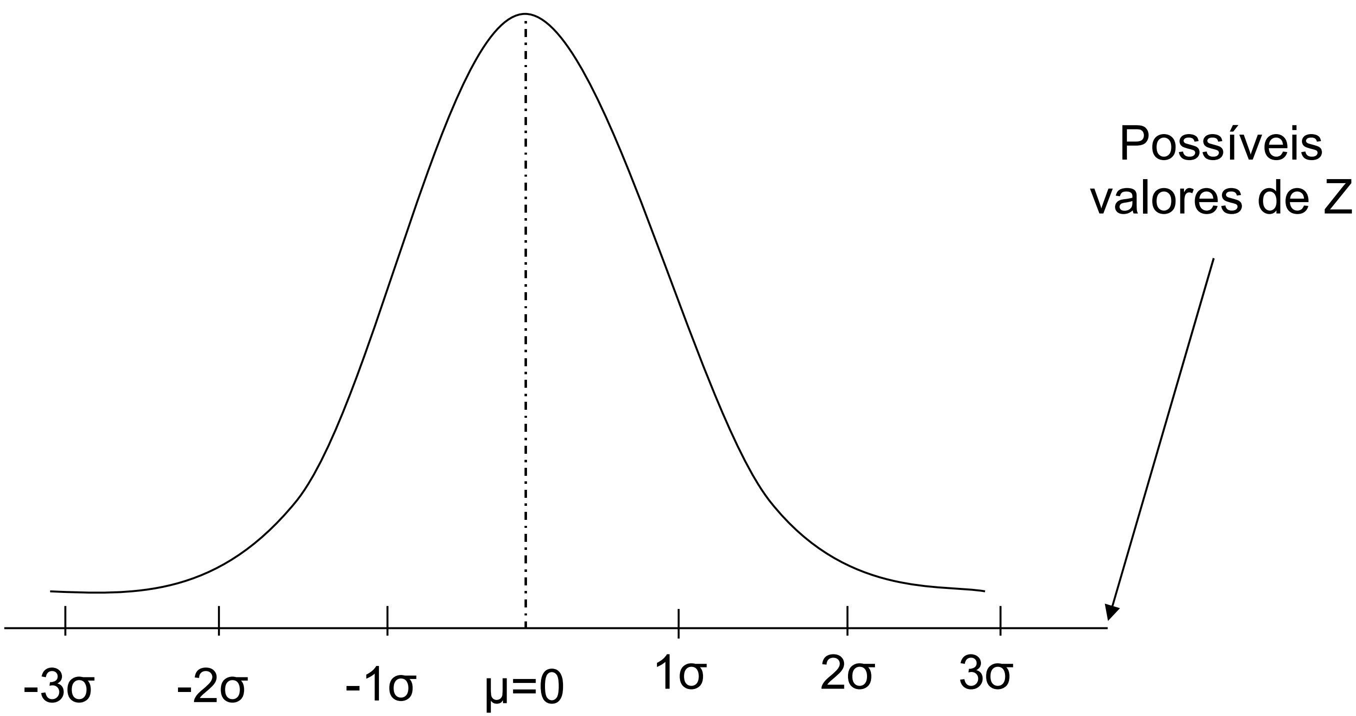
Análise Gráfica

- A área sob a curva corresponde à probabilidade de a variável aleatória assumir qualquer valor real, deve ser um valor entre 0 e 1
- Valores maiores que a média e os valores menores têm a mesma probabilidade, pois a curva é simétrica



Análise Gráfica

- 68% dos valores de Z estão entre -1σ e 1σ
- 95,5% dos valores de Z estão entre -2σ e 2σ
- 99,7% dos valores de Z estão entre -3σ e 3σ



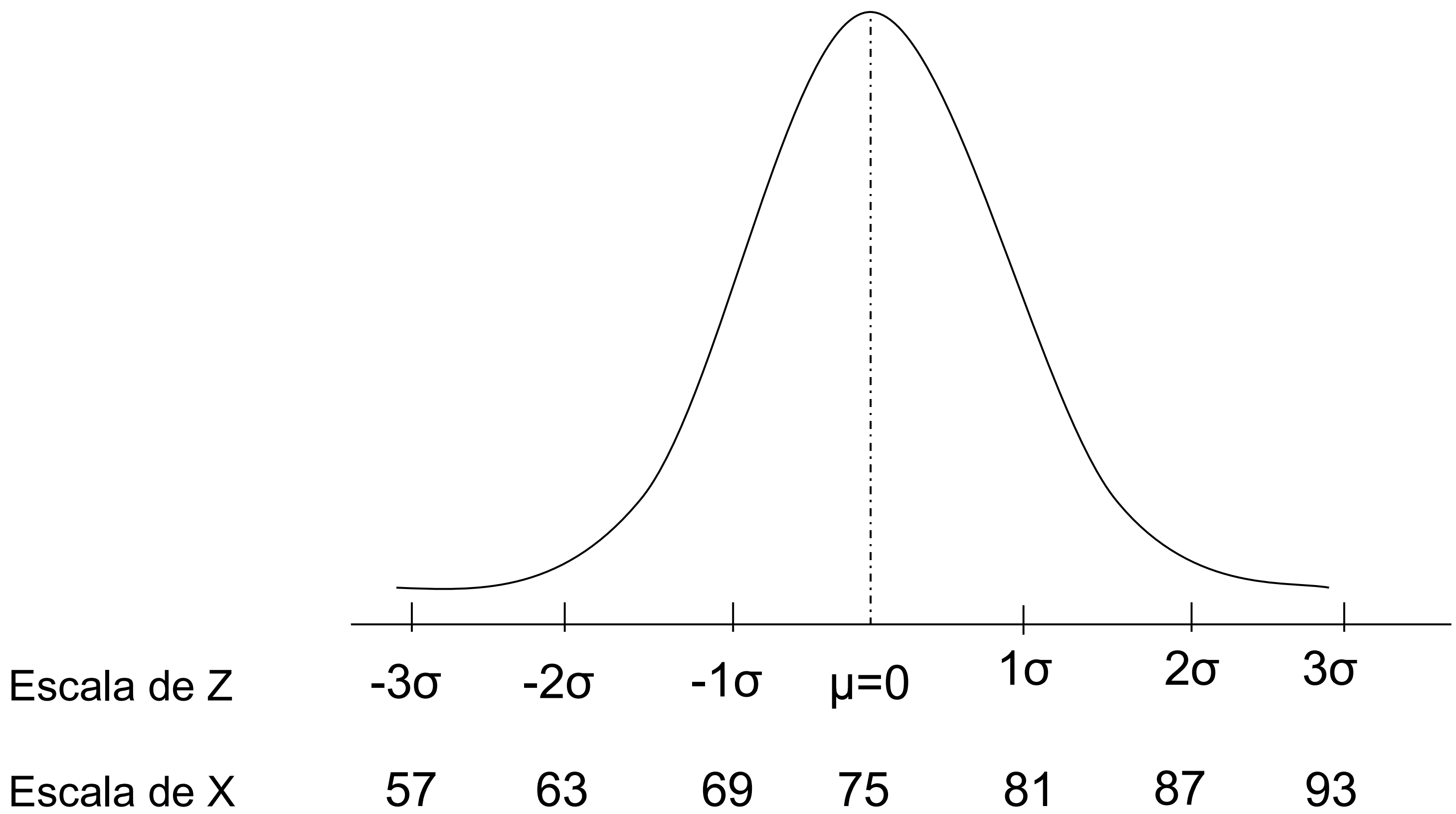
Aplicação - Um significado prático para o que aprendemos

1. Suponha um consultor investigando o tempo que os trabalhadores de uma fábrica levam para montar determinada peça.

3. Suponha que análises da linha de produção tenham calculado tempo médio de 75 segundos e desvio padrão de 6 segundos

5. O que isto significa graficamente?

Aplicação - Um significado prático para o que aprendemos



Aplicação - Um significado prático para o que aprendemos

- Ainda na Escala de X, o tempo central é a média de 75 segundos

- Na Escala de Z, a média é 0 e os intervalos tem como base o desvio padrão. Mas, assim como X, a variável Z é contínua

- Pergunta: como 87, na Escala de X, pode ser relacionado a 2σ, na Escala de Z?

- Conseguiram responder? A seguir temos duas explicações.

Aplicação - Um significado prático para o que aprendemos

- Na Escala de Z, 2σ significa dois desvios padrões a partir da média (0+ 2σ= 2σ), na Escala de X, este deslocamento é análogo (75+2*6=87)

- Outra forma de relacionar estes valores é através da fórmula de transformação apresentada anteriormente:

$$Z = \frac{X - \mu}{\sigma} = \frac{87 - 75}{6} = \frac{12}{6} = 2$$

Dúvidas?

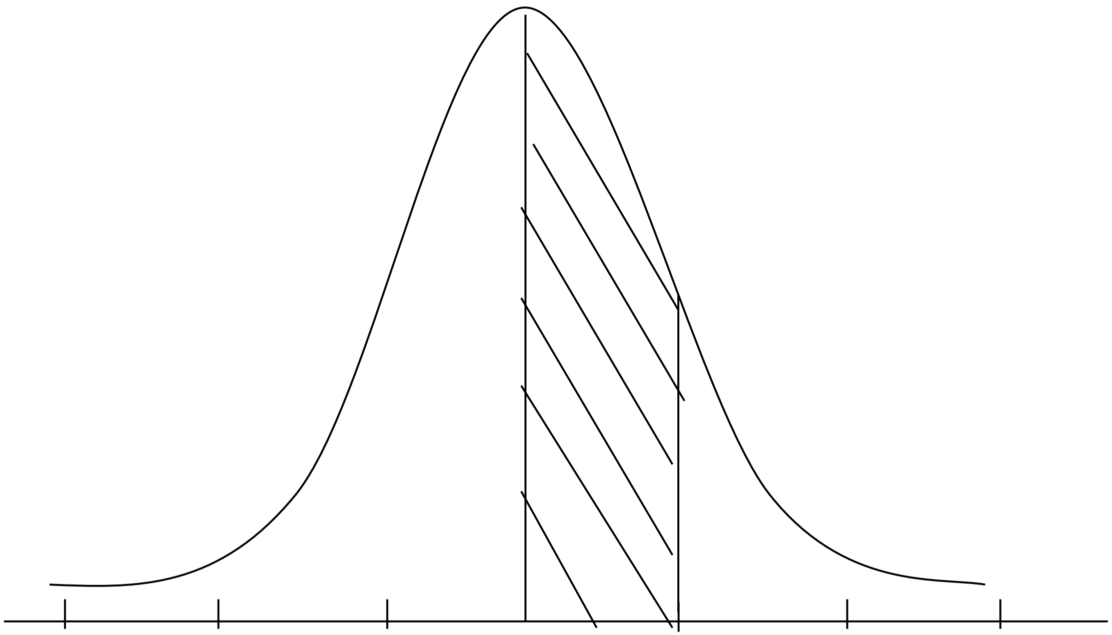
Aplicação - Um significado prático para o que aprendemos

- Suponha agora, que o consultor queira saber qual a probabilidade de um trabalhador levar um tempo entre 75 e 81 segundos para montar uma peça, ou seja, P(75≤X≤81). Como proceder?
  - Transformar as variáveis X em variáveis normais padronizadas Z:

$$Z = \frac{75 - 75}{6} = \frac{0}{6} = 0 \qquad Z = \frac{81 - 75}{6} = \frac{6}{6} = 1$$

Logo temos a probabilidade P(0≤Z≤1), que é ilustrada a seguir, e cujo valor é determinado consultando a tabela no slide seguinte.

Aplicação - Um significado prático para o que aprendemos



Escala de Z -3σ -2σ -1σ μ=0 1σ 2σ 3σ

Escala de X 57 63 69 75 81 87 93

Área sob a Curva Normal

(tabela parcial)

Z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1	**0,3413**	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621

Aplicação - Um significado prático para o que aprendemos

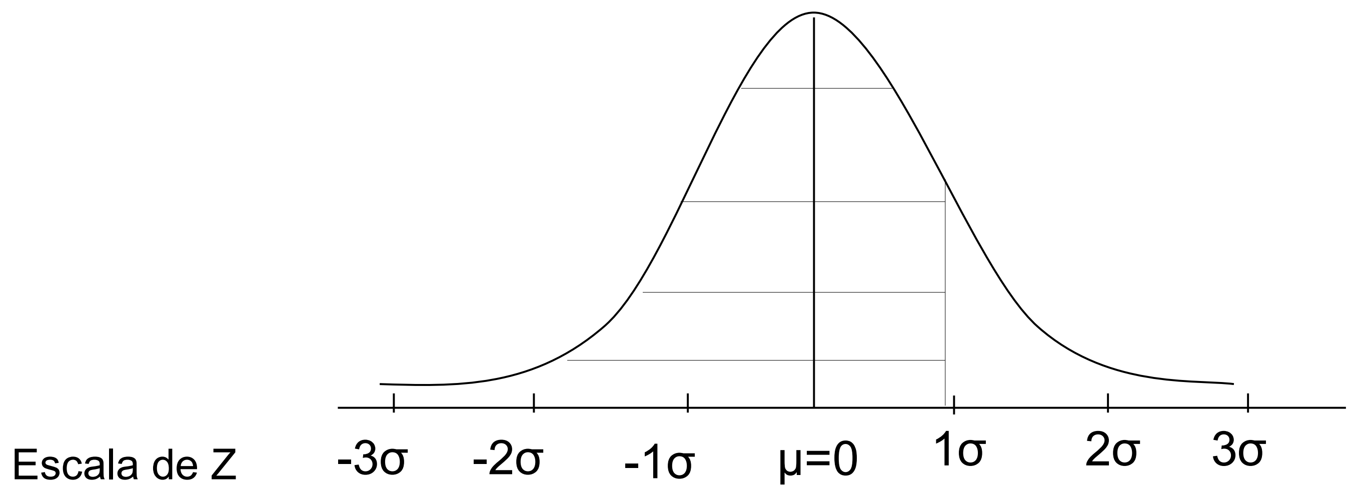
Consultando a tabela, encontramos o valor da área indicada, que significa a probabilidade
$P(75 \leq X \leq 81) = P(0 \leq Z \leq 1) = 0{,}3413$

Este resultado nos informa que há probabilidade de 0,3413 de um trabalhador levar um tempo entre 75 e 81 segundos para montar uma peça.

Outra interpretação é que 34,13% dos trabalhadores levarão um tempo dentro do intervalo de 75 e 81 segundos

Distribuição Normal em Planilhas

- Função: DIST.NORMP( )
  - Parâmetro: valor de z
  - Resultado: área até o valor de z



Exercícios

A aplicação da distribuição normal só se aprende com muita prática:

- Qual a probabilidade de um trabalhador montar uma peça entre 69 e 81 segundos?(0,6826)
- Qual a probabilidade de um trabalhador montar uma peça em menos de 62 segundos?(0,0150)
- Qual a probabilidade de um trabalhador montar uma peça entre 62 e 69 segundos?(0,1437)
- Em qual intervalo de tempo 99,7% dos trabalhadores montam um peça?(57 e 93 segundos)

Exercícios

- Um marinheiro recebe um telegrama avisando que sua esposa deu a luz naquele dia, 308 dias após sua última visita. Sendo que os prazos de gravidez têm distribuição normal com média de 268 dias e desvio padrão de 15 dias, pergunta-se: o marinheiro deve se preocupar...?

Calculando valores a partir de probabilidades

- Considerando o exemplo da fábrica, qual o tempo que separa os 90% mais rápidos dos 10% mais lentos ?
- Para fazer este cálculo deve-se:
  - Lembrar que porcentagem ou probabilidades são áreas do gráfico, e não valores de z
  - A leitura da tabela é invertida (pela área descobre-se)
  - Escolher o lado correto do gráfico (os maiores tempos estão do lado direito)
  - Aplicar a variação da fórmula de padronização $x = \mu + z * \sigma$

Calculando valores a partir de probabilidades

- A área a ser analisada é corresponde à 40%, ou 0,4 (já que o lado direito possui 50% dos tempos) – represente esta situação graficamente
- Recorrendo à tabela de distribuição normal, têm-se z=1,28
- Aplicando a fórmula:

$$x = \mu + z * \sigma = 75 + 1{,}28 * 6 = 82{,}68$$

- Logo, o tempo que separa os 90% mais rápidos dos 10% mais lentos é de 82,68 segundos

Teorema Central do Limite

- Na medida em que o tamanho da amostra aumenta, a distribuição das médias amostrais tende para uma distribuição normal...

- ...independente do tipo de distribuição da população

- Logo, a média das médias das amostras poderá ser considerada como a média da população

- Porém o desvio padrão será: $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

Teorema Central do Limite

- Para amostras de tamanho n > 30, a distribuição das médias amostrais pode ser aproximada satisfatoriamente por uma distribuição normal
- A aproximação melhora à medida em que aumenta o tamanho da amostra *n*
- Se a população tem distribuição normal, as médias amostrais terão distribuição normal para qualquer tamanho amostral *n*

Teorema Central do Limite

- Considerando ainda o exemplo da fábrica, calcule e interprete os resultados obtidos:
  - A probabilidade de 1 tempo escolhido aleatoriamente ser inferior a 73 segundos (0,3707)
  - A probabilidade de 49 tempos escolhidos aleatoriamente serem inferiores a 73 segundos (0,0099)
  - A probabilidade de 100 tempos escolhidos aleatoriamente serem inferiores a 73 segundos (0,001

Obrigado!

Até a próxima aula.