🧠 O que é o **Z da distribuição normal**?

O **Z**, também chamado de **Z-score**, é um número que diz **quantos desvios padrão** um valor está **acima ou abaixo da média** em uma **distribuição normal**.

🧾 A fórmula é:

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

Onde:

- X = valor observado
- μ = média da população
- σ = desvio padrão

📌 O que o valor de Z nos diz?

Z-score	Interpretação
0	valor **igual à média**
+1	**1 desvio acima** da média
-1	**1 desvio abaixo** da média
+2	2 desvios acima
-2	2 desvios abaixo
...	e assim por diante

📚 Um pouco da história

O conceito do Z-score vem da **estatística clássica**, com origens nos estudos de **Carl Friedrich Gauss** e **Abraham de Moivre** sobre distribuições normais no século XVIII.

- **Gauss** descreveu a famosa **"curva em forma de sino"**, que mostra como variáveis naturais (como altura, QI, tempo de reação etc.) se distribuem ao redor de uma média.
- A **padronização com Z** foi criada para **comparar diferentes conjuntos de dados** de forma justa, mesmo que tenham escalas diferentes.

🧪 Como e onde aplicar?

O Z é **fundamental** em diversas aplicações:

✅ 1. Comparar valores de diferentes distribuições

Exemplo: comparar notas de provas com médias e desvios diferentes:

- Prova A: nota 80, média 70, desvio 5 → Z = (80 - 70)/5 = 2
- Prova B: nota 88, média 85, desvio 2 → Z = (88 - 85)/2 = 1.5

> A nota da **Prova A** está **mais distante da média**, ou seja, é melhor relativamente.

✅ 2. Encontrar **probabilidades**

Usando a **tabela Z** (ou função em Python/Excel), você pode descobrir:

- Qual a chance de um valor estar **abaixo/acima** de um ponto
- Qual a **área sob a curva normal** até um certo valor

> Exemplo: Z = 1.96 → ~97,5% dos valores estão abaixo desse ponto.

✅ 3. Criar **intervalos de confiança**

Confiança	Valor de Z
90%	1.645
95%	1.96
99%	2.576

Você usa esses valores para dizer coisas como:

> "Com 95% de confiança, a média está entre X e Y."

✅ 4. Fazer **testes de hipótese**

Você compara o Z obtido com o Z crítico:

- Se Z calculado for muito extremo, **rejeita a hipótese nula**
- Serve para ver se uma média **realmente mudou**, se dois grupos **são diferentes**, etc.

🔢 Exemplo prático

Imagine uma turma com:

- Média de altura: 1.70 m
- Desvio padrão: 0.05 m

Aluno com 1.80 m de altura:

$
Z = \frac{1.80 - 1.70}{0.05} = 2.0
$

Conclusão: está **2 desvios acima da média**, ou seja, é bem mais alto que o típico aluno da turma.

💡 Curiosidade: Regra Empírica

Na distribuição normal, temos:

- **68%** dos dados entre Z = -1 e +1
- **95%** dos dados entre Z = -2 e +2
- **99.7%** dos dados entre Z = -3 e +3

Esse é o famoso "**empirical rule**" ou "**68-95-99.7 rule**".

⚙️ Ferramentas que usam o Z

- **Excel:** `NORM.S.DIST(Z, TRUE)` → retorna a área até o Z
- **Python (scipy):**

```
from scipy.stats import norm
norm.cdf(1.96)  # ~0.975
```

- **R:** `pnorm(1.96)` → também ~0.975

📌 Fórmula do Tamanho da Amostra (para proporções)

$
n = \left(\frac{Z^2 \cdot p \cdot (1 - p)}{E^2} \right)
$

🧠 Onde:

- n = tamanho da amostra necessário
- Z = valor da distribuição normal padrão associado ao nível de confiança (ex: 1.96 para 95%, 2.576 para 99%)
- p = proporção estimada da população (use 0.5 se não souber, pois gera o pior caso)
- E = erro amostral tolerado (margem de erro), em decimal (ex: 5% → 0.05)

✅ Exemplo prático:

Você quer:

- 99% de confiança → $Z = 2.576$
- Margem de erro de 3% → $E = 0.03$
- Não conhece a proporção populacional → usa $p = 0.5$

$$
n = \frac{(2.576)^2 \cdot 0.5 \cdot (1 - 0.5)}{(0.03)^2} = \frac{6.635 \cdot 0.25}{0.0009} \approx 1843
$$

👉 **Você precisa de cerca de 1843 pessoas na amostra.**

🔄 Se a população for pequena:

Use a **correção de população finita**:

$$
n_{corrigido} = \frac{n}{1 + \frac{n - 1}{N}}
$$

- N = tamanho total da população

📏 O que é **Tamanho de Amostra Proporcional**?

O **tamanho da amostra proporcional** é uma técnica usada quando você quer garantir que **cada grupo** ou **segmento** de uma população esteja **representado proporcionalmente** na amostra final.

✅ Exemplo prático:

Imagine uma escola com 1000 alunos divididos por séries:

Série	Número de Alunos	Proporção (%)
1ª	200	20%
2ª	300	30%
3ª	500	50%

Você quer fazer uma pesquisa com **200 alunos** (sua amostra total).

👉 Para manter a **proporcionalidade**, calcula-se:

- 1ª série: 200 × 20% = **40 alunos**
- 2ª série: 200 × 30% = **60 alunos**
- 3ª série: 200 × 50% = **100 alunos**

> 🎯 Isso garante que a amostra represente bem a estrutura da população.

🧮 Como calcular o **tamanho da amostra** com base na **confiabilidade (nível de confiança)**?

Como vimos antes, a fórmula é:

$
n = \frac{Z^2 \cdot p \cdot (1 - p)}{E^2}
$

Onde:

- n: Tamanho da amostra
- Z: Valor z da distribuição normal (depende da confiabilidade)
- p: Proporção esperada (use 0.5 se não souber)
- E: Margem de erro (em decimal)

📊 Tabela de valores Z (nível de confiança):

Nível de Confiança	Valor Z
90%	1.645
95%	1.96
99%	2.576

✅ Exemplo com diferentes níveis de confiança:

Vamos supor:

- $p = 0.5$ (sem conhecimento da proporção)
- $E = 0.05$ (5% de margem de erro)

90% de confiança:

$
n = \frac{(1.645)^2 \cdot 0.5 \cdot 0.5}{(0.05)^2} = \frac{0.6765}{0.0025} \approx 271
$

95% de confiança:

$
n = \frac{(1.96)^2 \cdot 0.5 \cdot 0.5}{(0.05)^2} = \frac{0.9604}{0.0025} \approx 384
$

99% de confiança:

$
n = \frac{(2.576)^2 \cdot 0.5 \cdot 0.5}{(0.05)^2} = \frac{1.658}{0.0025} \approx 664
$

Para calcular o **tamanho da amostra** a partir de uma população, levando em consideração os **níveis de confiança** (90%, 95% ou 99%), usamos fórmulas estatísticas baseadas em **amostragem**

probabilística, especialmente a **amostragem aleatória simples**.

✅ Fórmula básica para tamanho da amostra (população infinita)

$$
n = \frac{Z^2 \cdot p \cdot (1 - p)}{e^2}
$$

Onde:

- n: tamanho da amostra
- Z: valor da **distribuição normal padrão** associado ao nível de confiança
- p: proporção esperada (suponha 0,5 se desconhecida – maximiza o tamanho da amostra)
- e: margem de erro (erro amostral tolerável, geralmente 0,05 = 5%)

🔢 Valores de Z para os principais níveis de confiança:

Nível de confiança	Valor de Z
90%	1,645
95%	1,960
99%	2,576

🔍 Exemplo com população infinita

Suponha que queremos estimar uma proporção com:

- Nível de confiança: **95%**
- Proporção esperada: **0,5** (p = 50%)
- Margem de erro: **5%** (e = 0,05)

$$
n = \frac{(1,96)^2 \cdot 0,5 \cdot (1 - 0,5)}{(0,05)^2}
$$

$$
n = \frac{3,8416 \cdot 0,25}{0,0025} = \frac{0,9604}{0,0025} = \boxed{384,16}
$$

Resultado: Você precisaria de **aproximadamente 385 pessoas** na amostra.

👥 População finita (com correção)

Se você **conhece o tamanho da população (N)**, use a **correção de população finita**:

$$
n_{ajustada} = \frac{n}{1 + \left(\frac{n - 1}{N}\right)}
$$

🔧 Exemplo com população finita:

População total $N = 1.000$
Amostra anterior $n = 385$

$$
n_{ajustada} = \frac{385}{1 + \left(\frac{384}{1000}\right)} = \frac{385}{1 + 0,384} = \frac{385}{1,384} = \boxed{278}
$$

Resultado: Para uma população de 1.000 pessoas, bastam **278 indivíduos** na amostra para os mesmos parâmetros.

📊 Comparação rápida de tamanhos de amostra para populações grandes:

Margem de erro	90% (Z=1,645)	95% (Z=1,96)	99% (Z=2,576)
10%	68	97	166
5%	271	385	666
3%	752	1.067	1.843

🧠 Dica:

- Se você **não sabe a proporção esperada (p)**, use **0,5**.
- Se quiser **diminuir o tamanho da amostra**, aumente a **margem de erro** ou reduza o **nível de confiança**.

🛠️ Aplicando na prática

Se você quer fazer uma **pesquisa proporcional** com um nível de confiança de 99% e margem de erro de 5%, e sua população tem **10 mil pessoas em 4 regiões diferentes**, faça assim:

1. Calcule o tamanho da amostra total com a fórmula.
2. Distribua proporcionalmente esse valor pelas regiões de acordo com o percentual da população de cada uma.

✅ O que é **Proporção Esperada (p)**?

A **proporção esperada** é a **estimativa da proporção da população** que tem determinada característica que você quer estudar.

🧠 Exemplo prático 1:

Você quer saber **quantos alunos da escola usam transporte público**.

Se você **já sabe** (por uma pesquisa anterior) que **60% usam**, então:

- $p = 0.6$

Se quer saber **quantos usam celular na sala de aula**, e não tem nenhuma ideia ou dado anterior, então:

- $p = 0.5$ (valor mais conservador, explico abaixo!)

🤔 Por que usar **p = 0.5** se não sei?

Porque **0.5 é o pior caso possível** em termos de variabilidade. Isso significa que:

- Maximiza a incerteza
- Garante um **tamanho de amostra suficientemente grande**
- Funciona como uma **estimativa conservadora** e segura

🔍 Comparação:

Proporção esperada (p)	Variabilidade $p(1-p)$
0.1	0.09
0.3	0.21
0.5	**0.25** ← maior
0.7	0.21
0.9	0.09

> 📌 Quanto maior $p(1-p)$, maior a variabilidade — então o tamanho da amostra será maior para garantir a precisão.

✅ Regra prática:

- **Se você sabe algo sobre a população** (dados anteriores ou piloto), use esse valor como p.
- **Se não sabe nada**, use **p = 0.5** — isso garante segurança no cálculo.

📊 Exemplo comparando:

Imagine uma pesquisa com:

- Erro: 5% (E = 0.05)

- Confiança: 95% (Z = 1.96)

Com p = 0.5:

$$
n = \frac{1.96^2 \cdot 0.5 \cdot 0.5}{0.05^2} \approx 384
$$

Com p = 0.7 (sabendo que 70% têm a característica):

$$
n = \frac{1.96^2 \cdot 0.7 \cdot 0.3}{0.05^2} \approx 323
$$

> ✅ Usar o valor real de p (quando disponível) **pode reduzir** o tamanho necessário da amostra!

Perfeito! Vamos continuar e **aprofundar** mais no conceito do **Z-score** (ou escore Z), explorando:

1. 🧠 Como interpretar **áreas da curva normal** com o Z
2. 🧪 Como usar o Z-score para **tomar decisões em testes estatísticos**
3. 🎯 Como aplicar o Z na vida real (exemplos do cotidiano)
4. 🛠️ Como calcular e interpretar no Excel e Python
5. 📊 Relação com outras distribuições (t, chi², etc.)

---

1. 🧠 Áreas da curva normal e o Z-score

A **distribuição normal padronizada** (média 0, desvio padrão 1) é usada como **modelo universal**.

O que isso significa?

Ao calcular um Z-score, você pode olhar numa tabela ou função e **descobrir a probabilidade de um valor acontecer**.

Por exemplo:

- Se você obtém **Z = 1.96**, isso significa que seu valor está **acima de 97.5% da distribuição normal** (ou seja, só 2.5% estão acima dele).
- Se **Z = -1.96**, ele está **abaixo de 2.5% dos valores**.

🟩 **Curva normal:** a área **embaixo da curva** representa **probabilidades**.

---

2. 🧪 Testes estatísticos com Z-score

O Z é a **base dos testes z**, usados quando:

- Você conhece a média e o desvio padrão da população.
- Tem uma amostra grande (n > 30).

Exemplo prático:

Você é professor e sabe que a média histórica da sua turma em uma prova é **70 pontos** com desvio de **10 pontos**. Esse ano, uma turma tirou média **73** com **n = 100 alunos**.

Você quer saber: **isso é estatisticamente diferente?**

Use o **Z-teste da média:**

$$
Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{73 - 70}{\frac{10}{\sqrt{100}}} = \frac{3}{1} = 3.0
$$

- Z = 3 → **muito improvável ser por acaso** (apenas 0.3% dos casos)
- Você **rejeita a hipótese nula** de que "nada mudou"

---

3. 🎯 Aplicações reais do Z-score

Z é super útil em diversas áreas:

💼 RH e Seleção

Comparar candidatos em provas diferentes:

- João tirou 85 numa prova com média 75 e desvio 5 → Z = 2.0
- Maria tirou 90 numa com média 88 e desvio 1 → Z = 2.0

> Ambos estão igualmente bem, **relativamente** às suas turmas.

🏥 Medicina

Medições como:

- Pressão arterial
- Colesterol
- Peso de recém-nascidos

Exemplo:

> "Seu colesterol está 2 desvios acima do normal"
> → Indica que está fora do padrão e precisa de atenção.

🏦 Finanças

- Em **controle de risco**, o Z-score ajuda a prever **quão anormal** é um retorno financeiro.
- Em **credit scoring**, pode indicar **probabilidade de inadimplência**.

---

4. 🛠️ Z-score no Excel e Python

Excel

Fórmula	Significado
`=STANDARDIZE(x, média, desvio)`	Calcula o Z de `x`
`=NORM.S.DIST(z, TRUE)`	Área até o Z-score
`=NORM.S.INV(0.975)`	Dá o Z para uma área (ex: 0.975 → 1.96)

Python (SciPy)

```
from scipy.stats import norm

# Z-score de um valor
z = (valor - media) / desvio

# Probabilidade até Z
prob = norm.cdf(z)

# Z correspondente a uma probabilidade
z_critico = norm.ppf(0.975)  # → 1.96
```

---

5. 📊 Relação com outras distribuições

- A **distribuição t** é parecida com a normal, mas usada quando o **n é pequeno** e **desvio da população é desconhecido**.
- A distribuição **chi-quadrado** (χ^2) é usada para variâncias e tabelas de frequência.
- A distribuição **F** é usada para comparar **duas variâncias**.

O **Z é o ponto de partida** para entender essas distribuições.

📌 Dica final

Se você lembrar só de uma coisa sobre Z-score, lembre-se disso:

> O Z transforma qualquer valor em uma **medida padronizada**, permitindo **comparações justas e precisas**, além de **calcular probabilidades** e **decidir estatisticamente se algo é relevante ou não**.

Exemplo em python

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm
```

```python
# Cria uma faixa de valores de -4 a 4
x = np.linspace(-4, 4, 1000)

# Distribuição normal padrão: média = 0, desvio padrão = 1
y = norm.pdf(x, 0, 1)

# Pontos da regra empírica (Z-scores)
z_scores = [-3, -2, -1, 0, 1, 2, 3]
labels = ['-3σ', '-2σ', '-1σ', '0', '+1σ', '+2σ', '+3σ']

# Criação do gráfico
plt.figure(figsize=(12, 6))
plt.plot(x, y, label='Distribuição Normal Padrão', color='black')

# Áreas da regra empírica
plt.fill_between(x, y, where=(x > -1) & (x < 1), color='#d0e1f9',
alpha=0.8, label='68% dos dados')
plt.fill_between(x, y, where=(x > -2) & (x < 2), color='#a9d0f5',
alpha=0.5, label='95% dos dados')
plt.fill_between(x, y, where=(x > -3) & (x < 3), color='#74c0fc',
alpha=0.3, label='99.7% dos dados')

# Linhas verticais com os Z-scores
for i, z in enumerate(z_scores):
    plt.axvline(z, linestyle='--', color='gray', alpha=0.6)
    plt.text(z, norm.pdf(z) + 0.01, labels[i], ha='center', fontsize=9)

# Ajustes finais do gráfico
plt.title('Distribuição Normal Padrão e Regra Empírica (Z-score)',
fontsize=14)
plt.xlabel('Z-score')
plt.ylabel('Densidade de Probabilidade')
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()
```