CHAPTER 4: NORMAL (GAUSSIAN) DISTRIBUTION AND ERROR THEORY

(Week 7: Lecture Notes)

---

1. INTRODUCTION: WHY "NORMAL"?

In nature and production processes, if an event is influenced by many small, independent, and random factors, the results surprisingly always take the same shape: **The Bell Curve**.

In mathematics, this is called the **Normal Distribution** or **Gaussian Distribution**.

Examples:

- Heights of people.
- Diameters of screws produced by a machine.
- Exam grades.
- GPS measurement errors.

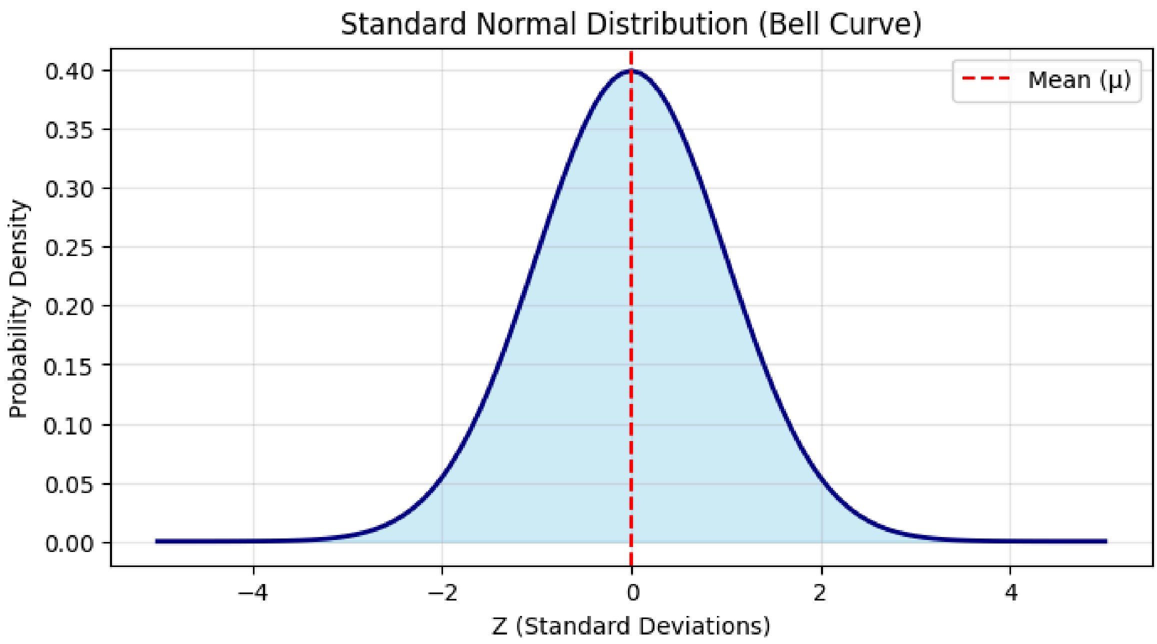
Feature of this distribution: Most data clusters around the **Mean**, and the probability decreases as you go to the extremes (very small or very large values).

In [1]:
```python
import matplotlib.pyplot as plt
import math

# Manual Normal Distribution Function (PDF)
def normal_pdf(x, mu, sigma):
    return (1 / (sigma * math.sqrt(2 * math.pi))) * math.exp(-0.5 * ((x - mu) /

# Generate Data points
mu = 0
sigma = 1
x_values = [i * 0.1 for i in range(-50, 51)] # -5.0 to 5.0
y_values = [normal_pdf(x, mu, sigma) for x in x_values]

plt.figure(figsize=(8, 4))
plt.plot(x_values, y_values, color='navy', linewidth=2)
plt.fill_between(x_values, y_values, color='skyblue', alpha=0.4)
plt.title("Standard Normal Distribution (Bell Curve)")
plt.xlabel("Z (Standard Deviations)")
plt.ylabel("Probability Density")
plt.axvline(mu, color='red', linestyle='--', label='Mean (μ)')
plt.legend()
plt.grid(True, alpha=0.3)
plt.show()
```



2. PARAMETERS OF THE DISTRIBUTION (Mu and Sigma)

There are only two numbers that determine the shape of the Normal distribution. If you know these two numbers, you know everything.

2.1. Mean (μ - Mu)

It is the peak (center) of the curve.

- Axis of symmetry of the curve.
- **If μ increases:** The curve shifts to the right without changing shape.
- **If μ decreases:** The curve shifts to the left.

2.2. Standard Deviation (σ - Sigma)

It is the width (spread) of the curve.

- **If σ is small:** Data is very close to the mean. The curve is peaked and narrow. (Precision manufacturing).
- **If σ is large:** Data is scattered. The curve is flat and wide. (Low quality manufacturing).

Formula (Probability Density Function - PDF):

(You don't need to memorize this, just know it exists)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

3. STANDARD NORMAL DISTRIBUTION (Z-Transform)

There are infinite Normal Distributions in the world.

- For concrete strength: $\mu = 30, \sigma = 5$
- For human height: $\mu = 175, \sigma = 10$

It is impossible to calculate integrals for each one separately. So, we use a math trick called **"Standardization"**. We convert all distributions into a single common language: **The Z Distribution**.

3.1. Features of Standard Normal Distribution

In this special distribution:

- **Mean (μ) = 0**
- **Standard Deviation (σ) = 1**

3.2. Z-Score Formula (Transformation)

To convert any data X into a Z-score:

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

What does this formula mean? The Z-score tells us: "*How many standard deviations away is your value (X) from the mean?*"

- **Z = 1:** 1 sigma to the right (above) of the mean.
- **Z = -2:** 2 sigmas to the left (below) of the mean.
- **Z = 0:** Exactly on the mean.

4. HOW TO USE THE Z-TABLE?

We cannot take the integral to calculate the probabilities of the Normal distribution (because it has no analytical integral). Instead, we use the pre-calculated **Z-Table**.

4.1. Table Logic

The Z-Table usually gives the **"Area to the Left"** (Cumulative Probability, $P(Z < z)$). Total area is always 1.

4.2. Reading Practice

Question: What is the probability of $Z < 1.25$?

1. Find **1.2** from the left column.

2. Find **0.05** from the top row. (Because 1.2 + 0.05 = 1.25).
3. Read the number at the intersection: **0.8944**. **Result:** $P(Z < 1.25) = 89.44\%$.

4.3. Critical Rules (For Exams)

The table usually gives only positive Zs and the left side. We use symmetry for other cases:

1. Greater Than Case $P(Z > a)$: The table gives "less than". Since total area is 1:

$$P(Z > a) = 1 - P(Z < a)$$

2. Negative Case $P(Z < -a)$: Due to symmetry, left of -a is same as right of +a.

$$P(Z < -a) = P(Z > a) = 1 - P(Z < a)$$

3. Interval Case $P(a < Z < b)$: Subtract the smaller area from the larger area.

$$P(a < Z < b) = P(Z < b) - P(Z < a)$$

5. NUMERICAL EXAMPLE: ENGINEERING APPLICATION

The breaking strength of steel ropes produced in a factory follows a Normal Distribution.

- **Mean (μ):** 1000 kg
- **Standard Deviation (σ):** 50 kg

Question 1: What is the probability that a randomly selected rope is stronger than **1075 kg**?

Solution:

1. Define X: $X = 1075$

2. Convert to Z-Score:

$$Z = \frac{1075 - 1000}{50} = \frac{75}{50} = 1.5$$

3. Translate Question to Z-Language: We wanted $P(X > 1075)$. Now we want $P(Z > 1.5)$.

4. Look at the Table: Look at 1.5 in the table -> **0.9332** (This is the left side).

5. Find Result: Since "Greater Than" is asked, subtract from 1.

$$P(Z > 1.5) = 1 - 0.9332 = 0.0668$$

Interpretation: Only **6.68%** of the ropes are this strong.

In [2]:
```
import math

# Python's cumulative distribution function (CDF) implementation
def normal_cdf(x, mu, sigma):
    return 0.5 * (1 + math.erf((x - mu) / (sigma * math.sqrt(2))))

mu = 1000
sigma = 50
x_target = 1075

# Calculate P(X < 1075) (Left side / Table value)
prob_left = normal_cdf(x_target, mu, sigma)

# Calculate P(X > 1075) (Right side / Greater than)
prob_result = 1 - prob_left

print(f"Engineering Problem Solution:")
print(f"Mean (μ) = {mu}, Std Dev (σ) = {sigma}")
print(f"Target X = {x_target}")
print(f"Z-Score Calculation: ({x_target} - {mu}) / {sigma} = {(x_target - mu)/si
print(f"P(X > {x_target}) = {prob_result:.4f} ({prob_result*100:.2f}%)")
```

```
Engineering Problem Solution:
Mean (μ) = 1000, Std Dev (σ) = 50
Target X = 1075
Z-Score Calculation: (1075 - 1000) / 50 = 1.5
P(X > 1075) = 0.0668 (6.68%)
```

6. INVERSE PROBLEM: GOING FROM PROBABILITY TO VALUE

Sometimes we are given the % and asked for the X value.

Question 2: We want to scrap the **weakest 10%** of the ropes. Below what strength should we discard?

Solution:

1. Probability: Weakest 10% means the area in the **left tail** is **0.1000**.

2. Finding Z from Table: We look for the number closest to 0.1000 inside the table. *(Usually, if there is no negative table, we look for 0.9000 and flip the sign)*. The Z value for area 0.90 is approx **1.28**. Since ours is the left tail (weakest), $Z = -1.28$.

3. Finding X (Return): Let's rewrite the Z formula in reverse: $X = \mu + (Z \cdot \sigma)$

$$X = 1000 + (-1.28 \cdot 50)$$

$$X = 1000 - 64 = 936 \text{ kg}$$

Decision: We will throw away all ropes showing strength below **936 kg**.

In [3]:
```
# Inverse Problem Check
mu = 1000
sigma = 50
```

```python
z_score = -1.28

x_threshold = mu + (z_score * sigma)

print(f"Inverse Problem Solution:")
print(f"Target Probability: Weakest 10%")
print(f"Corresponding Z-Score: {z_score}")
print(f"Calculated Threshold (X): {x_threshold} kg")
```

```
Inverse Problem Solution:
Target Probability: Weakest 10%
Corresponding Z-Score: -1.28
Calculated Threshold (X): 936.0 kg
```

7. THE 3-SIGMA RULE (EMPIRICAL RULE)

The rule of thumb engineers use to make quick estimates without looking at the table.

In a Normal Distribution:

- **68%:** Between Mean ± 1 Standard Deviation. ($\mu \pm 1\sigma$)
- **95%:** Between Mean ± 2 Standard Deviations. ($\mu \pm 2\sigma$)
- **99.7%:** Between Mean ± 3 Standard Deviations. ($\mu \pm 3\sigma$)

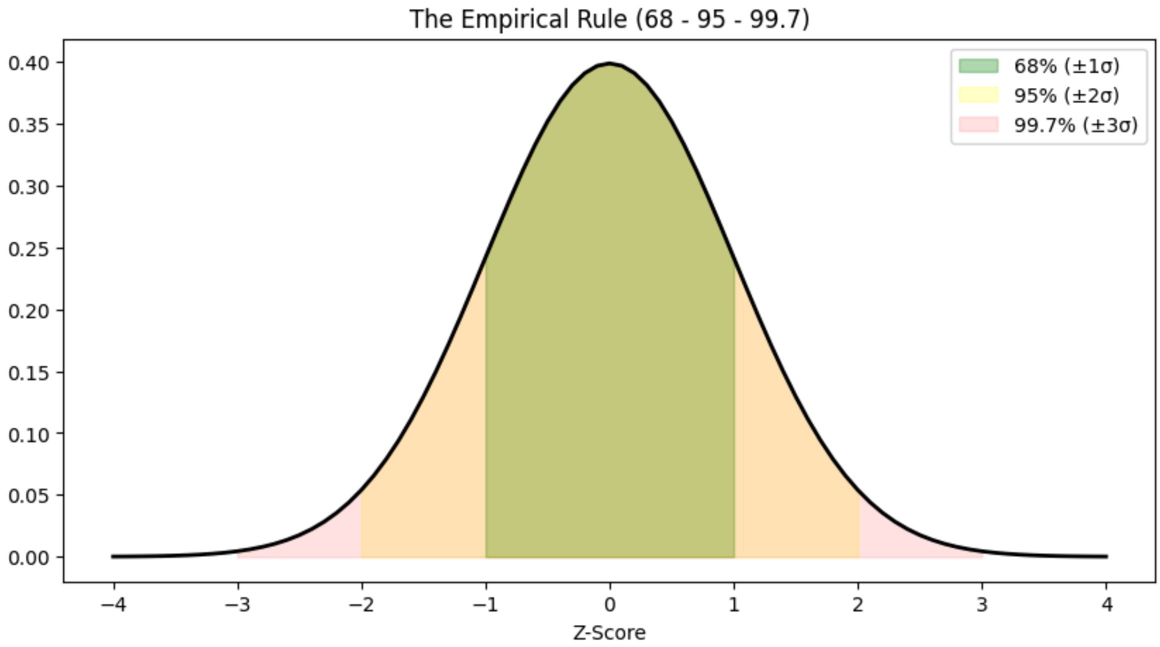
Engineering Meaning (Six Sigma): If you keep your production tolerances within $\pm 3\sigma$, **997 out of 1000** parts you produce will be good. (Only 3 will be defects). This is the basis of quality control.

In [4]:
```python
# Visualization of Empirical Rule
mu = 0
sigma = 1
x = [i * 0.1 for i in range(-40, 41)]
y = [normal_pdf(val, mu, sigma) for val in x]

plt.figure(figsize=(10, 5))
plt.plot(x, y, color='black', linewidth=2)

# Fill areas
plt.fill_between(x, y, where=[(val >= -1 and val <= 1) for val in x], color='gre
plt.fill_between(x, y, where=[(val >= -2 and val <= 2) for val in x], color='yel
plt.fill_between(x, y, where=[(val >= -3 and val <= 3) for val in x], color='red

plt.title("The Empirical Rule (68 - 95 - 99.7)")
plt.xlabel("Z-Score")
plt.legend()
plt.show()
```



8. Lecture Summary

1. **Use Symmetry:** If there is no negative value in the table, don't panic. $P(Z < -1.5)$ is the same as $P(Z > 1.5)$. Which is $(1 - \text{Table Value})$.
2. **Watch the Units:** Z-value is unitless (kg / kg). Therefore, you can compare different units (kg vs meters) using Z-scores.
3. **Continuity Correction:** Normal distribution is continuous. $P(X = 100)$ is **0**. In questions, "100 or more" and "more than 100" are the **same thing**. (This was different in discrete distributions, don't confuse them).