# Normal Distribution

Many biological and physical processes exhibit a distribution called the **normal** or **Gaussian distribution**. Values tend to cluster around a mean and extreme values are relatively rare. Each normal distribution has different units of measure but they can be normalized using the *z*-score transform (***z*** = (**X -** m**)**/s). This defines the **standard normal distribution** or ***z*-distribution.**

## Definition:

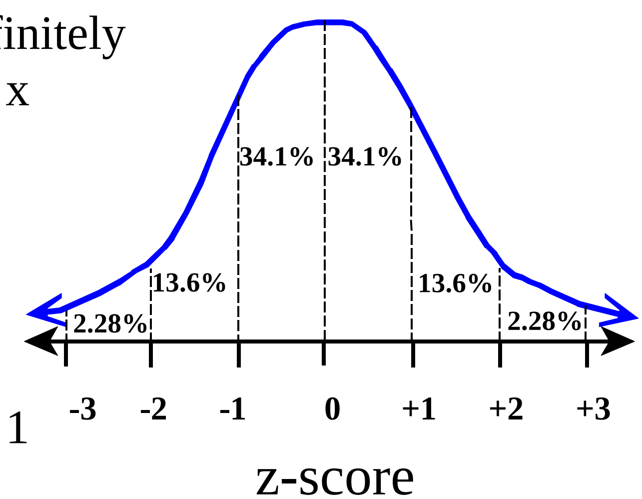
$$y = \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}}$$

**e**$^x$ is called the **exponential function**. **e**$^{1}$, called Euler's number, is a transcendental number equal to:

S **1/n!** @ **2.718 281 828 459 045 1...**

## Properties of Standard Normal Distribution:

1. "bell" shaped, unimodal distribution
2. mean, median and mode are the same and equal to 0
3. standard deviation is equal to 1
4. symmetric about the mean
5. continuous function (infinitely differentiable, for every x there is a single y)
6. asymptotic to the x-axis at both ends (y values approach but never become zero)
7. area under curve equals 1



# Applications of Normal Distribution 1

## Uses:

- **computing areas and percentiles of scores that are "normal distributed"**
- **testing hypotheses concerning means of different populations (are they the same or different?)**

## Examples:

1. **Find the percentage of scores between +/–1 standard deviations.**

    area between 0 and $+1\sigma = 0.3413$

    area between 0 and $-1\sigma = 0.3413$

    area between $-1\sigma$ and $+1\sigma = 2 \times 0.3413 = 0.6826 = 68.3\%$

2. **Find the percentage of scores between +/– 2 and +/–3 standard deviations.**

    area between 0 and $+2\sigma = 0.4772$

    area between $-2\sigma$ and $+2\sigma = 2 \times 0.4772 = 0.954 = 95.4\%$

    area between $-3\sigma$ and $+3\sigma = 2 \times 0.4987 = 0.997 = 99.7\%$

3. **Find the z-score that defines 95% of scores around the mean.**

$$\frac{95\%}{2} = 47.5\% = 0.475$$

$$\text{from table } 0.475 \rightarrow z = 1.96$$

$$\text{therefore } 95\% \rightarrow z = \pm 1.96$$

$$\text{or } -1.96 \leq z \leq +1.96$$

# Applications of Normal Distribution 2

**4. Find *z*-score that defines the lower 90$^{th}$ percent.**

$$90\% - 50\% = 40.0\% = 0.400$$
$$\text{from table: } 0.400 \rightarrow z \approx 1.29$$
$$-\infty \leq z \leq +1.29$$

**5. Find 25$^{th}$ and 75$^{th}$ percentile z-scores.**

$$25\% = 0.25$$
$$\text{from table: } 0.250 \rightarrow z \approx 0.68$$
$$z_{25} = -0.68 \text{ and } z_{75} = +0.68$$

**6. College wants top 15% of students who take a test which has a mean of 125 and standard deviation of 25.**

$$50\% - 15\% = 35\% = 0.35$$
$$\text{from table: } 0.35 \rightarrow z \approx 1.04$$
$$\text{Since, } X = m + zs$$
$$\text{therefore, } X = 125 + 1.04(25) = 151.0$$
$$\text{Thus, } X \geq 151.0$$

**7. Determine the 5$^{th}$ and 95$^{th}$ percentile heights of a population that has a mean of 150.0 ±20.0 cm.**

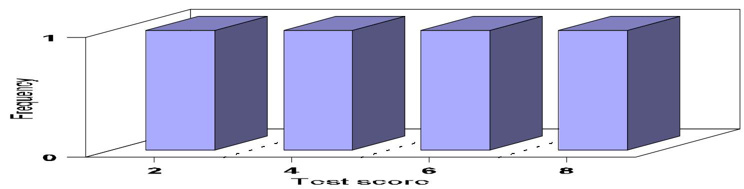
$$95\% - 50\% = 45\% = 0.450$$
$$\text{from table: } 0.450 \rightarrow z \approx 1.65$$
$$\text{Since, } X = m + zs$$
$$\text{therefore, } X_{95} = 150 + 1.65(20.0) = 183.0 \text{ cm}$$
$$\text{and } X_5 = 150 - 1.65(20.0) = 117.0 \text{ cm}$$
$$\text{Thus, } 117.0 \text{ cm} \leq X \leq 183.0 \text{ cm}$$

# Central Limit Theorem 1

## Sampling Distribution of Sample Means:

distribution based on the **means of random samples** of a specified size (***n***=constant) taken from a population

**Example:** Test scores from a class of four students. Scores were 2,4,6,8 (uniform distribution)

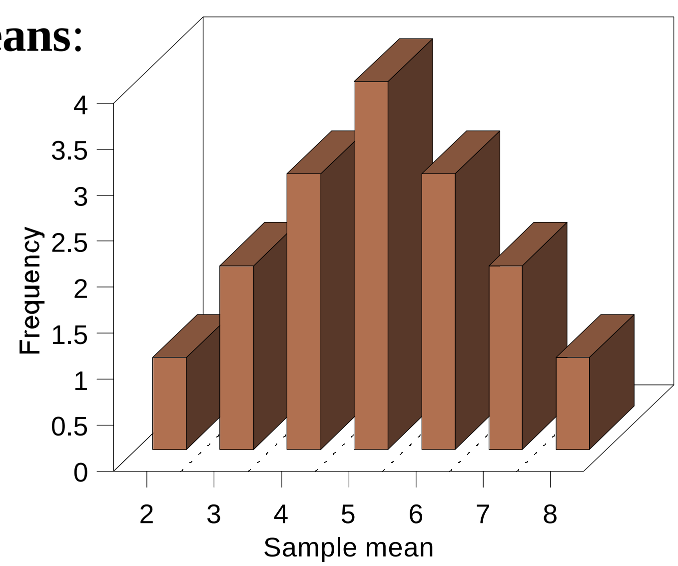



**List all possible samples of size *n*=2, allowing replacement.**

| Sample | Mean | Sample | Mean |
|---|---|---|---|
| 2,2 | 2 | 6,2 | 4 |
| 2,4 | 3 | 6,4 | 5 |
| 2,6 | 4 | 6,6 | 6 |
| 2,8 | 5 | 6,8 | 7 |
| 4,2 | 3 | 8,2 | 5 |
| 4,4 | 4 | 8,4 | 6 |
| 4,6 | 5 | 8,6 | 7 |
| 4,8 | 6 | 8,8 | 8 |

**Sampling distribution of means**:

| Mean | frequency |
|---|---|
| 2 | 1 |
| 3 | 2 |
| 4 | 3 |
| 5 | 4 |
| 6 | 3 |
| 7 | 2 |
| 8 | 1 |



# Central Limit Theorem 2

As sample size (***n***) increases the shape of the sampling distribution of sample means taken from a population with mean, m, and standard deviation, s, will approach a **normal distribution**, with mean, m, and standard deviation, $\frac{\sigma}{\sqrt{n}}$.

- the standard deviation of the sampling distribution is called the **standard error of the mean** (notice, by definition, it is always less than sample's standard deviation when $n > 1$)

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

Note, whenever the sample size (***n***) exceeds 5% of the population size (***N***) the standard error must be adjusted by the **Finite Population Correction Factor:**

$$\sqrt{\frac{N-n}{N-1}}$$

That is,

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

## Example:

What is the standard error of the mean for a sampling distribution given a sample of size of 100 and s.d. of 5.00 taken from a population of size, 1000.

$$\sigma_{\overline{x}} = \frac{5.00}{\sqrt{100}} \times \sqrt{\frac{1000-100}{1000-1}} = 0.500 \times \sqrt{\frac{900}{999}} = 0.475$$