

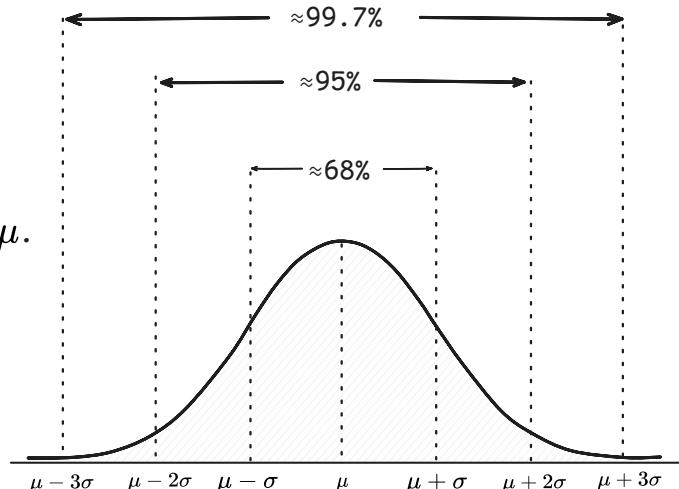
3 - Position in distribution [ES 2.5, PS 2.1, 2.2]

Normal distribution and Empirical Rule [ES 2.4, PS 2.1]

Many datasets naturally form a bell-shaped distribution called the **normal distribution** whose shape is completely determined by the mean and the standard deviation.

Examples:

- scores or tests taken by many people (SAT exams, IQ tests)
- repeated careful measurements of the same quantity
- characteristics of many biological populations (cricket length, corn yields)



68-95-99.7 Rule.

In a normal distribution with mean μ and standard deviation σ :

$\approx 68\%$ of the data fall within σ of the mean μ .

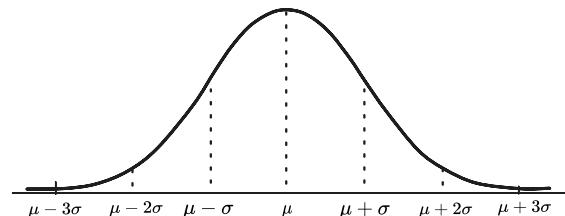
$\approx 95\%$ of the data fall within 2σ of μ .

$\approx 99.7\%$ of the data fall within 3σ of μ .

Example 1. NCHS survey finds women (US, age 20-29) have mean height of 64.2 inches and standard deviation of 2.9 inches. Estimate the percent of women whose heights are between 58.4-64.2 inches.

Answer. Distribution of women's heights is roughly bell-shaped.

Also, 58.4 inches = mean - 2 pop.sd
So we want the region from $\mu - 2\sigma$ to μ .



This is $95/2 = 47.5\%$ of women by the 68-95-99.7 rule.

Example 2. Men (US, age 20-29) heights follow a rough bell-shape with mean 69.4 inches and standard deviation of 2.9 inches.

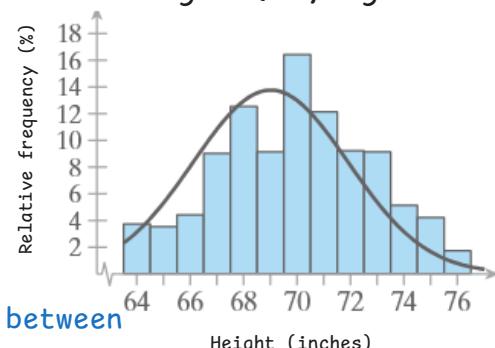
- Estimate the two heights containing the middle 95% of the data.
- Is a 25 year old with height 74 inches unusual?

Answer(a): By 68-95-99.7 rule, 95% of data are between

$$\mu + 2\sigma = 69.4 + 2 \times 2.9 = 75.2 \text{ inches}$$

$$\mu - 2\sigma = 69.4 - 2 \times 2.9 = 63.6 \text{ inches}$$

Men's height (US, age 20-29)



Standard normal distribution, z-score [ES 2.5, PS 2.1]

We can standardize datasets to compare them.

\vec{x} Centering $\Rightarrow \vec{x} - \mu$ Mean = 0 Standardizing $\Rightarrow \frac{1}{\sigma}(\vec{x} - \mu)$
 Original dataset Mean = 0
pop.sd = 1

The **standard score** or **z-score** of a data point x is $z = \frac{x - \mu}{\sigma}$.

Example 3. The dataset $\vec{x} = (x_1, x_2) = (-1, 3)$ has mean 1 and pop.sd 2. So its standardization is $\frac{1}{2}(-1) + \frac{1}{2}(3) = (0, 1)$.

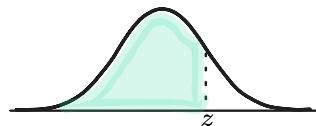
$$\frac{1}{\sigma}(\vec{x} - \mu) = \frac{1}{2}[(-1, 3) - (1, 1)] = \frac{1}{2}(-2, 2) = (-1, 1)$$

z-score of x_1 z-score of x_2

Standardizing a normal distribution gives the standard normal distribution.

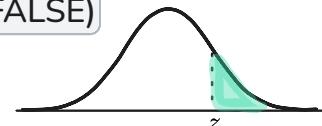
`pnorm(z)`

finds the area to
the left of z .



```
pnorm(z, lower.tail=FALSE)
```

finds the area to
the right of z .



Example 4. Estimate $\text{pnorm}(2)$ without electronics.

In the standard normal distribution,

$z=2$ is 2 standard deviations out.

The area to the left of $z=0$ is 50% of area.

The area from $z=0$ to $z=2$ is $(95)/2 = 47.5$ by the 68-95-99.7 rule.

Answer: $50\% \pm 47.5\% = 97.5\%$ or 0.975 .

