Module 5D: A Parametric Test

Z-scores & RNAseq analysis

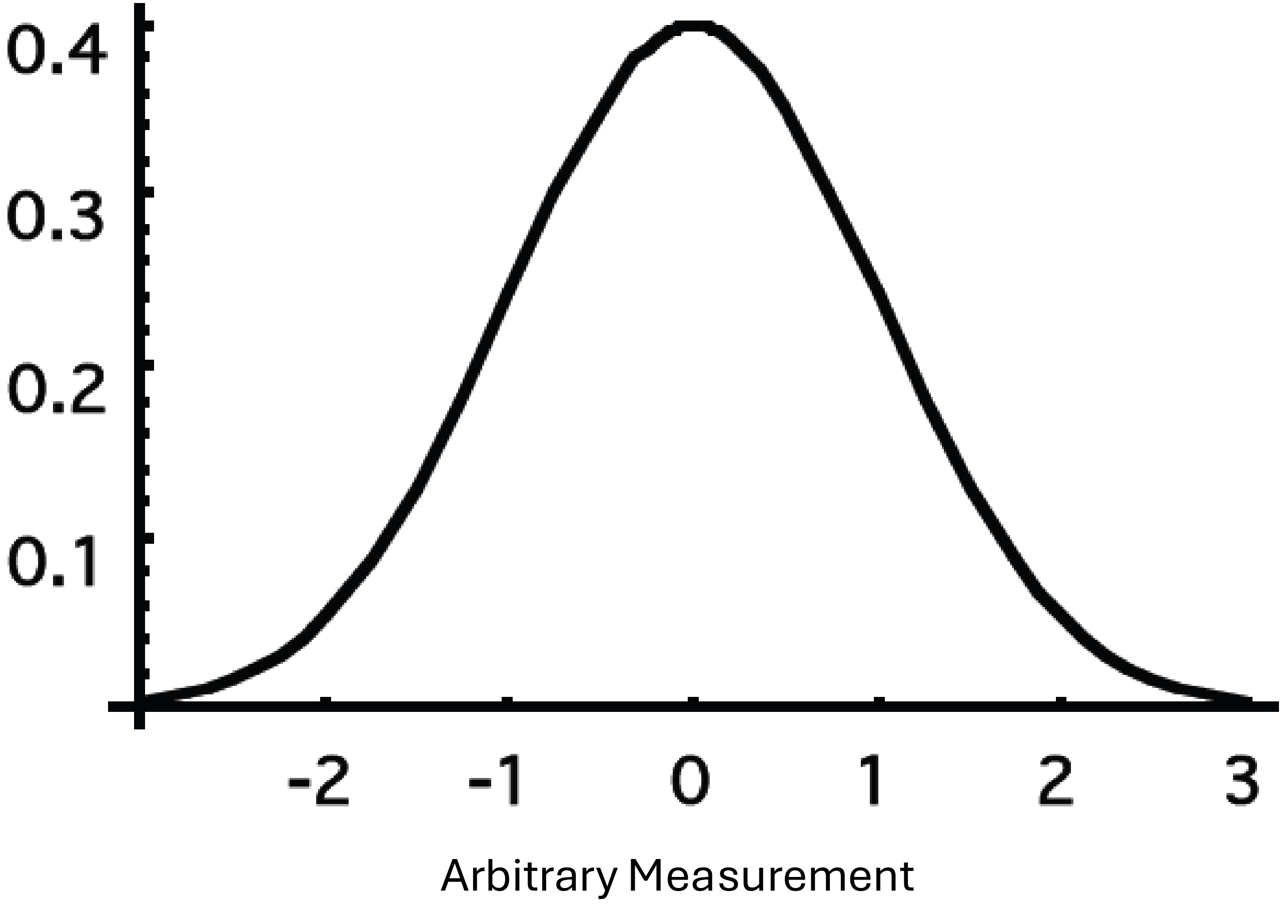
<u>Agenda:</u>

1. Z-scores
2. RNASeq

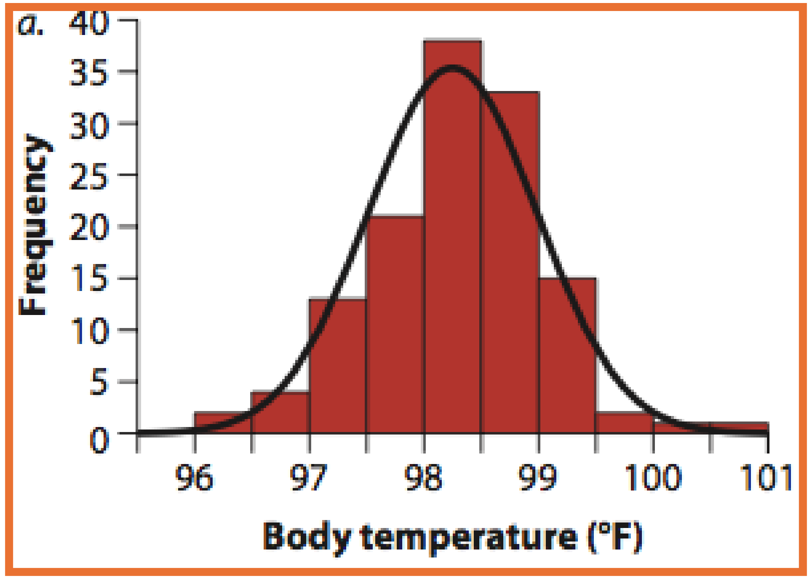
The Normal Distribution

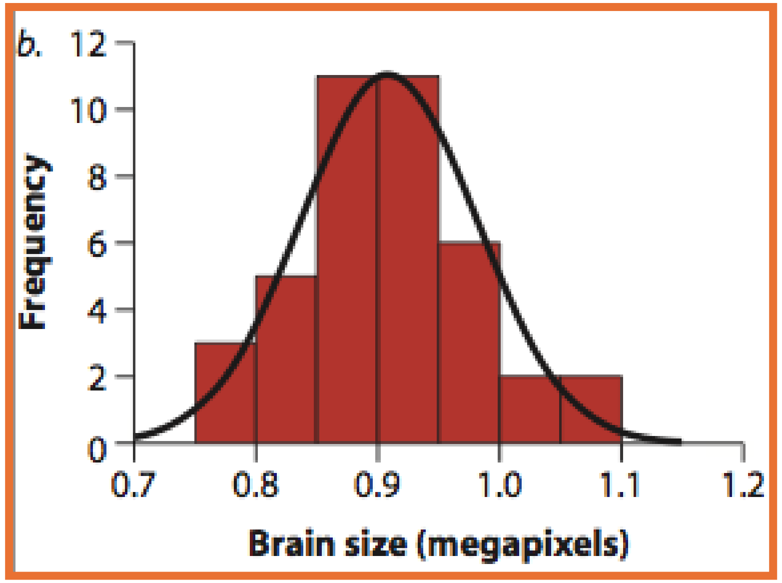
The Normal Distribution:

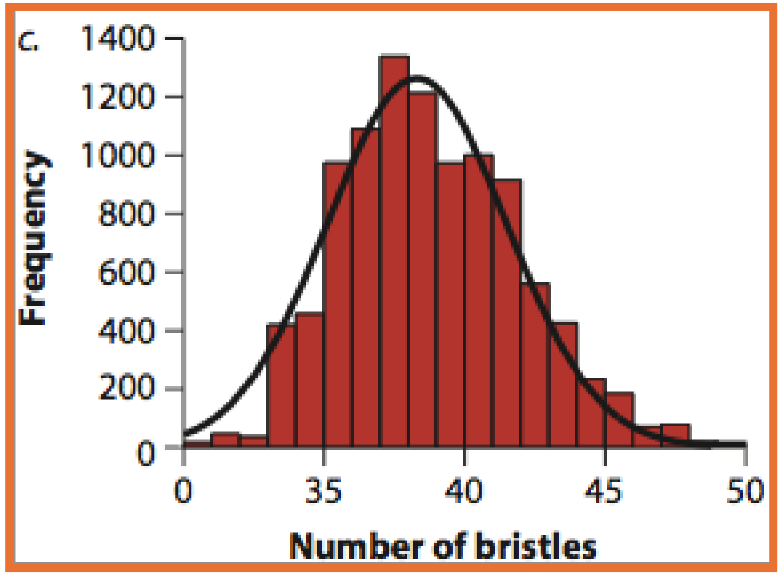
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



The Normal Distribution is common in nature:

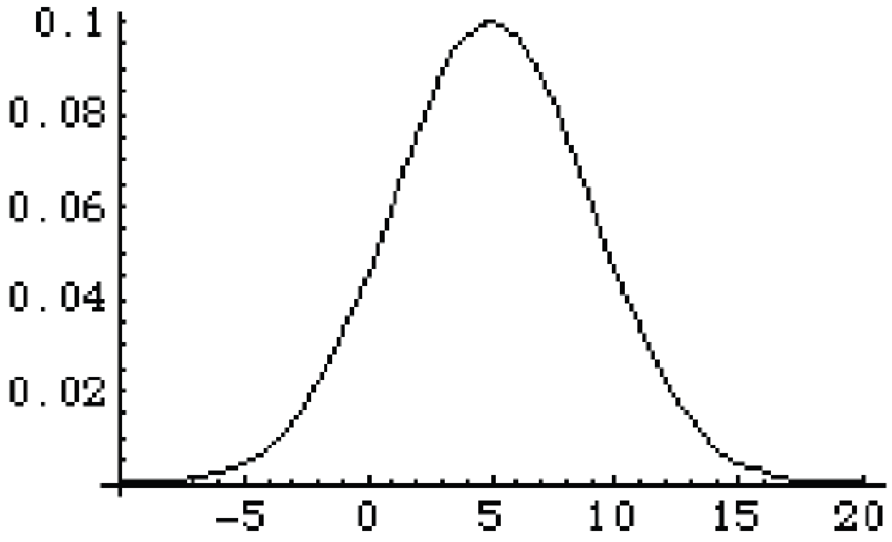




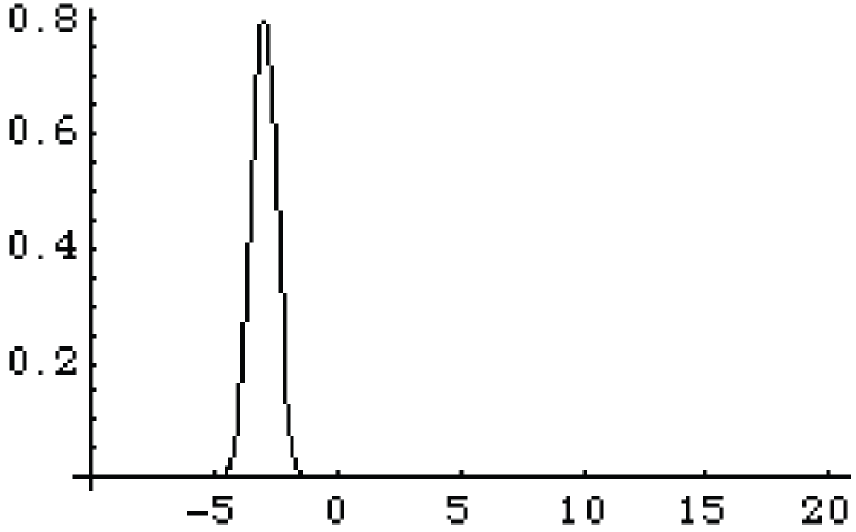


Properties of the Normal Distribution:

1. Fully described by its mean and standard deviation



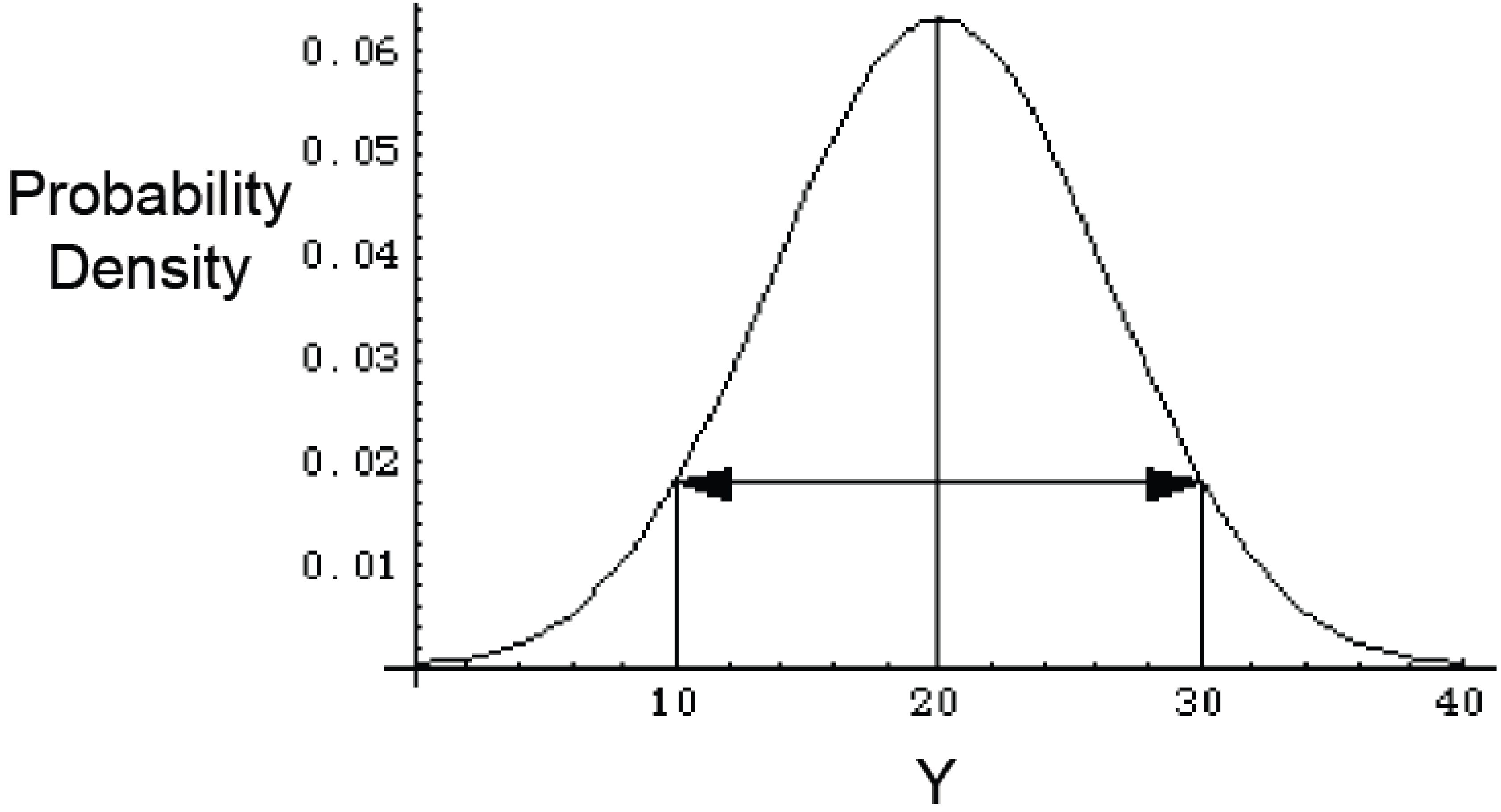
$\mu = 5; \sigma = 4$



$\mu = -3; \sigma = 1/2$

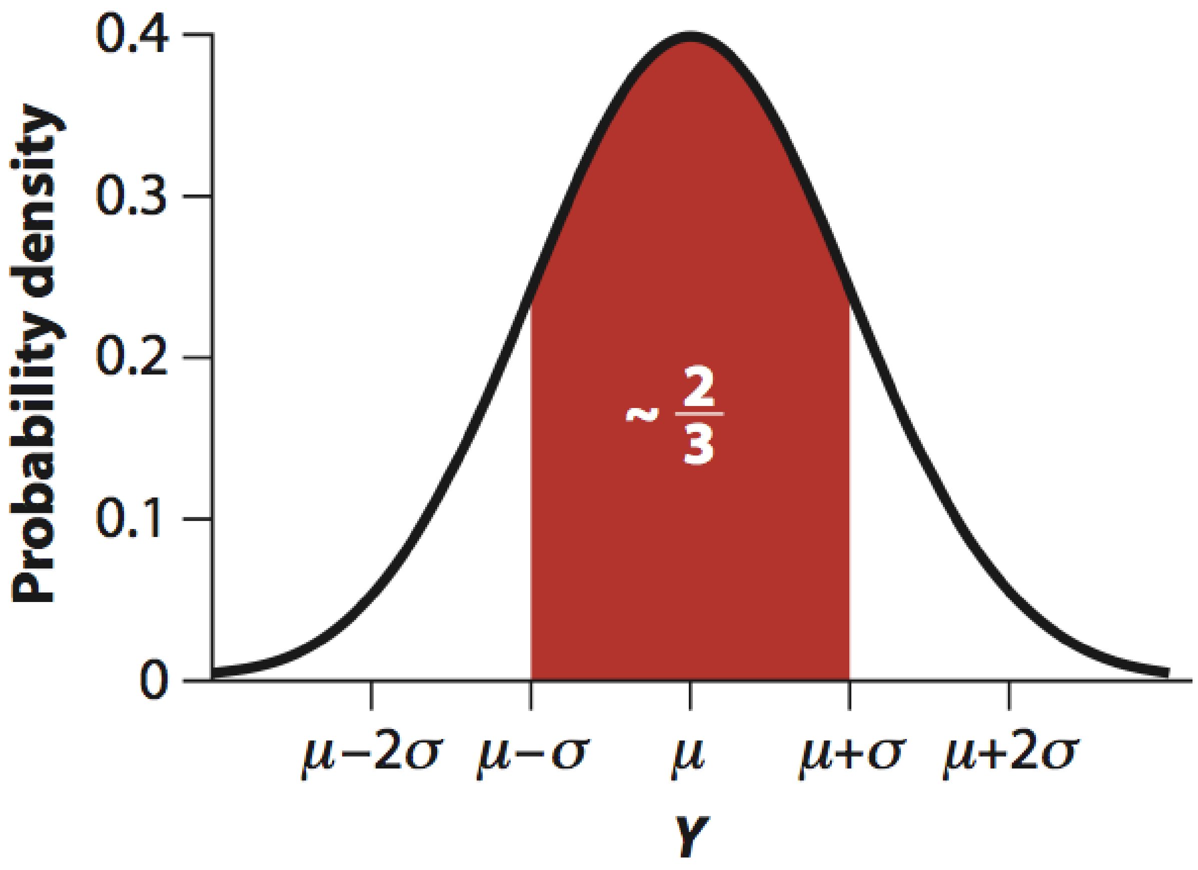
Properties of the Normal Distribution:

1. Fully described by its mean and standard deviation
2. Symmetric around its mean



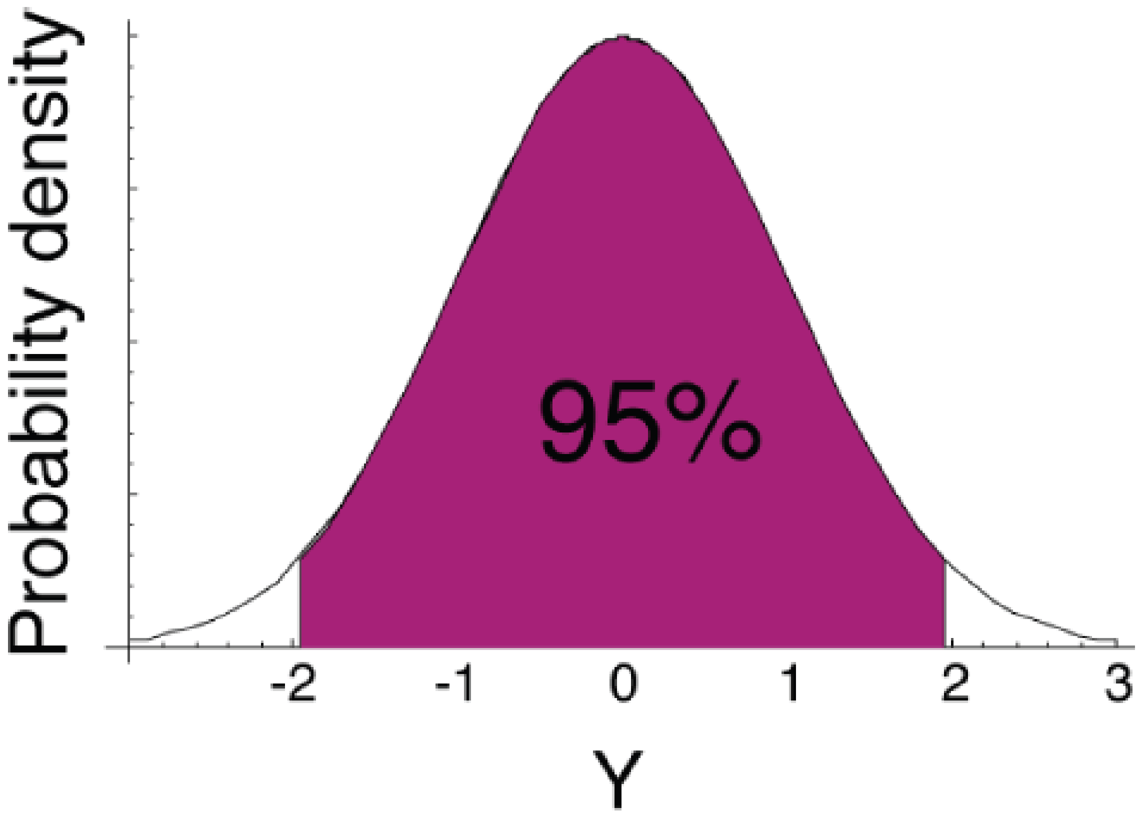
Properties of the Normal Distribution:

1. Fully described by its mean and standard deviation
2. Symmetric around its mean
3. ~ 2/3 of random draws are within one standard deviation of the mean

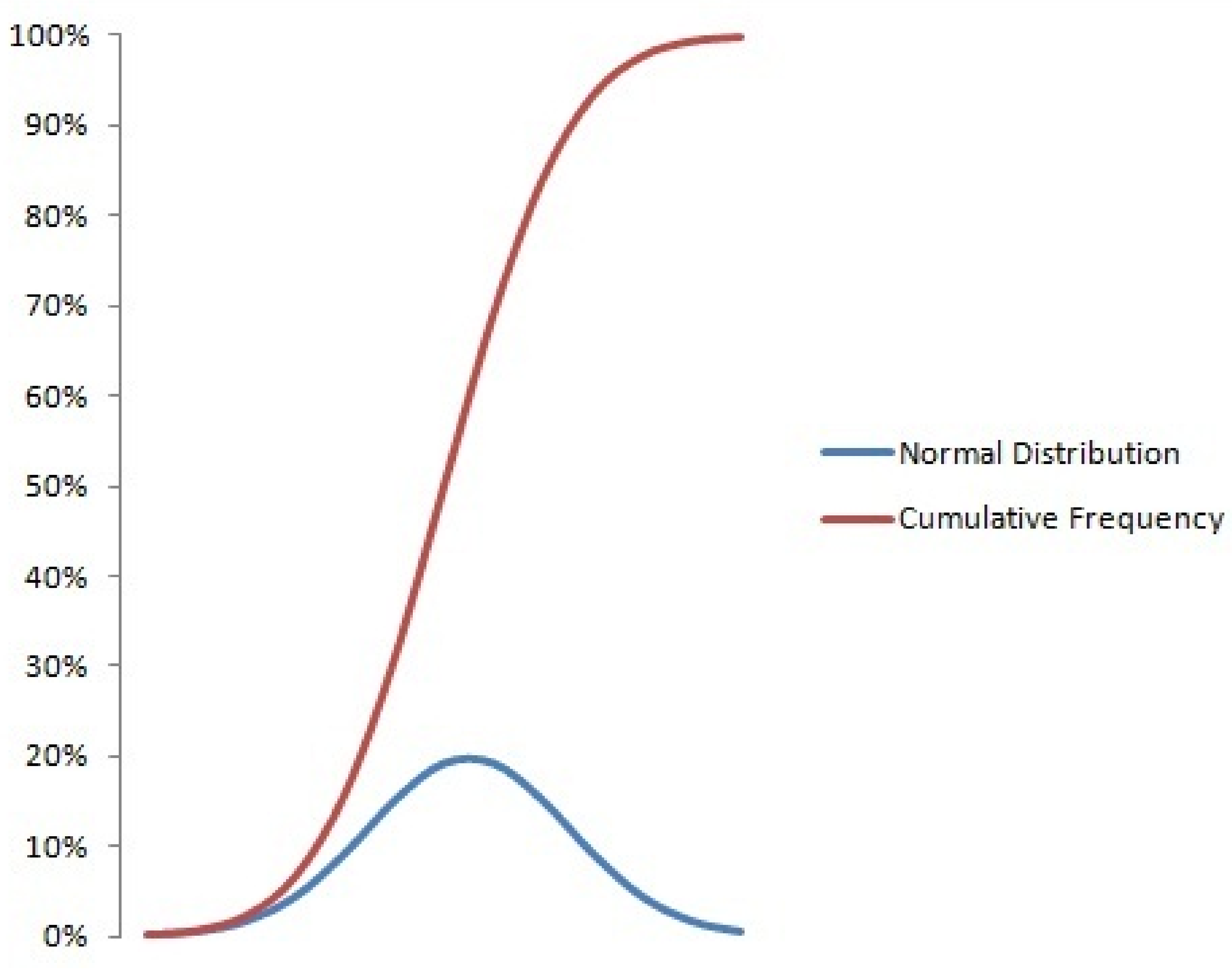


Properties of the Normal Distribution:

1. Fully described by its mean and standard deviation
2. Symmetric around its mean
3. ~ 2/3 of random draws are within one standard deviation of the mean
4. ~ 95% of random draws are within two standard deviations of the mean (*really, it is 1.96 SD*)

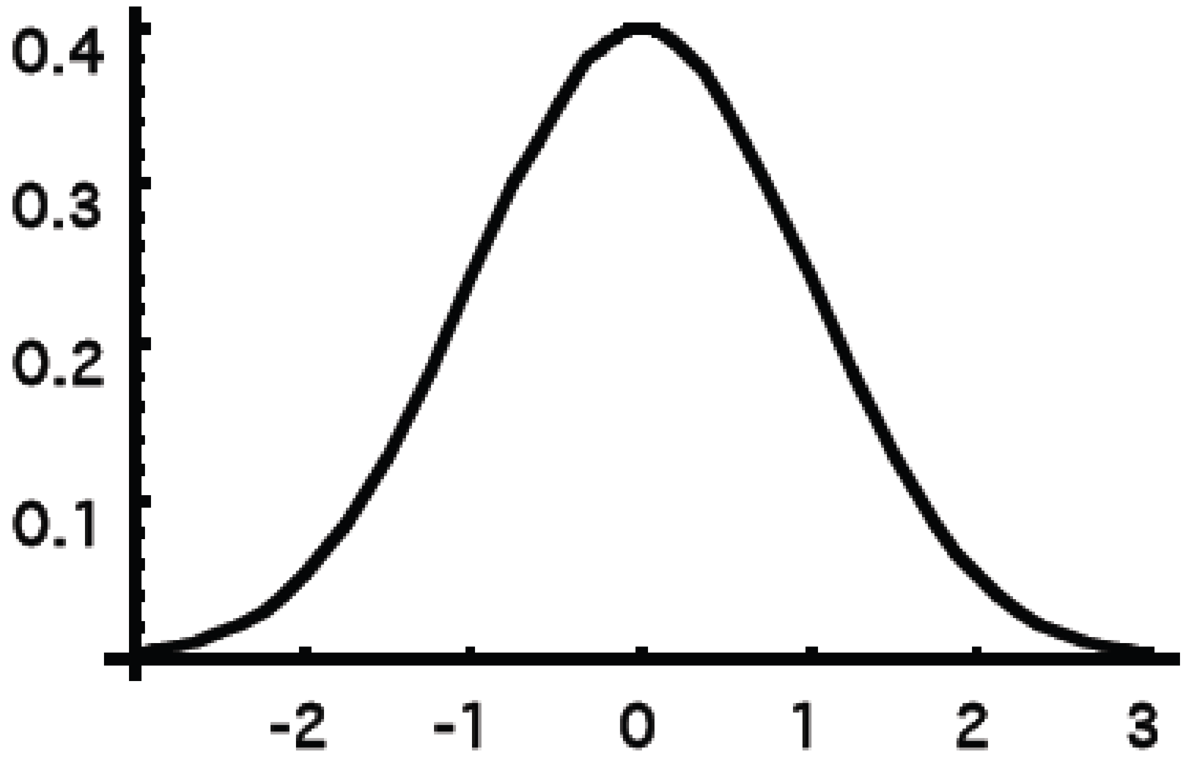


Properties of the Normal Distribution:



The Standard Normal Distribution:

- Mean is zero ($\mu = 0$)
- Standard deviation is 1 ($\sigma = 1$)



Z-scores:

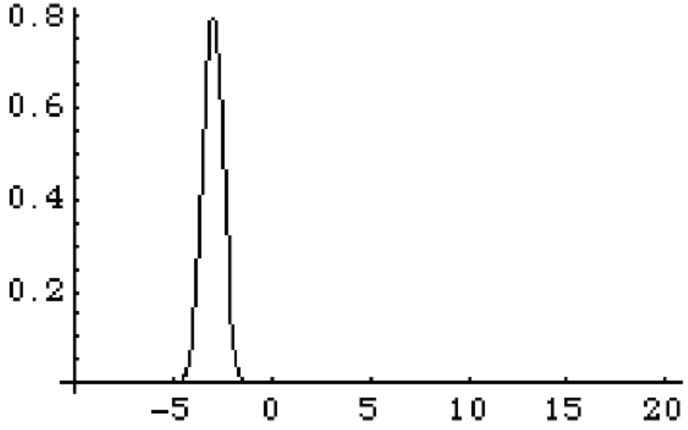
- converts **raw** normally distributed scores into standard deviation units
  - useful for comparing distributions with different scales, for instance.
  - percentiles

- allows calculation of probability of variable value

- z-score indicates how far above or below the mean a value is in standard deviation units
  - how large/small is individual score **relative** to others in the distribution

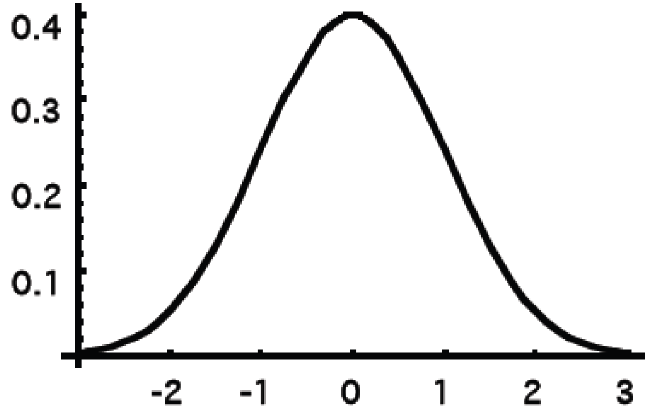
The Standard Normal Distribution:

- Mean is zero ($\mu = 0$)
- Standard deviation is 1 ($\sigma = 1$)



$\mu = -3$; $\sigma = 1/2$

$$\boxed{Z = \frac{X_i - \mu}{\sigma}}$$



Interpret the following statements:

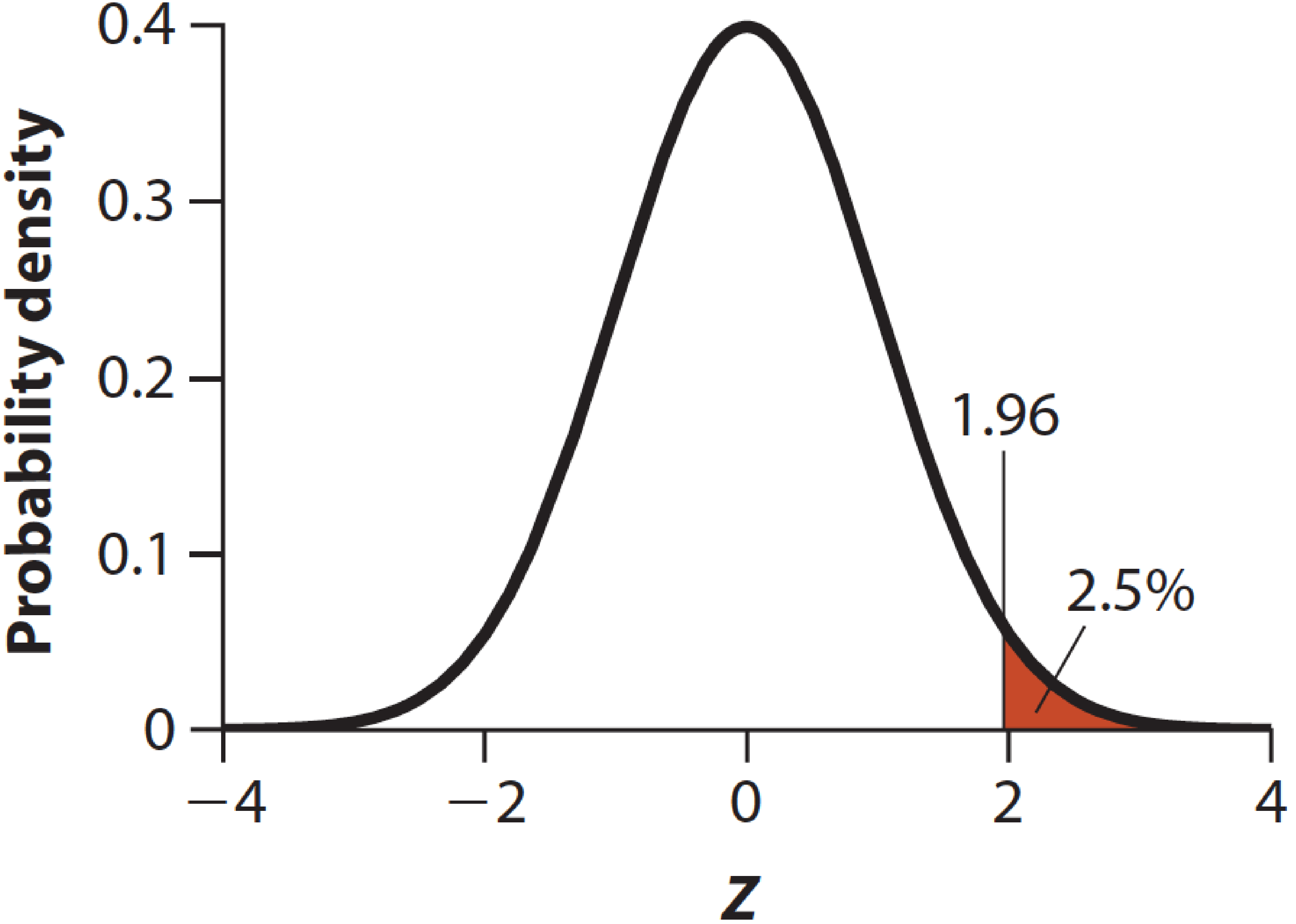
- **Student A gets a *z-score* of -1.5 on an exam**

- **Student B received a *z-score* of 0.29 on the exam**

  - Does the *z –score* tell you sample size?

  - What the mean score on the test was?

  - The percentage of answers Student B got right?



The probability of getting a random draw from a standard normal distribution greater than a given value which is the area under the curve.

Mechanics of Z tables:

The table works for P[Z>a.bc]. Not all tables are set up the same way!

First two digits of *a.bc*	Second digit after decimal (*c*) 0	1	2	3	4	5	6	7	8	9
1.6	0.05480	0.05370	0.05262	0.05155	0.05050	0.04947	0.04846	0.04746	0.04648	0.04551
1.7	0.04457	0.04363	0.04272	0.04182	0.04093	0.04006	0.03920	0.03836	0.03754	0.03673
1.8	0.03593	0.03515	0.03438	0.03362	0.03288	0.03216	0.03144	0.03074	0.03005	0.02938
1.9	0.02872	0.02807	0.02743	0.02680	0.02619	0.02559	$\boxed{0.02500}$	0.02442	0.02385	0.02330
2.0	0.02275	0.02222	0.02169	0.02118	0.02068	0.02018	0.01970	0.01923	0.01876	0.01831
2.1	0.01786	0.01743	0.01700	0.01659	0.01618	0.01578	0.01539	0.01500	0.01463	0.01426

For Z = 1.96 → P[Z>1.96]=0.025

https://www.z-table.com/

Since the standard normal is symmetric:

P[Z > x] = P[Z < -x]

P[Z < x] = 1 - P[Z > x]

remember: instead of α, you have $\alpha/2$ at each tail

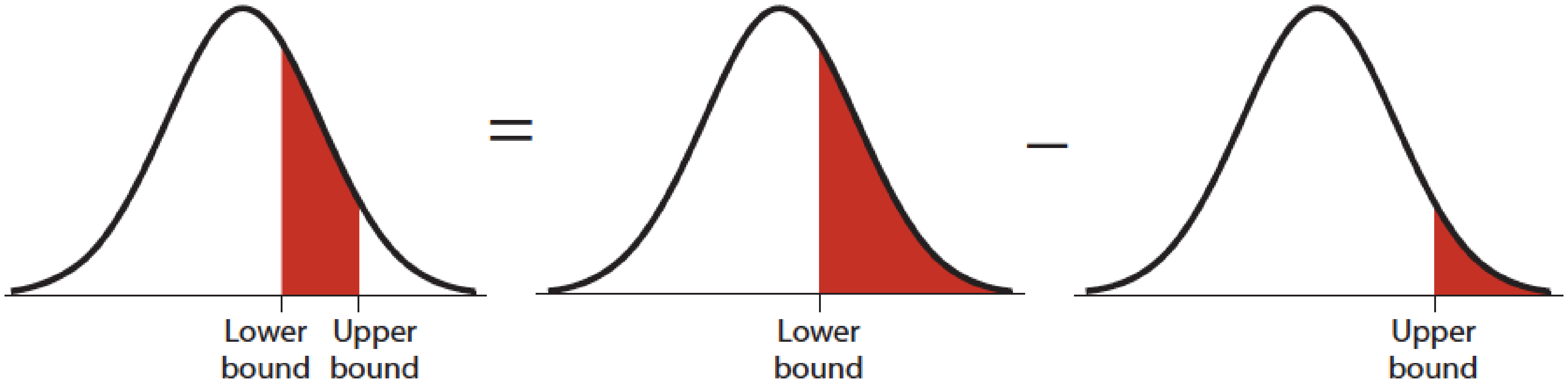
Example:

P[Z < -1.96] = P[Z > 1.96]

Example:

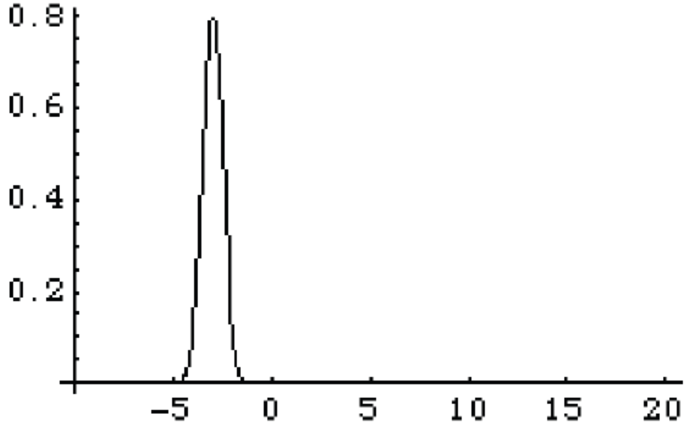
P*[lower bound < Z < Upper bound]* = P[Z > *lower bound*] - P[Z > *Upper bound*]

$$P(a<Z<b)=F\left(\frac{b-\mu}{\sigma}\right)-F\left(\frac{a-\mu}{\sigma}\right)$$



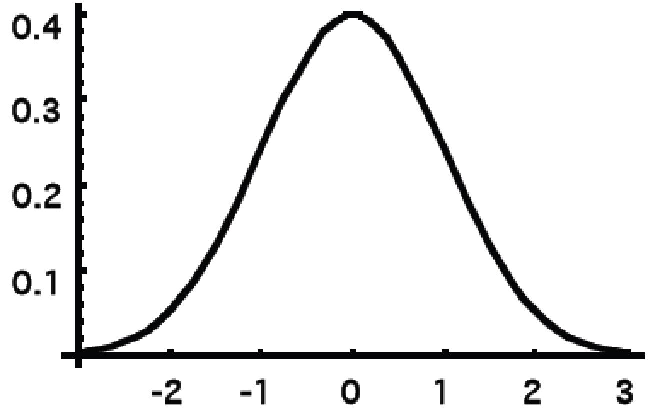
The Standard Normal Distribution:

- Mean is zero (μ = 0)
- Standard deviation is 1 (σ = 1)



μ = -3; σ = 1/2

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



3 major motivations:

- Z score tells us how many standard deviations our normally distributed variable is from the mean

$$Z = \frac{\text{Raw score - Mean}}{\text{Standard deviation}}$$

- Confidence interval: $(a,b) = \bar{x} \pm z_{\alpha/2}(\sigma / \sqrt{2})$

- Determine proportion of scores that fall between two raw scores

- Allows us to use standard normal table