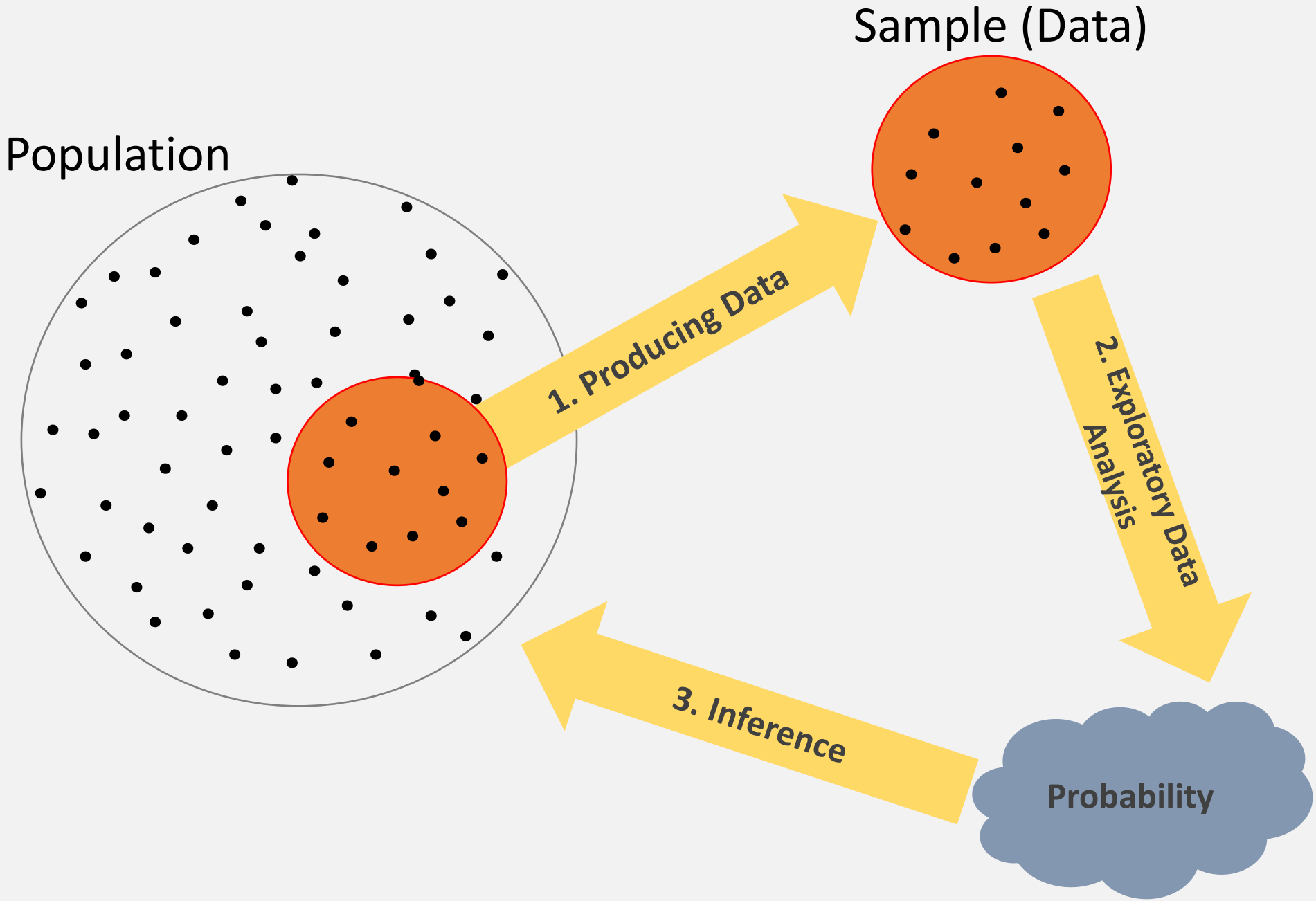
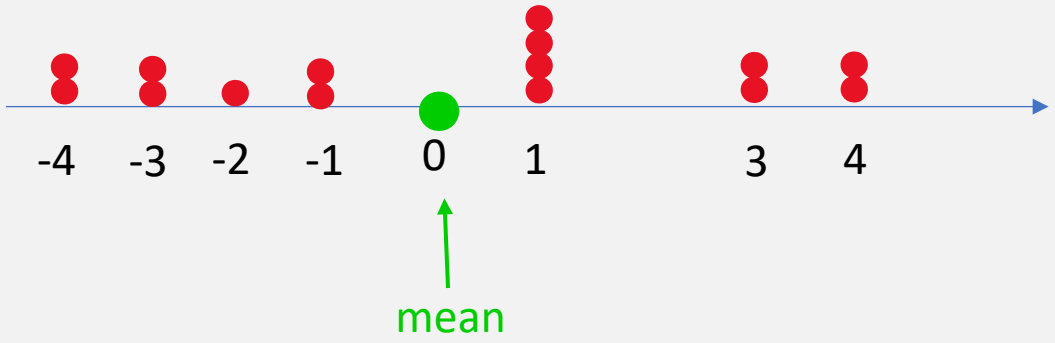

Sample (Data)
Population
1. Producing Data
2. Exploratory Data Analysis
3. Inference
Probability

Measures of Central Tendency

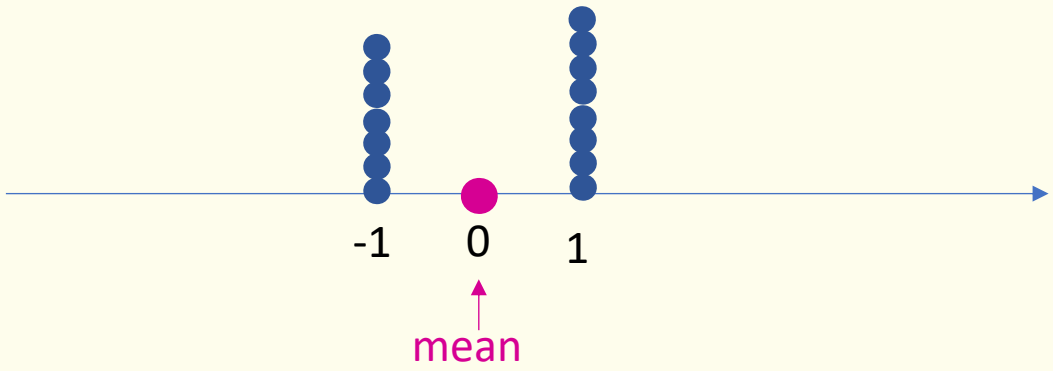
$v = [-4, -4, -3, -3, -2, -1, -1, 1, 1, 1, 1, 3, 3, 4, 4]$

$meav(v) = 0, \qquad median = 1$

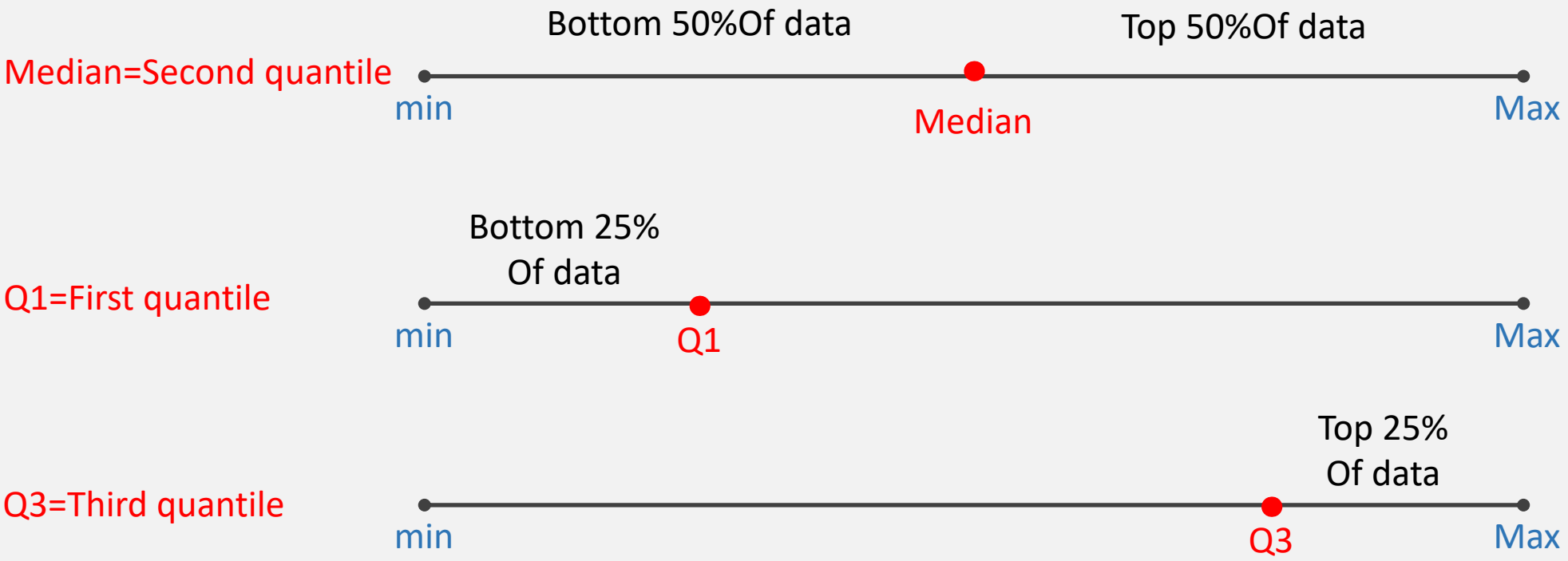


$w = [-1, -1, -1, -1, -1, -1, -1, 1, 1, 1, 1, 1, 1, 1, 1]$

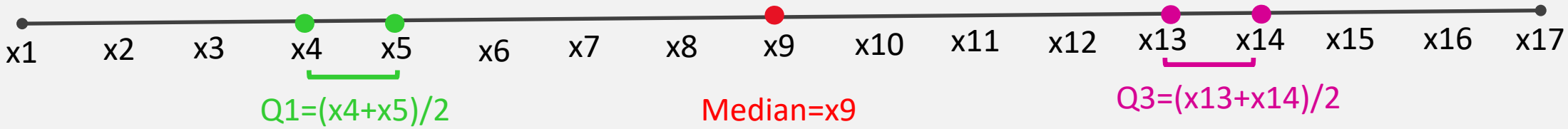
$meav(w) = 0, \qquad median(w) = 1$

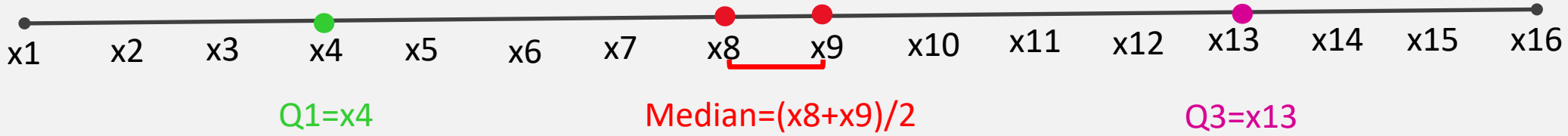


Quantiles

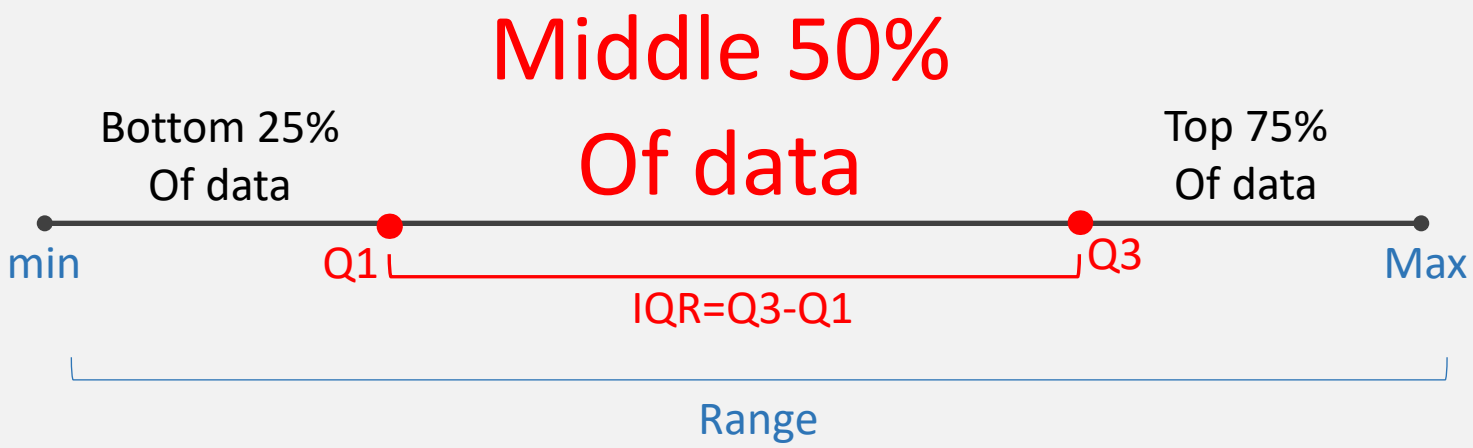

Median=Second quantile
Bottom 50%Of data
Top 50%Of data
min
Median
Max
Q1=First quantile
Bottom 25%
Of data
min
Q1
Max
Q3=Third quantile
Top 25%
Of data
min
Q3
Max

Quantiles

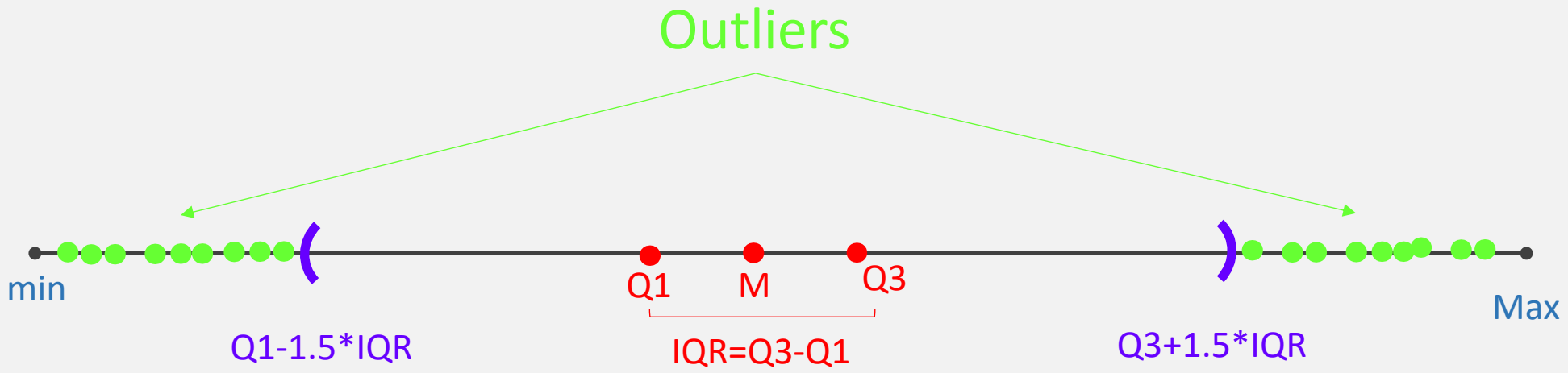

x1
x2
x3
x4
x5
x6
x7
x8
x9
x10
x11
x12
x13
x14
x15
x16
x17
Q1=(x4+x5)/2
Median=x9
Q3=(x13+x14)/2

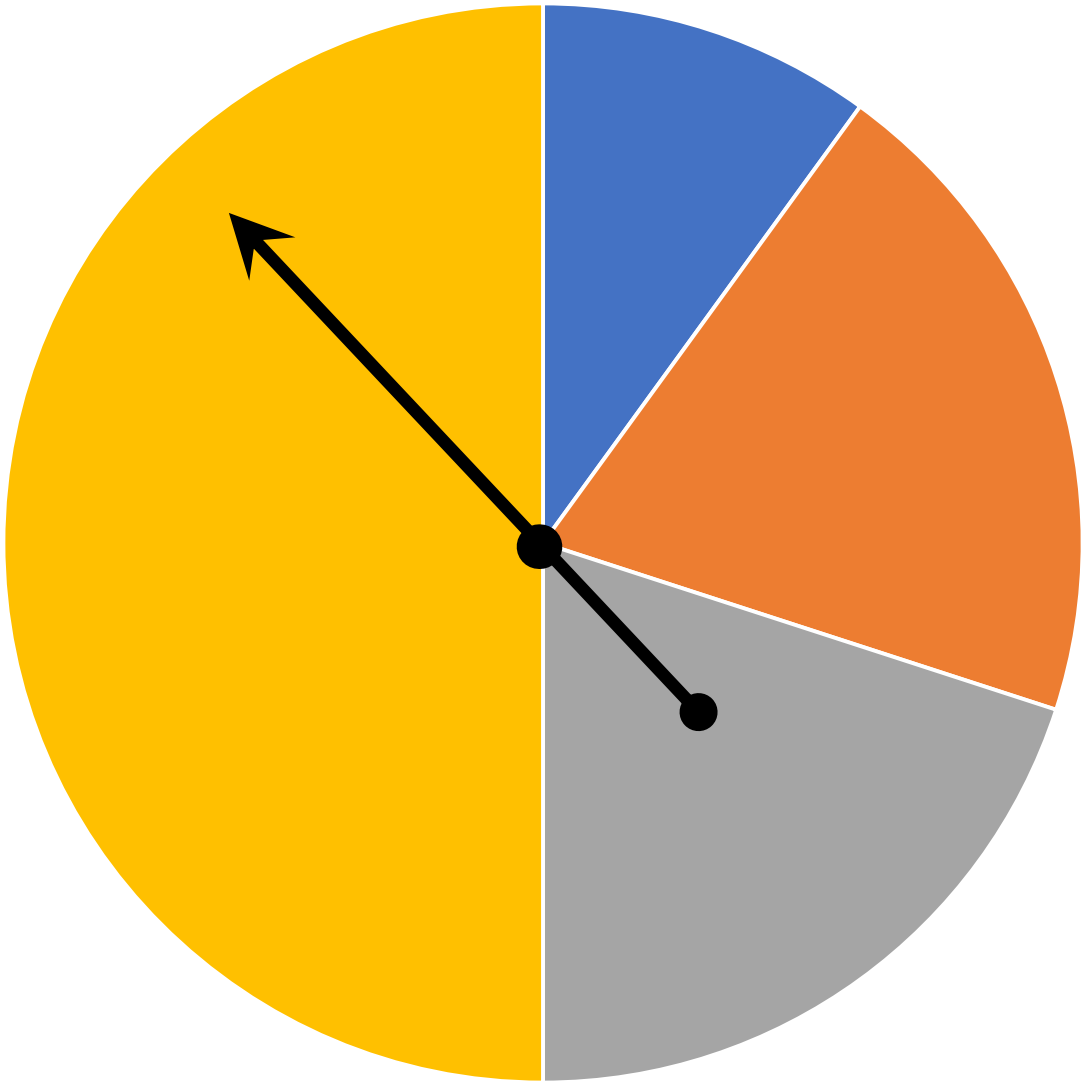

x1
x2
x3
x4
x5
x6
x7
x8
x9
x10
x11
x12
x13
x14
x15
x16
Q1=x4
Median=(x8+x9)/2
Q3=x13

Interquartile Range (IQR)

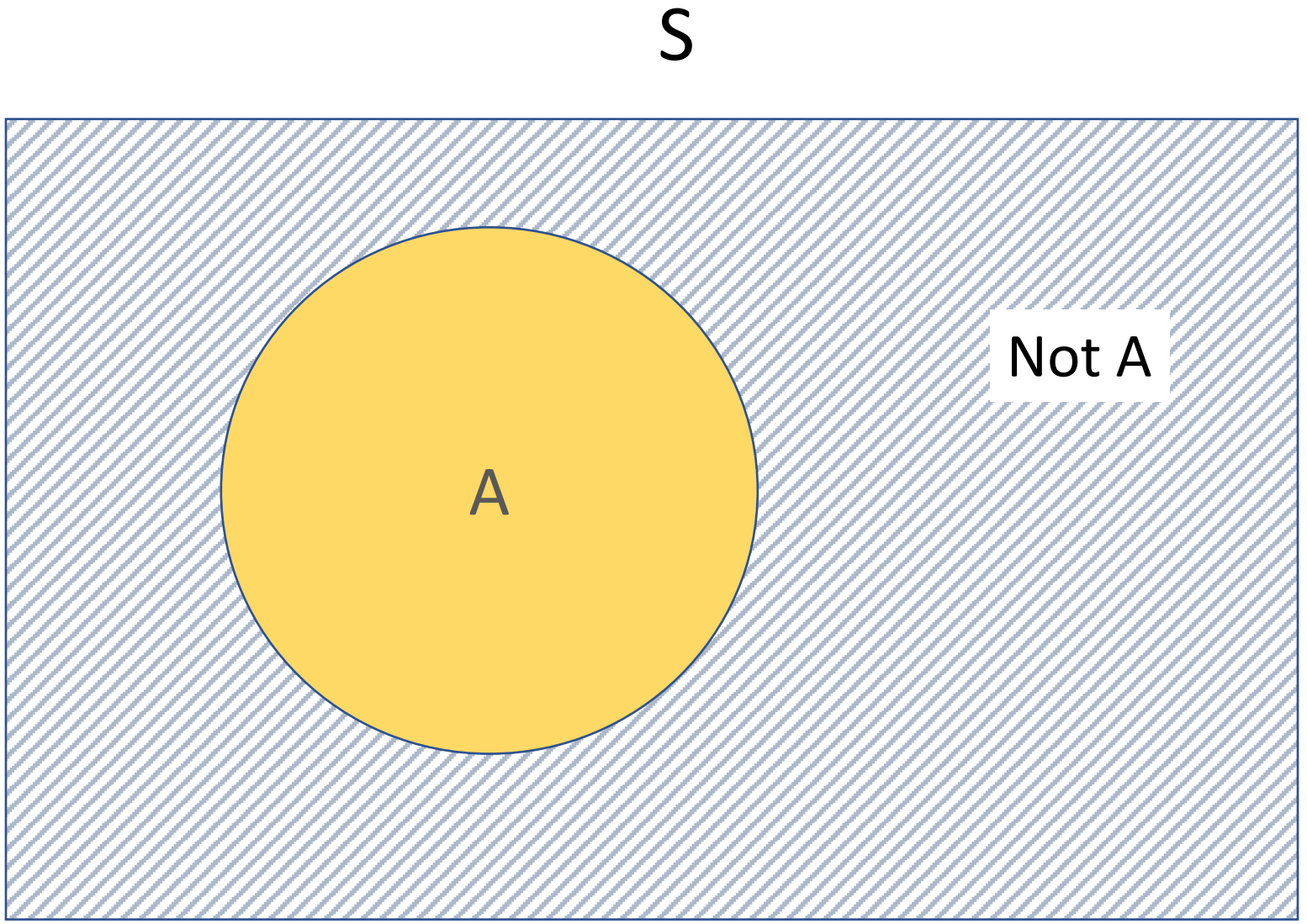


Outliers

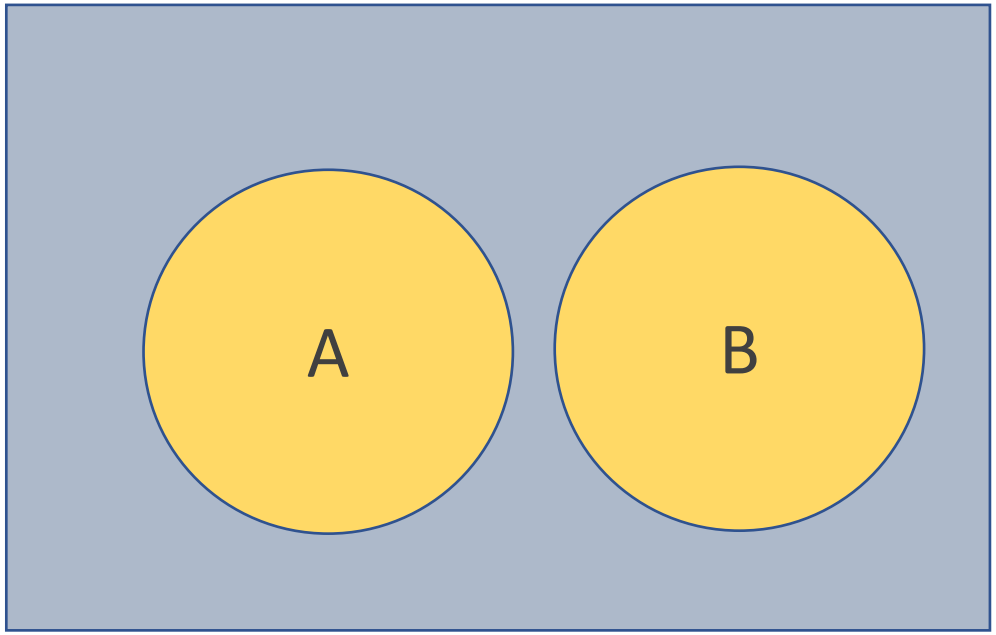




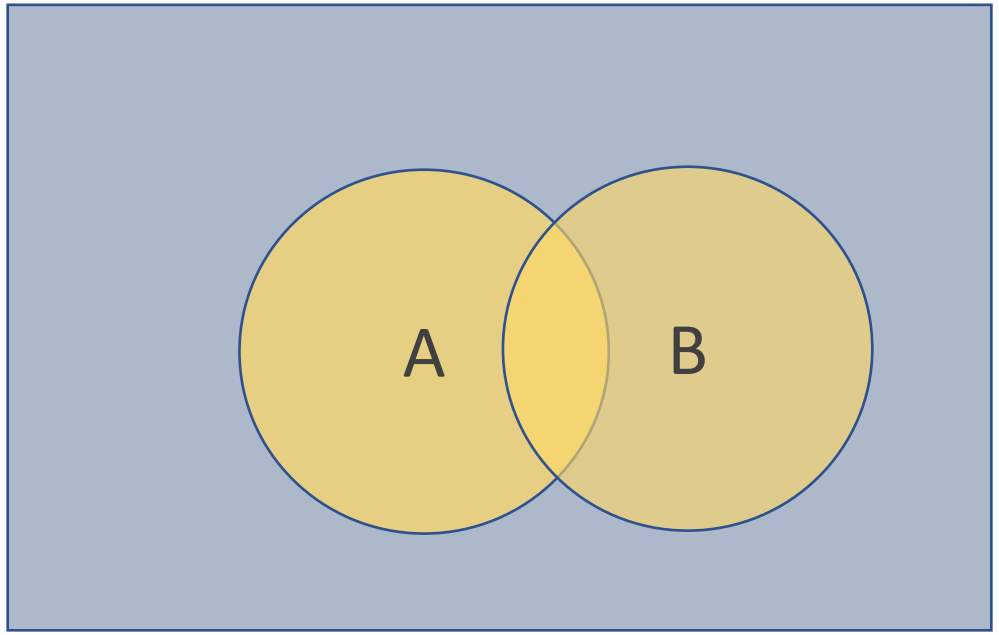
Highest Level of Education Attained	Probability
Below high school	0.063
Some high school	0.085
High school degree	0.322
Some college	0.168
College degree	0.181
Graduate or professional degree	0.095


S
A
Not A

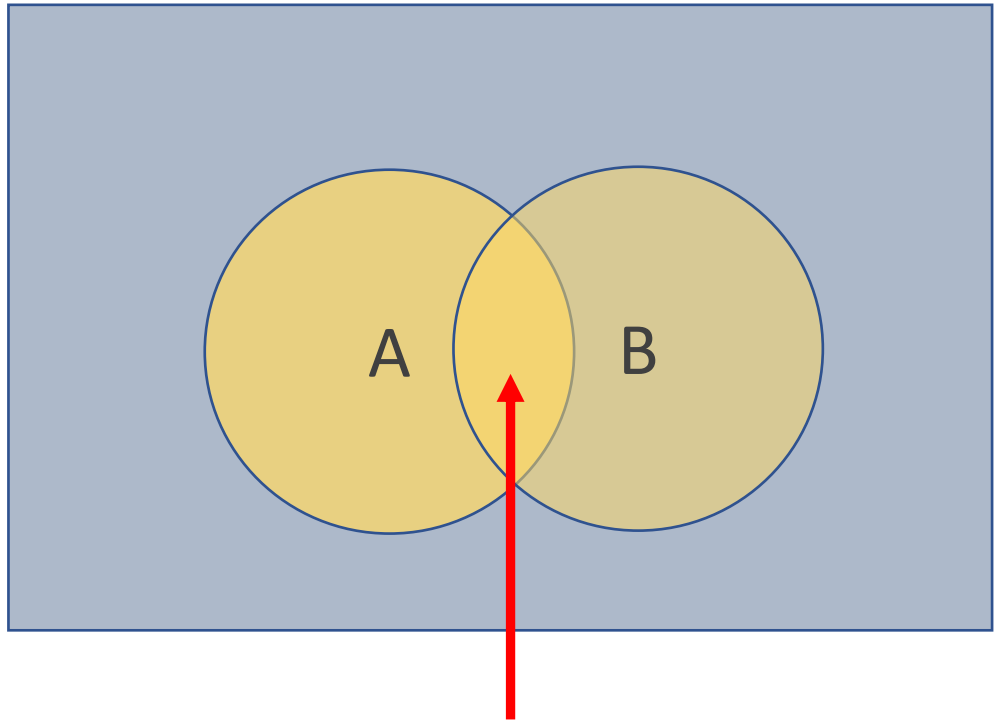
Mutually Exclusive



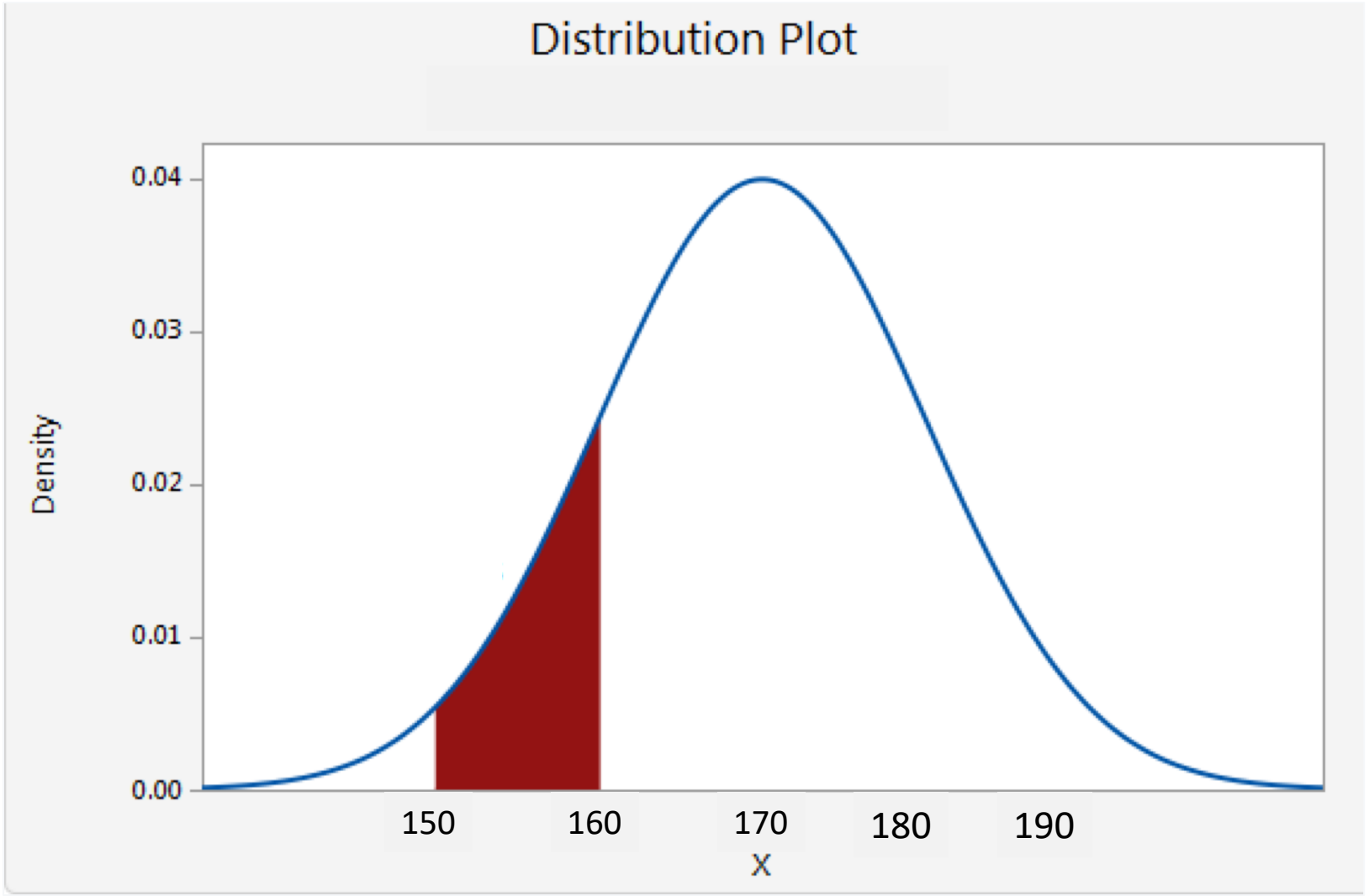
Non-Mutually Exclusive

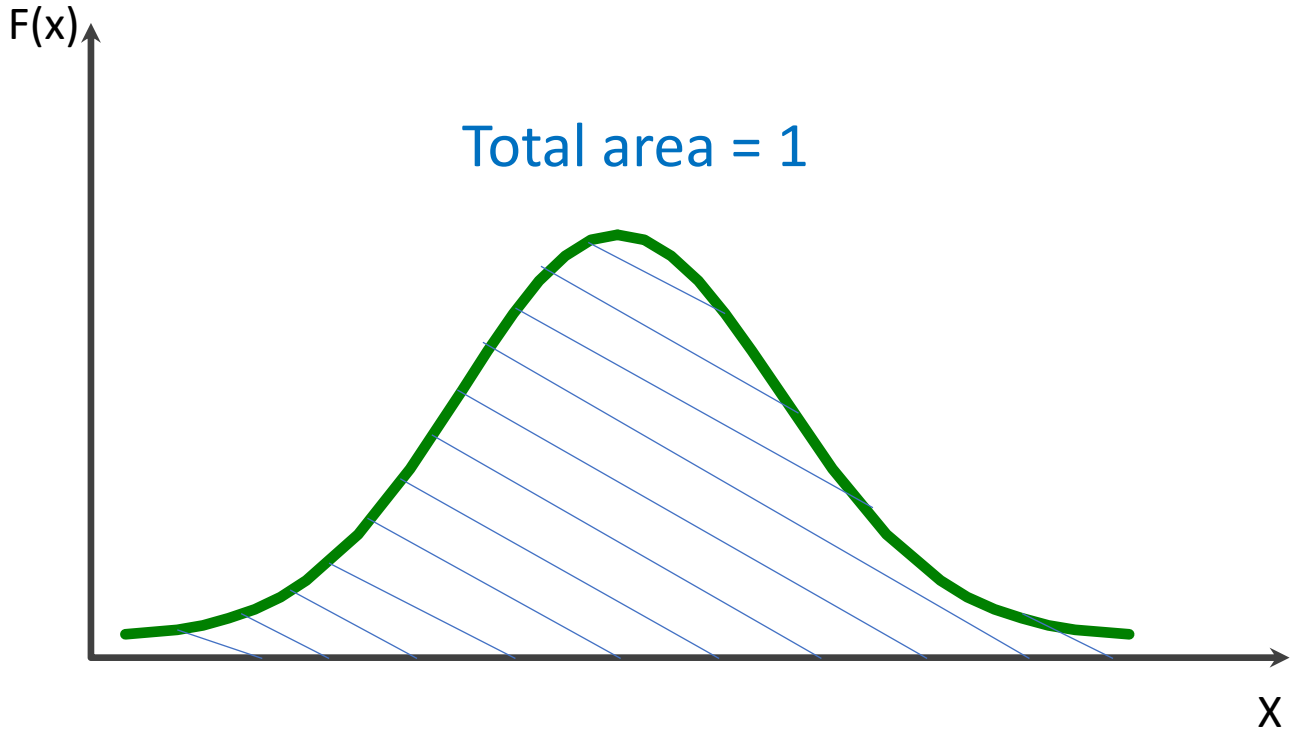


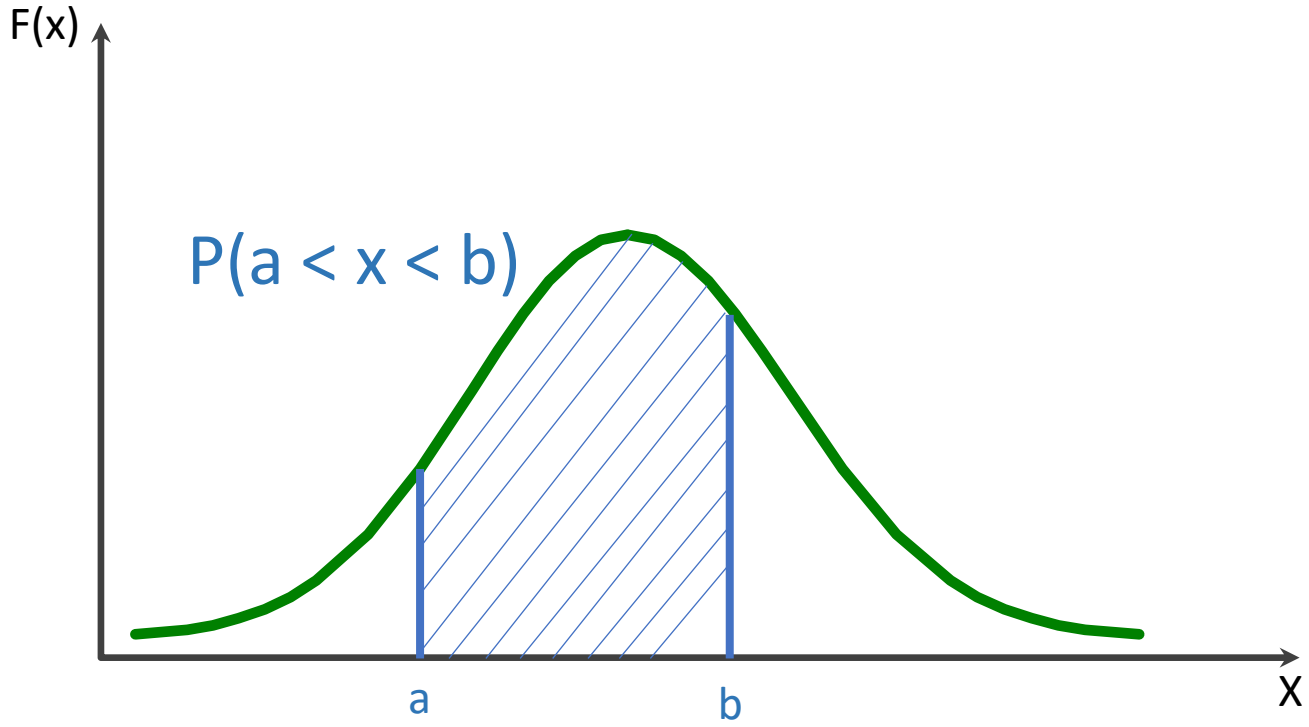
Non-Mutually Exclusive

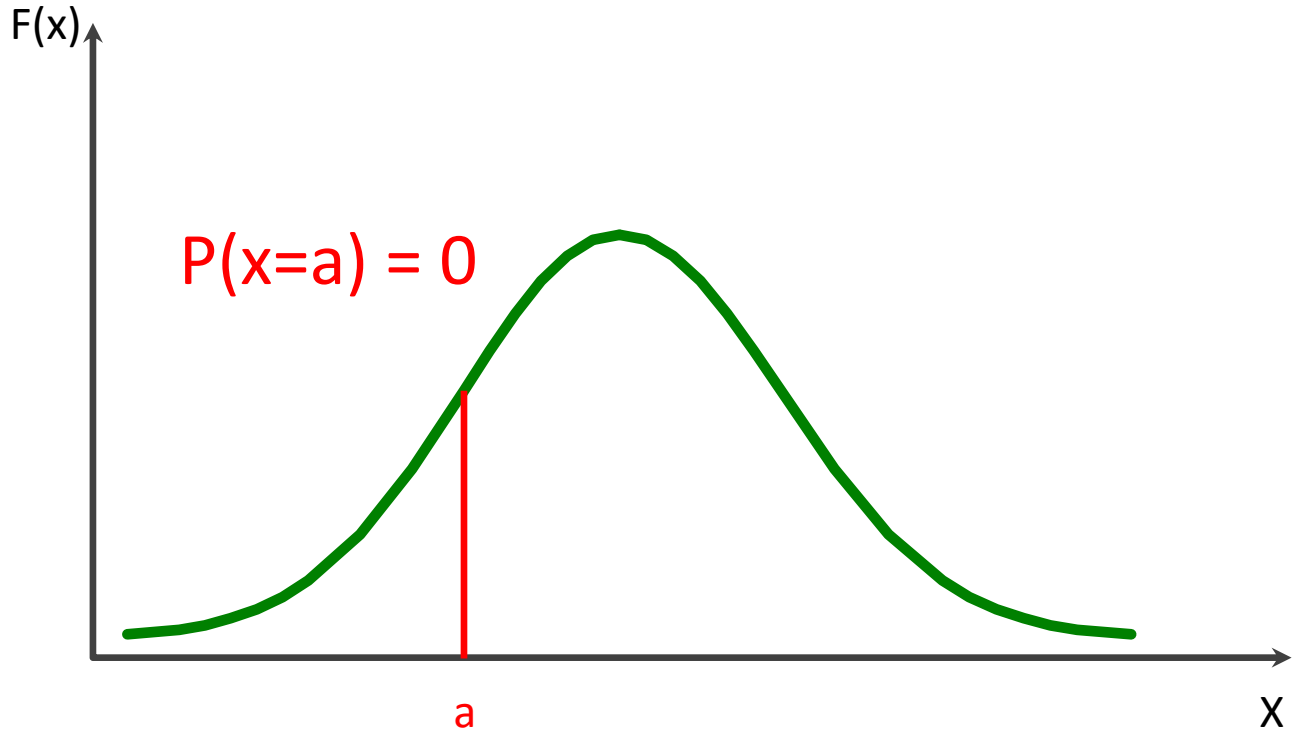


	B	not B	Total
A	0.75	0.15	0.90
not A	0.05	0.05	0.10
Total	0.80	0.20	1.00

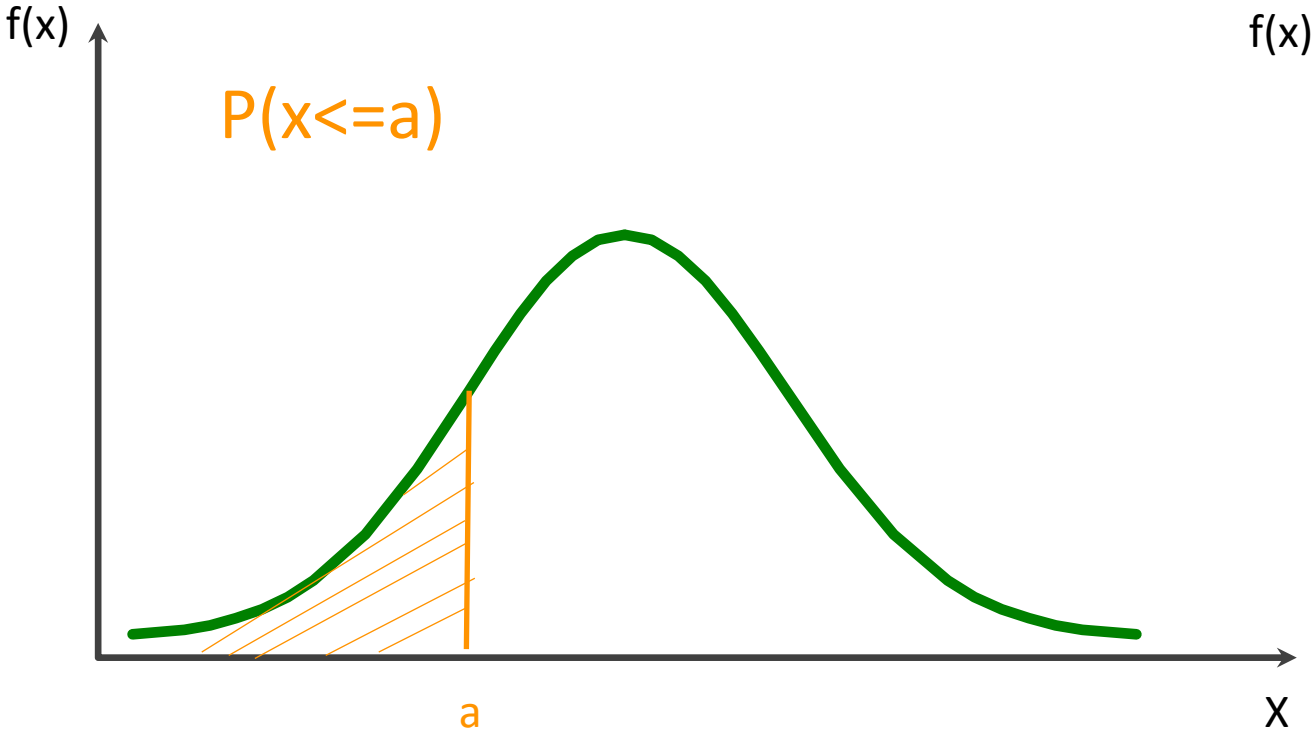

Distribution Plot
Density
0.04
0.03
0.02
0.01
0.00
150
160
170
180
190
X

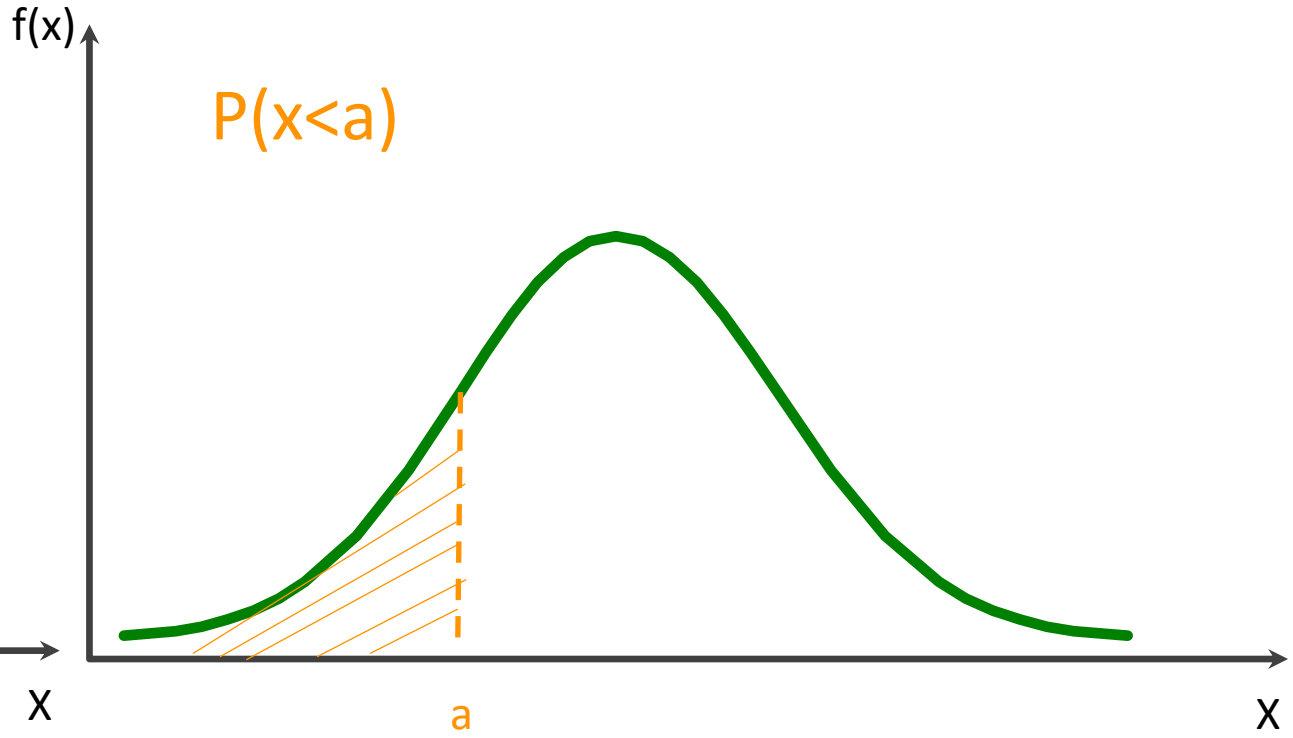

F(x)
Total area = 1
X

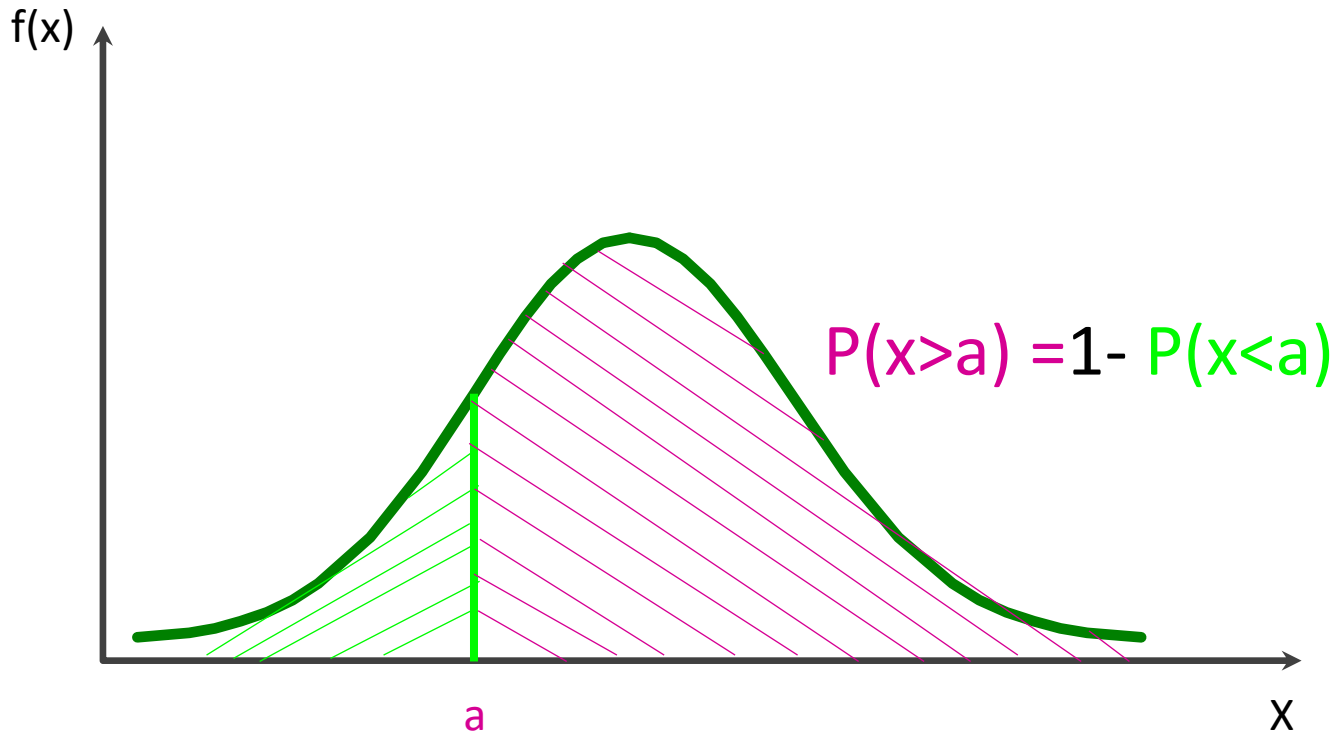

F(x)
P(a < x < b)
a
b
X


F(x)
P(x=a) = 0
a
X

P(x<=a) = P(x<a)






f(x)
P(x>a) =1- P(x<a)
a
X

		Population (parameter)	Sample (Statistic)
Categorical Variable	**Proportion**	**P=**population proportion	$\hat{p}$=sample proportion
Numerical Variable	**Mean**	μ=population mean	$\bar{x}$=sample mean
	Standard Deviation	σ=population standard deviation	**s=**sample standard deviation