Statistics for Dummies

Fred Moyer
@phredmoyer
@circonus

Data

“Without data, you’re just another person with an opinion”

W. Edwards Deming

Average

“Arithmetic Mean”

Avg = sum / # samples

avg(1,2,3) = 6/3 = 2

Median

Midpoint of a data set

Value	11	22	**33**	44	55
Sample	1	2	3	4	5

50th percentile q(0.5) = 33

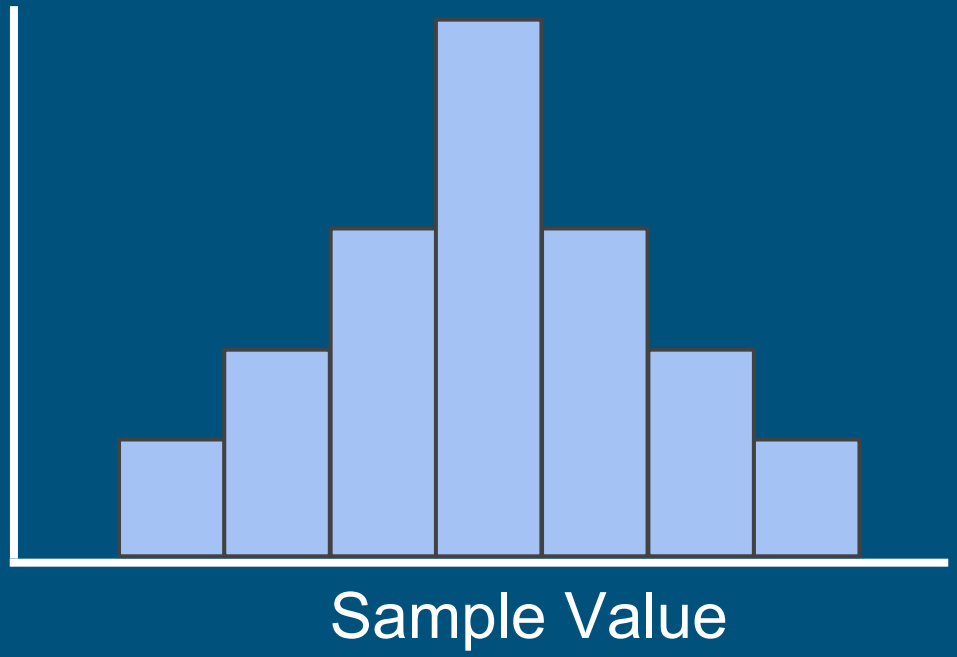
90th Percentile

90% of the values are below it

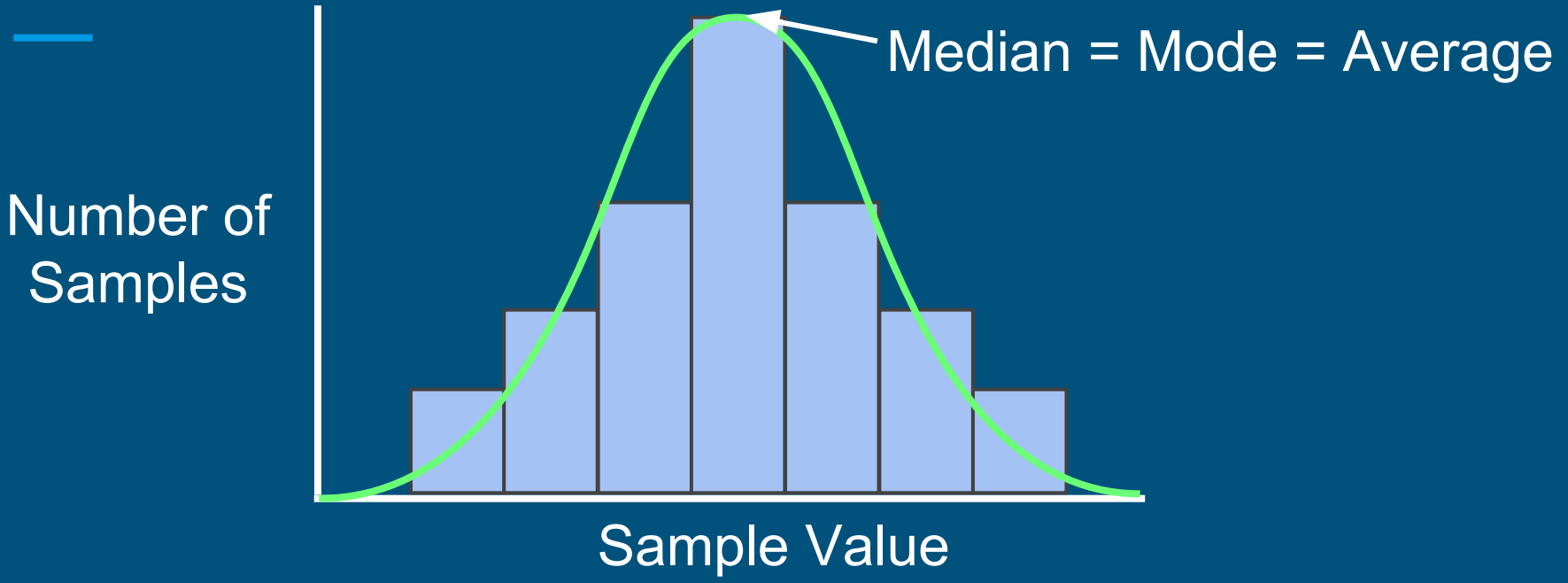
Value	11	22	33	44	55	66	77	88	99	**100**	111
Sample	1	2	3	4	5	6	7	8	9	10	11

$q(0.9) = 100$

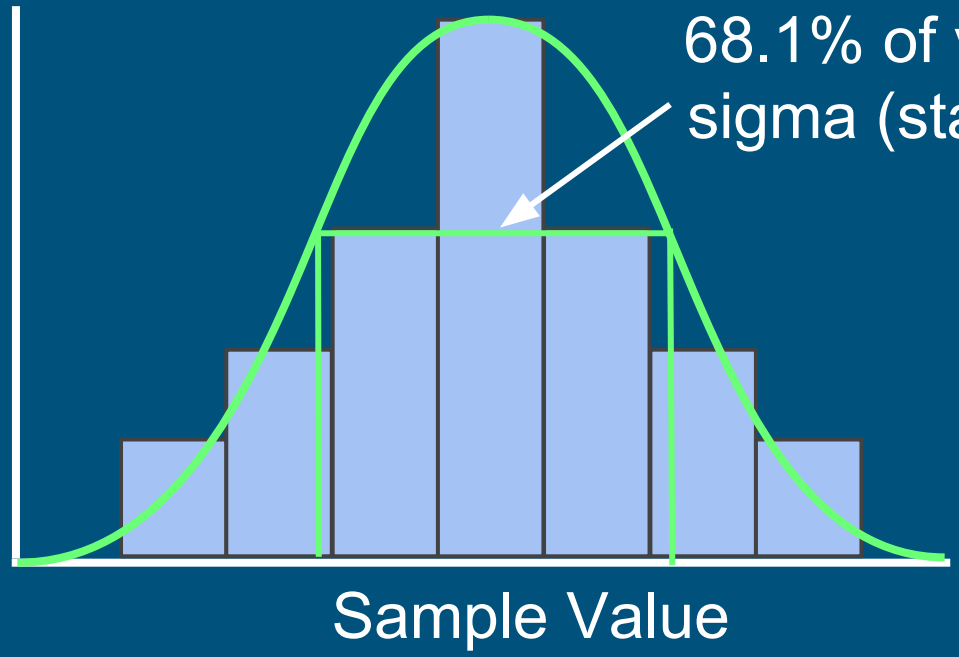
Histogram



Normal Distribution



Standard Deviation

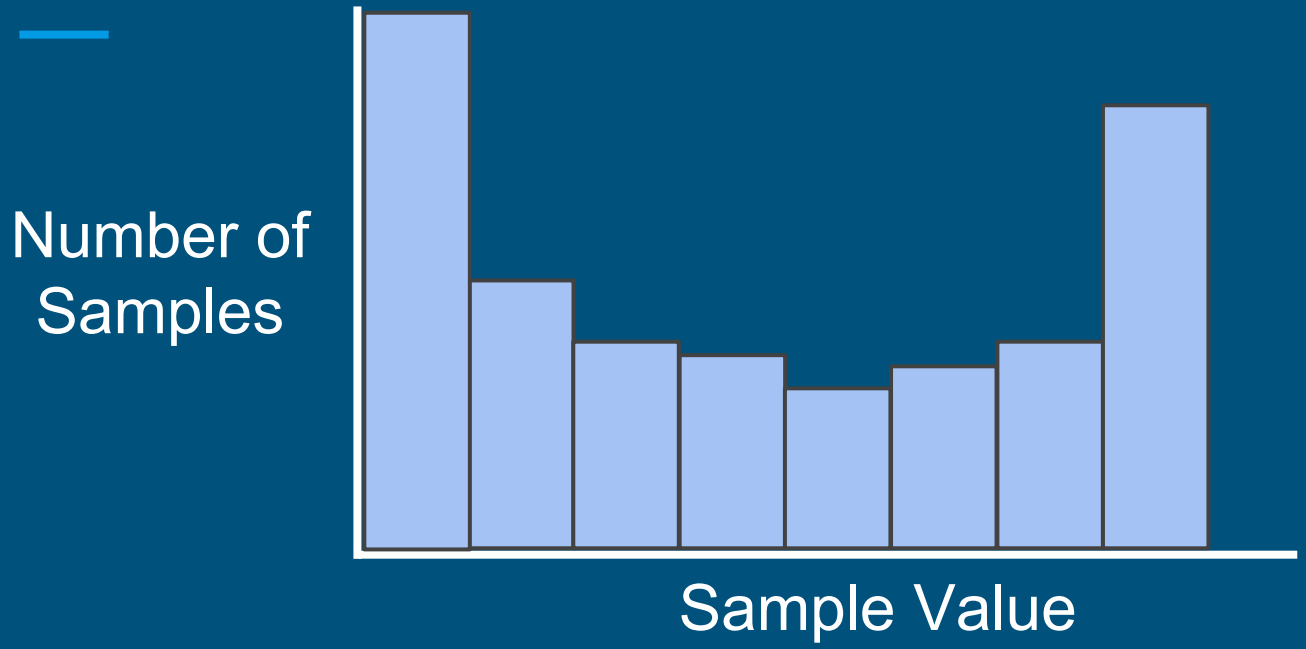


Standard Deviation (σ)

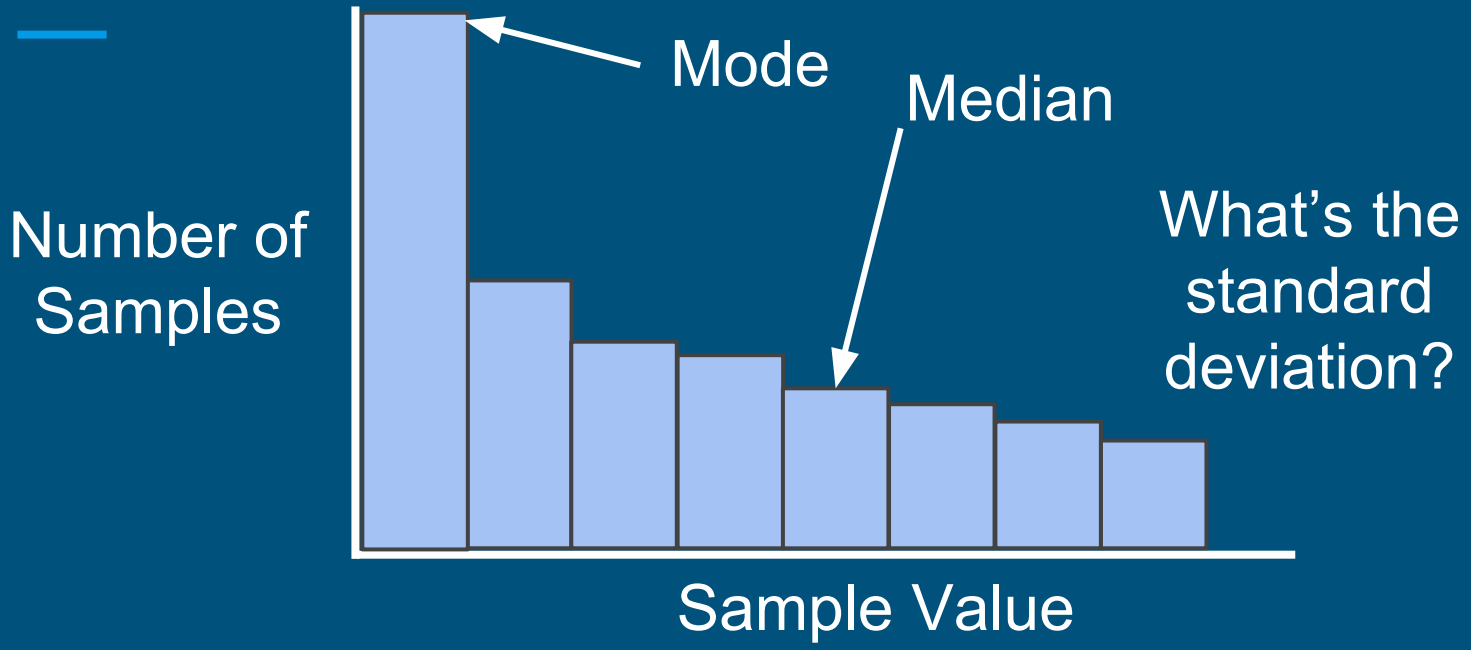
$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2}$$

Subtract the mean μ from all the samples, square the difference, sum and divide by number of samples

Non Normal Distribution



Non Normal Distribution



Percentiles

Sample set A q(0.95) = 10

Sample set B q(0.95) = 20

What is q(0.95) for A U B?

Percentiles

a(0.95){A U B} !=

avg(q(0.95){A} + q(0.95){B})

q(0.95){A U B } needs raw data

Null Hypothesis

“The Double” = 2012

Determine who is Cassius the Assassin

Null Hypothesis

“Stephen Hawking is Cassius”

Formulate a null hypothesis

“Stephen Hawking is not Cassius”

Null Hypothesis

Try to disprove the null hypothesis

Hawking wasn’t in any crime scene photos

Null hypothesis proved, Hawking != Cassius

Large p-value, > 0.05, weak evidence against

Null Hypothesis

“Richard Gere is Cassius”

Formulate a null hypothesis

“Richard Gere is not Cassius”

Null Hypothesis

Try to disprove null hypothesis

Richard Gere was in ALL of the crime scene photos

Small p-value, <= 0.05, weak evidence for

Null Hypothesis

Null hypothesis disproved

Richard Gere is Cassius

Don’t be a Dummy

Go learn some statistics!