

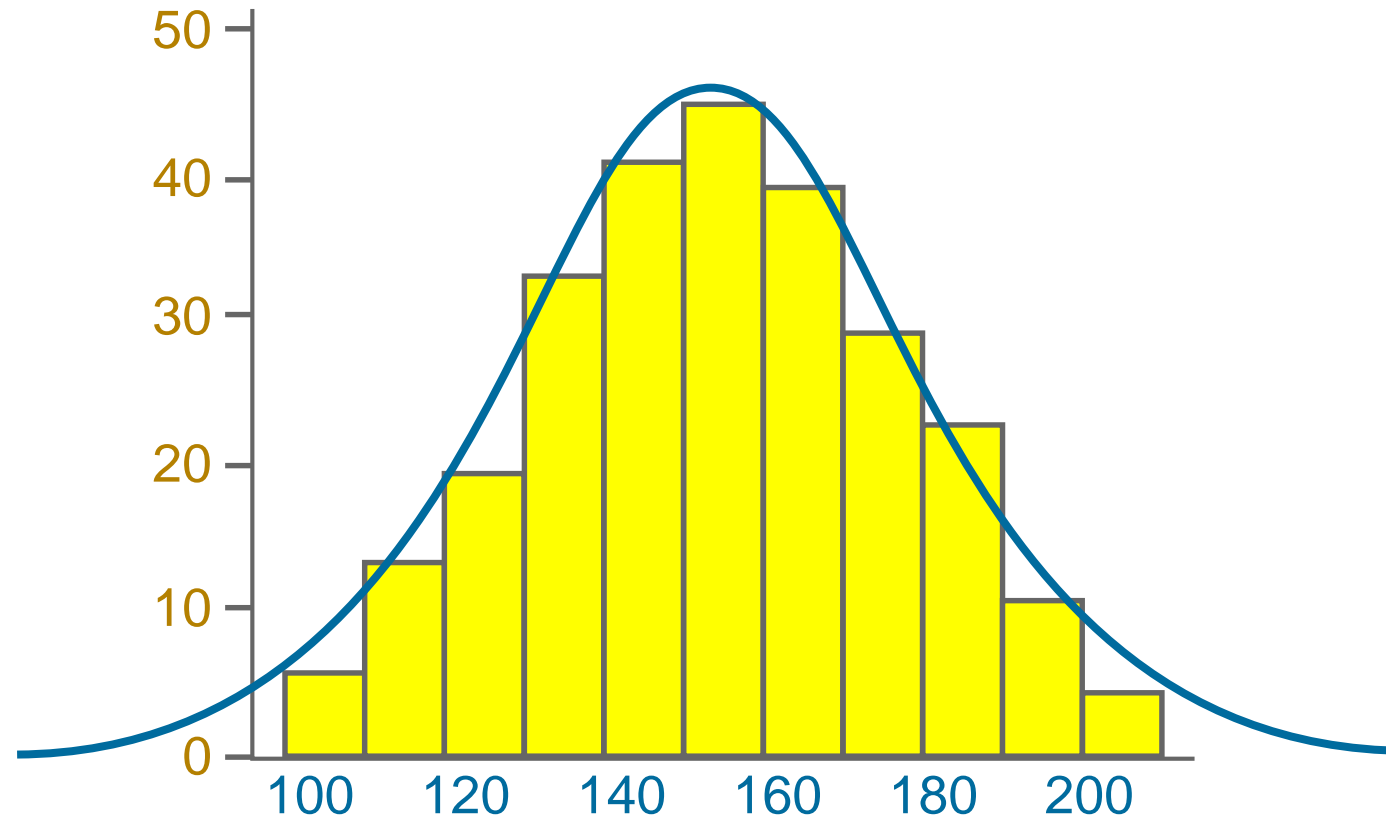


Week 5: Probability Distributions

Unit 2: The Normal Distribution

The Normal Distribution

Introduction



The Normal Distribution

<https://www.mathsisfun.com/data/standard-normal-distribution.html>

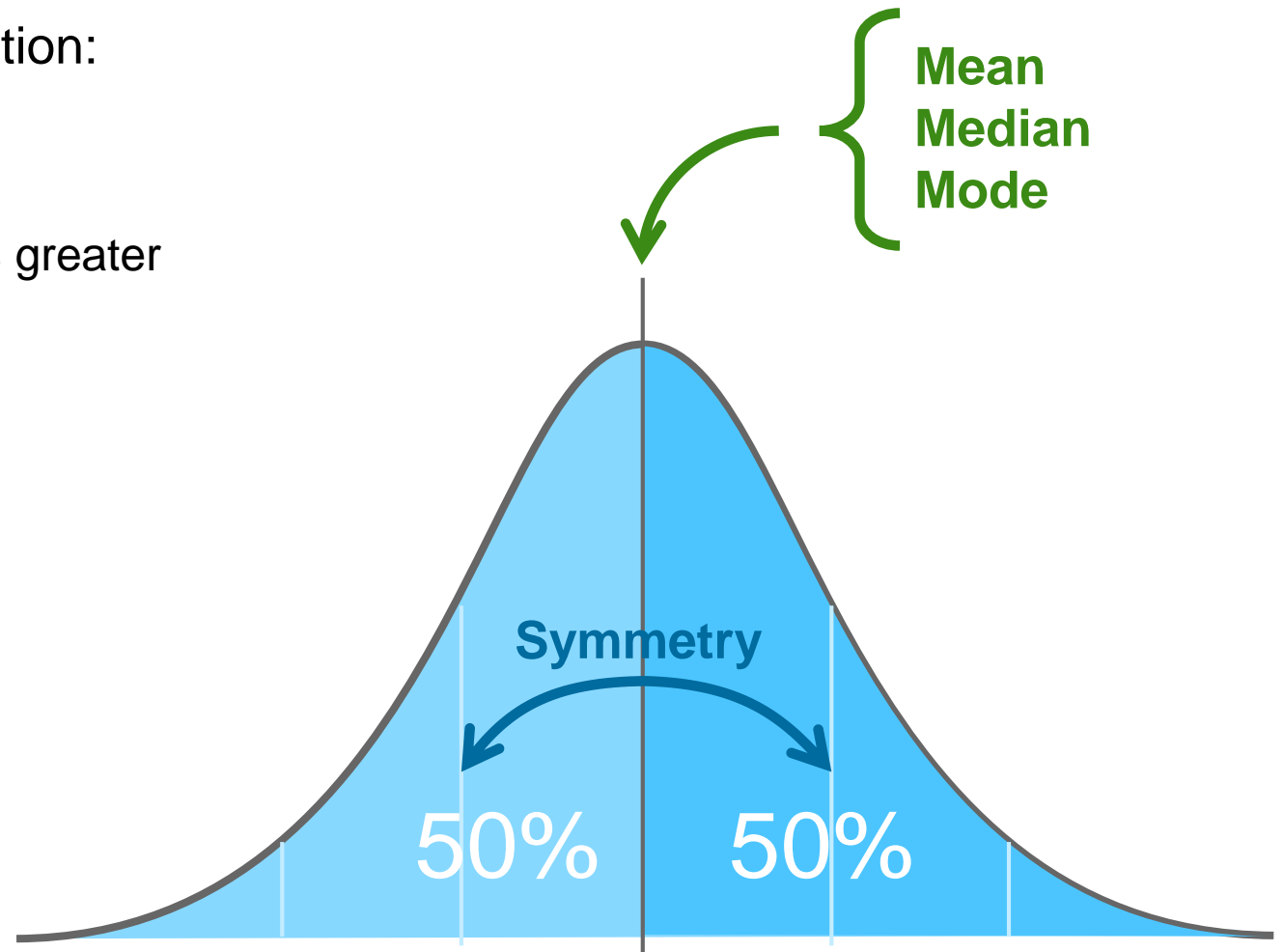
https://en.wikipedia.org/wiki/Normal_distribution

The Normal Distribution

Characteristics

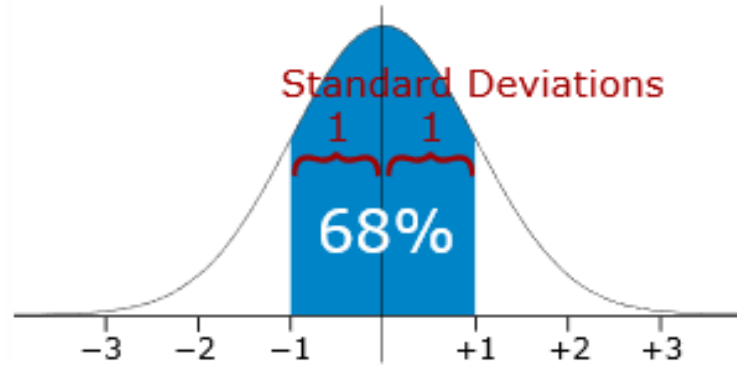
The characteristics of the normal distribution:

- mean = median = mode
- symmetry about the centre
- 50% of values less than the mean and 50% greater than the mean

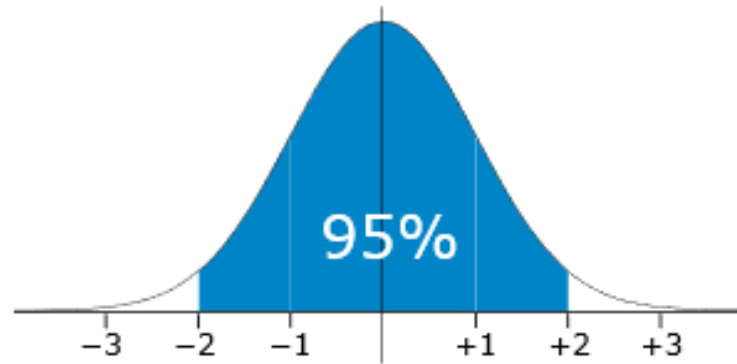


The Normal Distribution

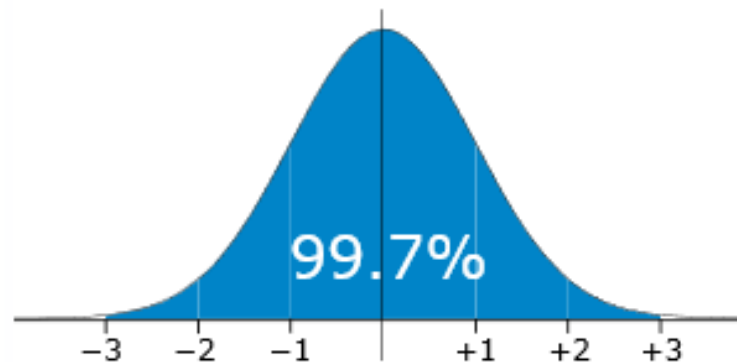
Standard deviation



68% of values are within 1 standard deviation of the mean



95% of values are within 2 standard deviations of the mean

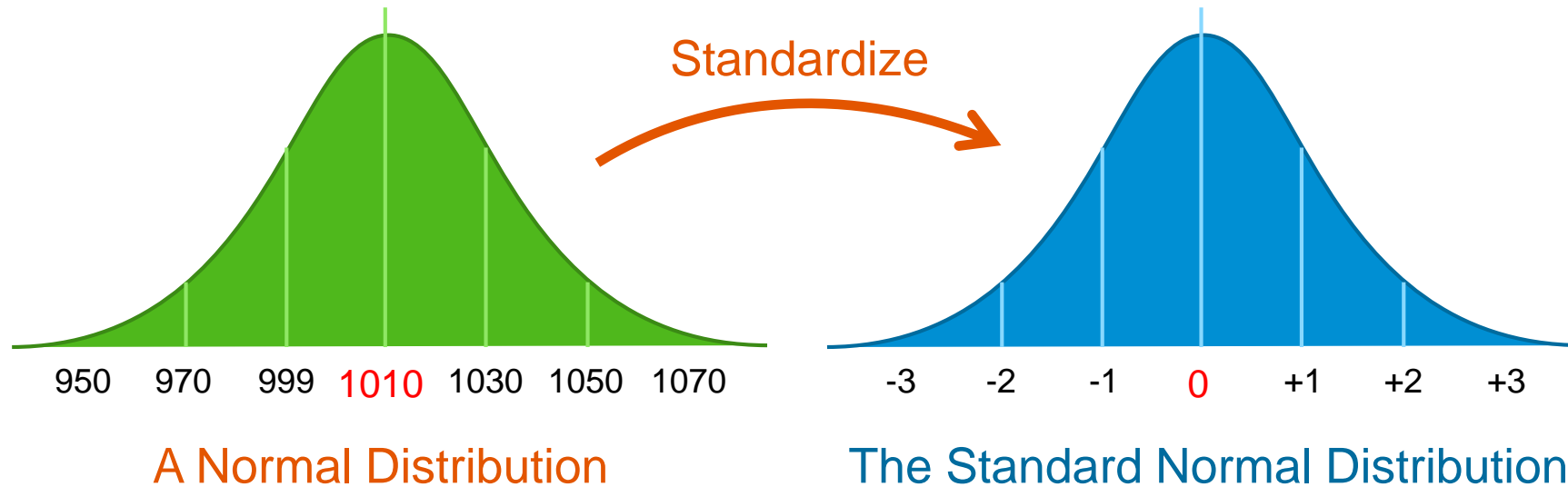


99.7% of values are within 3 standard deviations of the mean

For a standard deviation calculator, see:
<https://www.mathsisfun.com/data/standard-deviation-calculator.html>

The Normal Distribution

Standard normal distribution

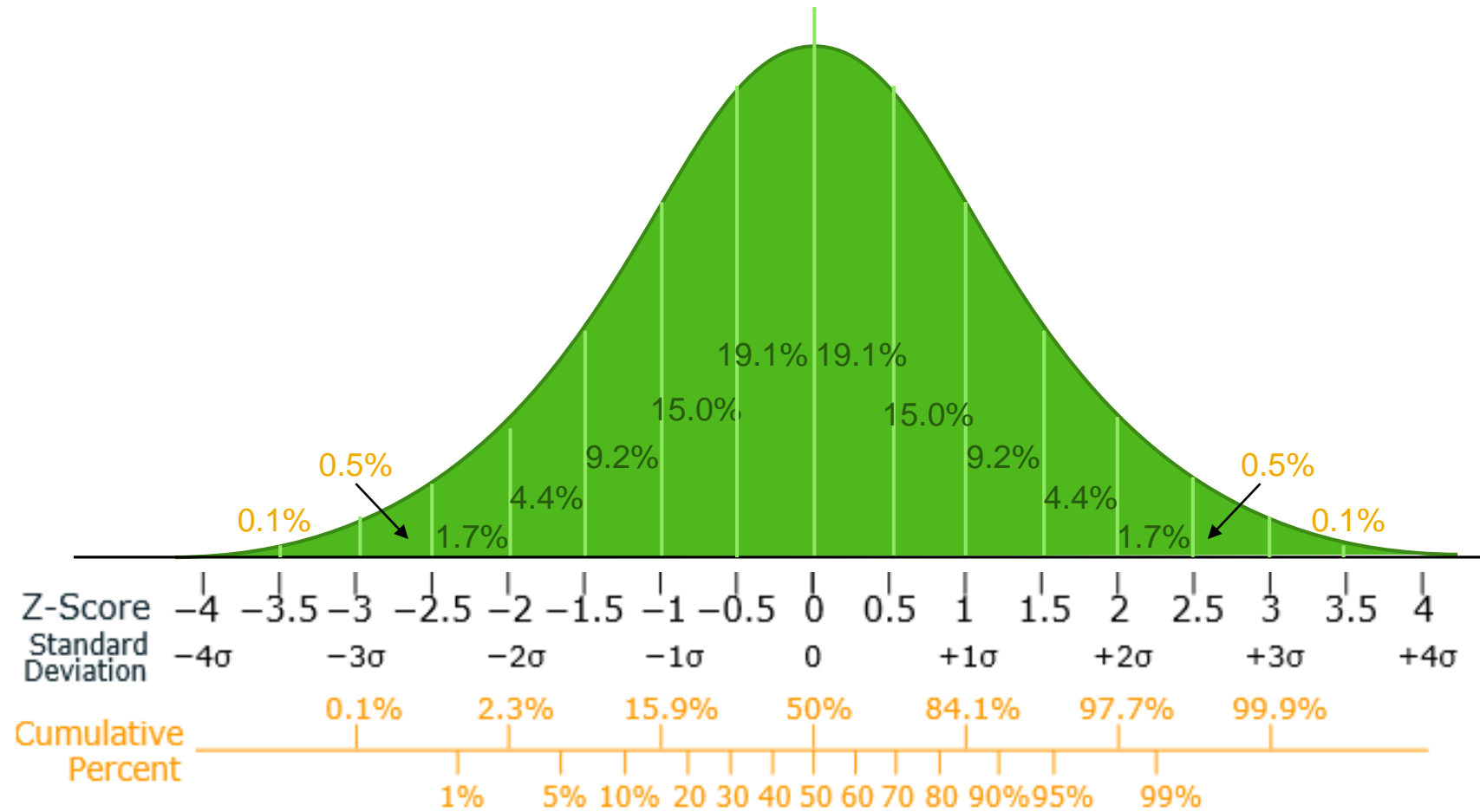


The formula for the z-score:
$$z = \frac{x - \mu}{\sigma}$$

z is the "z-score" (standard score)
x is the value to be standardized
μ is the mean
σ is the standard deviation

For an interactive standard normal distribution calculator, see:
<https://www.mathsisfun.com/data/standard-normal-distribution-table.html>

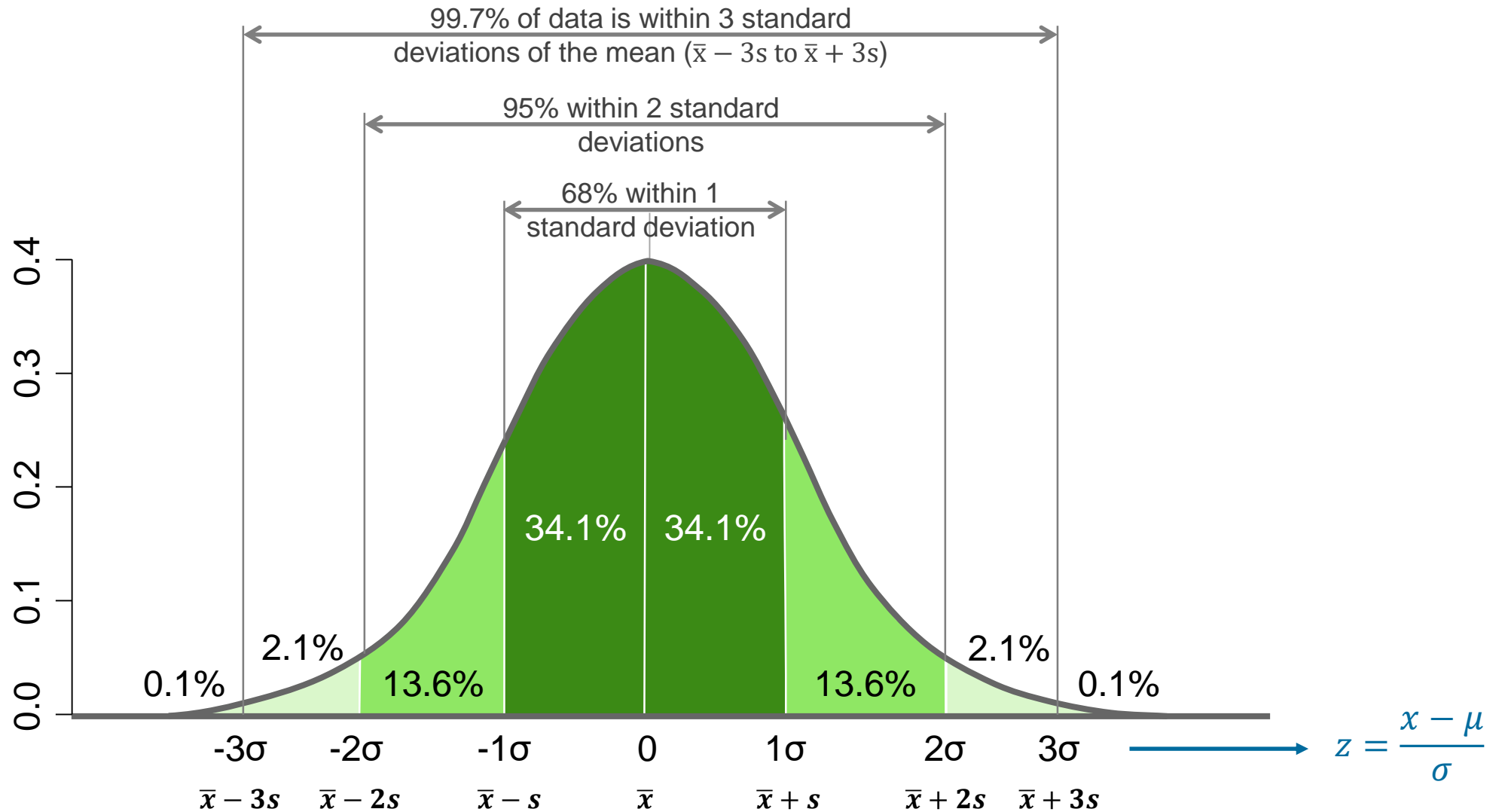
Standard normal distribution example



Standard Normal Distribution

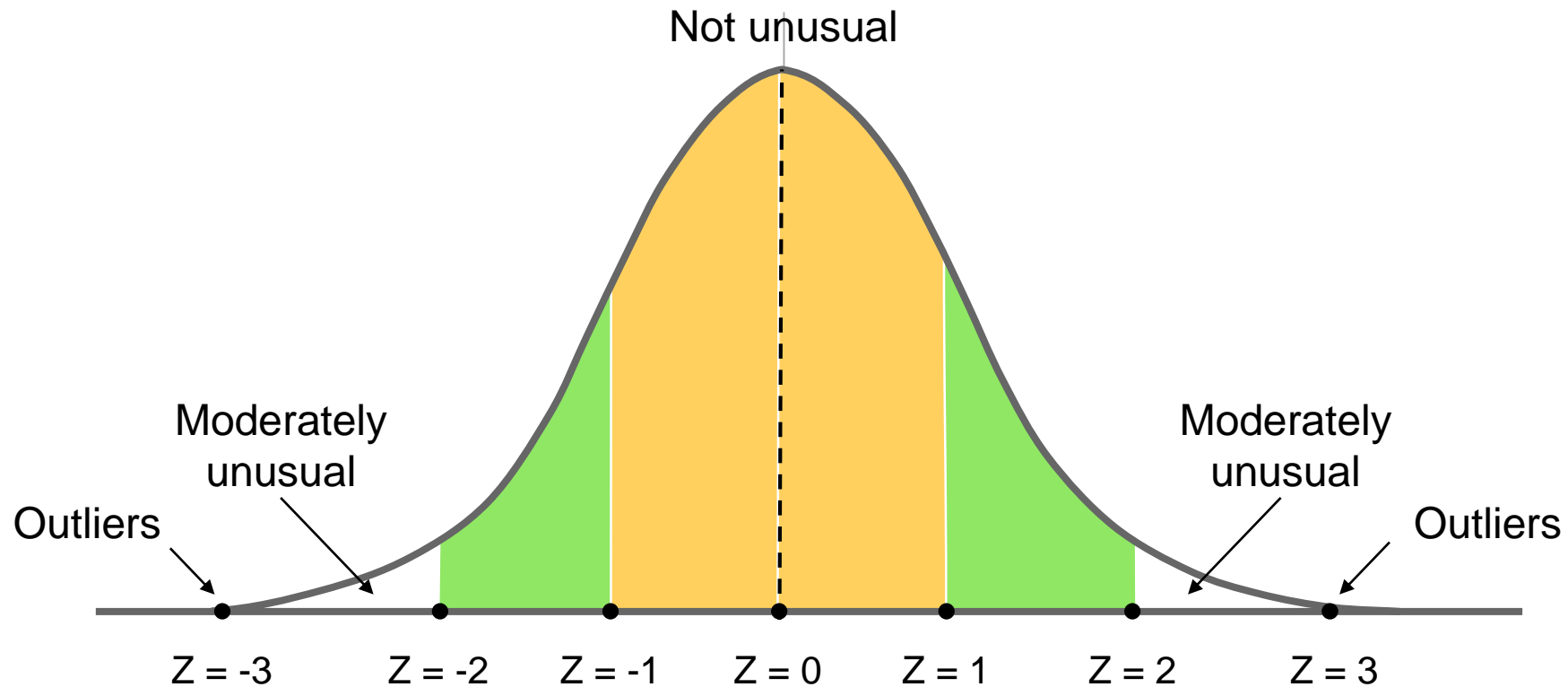
The Normal Distribution

The empirical rule



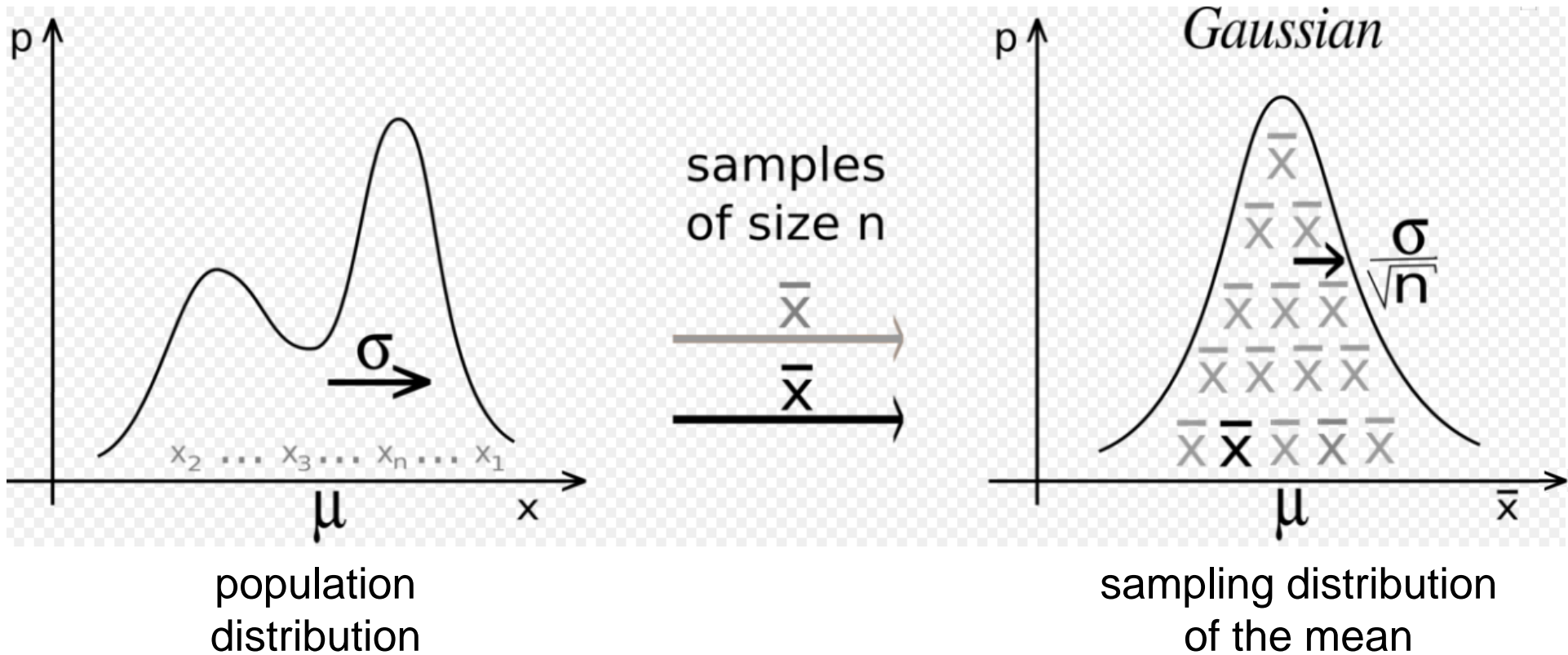
Rules of thumb for detecting outliers

<i>Possible Outliers</i>	<i>Outliers</i>
$ z > 2$	$ z > 3$



The Normal Distribution

Central limit theorem



https://en.wikipedia.org/wiki/Central_limit_theorem

<https://machinelearningmastery.com/a-gentle-introduction-to-the-central-limit-theorem-for-machine-learning/>

The Normal Distribution

Summary

- The **normal** distribution is a very commonly encountered continuous probability distribution.
- The characteristics of the normal distribution are:
 - mean = median = mode
 - symmetry about the centre
 - 50% of values less than the mean and 50% greater than the mean
- When we calculate the standard deviation, we find that generally:
 - 68% of values are within 1 standard deviation of the mean
 - 95% of values are within 2 standard deviations of the mean
 - 99.7% of values are within 3 standard deviations of the mean
- The **empirical rule** states that for a normal distribution, nearly all of the data will fall within three standard deviations of the mean.
- The **central limit theorem** (CLT) establishes that when independent random variables are added, their properly normalized sum tends towards a normal distribution even if the original variables themselves are not normally distributed.



Thank you.

Contact information:

open@sap.com

Follow all of SAP



www.sap.com/contactsap

© 2019 SAP SE or an SAP affiliate company. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP SE or an SAP affiliate company.

The information contained herein may be changed without prior notice. Some software products marketed by SAP SE and its distributors contain proprietary software components of other software vendors. National product specifications may vary.

These materials are provided by SAP SE or an SAP affiliate company for informational purposes only, without representation or warranty of any kind, and SAP or its affiliated companies shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP or SAP affiliate company products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.

In particular, SAP SE or its affiliated companies have no obligation to pursue any course of business outlined in this document or any related presentation, or to develop or release any functionality mentioned therein. This document, or any related presentation, and SAP SE's or its affiliated companies' strategy and possible future developments, products, and/or platforms, directions, and functionality are all subject to change and may be changed by SAP SE or its affiliated companies at any time for any reason without notice. The information in this document is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. All forward-looking statements are subject to various risks and uncertainties that could cause actual results to differ materially from expectations. Readers are cautioned not to place undue reliance on these forward-looking statements, and they should not be relied upon in making purchasing decisions.

SAP and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP SE (or an SAP affiliate company) in Germany and other countries. All other product and service names mentioned are the trademarks of their respective companies.

See www.sap.com/copyright for additional trademark information and notices.