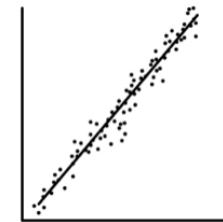
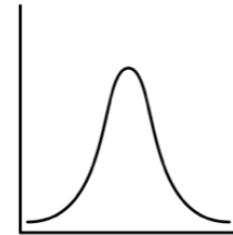
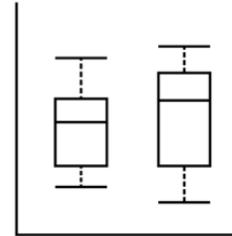


PSYC 2300

Introduction to Statistics



Lecture 06: Probability and the Normal Curve

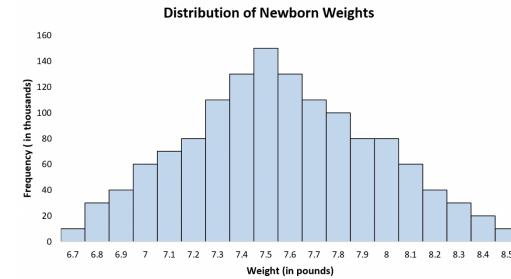
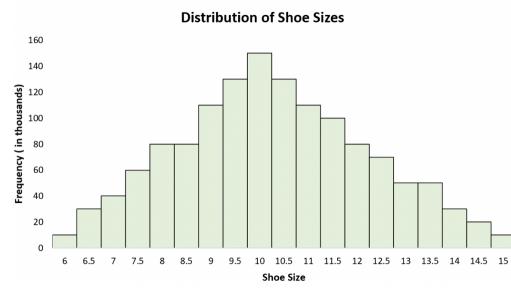
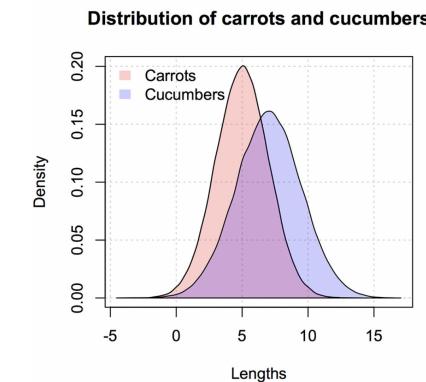
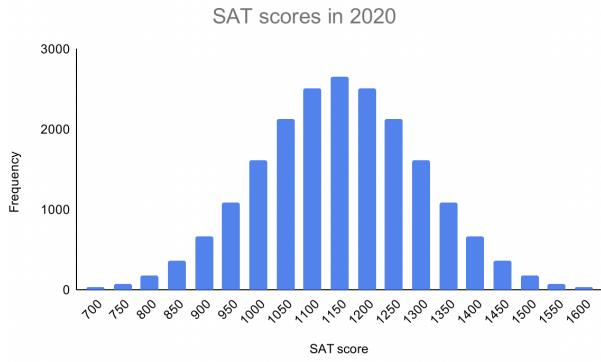
Outline for today

- Understanding the normal curve
- Standardization via z-scores
- z-score practice problem
 - Have  and  ready
- Computing z-scores in JASP
 - Download [Stats Class 7 Dataset \(Z-Scores\).jasp](#)



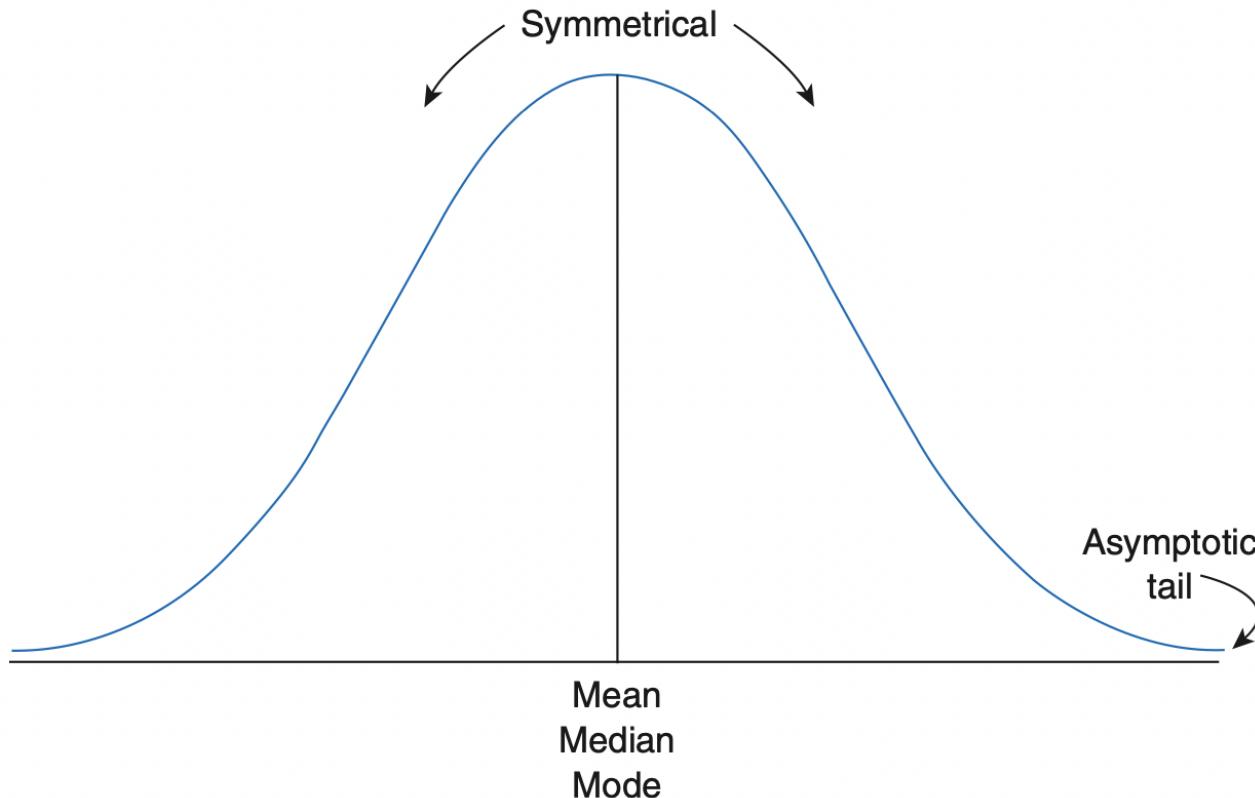
Normal Distribution

Normal Distribution



The normal distribution comes up everywhere

Normal Distribution

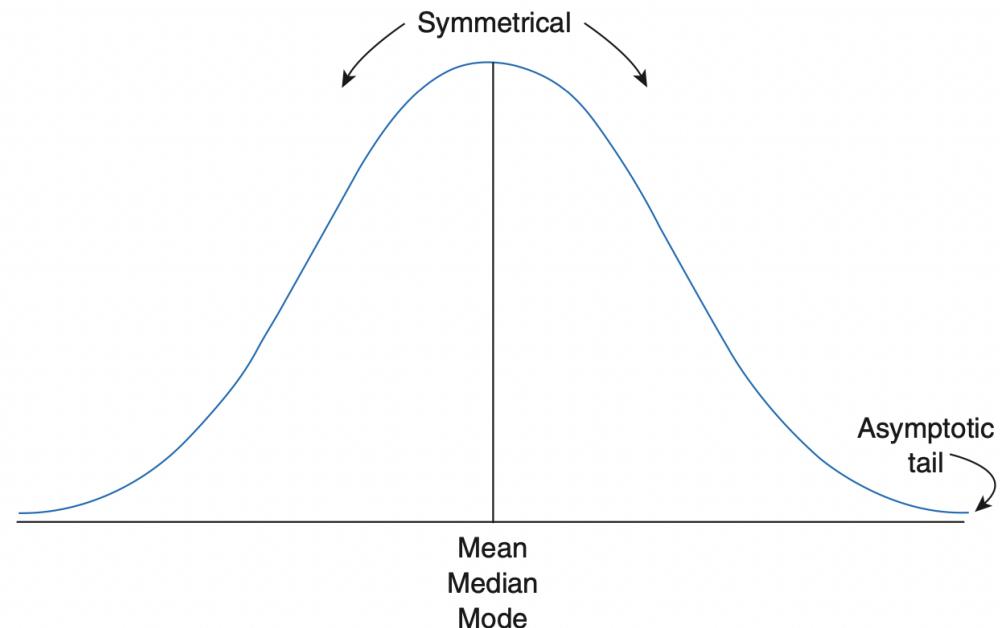


Also called the *Gaussian* distribution or *bell-curve* distribution

Normal Distribution

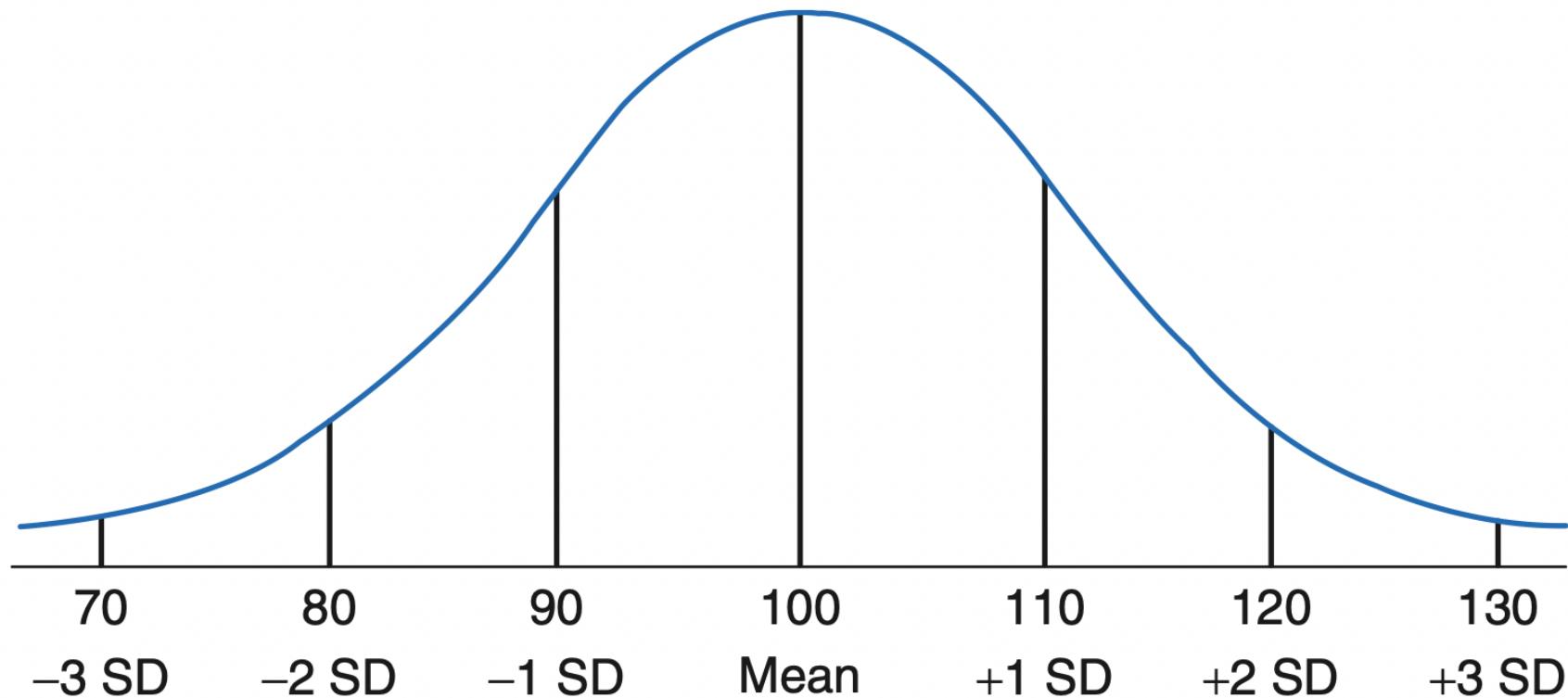
Three Central Characteristics

- Mean = Median = Mode
- Symmetrical around the mean
- Asymptotic tails: they approach but never touch the x-axis



Normal Distribution

We know a great deal about the normal distribution, particularly the probabilities associated with possible outcomes in a normal distribution



Probability

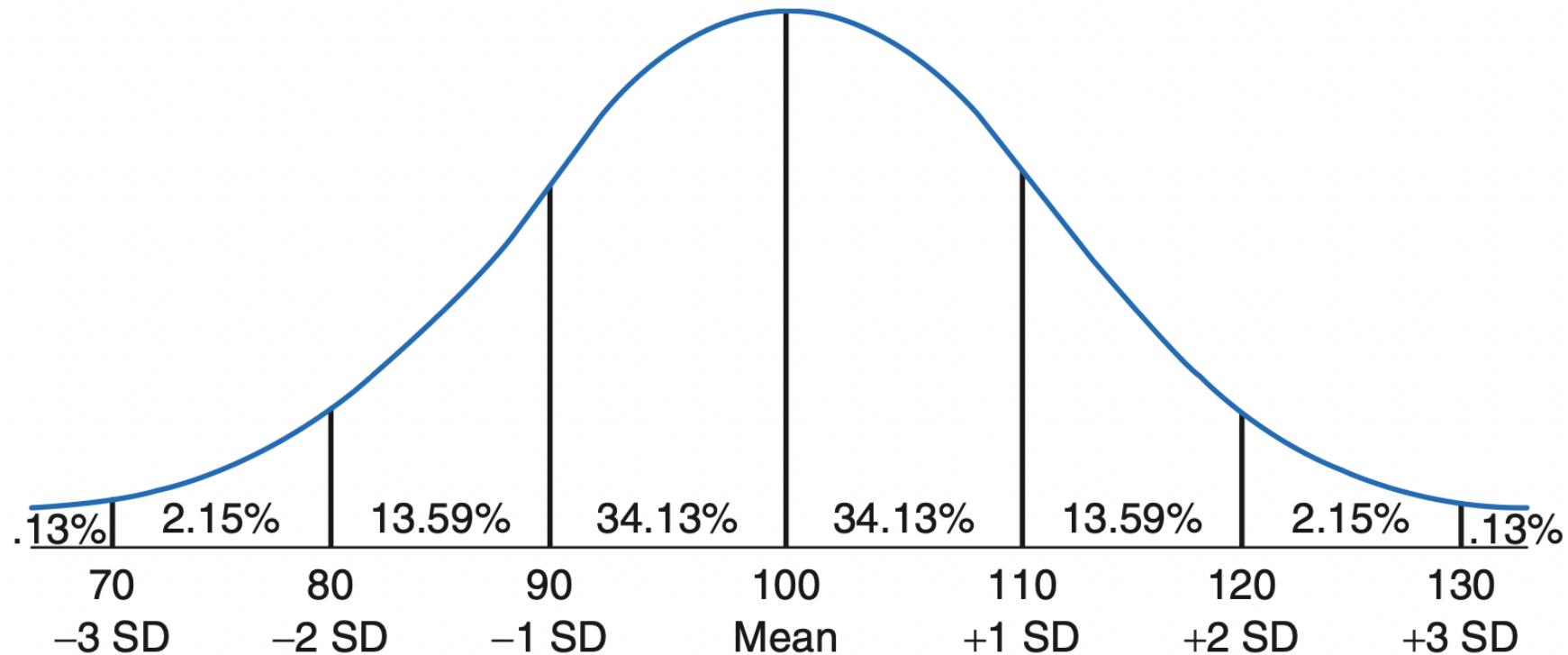
Probability is a huge topic that extends far beyond the limits of introductory statistics, and we do not attempt to examine it all here

Probability: For a situation in which several different outcomes are possible, the probability for any specific outcome is defined as a *fraction* or a *proportion* of all the possible outcomes. If the possible outcomes are identified as A, B, C, D, and so on, then

$$\text{probability of } A = \frac{\text{number of outcomes classified as } A}{\text{total number of possible outcomes}}$$

Normal Distribution

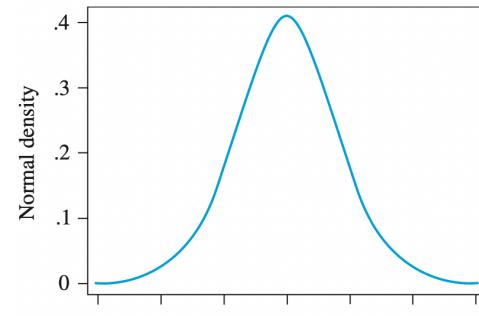
Approximate probabilities under the curve



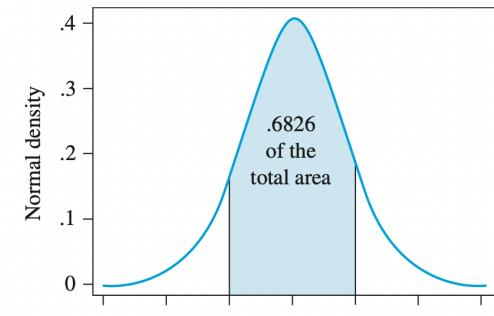
Normal Distribution

- 68.26% is within \pm **one** standard deviation of the mean
- 95.44% is within \pm **two** standard deviations of the mean
- 99.74% is within \pm **three** standard deviations of the mean

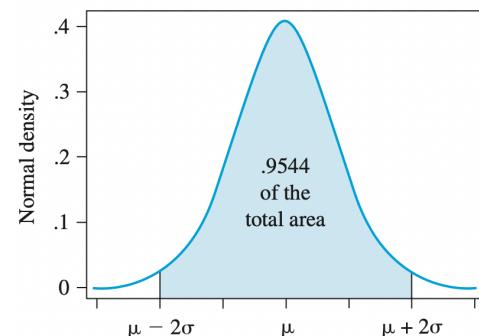
Also known as the "68-95-99.7 rule"



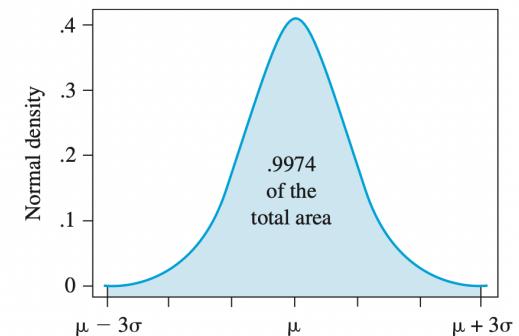
(a) Density of the normal distribution



(b) Area under normal curve within 1 standard deviation of mean



(c) Area under normal curve within 2 standard deviations of mean



(d) Area under normal curve within 3 standard deviations of mean

Normal Distribution

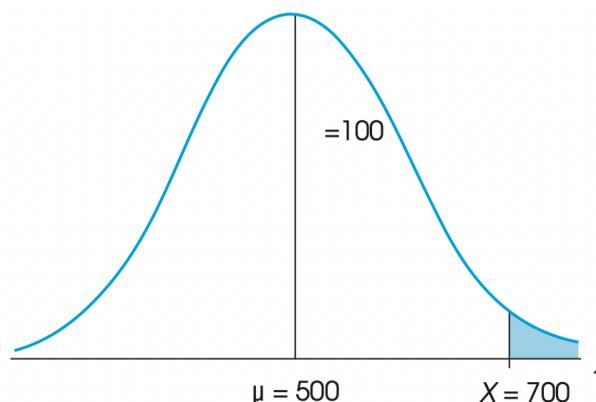
SAT scores is normal with a mean of $\mu = 500$ and a standard deviation of $\sigma = 100$

What is the probability of randomly selecting an individual from this population who has an SAT score greater than $x_i = 700$? In other words, $p(x_i > 700) = ?$

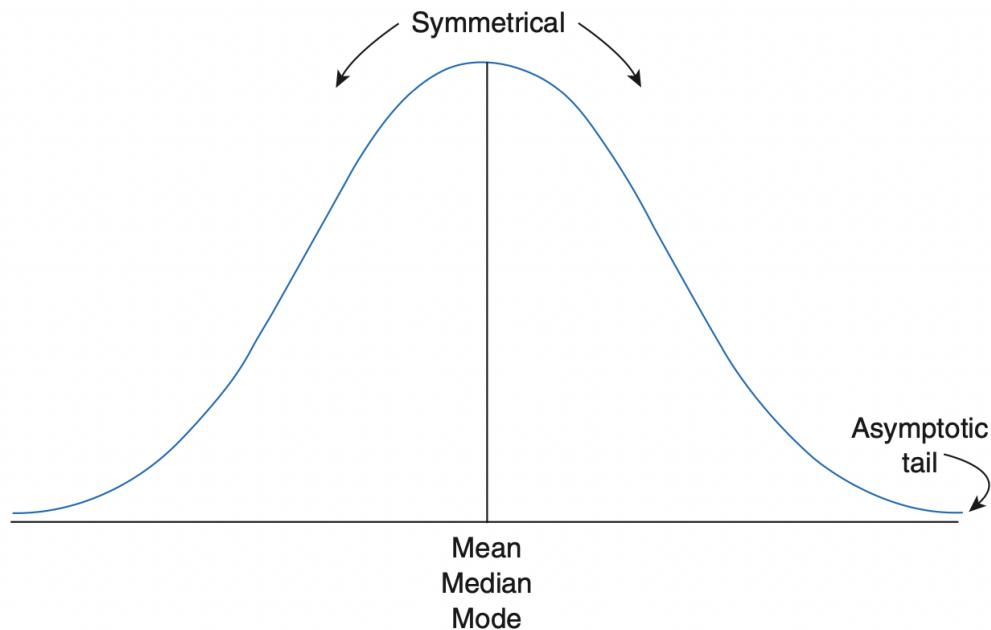
95.44% is within \pm **two** standard deviations of the mean

$$100 - 95.44 = 4.56$$

$$\frac{4.56}{2} = 2.28$$



Normal Distribution



Also referred to as the *probability density function* of the normal distribution

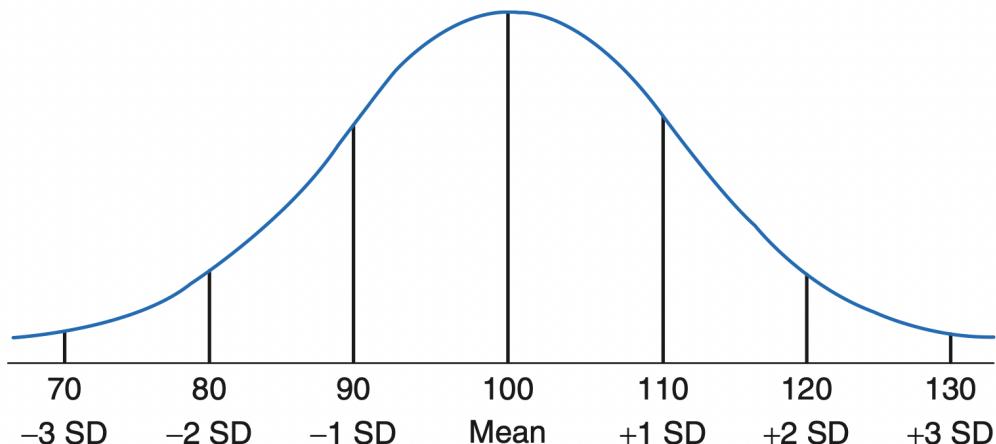
Normal Gaussian

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

The exact shape of the normal distribution is specified by the above equation relating to each x value (score) with each y value (frequency)

Normal Distribution

Example *normal probability density function*



$$\mu = 100$$

$$\sigma = 10$$

Normal Gaussian

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

$$\frac{1}{\sqrt{2\pi10^2}} e^{-\frac{(x-100)^2}{2*10^2}}$$

Normal Distribution

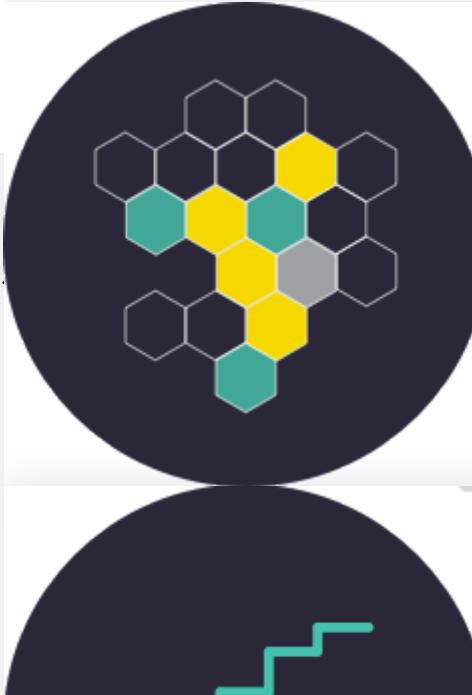
≡ Seeing Theory

English

Chapter 3

Probability Distribution

A probability distribution specifies the relative likelihoods of all possible outcomes.

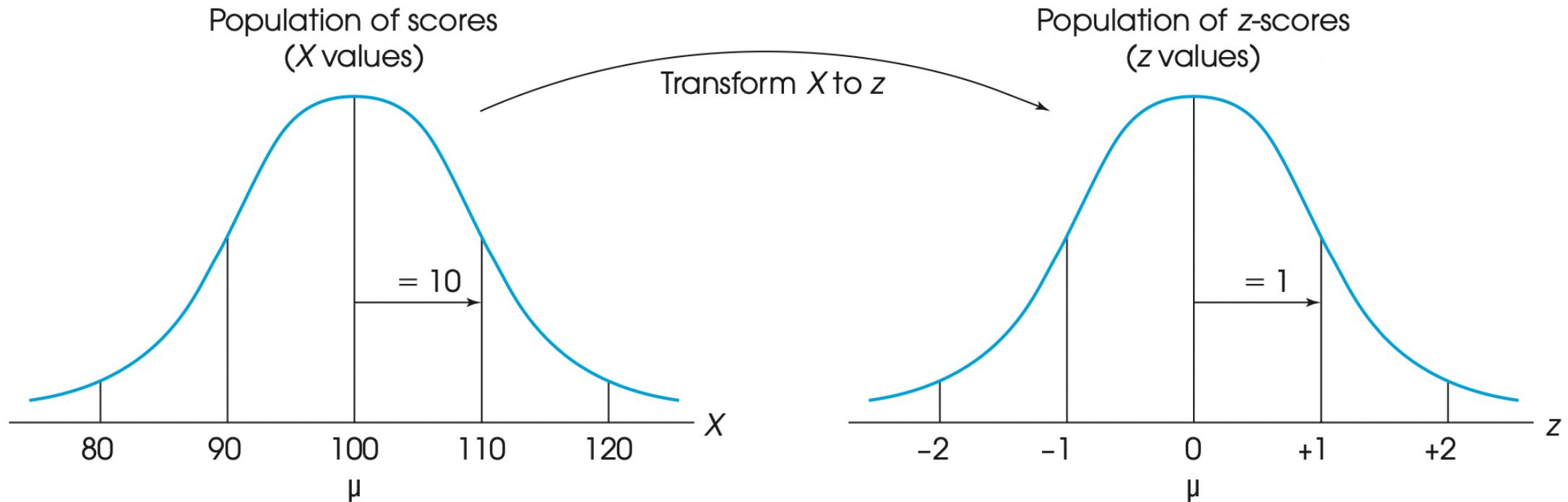


Random Variable

[Link to visualizing the probability distribution function of the normal distribution](#)

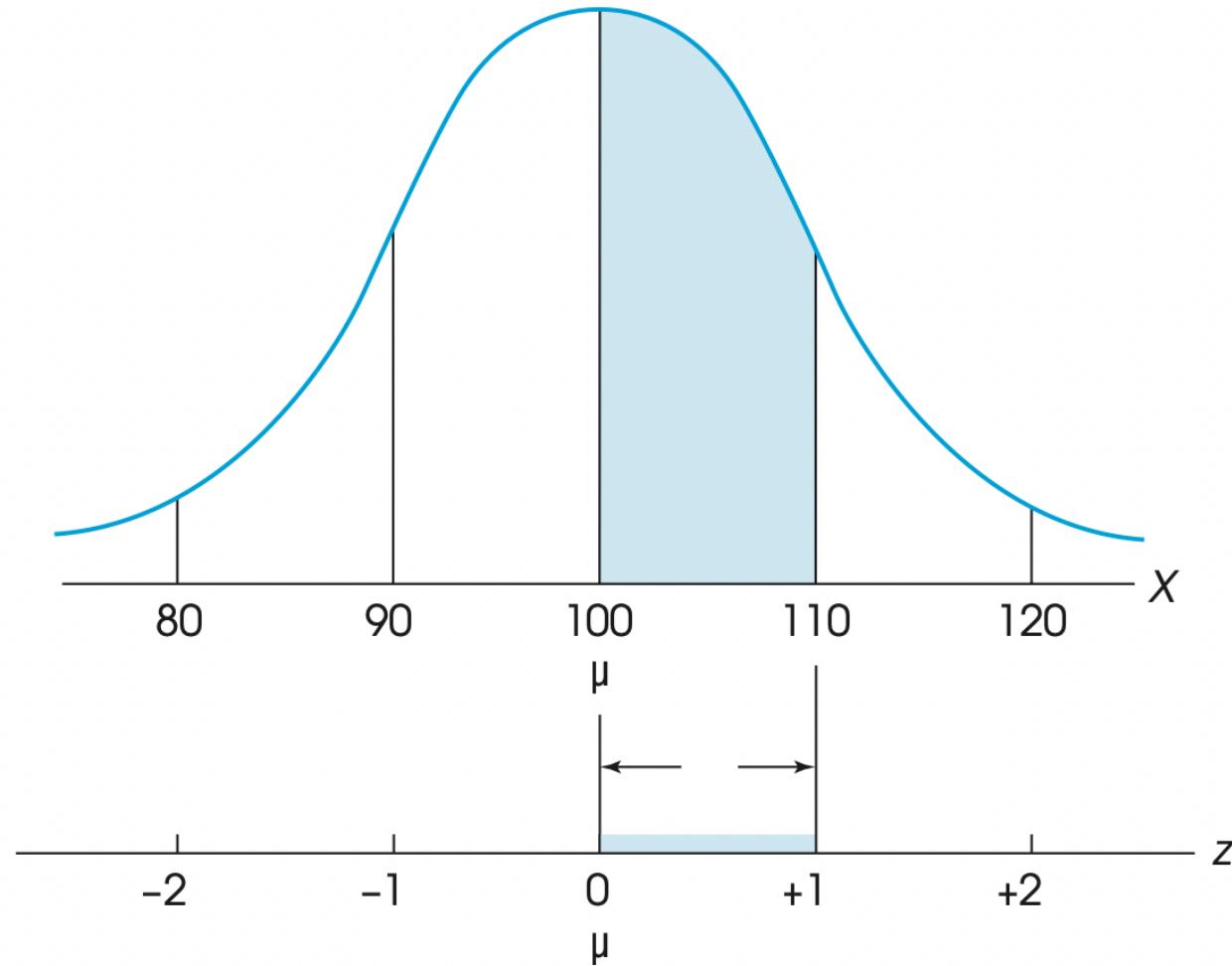
z-scores

Standardizing Scores



Standardized scores are comparable because they are standardized in units of standard deviations

Standardizing Scores



Developing z-scores

Two hypothetical students in my class:

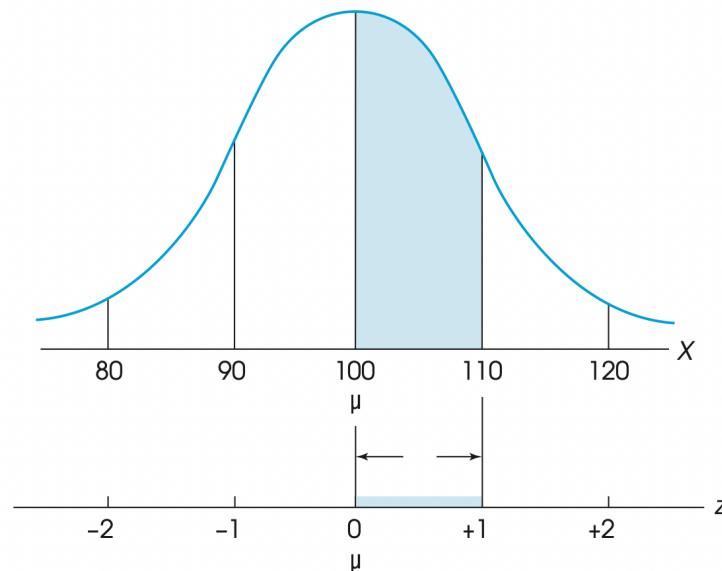
Name	Quiz 1 raw scores	Quiz 2 raw scores
Jessie	76	88
Sam	81	85

Who's doing better in the course?

We don't know from the raw scores alone

Developing z-scores

z-score: describes the position of a raw score in terms of its distance from the mean, when measured in *standard deviation units*. The z-score is positive if the value lies above the mean, and negative if it lies below the mean



Developing z-scores

Step 1: Calculate the mean for Quiz 1, μ_1 , and Quiz 2, μ_2

Name	Quiz 1 raw scores	Quiz 2 raw scores
Jessie	76	88
Sam	81	85

	Quiz 1	Quiz 2
μ	81.40	87.90

Developing z-scores

Step 2: subtract raw scores from the mean (*deviation scores*)

Name	Quiz 1 <i>raw scores</i>	Quiz 2 <i>raw scores</i>
Jessie	$76 - \mu_1$	$88 - \mu_2$
Sam	$81 - \mu_1$	$85 - \mu_2$

	Quiz 1	Quiz 2
μ	81.40	87.90

- Subtract raw scores from Quiz 1, x_1 , from the mean of Quiz 1, μ_1
- Subtract raw scores from Quiz 2, x_2 , from the mean of Quiz 2, μ_2

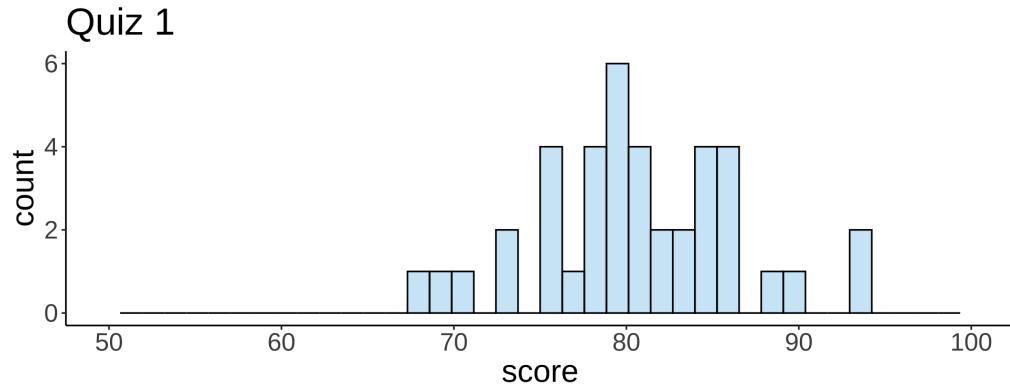
Developing z-scores

Name	Quiz 1 <i>deviation scores</i>	Quiz 2 <i>deviation scores</i>
Jessie	-5.40	0.10
Sam	-0.40	-2.90

	Quiz 1	Quiz 2
μ	81.40	87.90

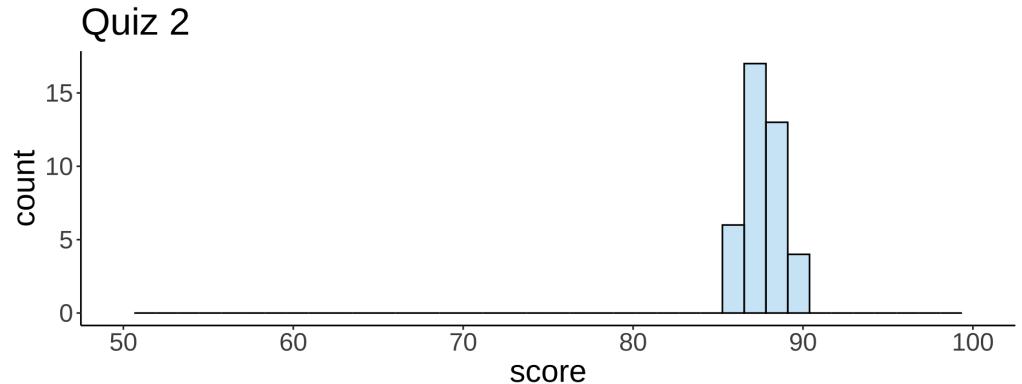
We now know how far away each score is from the mean, but we still have one more thing about distributions to consider

Developing z-scores



$$\mu = 81.40$$

$$\sigma = 8.40$$



$$\mu = 87.90$$

$$\sigma = 1.10$$

The variability for each Quiz is different

Developing z-scores

Name	Quiz 1 <i>deviation scores</i>	Quiz 2 <i>deviation scores</i>
Jessie	-5.40	0.10
Sam	-0.40	-2.90

	Quiz 1	Quiz 2
μ	81.40	87.90
σ	8.40	1.10

- We need to consider the standard deviation, σ , of each of the Quizzes

Developing z-scores

Step 3: divide each deviation score, $deviation_i$, by their respective standard deviations

Name	Quiz 1 <i>deviation scores</i>	Quiz 2 <i>deviation scores</i>
Jessie	$\frac{-5.40}{\sigma_1}$	$\frac{0.10}{\sigma_2}$
Sam	$\frac{-0.40}{\sigma_1}$	$\frac{-2.90}{\sigma_2}$

	Quiz 1	Quiz 2
μ	81.40	87.90
σ	8.40	1.10

Developing z-scores

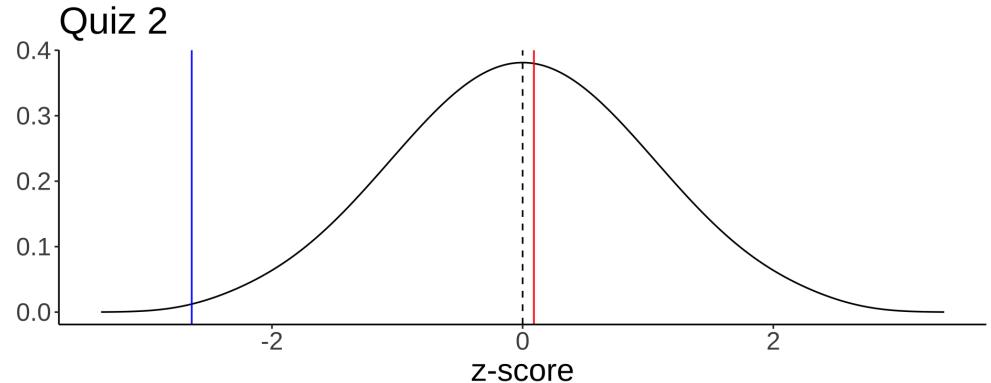
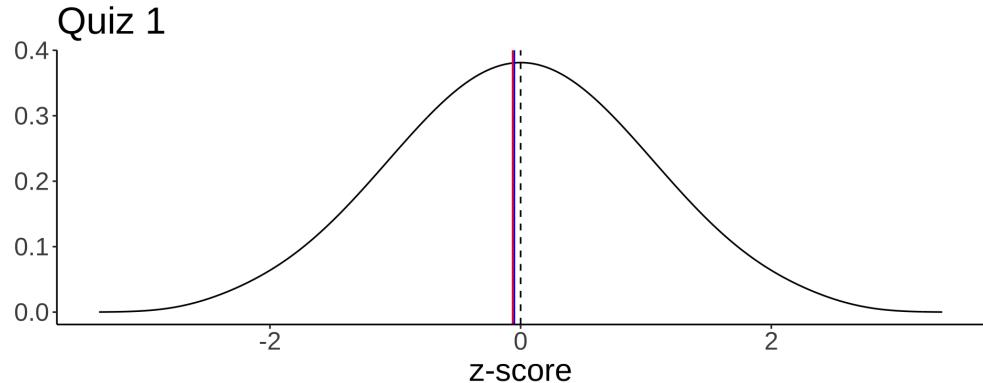
Name	Quiz 1 z-scores	Quiz 2 z-scores
Jessie	-0.64	0.09
Sam	-0.05	-2.64

	Quiz 1	Quiz 2
μ	81.40	87.90
σ	8.40	1.10

On Quiz 1, Jessie and Sam scored similarly.

On Quiz 2, Jessie did well whereas Sam did not.

Developing z-scores



Name	Quiz 1 z-scores	Quiz 2 z-scores
Jessie	-0.64	0.09
Sam	-0.05	-2.64

Jessie in red
Sam in blue

Developing z-scores

The result of the steps we just did is called calculating z-scores

Z-score

$$z_i = \frac{x_i - \mu}{\sigma}$$

$$z_i = \frac{\text{deviation}_i}{\sigma}$$

The number of standard deviations a particular score deviates from its corresponding mean

z-scoring is useful because a z-score of 1
means the same thing in any distribution

Developing z-scores

Z-score

$$z_i = \frac{x_i - \mu}{\sigma}$$

x score from z-score

$$x_i = \mu + z_i \sigma$$

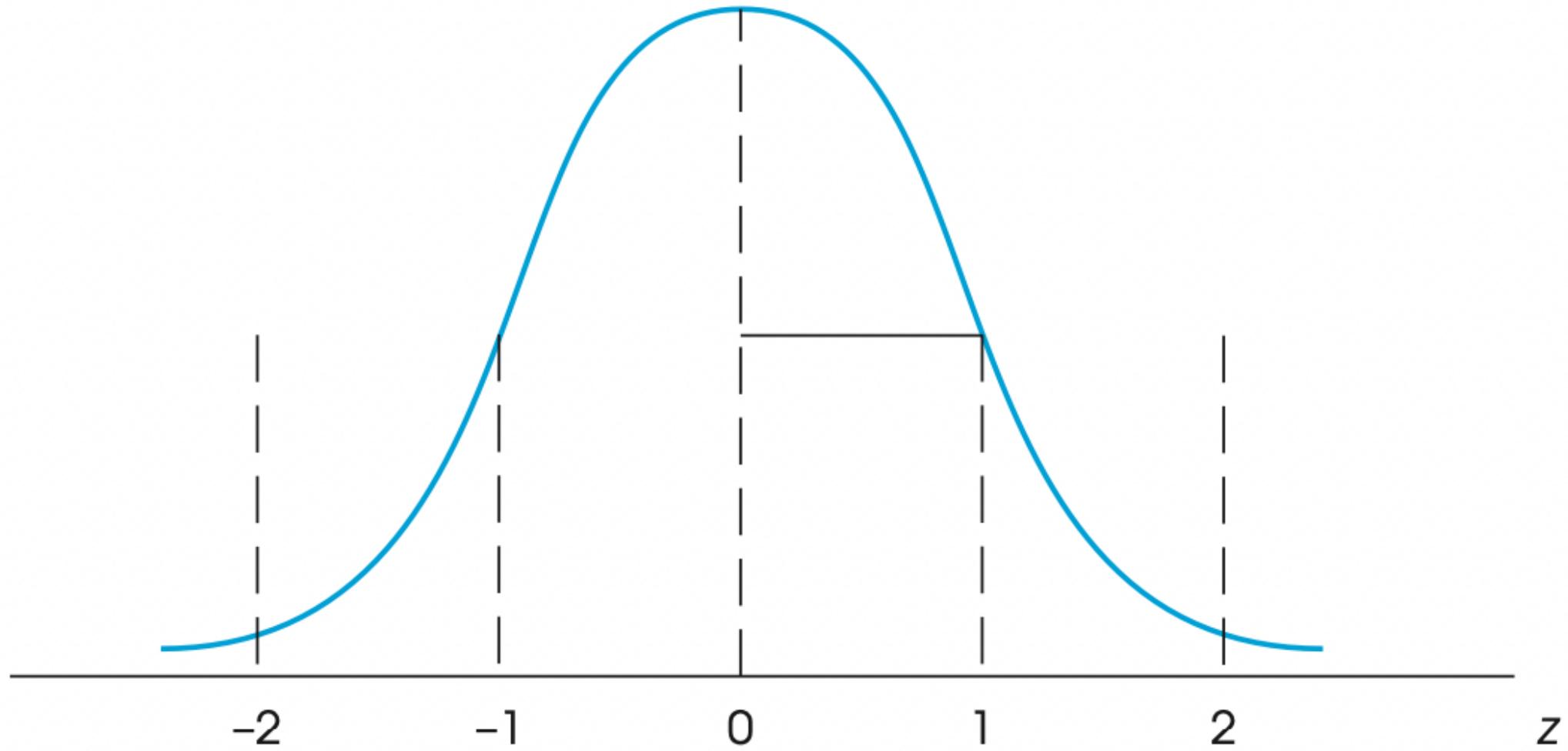
Example: Your IQ z-score is 1.4 and the distribution of IQ is $\mu = 100$ and $\sigma = 15$

$$x_i = 100 + (1.4)(15) = 121$$

Uses of z-scores

- (1) Describing scores in distributions with a single number
- (2) Equating and rescaling entire distributions
 - Z-scoring an entire distribution (or dataset) gets you a new distribution with a mean of 0, and a standard deviation of 1
 - But, the distribution **maintains** its shape
- (3) Comparing scores from non-equivalent datasets

Uses of z-scores



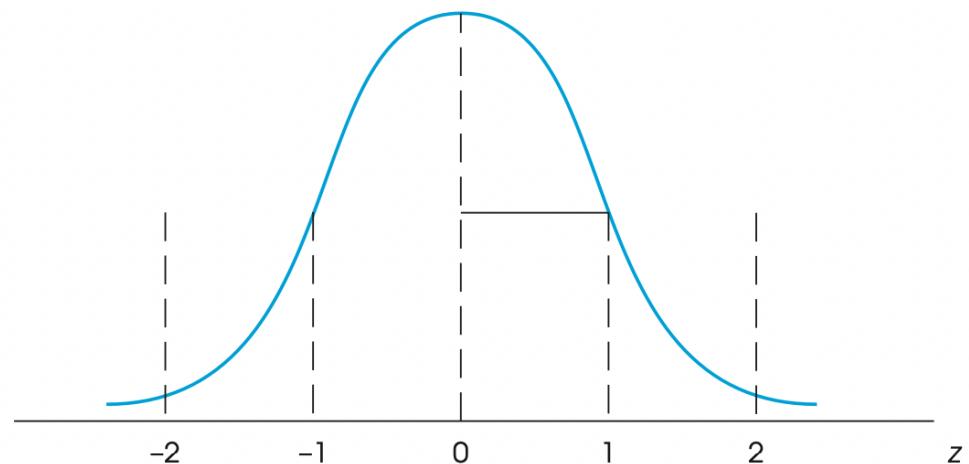
Uses of z-scores

Z-scoring a distribution that is already normal yields the **unit normal distribution**

$$\mu = 0$$

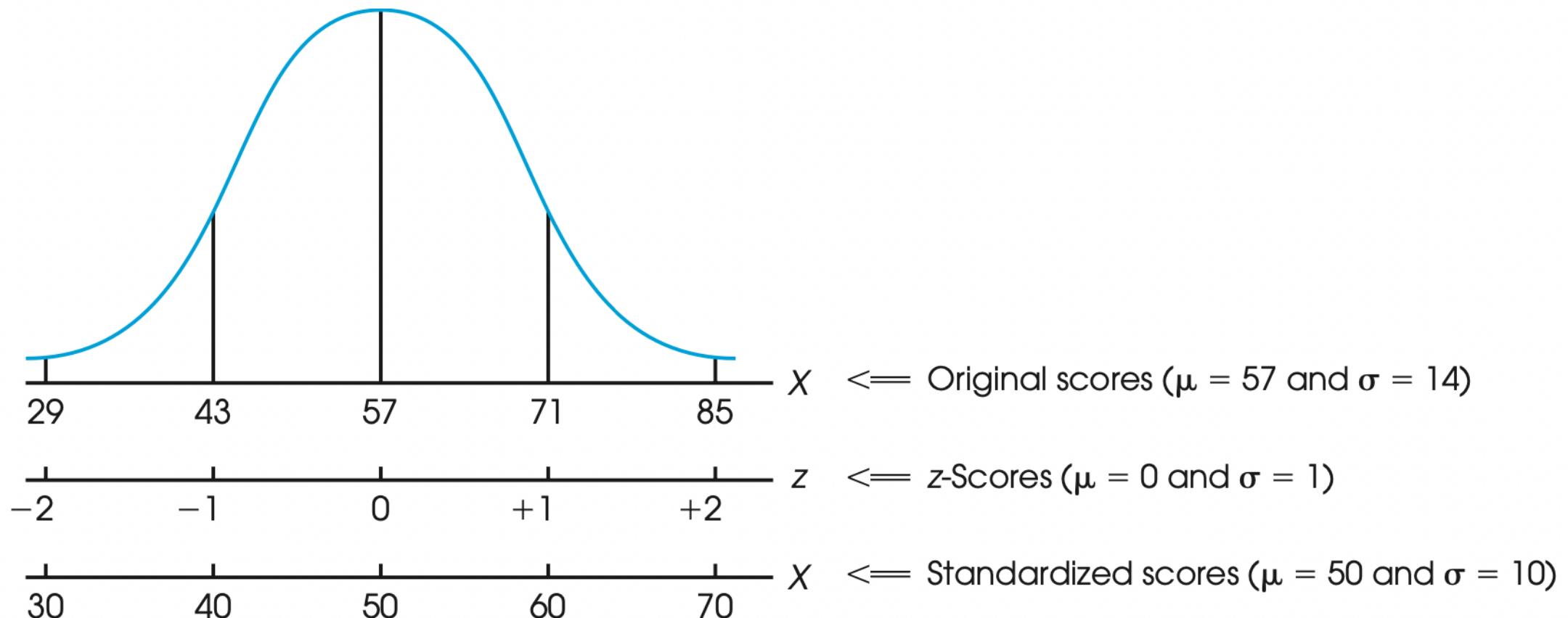
$$\sigma = 1$$

$$\frac{1}{\sqrt{2\pi 1^2}} e^{-\frac{(x-0)^2}{2*1^2}}$$



Uses of z-scores

Equating and rescaling entire distributions



z-score Practice

z-score Practice

Suppose, for example, Alex received a score of $x_1 = 60$ on a psychology exam and a score of $x_2 = 56$ on a biology test.

Suppose the psychology scores had $\mu_1 = 50$ and $\sigma_1 = 10$, and the biology scores had $\mu_2 = 48$ and $\sigma_2 = 4$.

Compute the z-score for each class and from the z-score, determine which class Alex did better in.

Z-score

$$z_i = \frac{x_i - \mu}{\sigma}$$

z-score Practice

Psychology

$$x_1 = 60$$

$$\mu_1 = 50$$

$$\sigma_1 = 10$$

$$z_1 = \frac{60 - 50}{10} = 1.00$$

Biology

$$x_2 = 56$$

$$\mu_2 = 48$$

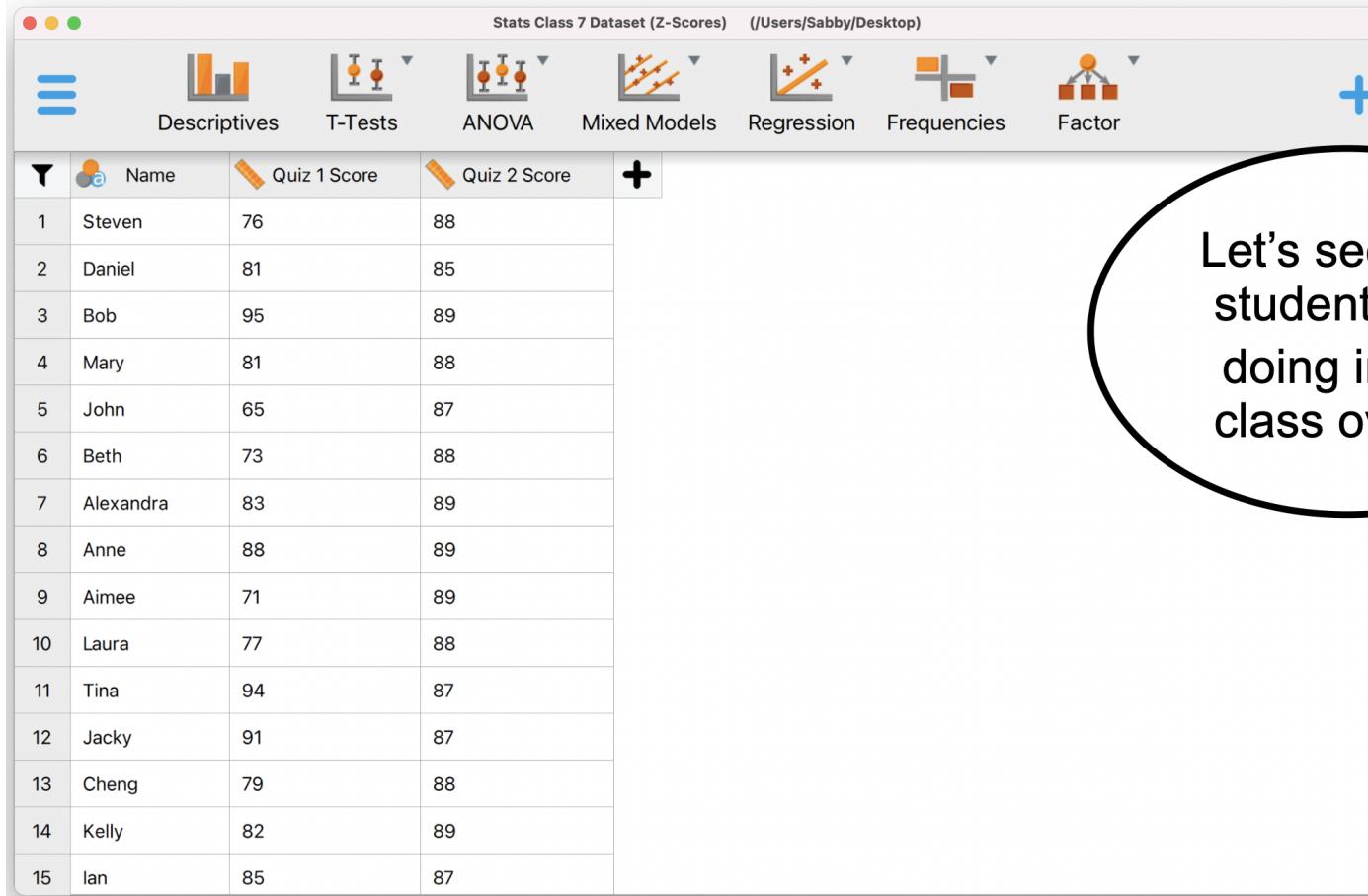
$$\sigma_2 = 4$$

$$z_2 = \frac{56 - 48}{4} = 2.00$$

Alex did better in biology compared to psychology

Computing z-scores in JASP

JASP: z-scores

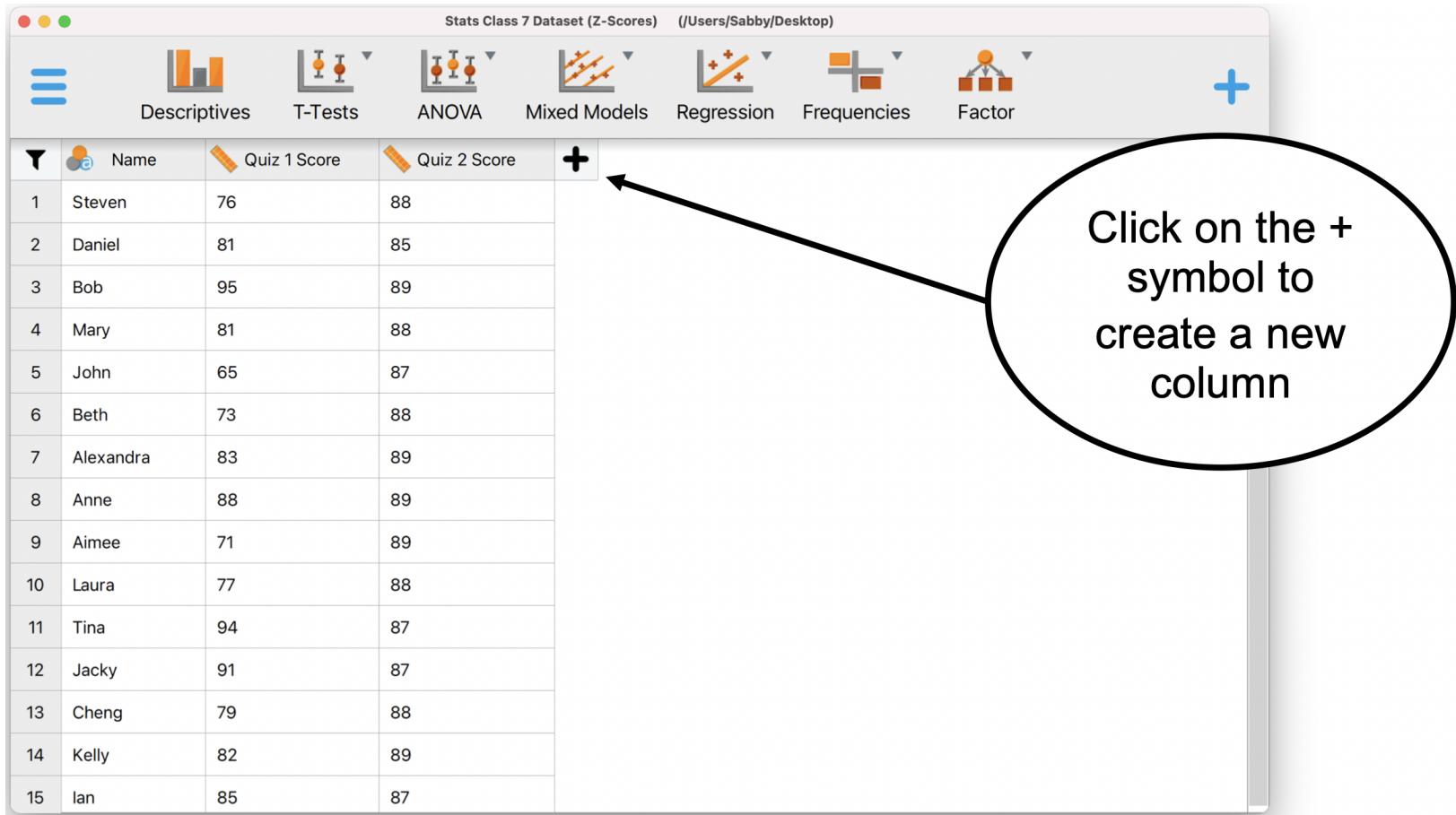


The screenshot shows the JASP software interface with the title "Stats Class 7 Dataset (Z-Scores) (/Users/Sabby/Desktop)". The top menu bar includes Descriptives, T-Tests, ANOVA, Mixed Models, Regression, Frequencies, Factor, and a plus sign icon. Below the menu is a table with columns: Name, Quiz 1 Score, and Quiz 2 Score. The table contains 15 rows of student data.

	Name	Quiz 1 Score	Quiz 2 Score
1	Steven	76	88
2	Daniel	81	85
3	Bob	95	89
4	Mary	81	88
5	John	65	87
6	Beth	73	88
7	Alexandra	83	89
8	Anne	88	89
9	Aimee	71	89
10	Laura	77	88
11	Tina	94	87
12	Jacky	91	87
13	Cheng	79	88
14	Kelly	82	89
15	Ian	85	87

Let's see how students are doing in the class overall

JASP: z-scores



Stats Class 7 Dataset (Z-Scores) (/Users/Sabby/Desktop)

Descriptives T-Tests ANOVA Mixed Models Regression Frequencies Factor +

	Name	Quiz 1 Score	Quiz 2 Score	+ [New Column]
1	Steven	76	88	
2	Daniel	81	85	
3	Bob	95	89	
4	Mary	81	88	
5	John	65	87	
6	Beth	73	88	
7	Alexandra	83	89	
8	Anne	88	89	
9	Aimee	71	89	
10	Laura	77	88	
11	Tina	94	87	
12	Jacky	91	87	
13	Cheng	79	88	
14	Kelly	82	89	
15	Ian	85	87	

Click on the + symbol to create a new column

JASP: z-scores

The screenshot shows the JASP software interface with a dataset titled "Stats Class 7 Dataset (Z-Scores) (/Users/Sabby/Desktop)". The dataset contains 15 rows of data with columns: Name, Quiz 1 Score, and Quiz 2 Score. A "Create Computed Column" dialog box is open, prompting the user to "Create Computed Column". The "Name:" field contains "Quiz 1 Z-Scores" with an arrow pointing to it. Below the name are buttons for "R" and a hand cursor icon. At the bottom of the dialog are buttons for "Scale", "Ordinal", "Nominal", "Text", "i" (info), "Create Column", and "x" (close). A large callout bubble on the right side of the dialog contains the text: "Create a title and then click on “Create Column”".

	Name	Quiz 1 Score	Quiz 2 Score
1	Steven	76	88
2	Daniel	81	85
3	Bob	95	89
4	Mary	81	88
5	John	65	87
6	Beth	73	88
7	Alexandra	83	89
8	Anne	88	89
9	Aimee	71	89
10	Laura	77	88
11	Tina	94	87
12	Jacky	91	87
13	Cheng	79	88
14	Kelly	82	89
15	Ian	85	87

JASP: z-scores

The screenshot shows the JASP software interface with the title "Stats Class 7 Dataset (Z-Scores)* (/Users/Sabby/Desktop)". The top menu bar includes Descriptives, T-Tests, ANOVA, Mixed Models, Regression, Frequencies, Factor, and a plus sign for new files. Below the menu is a section titled "Computed Column: Quiz 1 Z-Scores" with a formula editor. The formula is $\frac{\text{Quiz 1 Score} - \text{mean}(\text{Quiz 1 Score})}{\sigma \text{Quiz 1 Score}}$. A callout bubble with a black border contains the text: "Re-create the z-score formula, then 'Compute column'". Two arrows point from this text to the formula editor and the "Compute column" button in the toolbar below it. The main area shows a table with columns: Name, Quiz 1 Score, Quiz 2 Score, and Quiz 1 Z-Scores. The data rows are:

	Name	Quiz 1 Score	Quiz 2 Score	Quiz 1 Z-Scores
1	Steven	76	88	-0.639705
2	Daniel	81	85	-0.0473855
3	Bob	95	89	1.61111
4	Mary	81	88	-0.0473855
5	John	65	87	-1.94281
6	Beth	73	88	-0.995096
7	Alexandra	83	89	0.189542

JASP: z-scores

The screenshot shows the JASP interface with the title "Stats Class 7 Dataset (Z-Scores)* (/Users/Sabby/Desktop)". The top menu bar includes Descriptives, T-Tests, ANOVA, Mixed Models, Regression, Frequencies, Factor, and a plus sign for new files. A sidebar on the left lists variables: Name, Quiz 1 Score, Quiz 2 Score, Quiz 1 Z-Scores, and Quiz 2 Z-Scores. The main area displays a "Computed Column: Quiz 2 Z-Scores" formula: $\frac{\text{Quiz 2 Score} - \text{mean}(\text{Quiz 2 Score})}{\sigma \text{Quiz 2 Score}}$. Below this, a message says "Computed columns code applied". A table below shows the data with computed columns for Quiz 1 and Quiz 2 Z-Scores.

Computed Column: Quiz 2 Z-Scores

$$\frac{\text{Quiz 2 Score} - \text{mean}(\text{Quiz 2 Score})}{\sigma \text{Quiz 2 Score}}$$

Computed columns code applied

	Name	Quiz 1 Score	Quiz 2 Score	f _x Quiz 1 Z-Scores	f _x Quiz 2 Z-Scores
1	Steven	76	88	-0.639705	0.11847
2	Daniel	81	85	-0.0473855	-2.5471
3	Bob	95	89	1.61111	1.00699
4	Mary	81	88	-0.0473855	0.11847
5	John	65	87	-1.94281	-0.770054
6	Beth	73	88	-0.995096	0.11847
7	Alexandra	83	89	0.189542	1.00699

Use the same procedure to find z-scores for the second quiz

JASP: z-scores

The screenshot shows the JASP interface with the title "Stats Class 7 Dataset (Z-Scores)* (/Users/Sabby/Desktop)". The top menu bar includes Descriptives, T-Tests, ANOVA, Mixed Models, Regression, Frequencies, Factor, and a plus sign for new files. Below the menu is a toolbar with icons for each function. A large callout bubble points to the "Computed Column: Standardized Average" section.

Computed Column: Standardized Average

$+ - * \div / ^ \sqrt \% = \neq < \leq > \geq \wedge \vee | -$

Code applied:

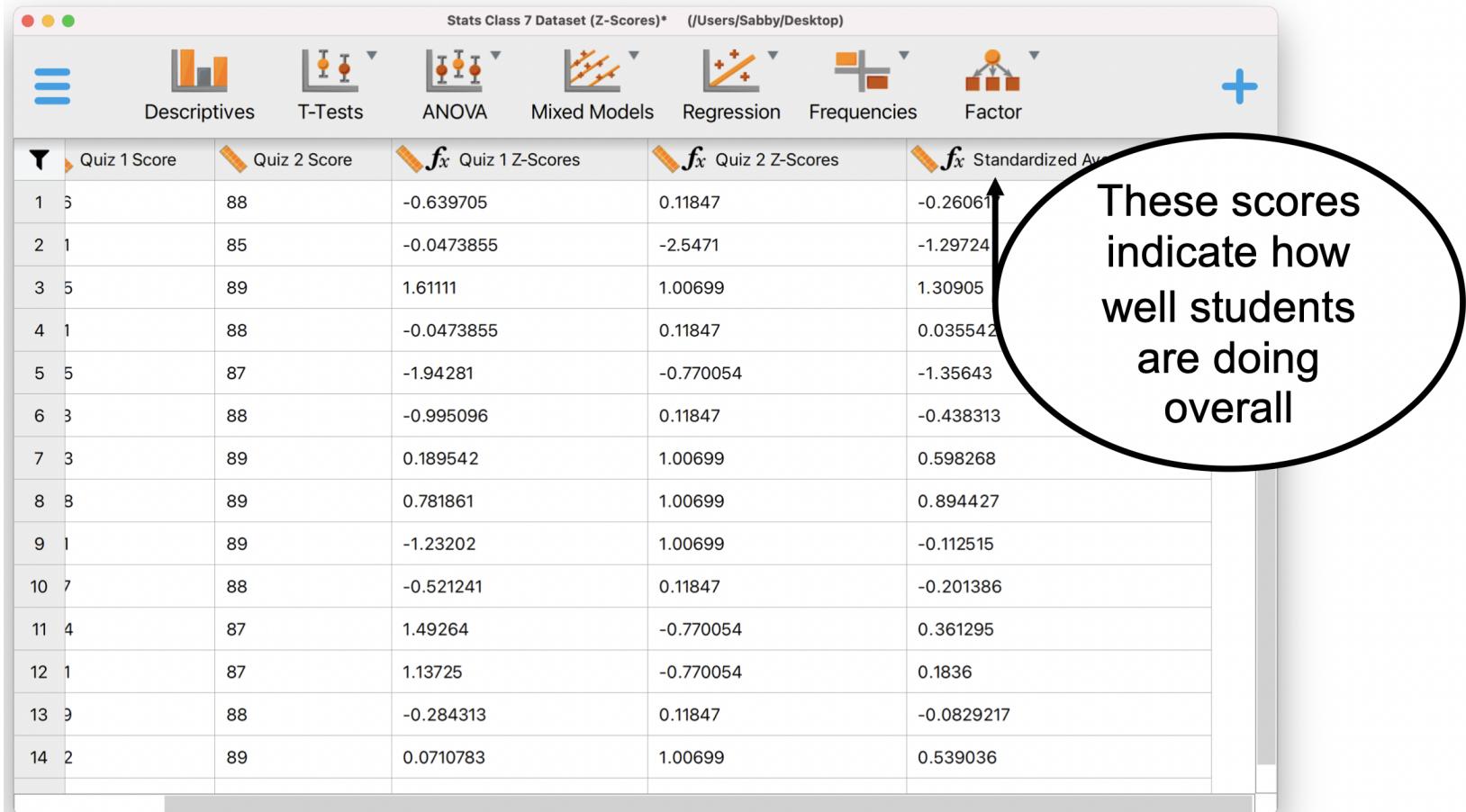
$$\frac{\text{Quiz 1 Z-Scores} + \text{Quiz 2 Z-Scores}}{2}$$

Computed columns code applied

The "Compute column" section shows the formula being typed. The data table below has columns for Quiz 1 Score, Quiz 2 Score, Quiz 1 Z-Scores, Quiz 2 Z-Scores, and Standardized Average. The first few rows of data are:

	Quiz 1 Score	Quiz 2 Score	f _x Quiz 1 Z-Scores	f _x Quiz 2 Z-Scores	f _x Standardized Average
1	3	88	-0.639705	0.11847	-0.260617
2	1	85	-0.0473855	-2.5471	-1.29724
3	5	89	1.61111	1.00699	1.30905
4	1	88	-0.0473855	0.11847	0.0355421
5	5	87	-1.94281	-0.770054	-1.35643
6	3	88	-0.995096	0.11847	-0.438313
7	3	89	0.189542	1.00699	0.598268

JASP: z-scores



The screenshot shows the JASP statistical software interface with a dataset titled "Stats Class 7 Dataset (Z-Scores)*". The dataset contains 14 rows of data with columns for Quiz 1 Score, Quiz 2 Score, Quiz 1 Z-Scores, Quiz 2 Z-Scores, and Standardized Average. An annotation with a callout bubble highlights the Standardized Average column, which represents the overall performance of each student relative to the class mean.

	Quiz 1 Score	Quiz 2 Score	f_x Quiz 1 Z-Scores	f_x Quiz 2 Z-Scores	f_x Standardized Avg
1	6	88	-0.639705	0.11847	-0.26061
2	1	85	-0.0473855	-2.5471	-1.29724
3	5	89	1.61111	1.00699	1.30905
4	1	88	-0.0473855	0.11847	0.035542
5	5	87	-1.94281	-0.770054	-1.35643
6	3	88	-0.995096	0.11847	-0.438313
7	3	89	0.189542	1.00699	0.598268
8	8	89	0.781861	1.00699	0.894427
9	1	89	-1.23202	1.00699	-0.112515
10	7	88	-0.521241	0.11847	-0.201386
11	4	87	1.49264	-0.770054	0.361295
12	1	87	1.13725	-0.770054	0.1836
13	9	88	-0.284313	0.11847	-0.0829217
14	2	89	0.0710783	1.00699	0.539036

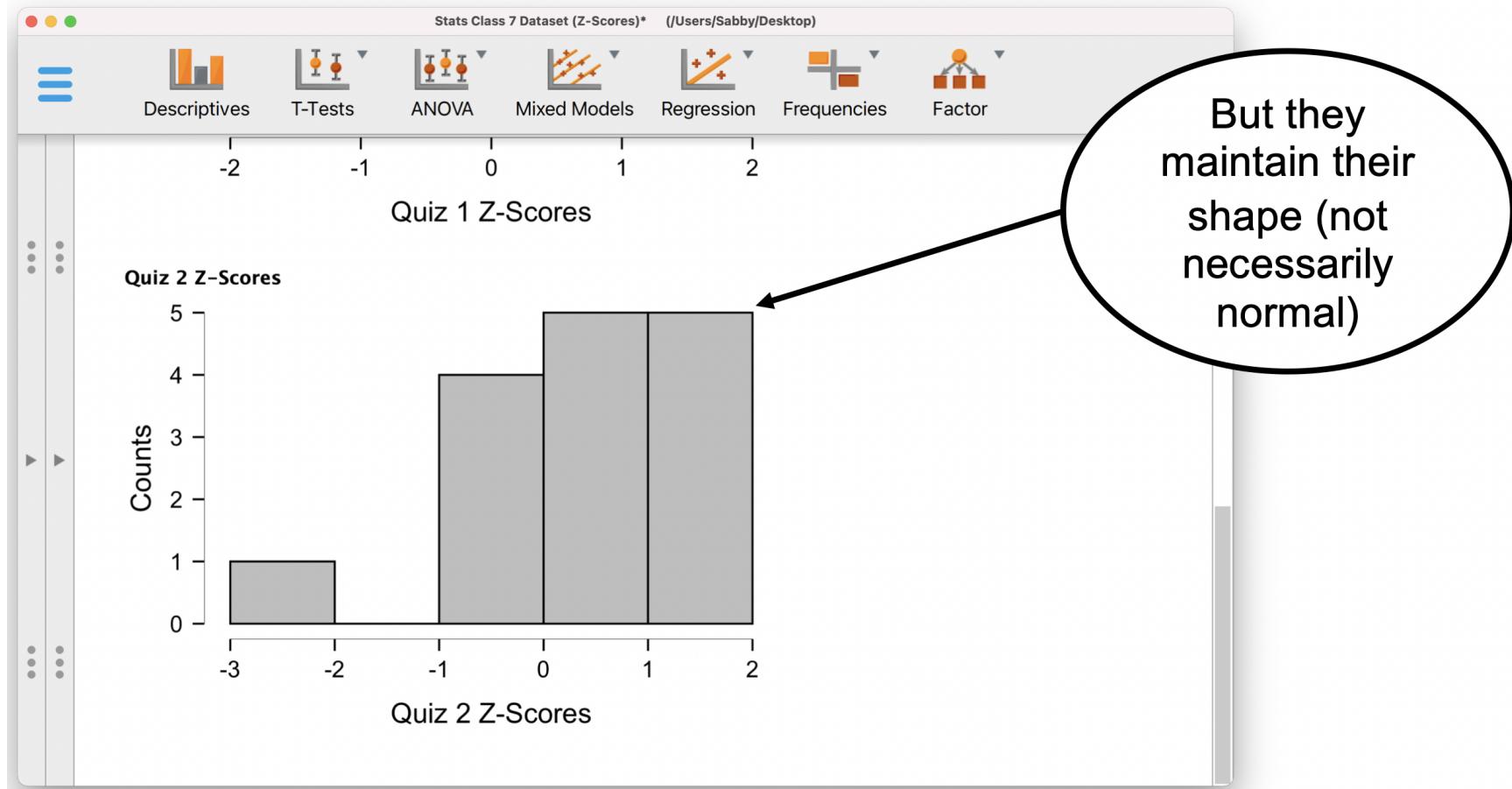
These scores indicate how well students are doing overall

JASP: z-scores

The screenshot shows the JASP interface with the title "Stats Class 7 Dataset (Z-Scores)* (/Users/Sabby/Desktop)". The top menu bar includes Descriptives, T-Tests, ANOVA, Mixed Models, Regression, Frequencies, and Factor. The left sidebar shows "Results" and "Descriptive Statistics". The main content displays a table of Descriptive Statistics for two columns: Quiz 1 Z-Scores and Quiz 2 Z-Scores. The table includes rows for Valid (15, 15), Missing (0, 0), Mean (-6.85e-16, 5.89e-15), Std. Deviation (1.00, 1.00), Minimum (-1.94, -2.55), and Maximum (1.61, 1.01). An annotation with a black arrow points from the text "Notice that these z-score columns have $\mu = 0, \sigma = 1$ " to the Mean row of the Quiz 2 Z-Scores table.

	Quiz 1 Z-Scores	Quiz 2 Z-Scores
Valid	15	15
Missing	0	0
Mean	-6.85e-16	5.89e-15
Std. Deviation	1.00	1.00
Minimum	-1.94	-2.55
Maximum	1.61	1.01

JASP: z-scores



Next time

Lecture

- Introduction to Statistical Significance

Reading

- Chapter Eight

