

EDP308: STATISTICAL LITERACY

The University of Texas at Austin, Fall 2020

RAZ: Rebecca A. Zárate, MA

Overview

- The Normal Distribution
 - ▣ AKA The Bell Curve or Gaussian Distribution
 - ▣ Historical Moment: Carl Gauss
 - ▣ Historical Moment: Maria-Sophia Germain
- Standardizing the Numbers
 - ▣ Z-Scores
- The 68%, 95%, 99% Rule
- Z-scores, Percentiles, and Probability
 - ▣ Looking up on the Unit Normal Table for Z-scores
 - ▣ Why Standardize?
- Z-scores in R

The Normal Distribution

The Normal Distribution



*Normal,
Bell-shaped Curve*

Percentage of cases in 8 portions of the curve

.13% 2.14% 13.59% 34.13% 34.13% 13.59% 2.14% .13%

Standard Deviations

-4σ -3σ -2σ -1σ 0 $+1\sigma$ $+2\sigma$ $+3\sigma$ $+4\sigma$

Cumulative Percentages

0.1% 2.3% 15.9% 50% 84.1% 97.7% 99.9%

Percentiles

1 5 10 20 30 40 50 60 70 80 90 95 99

Z scores

-4.0 -3.0 -2.0 -1.0 0 +1.0 +2.0 +3.0 +4.0

Also called the “Gaussian Curve”

Historical Moment: Carl Gauss

- **Johann Carl Friedrich Gauss** (April 30, 1777 – February 23, 1855)
- German mathematician and physicist born to poor, illiterate working class parents
 - ▣ They didn't record his date of birth...
- A teacher once told young Gauss to add all the numbers from 1 to 100, assuming it would take him a while. He solved it quickly by adding pairs...
 - ▣ $1+100 = 101$, $2+99 = 101$... You have 50 pairs of 101, so $50 \times 101 = 5050$
- His life was dedicated to learning and its process. He once stated, "It is not knowledge, but the act of learning, not possession but the act of getting there, which grants the greatest enjoyment."



Historical Moment: Maria-Sophia Germain

- **Maria-Sophia Germain** (April 1, 1776 – June 27, 1831)
- “Auguste Antoine Le Blanc” is the pseudonym Germain used while corresponding to her mathematician contemporaries.
- Taught herself Latin, mathematics, and physics in a time when women rarely obtained an education.
- Gauss was an informal mentor and correspondent.

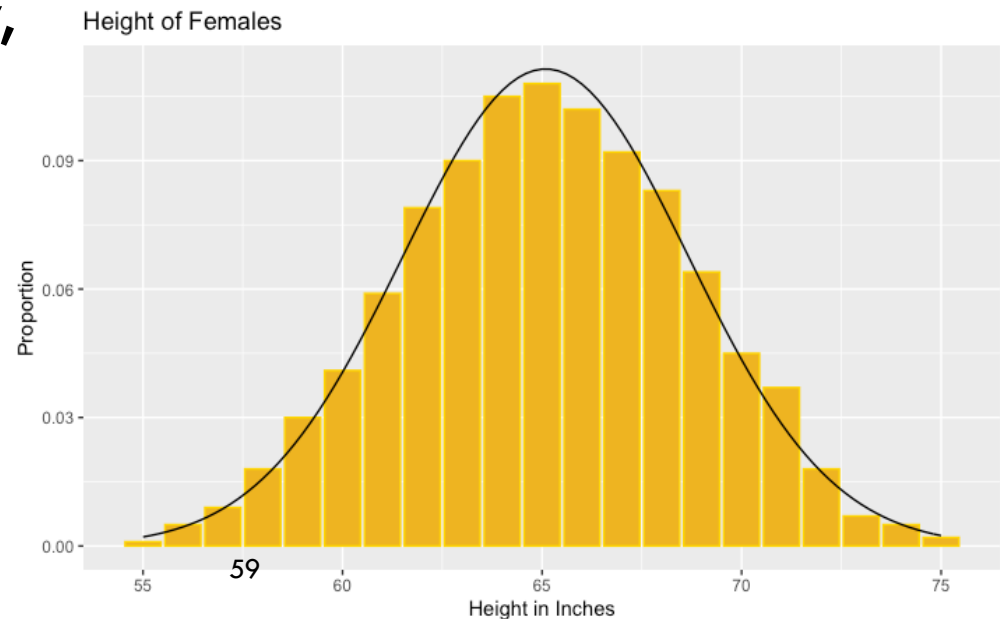


Sorbus: GERMAIN (portrait de), à l'âge de 21 ans.

The Normal Distribution

- The Normal Distribution (or Bell Curve) is a graphical representation of one variable's probability distribution.
- ▣ What is the probability a woman is 4'11 (59 inches)?
- ▣ Or said another way, what percent of the female population is 4'11?

Remember, proportions (which can be interpreted as probabilities) are just percentages in decimal form and range from 0-1 instead of 1%-100%.

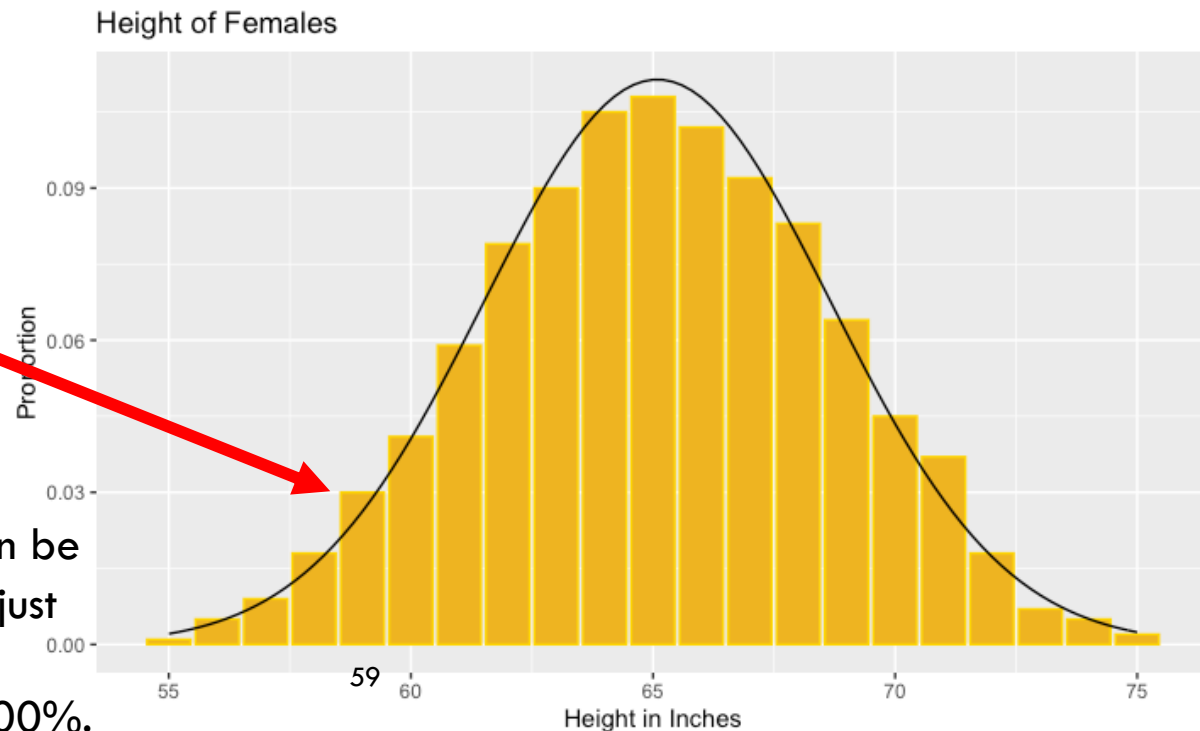


The Normal Distribution

- What is the probability a woman is 4'11 (59 inches)?
- Or said another way, what percent of the female population is 4'11?

If we look at the proportion on the Y-axis that corresponds to 59 inches (4'11), we see the proportion is about .03, or 3%. About 3% of the female population is 4'11. The probability of any given woman being 4'11 is .03

Remember, proportions (which can be interpreted as probabilities) are just percentages in decimal form and range from 0-1 instead of 1%-100%.



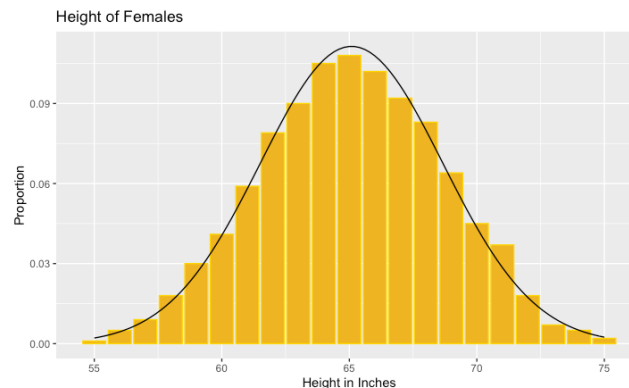
Probability and Distribution

- Why do we care?
 - Statistics is fraught with uncertainty... Due to sampling error, measurement error, and just the nature of sampling a subset
 - In statistics, we don't "prove" things, we can only really provide evidence for certain things happening or not happening... the likelihood of things...
 - How likely is it we see a woman who is 4'11 or 6'2?
 - How likely is it Flint, MI water is not contaminated with lead?
 - How likely is it UT-Austin will be having in person classes?
 - How likely is it you will catch COVID-19?
 - How likely is it you will make an A in this class?
 - Things like the normal distribution help us quantify some of that uncertainty.

Standardizing the Numbers

Translations

- Here is a lovely normal distribution depicting women's heights, but what if we want to see the distribution of some other variable? Or what if I don't really care about your actual height, but rather I want to know how tall you are relative to others of the same sex? How do we reduce it down or “translate it” to just the important core bits?



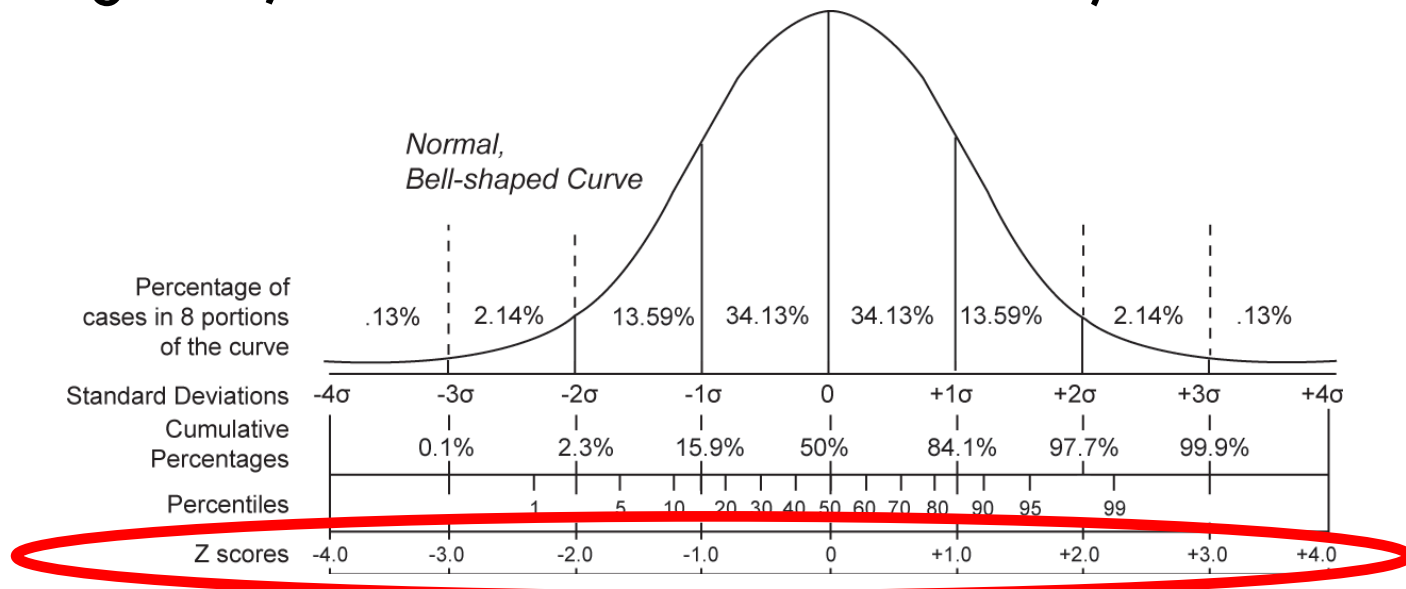
Zstandardizing with Z scale! (Standardizing)

- We can take any scale and transform it into simplified, standard scale... the Z-scale.
- *A z-score tells you how many standard deviations a person's score is away from the mean.*
- The Z-scale and the Z-scores associated with it allow us to make comparison across different scales, test, and situations.

Z

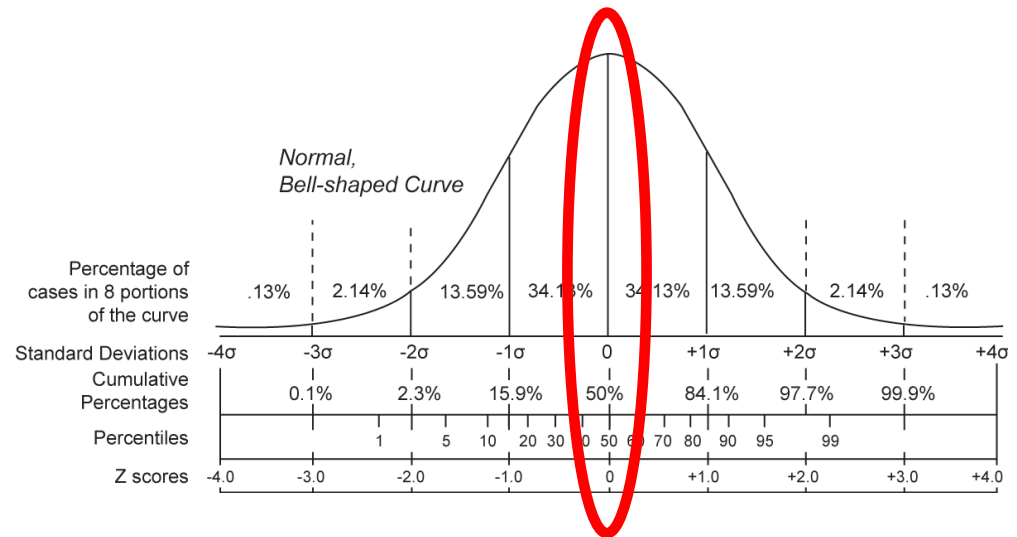
The Z-scale

- The Z-scale is a scale that has a mean of zero, and a standard deviation of 1
- This is very useful to us because we can take any normally distributed variable (ex. height, intelligence, even measurement errors!)



Different Languages, Same Message

- There is a lot going on with this curve... Lots of numbers each stacked on top of each other. But what is important to note is that each row of numbers is saying the same thing but in a slightly different way, another language.
- The mean of 1-100 is 50, the 50th percentile, or in the standard scale where we transform the data to make the mean 0, the z-score is 0. All the same message...



50th Percentile, 50%, z-score of 0, all are saying you are in the middle of everyone.

Different Languages, Same Message

- So how do we actually go about taking a scale like height, or intelligence, or your SAT score and turn into the simplified z-score?
- We need to know a few things...
 - ▣ The mean (μ)
 - ▣ The standard deviation (σ)
 - ▣ The score of interest (x , i.e. the person's score we want to translate)

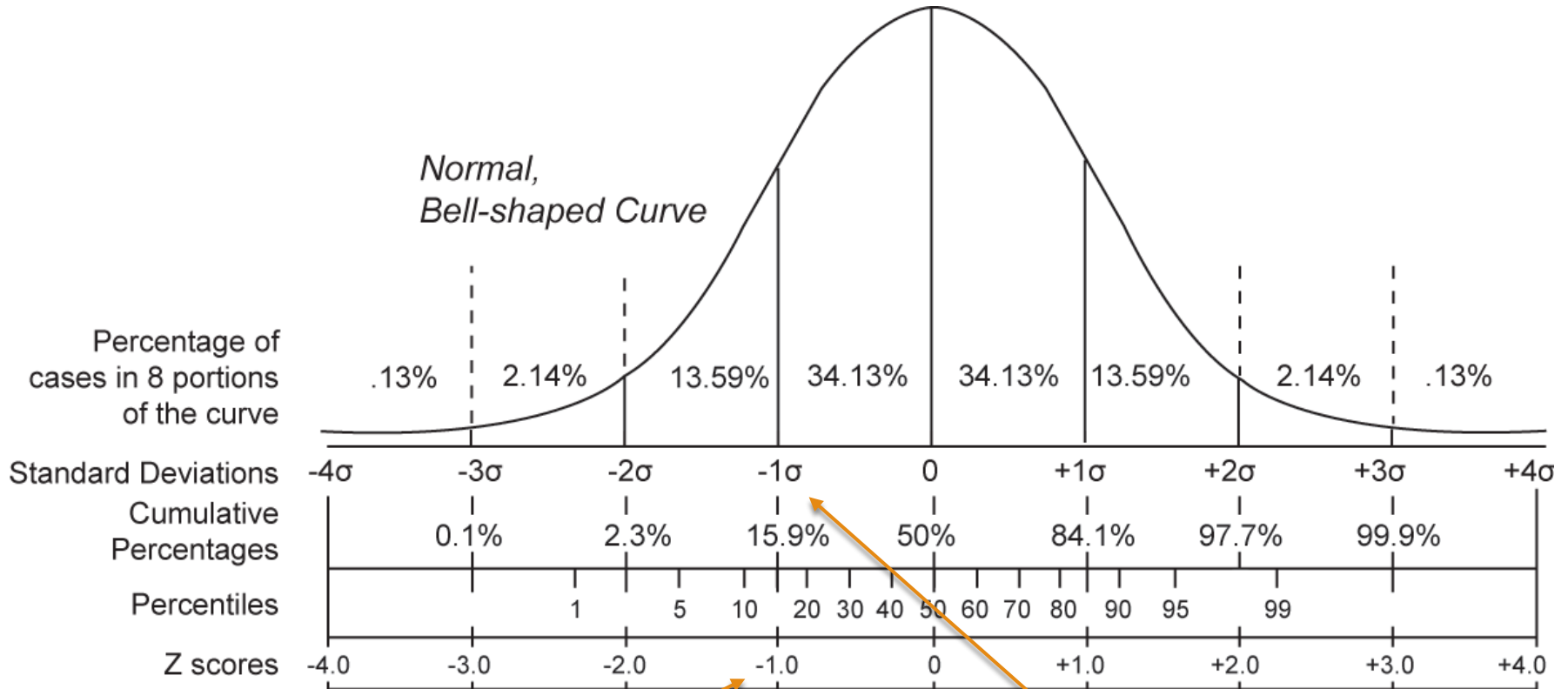
Z Standard Score

- Z-scores are standardized standard deviations, think of them as the “units of standard deviation”, the scale or “language” of standard deviants
- ▣ You can take a raw score, like an IQ test, and “translate it” into a z-score which makes interpretation easier and more comparable.

$$z - score = \frac{\text{raw score} - \text{mean}}{\text{standard deviation}}$$

$$z - score = \frac{x - \mu}{\sigma}$$

Z Standard Score



Notice that Z-scores are the same values as Standard Deviations.

SAT Z-Scores

Convert the raw scores
into Z-scores.

$$z - score = \frac{x - \mu}{\sigma}$$

$$\mu = 1000$$

$$\sigma = 210$$

SAT Score (x)	Z-Score
1555	
980	
585	
790	
1270	
1035	

*Data are modified for ease simplicity but are fairly close to 2019 SAT data.

SAT Z-Scores

Convert the raw scores
into Z-scores.

$$2.64 = \frac{1555 - 1000}{210}$$

$$\mu = 1000$$

$$\sigma = 210$$

SAT Score (x)	Z-Score
1555	2.64
980	
585	
790	
1270	
1035	

*Data are modified for ease simplicity but are fairly close to 2019 SAT data.

SAT Z-Scores

Convert the raw scores
into Z-scores.

$$z - score = \frac{x - \mu}{\sigma}$$

$$\mu = 1000$$

$$\sigma = 210$$

SAT Score (x)	Z-Score
1555	2.64
980	-0.10
1000	1.00
585	-1.98
790	-1.00
1270	1.29
1035	0.17

*Data are modified for ease simplicity but are fairly close to 2019 SAT data.

SAT Z-Scores

Convert the raw scores
into Z-scores.

$$z - score = \frac{x - \mu}{\sigma} \quad \begin{array}{l} \mu = 1000 \\ \sigma = 210 \end{array}$$

Knowing what you know
now, what would a z-score
of -3 tell you?

SAT Score (x)	Z-Score
1555	2.64
980	-0.10
1000	1.00
585	-1.98
790	-1.00
1270	1.29
1035	0.17

SAT Z-Scores

Convert the raw scores
into Z-scores.

$$z - score = \frac{x - \mu}{\sigma} \quad \begin{array}{l} \mu = 1000 \\ \sigma = 210 \end{array}$$

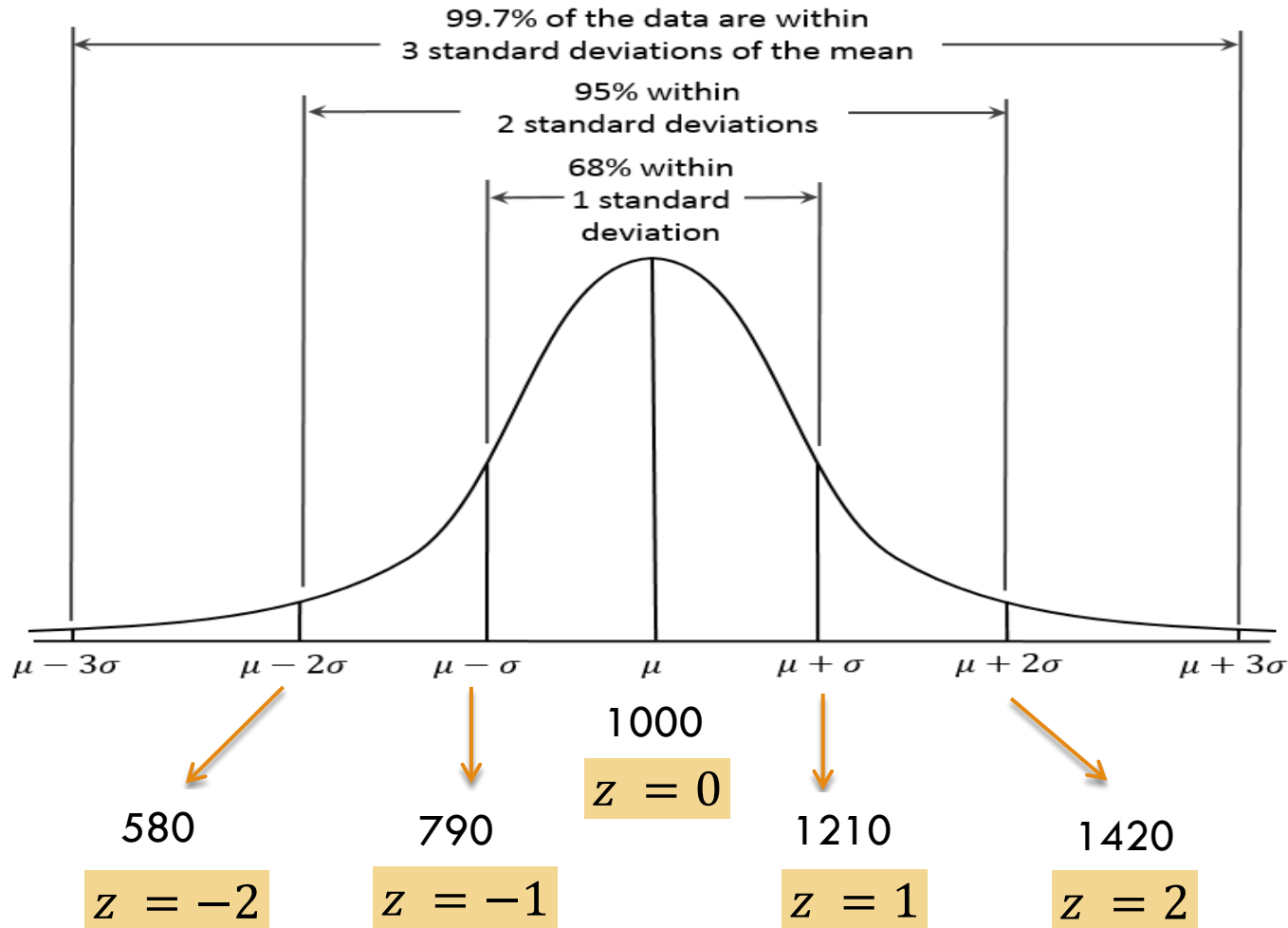
Knowing what you know
now, what would a z-score
of -3 tell you?

The person would have a
very low score of 370*...

SAT Score (x)	Z-Score
1555	2.64
980	-0.10
1000	1.00
585	-1.98
790	-1.00
1270	1.29
1035	0.17

*Lowest possible score on SAT actually 400

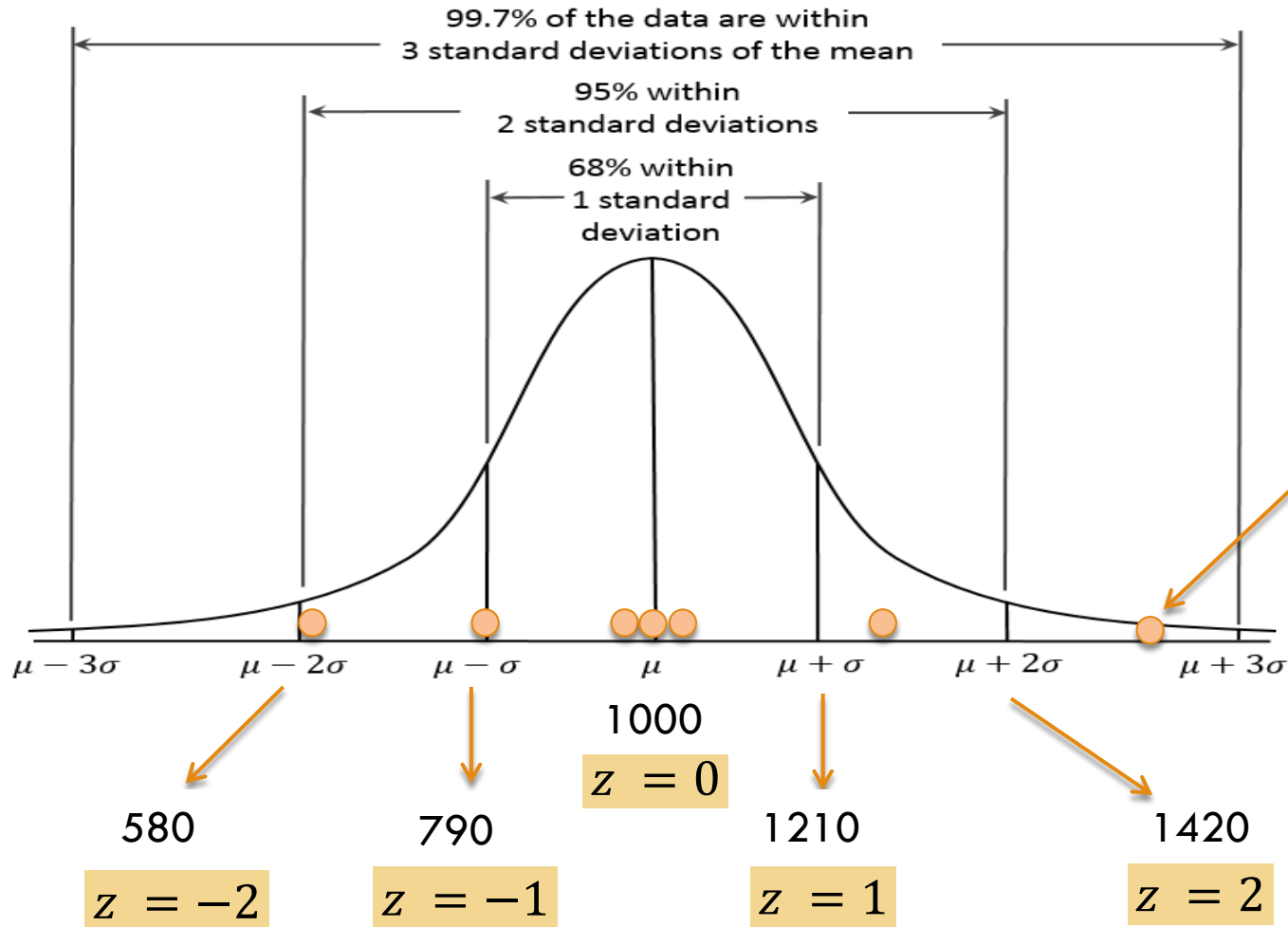
On the Curve



$$\mu = 1000$$
$$\sigma = 210$$

Here we have the z-scores and the SAT values associated with those z-scores mapped onto the normal curve.

On the Curve



$$\mu = 1000$$
$$\sigma = 210$$

SAT Score (x)	Z-Score
1555	2.64
980	-0.10
1000	0.00
585	-1.98
790	-1.00
1270	1.29
1035	0.17

Translating Back and Forth

$$z - score = \frac{x - \mu}{\sigma}$$

$$x \text{ (raw score)} = z * \sigma + \mu$$

$$\mu \text{ (mean)} = x - z * \sigma$$

$$\sigma \text{ (standard deviation)} = \frac{x - \mu}{z - score}$$

By rearranging the z-score formula (good ole algebra), you can determine any missing piece.

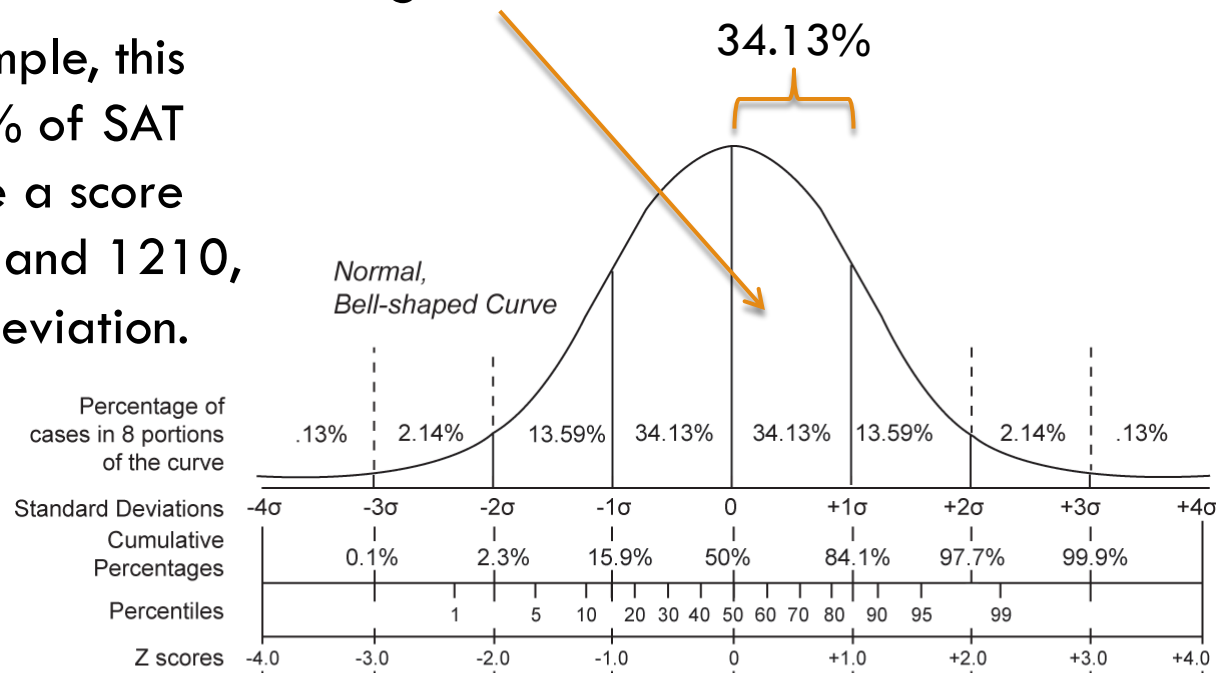
(Rather than memorizing them all, try to work them out with algebra, then you only have to know one equation.)

The 68, 95, 99 Rule

How Many in the Curve?

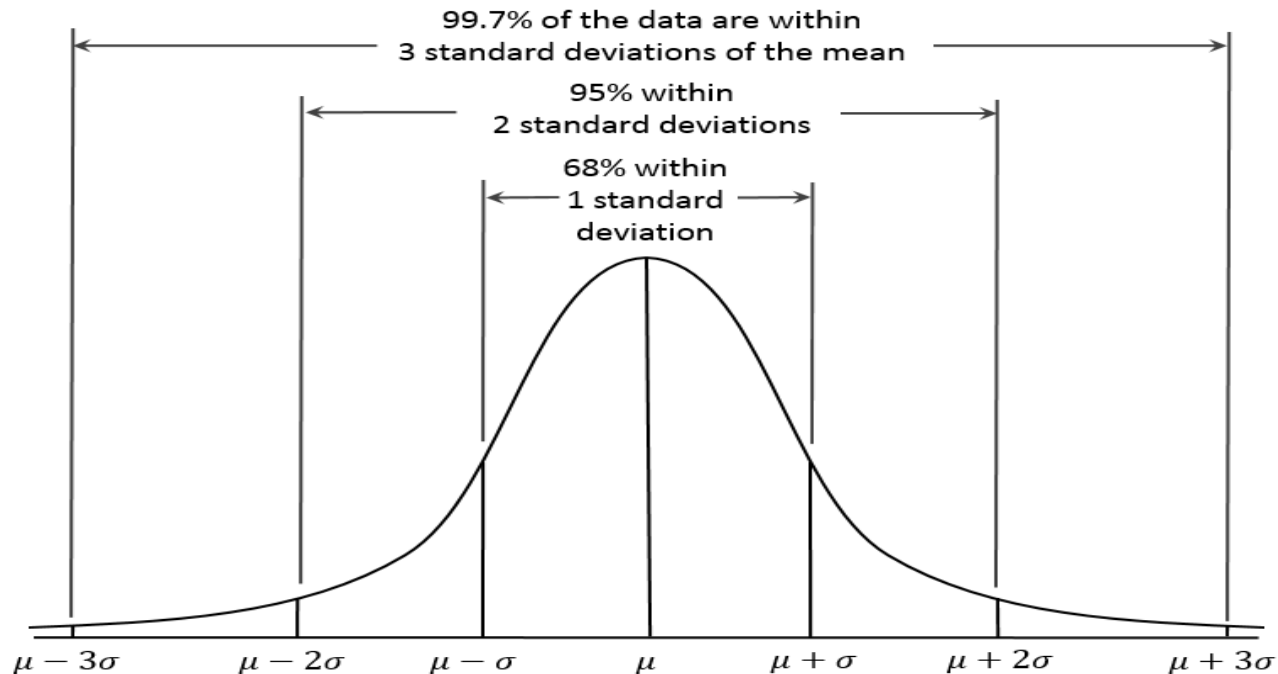
- Notice these percentages... This tells us what percent of observations (ex. people) fall within this range. From a z-score of 0 to +1, there are about 34% of people in this range.

- ▣ In our SAT example, this means that 34% of SAT test takers have a score between 1000 and 1210, one standard deviation.



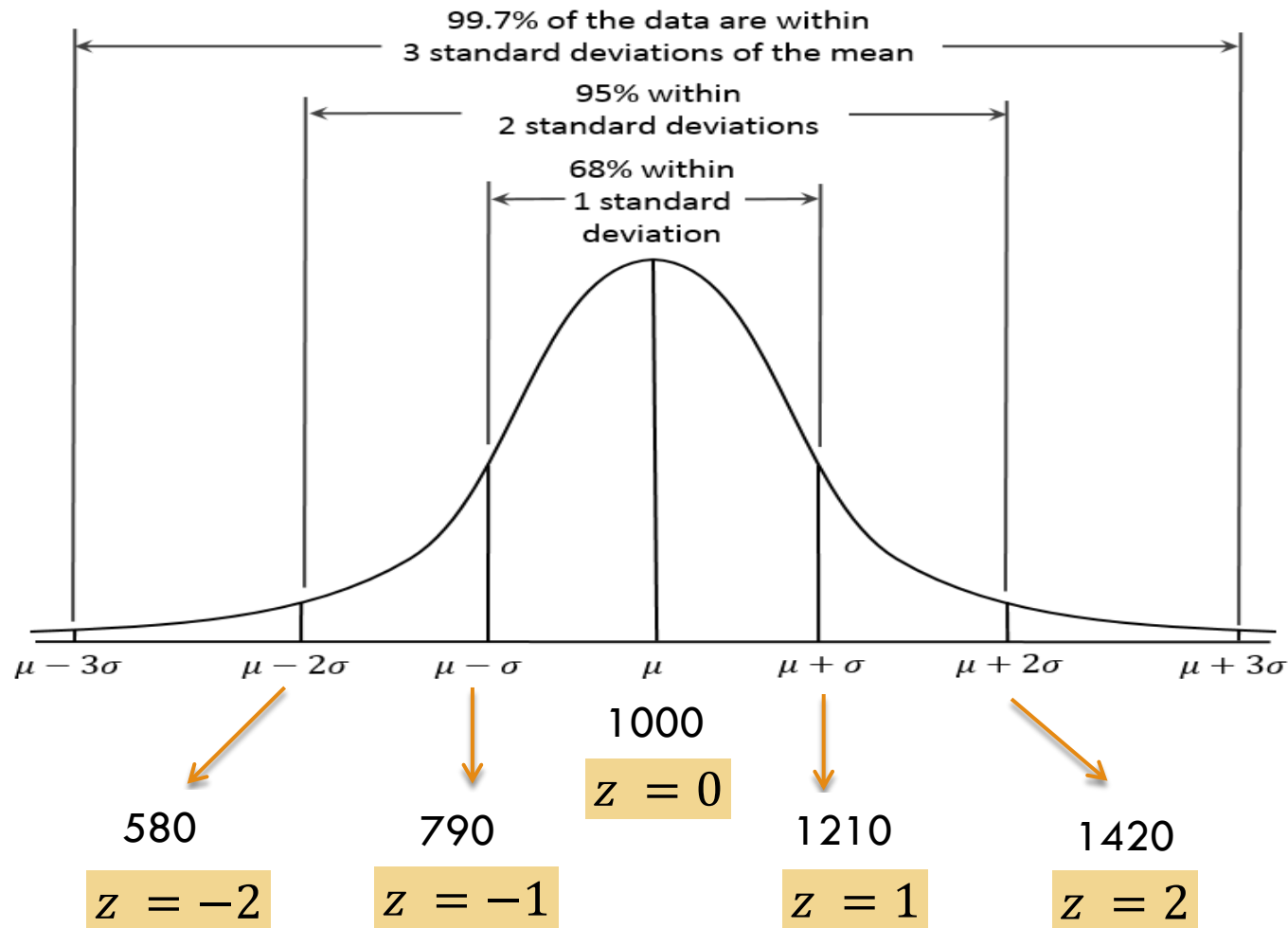
68, 95, 99 Rule

- For normally distributed population data where the mean (μ) and the standard deviation (σ) are known, the curve follows certain rules.



For normally distributed data: 68% of observations fall between ± 1 SDs
95% of observations fall between ± 2 SDs
99.7% of observations fall between ± 3 SDs

68, 95, 99 Rule with SAT Data



$$\mu = 1000$$
$$\sigma = 210$$

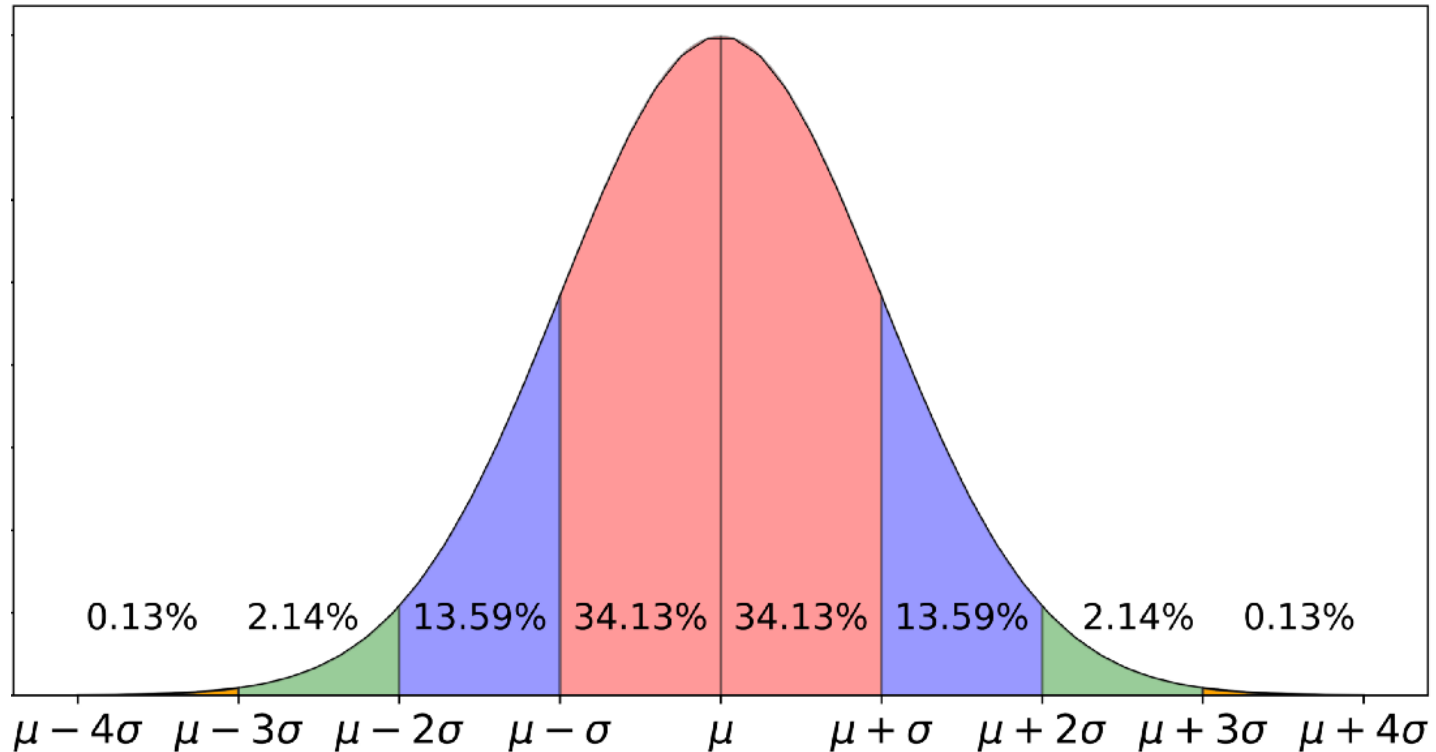
68% of people score between 790 to 1210

95% of people score between 580 and 1420

99.7% of people score between 370* and 1630*

*Actual SAT range 400-1600

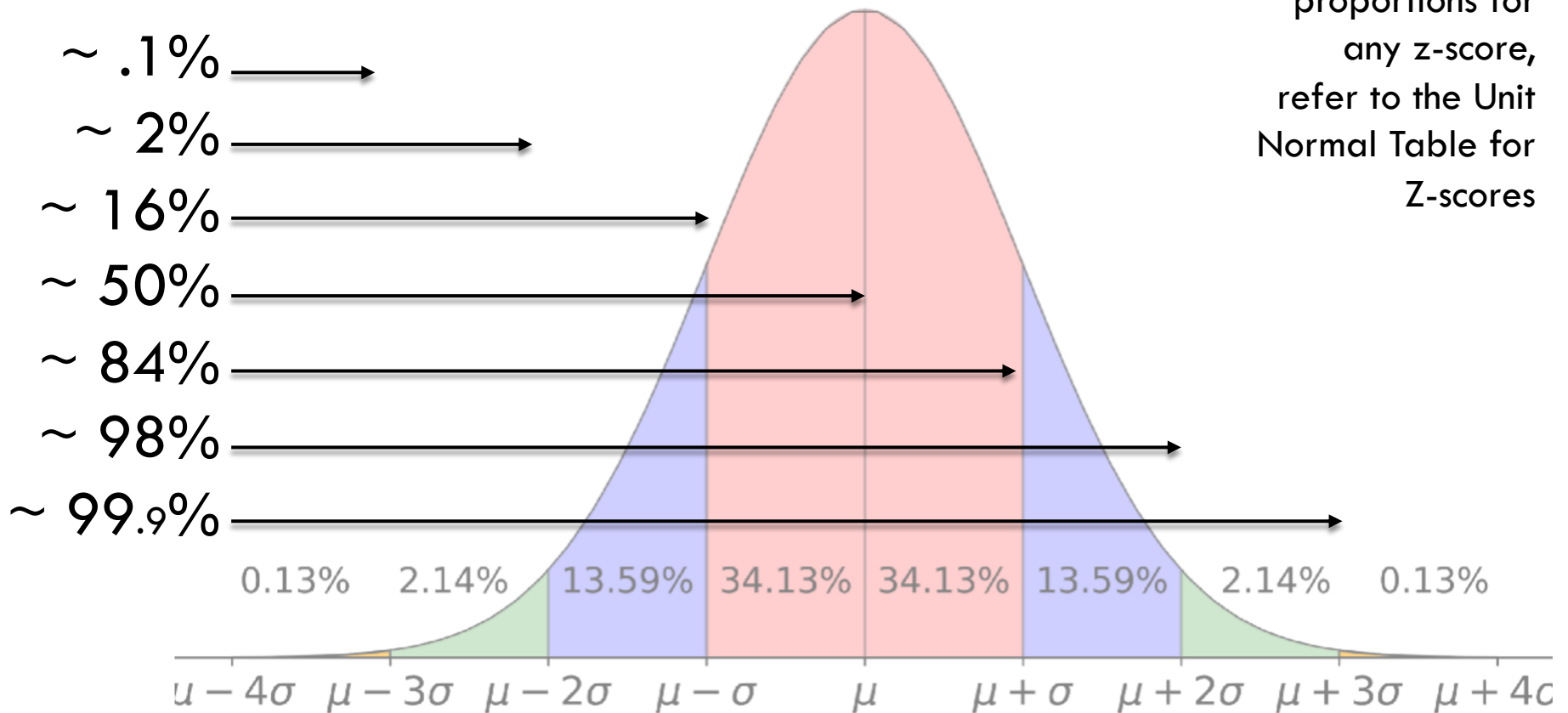
Proportions



For normally distributed data:

- $\sim .1\%$ of observations fall between ± 4 to ± 3 SDs
- $\sim 2\%$ of observations fall between ± 3 and ± 2 SDs
- $\sim 14\%$ of observations fall between ± 2 and ± 1 SDs
- $\sim 34\%$ of observations fall between ± 1 and 0 SDs

Cumulative



Z-Scores, Percentiles, and Probability

Z-scores and Percentiles

- You can look up any Z-score and determine the proportion that is equal to and above or equal to and below that given z-score
- This is also how you get percentile scores
 - ▣ Think back to your SAT or ACT
 - ▣ Or ahead to your GRE...

TABLE B.1 The Unit Normal Table*

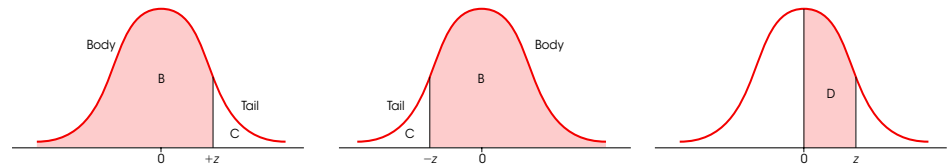
*Column A lists z-score values. A vertical line drawn through a normal distribution at a z-score location divides the distribution into two sections.

Column B identifies the proportion in the larger section, called the *body*.

Column C identifies the proportion in the smaller section, called the *tail*.

Column D identifies the proportion between the mean and the z-score.

Note: Because the normal distribution is symmetrical, the proportions for negative z-scores are the same as those for positive z-scores.



(A) z	(B) Proportion in Body	(C) Proportion in Tail	(D) Proportion Between Mean and z	(A) z	(B) Proportion in Body	(C) Proportion in Tail	(D) Proportion Between Mean and z
0.00	.5000	.5000	.0000	0.25	.5987	.4013	.0987
0.01	.5040	.4960	.0040	0.26	.6026	.3974	.1026
0.02	.5080	.4920	.0080	0.27	.6064	.3936	.1064
0.03	.5120	.4880	.0120	0.28	.6103	.3897	.1103
0.04	.5160	.4840	.0160	0.29	.6141	.3859	.1141
0.05	.5199	.4801	.0199	0.30	.6179	.3821	.1179
0.06	.5239	.4761	.0239	0.31	.6217	.3783	.1217
0.07	.5279	.4721	.0279	0.32	.6255	.3745	.1255
0.08	.5319	.4681	.0319	0.33	.6293	.3707	.1293
0.09	.5359	.4641	.0359	0.34	.6331	.3669	.1331
0.10	.5398	.4602	.0398	0.35	.6368	.3632	.1368
0.11	.5438	.4562	.0438	0.36	.6406	.3594	.1406
0.12	.5478	.4522	.0478	0.37	.6443	.3557	.1443
0.13	.5517	.4483	.0517	0.38	.6480	.3520	.1480
0.14	.5557	.4443	.0557	0.39	.6517	.3483	.1517
0.15	.5596	.4404	.0596	0.40	.6554	.3446	.1554
0.16	.5636	.4364	.0636	0.41	.6591	.3409	.1591
0.17	.5675	.4325	.0675	0.42	.6628	.3372	.1628
0.18	.5714	.4286	.0714	0.43	.6664	.3336	.1664
0.19	.5753	.4247	.0753	0.44	.6700	.3300	.1700
0.20	.5793	.4207	.0793	0.45	.6736	.3264	.1736
0.21	.5832	.4168	.0832	0.46	.6772	.3228	.1772
0.22	.5871	.4129	.0871	0.47	.6808	.3192	.1808
0.23	.5910	.4090	.0910	0.48	.6844	.3156	.1844
0.24	.5948	.4052	.0948	0.49	.6879	.3121	.1879

Z-scores and Percentiles

- For example, the percentile of someone who scored a 1035 on the SAT, $z\text{-score} = 0.17$, would be .5675. This means that ~57% of test takers scored below 1035.

TABLE B.1 The Unit Normal Table*

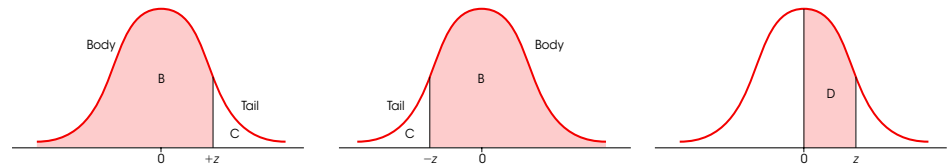
*Column A lists z -score values. A vertical line drawn through a normal distribution at a z -score location divides the distribution into two sections.

Column B identifies the proportion in the larger section, called the *body*.

Column C identifies the proportion in the smaller section, called the *tail*.

Column D identifies the proportion between the mean and the z -score.

Note: Because the normal distribution is symmetrical, the proportions for negative z -scores are the same as those for positive z -scores.



(A) z	(B) Proportion in Body	(C) Proportion in Tail	(D) Proportion Between Mean and z	(A) z	(B) Proportion in Body	(C) Proportion in Tail	(D) Proportion Between Mean and z
0.00	.5000	.5000	.0000	0.25	.5987	.4013	.0987
0.01	.5040	.4960	.0040	0.26	.6026	.3974	.1026
0.02	.5080	.4920	.0080	0.27	.6064	.3936	.1064
0.03	.5120	.4880	.0120	0.28	.6103	.3897	.1103
0.04	.5160	.4840	.0160	0.29	.6141	.3859	.1141
0.05	.5199	.4801	.0199	0.30	.6179	.3821	.1179
0.06	.5239	.4761	.0239	0.31	.6217	.3783	.1217
0.07	.5279	.4721	.0279	0.32	.6255	.3745	.1255
0.08	.5319	.4681	.0319	0.33	.6293	.3707	.1293
0.09	.5359	.4641	.0359	0.34	.6331	.3669	.1331
0.10	.5398	.4602	.0398	0.35	.6368	.3632	.1368
0.11	.5438	.4562	.0438	0.36	.6406	.3594	.1406
0.12	.5478	.4522	.0478	0.37	.6443	.3557	.1443
0.13	.5517	.4483	.0517	0.38	.6480	.3520	.1480
0.14	.5557	.4443	.0557	0.39	.6517	.3483	.1517
0.15	.5596	.4404	.0596	0.40	.6554	.3446	.1554
0.16	.5636	.4364	.0636	0.41	.6591	.3409	.1591
0.17	.5675	.4325	.0675	0.42	.6628	.3372	.1628
0.18	.5714	.4286	.0714	0.43	.6664	.3336	.1664
0.19	.5753	.4247	.0753	0.44	.6700	.3300	.1700
0.20	.5793	.4207	.0793	0.45	.6736	.3264	.1736
0.21	.5832	.4168	.0832	0.46	.6772	.3228	.1772
0.22	.5871	.4129	.0871	0.47	.6808	.3192	.1808
0.23	.5910	.4090	.0910	0.48	.6844	.3156	.1844
0.24	.5948	.4052	.0948	0.49	.6879	.3121	.1879

Z-scores and Percentiles

- If you have a z-score of .25, what is your percentile rank?
- Z-score of 0.0, what is your percentile rank?
- Z-score of .49, what is your percentile rank?

TABLE B.1 The Unit Normal Table*

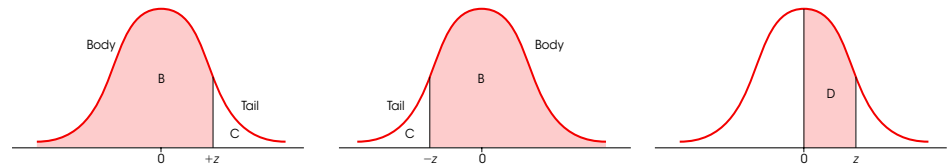
*Column A lists z-score values. A vertical line drawn through a normal distribution at a z-score location divides the distribution into two sections.

Column B identifies the proportion in the larger section, called the *body*.

Column C identifies the proportion in the smaller section, called the *tail*.

Column D identifies the proportion between the mean and the z-score.

Note: Because the normal distribution is symmetrical, the proportions for negative z-scores are the same as those for positive z-scores.



(A) z	(B) Proportion in Body	(C) Proportion in Tail	(D) Proportion Between Mean and z	(A) z	(B) Proportion in Body	(C) Proportion in Tail	(D) Proportion Between Mean and z
0.00	.5000	.5000	.0000	0.25	.5987	.4013	.0987
0.01	.5040	.4960	.0040	0.26	.6026	.3974	.1026
0.02	.5080	.4920	.0080	0.27	.6064	.3936	.1064
0.03	.5120	.4880	.0120	0.28	.6103	.3897	.1103
0.04	.5160	.4840	.0160	0.29	.6141	.3859	.1141
0.05	.5199	.4801	.0199	0.30	.6179	.3821	.1179
0.06	.5239	.4761	.0239	0.31	.6217	.3783	.1217
0.07	.5279	.4721	.0279	0.32	.6255	.3745	.1255
0.08	.5319	.4681	.0319	0.33	.6293	.3707	.1293
0.09	.5359	.4641	.0359	0.34	.6331	.3669	.1331
0.10	.5398	.4602	.0398	0.35	.6368	.3632	.1368
0.11	.5438	.4562	.0438	0.36	.6406	.3594	.1406
0.12	.5478	.4522	.0478	0.37	.6443	.3557	.1443
0.13	.5517	.4483	.0517	0.38	.6480	.3520	.1480
0.14	.5557	.4443	.0557	0.39	.6517	.3483	.1517
0.15	.5596	.4404	.0596	0.40	.6554	.3446	.1554
0.16	.5636	.4364	.0636	0.41	.6591	.3409	.1591
0.17	.5675	.4325	.0675	0.42	.6628	.3372	.1628
0.18	.5714	.4286	.0714	0.43	.6664	.3336	.1664
0.19	.5753	.4247	.0753	0.44	.6700	.3300	.1700
0.20	.5793	.4207	.0793	0.45	.6736	.3264	.1736
0.21	.5832	.4168	.0832	0.46	.6772	.3228	.1772
0.22	.5871	.4129	.0871	0.47	.6808	.3192	.1808
0.23	.5910	.4090	.0910	0.48	.6844	.3156	.1844
0.24	.5948	.4052	.0948	0.49	.6879	.3121	.1879

Z-scores and Percentiles

- Z-score of .25,
percentile = 60th
- Z-score of 0.0,
percentile = 50th
- Z-score of .49,
percentile = 69th
 - Equivalent to saying
69% scored below you

TABLE B.1 The Unit Normal Table*

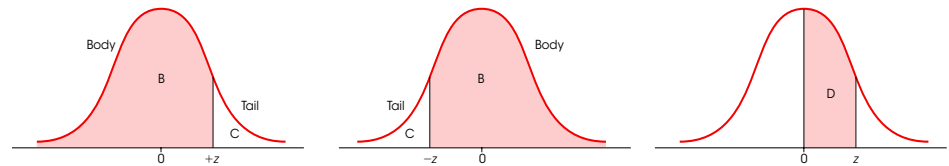
*Column A lists z -score values. A vertical line drawn through a normal distribution at a z -score location divides the distribution into two sections.

Column B identifies the proportion in the larger section, called the *body*.

Column C identifies the proportion in the smaller section, called the *tail*.

Column D identifies the proportion between the mean and the z -score.

Note: Because the normal distribution is symmetrical, the proportions for negative z -scores are the same as those for positive z -scores.



(A) z	(B) Proportion in Body	(C) Proportion in Tail	(D) Proportion Between Mean and z	(A) z	(B) Proportion in Body	(C) Proportion in Tail	(D) Proportion Between Mean and z
0.00	.5000	.5000	.0000	0.25	.5987	.4013	.0987
0.01	.5040	.4960	.0040	0.26	.6026	.3974	.1026
0.02	.5080	.4920	.0080	0.27	.6064	.3936	.1064
0.03	.5120	.4880	.0120	0.28	.6103	.3897	.1103
0.04	.5160	.4840	.0160	0.29	.6141	.3859	.1141
0.05	.5199	.4801	.0199	0.30	.6179	.3821	.1179
0.06	.5239	.4761	.0239	0.31	.6217	.3783	.1217
0.07	.5279	.4721	.0279	0.32	.6255	.3745	.1255
0.08	.5319	.4681	.0319	0.33	.6293	.3707	.1293
0.09	.5359	.4641	.0359	0.34	.6331	.3669	.1331
0.10	.5398	.4602	.0398	0.35	.6368	.3632	.1368
0.11	.5438	.4562	.0438	0.36	.6406	.3594	.1406
0.12	.5478	.4522	.0478	0.37	.6443	.3557	.1443
0.13	.5517	.4483	.0517	0.38	.6480	.3520	.1480
0.14	.5557	.4443	.0557	0.39	.6517	.3483	.1517
0.15	.5596	.4404	.0596	0.40	.6554	.3446	.1554
0.16	.5636	.4364	.0636	0.41	.6591	.3409	.1591
0.17	.5675	.4325	.0675	0.42	.6628	.3372	.1628
0.18	.5714	.4286	.0714	0.43	.6664	.3336	.1664
0.19	.5753	.4247	.0753	0.44	.6700	.3300	.1700
0.20	.5793	.4207	.0793	0.45	.6736	.3264	.1736
0.21	.5832	.4168	.0832	0.46	.6772	.3228	.1772
0.22	.5871	.4129	.0871	0.47	.6808	.3192	.1808
0.23	.5910	.4090	.0910	0.48	.6844	.3156	.1844
0.24	.5948	.4052	.0948	0.49	.6879	.3121	.1879

Z-scores and Percentiles

□ What percent scored ABOVE you if you had a z-score of .49?

▣ 31%

□ What percent scored ABOVE you if you had a z-score of .25?

▣ 40%

TABLE B.1 The Unit Normal Table*

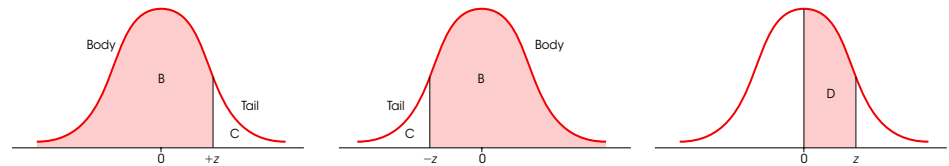
*Column A lists z-score values. A vertical line drawn through a normal distribution at a z-score location divides the distribution into two sections.

Column B identifies the proportion in the larger section, called the *body*.

Column C identifies the proportion in the smaller section, called the *tail*.

Column D identifies the proportion between the mean and the z-score.

Note: Because the normal distribution is symmetrical, the proportions for negative z-scores are the same as those for positive z-scores.



(A) z	(B) Proportion in Body	(C) Proportion in Tail	(D) Proportion Between Mean and z	(A) z	(B) Proportion in Body	(C) Proportion in Tail	(D) Proportion Between Mean and z
0.00	.5000	.5000	.0000	0.25	.5987	.4013	.0987
0.01	.5040	.4960	.0040	0.26	.6026	.3974	.1026
0.02	.5080	.4920	.0080	0.27	.6064	.3936	.1064
0.03	.5120	.4880	.0120	0.28	.6103	.3897	.1103
0.04	.5160	.4840	.0160	0.29	.6141	.3859	.1141
0.05	.5199	.4801	.0199	0.30	.6179	.3821	.1179
0.06	.5239	.4761	.0239	0.31	.6217	.3783	.1217
0.07	.5279	.4721	.0279	0.32	.6255	.3745	.1255
0.08	.5319	.4681	.0319	0.33	.6293	.3707	.1293
0.09	.5359	.4641	.0359	0.34	.6331	.3669	.1331
0.10	.5398	.4602	.0398	0.35	.6368	.3632	.1368
0.11	.5438	.4562	.0438	0.36	.6406	.3594	.1406
0.12	.5478	.4522	.0478	0.37	.6443	.3557	.1443
0.13	.5517	.4483	.0517	0.38	.6480	.3520	.1480
0.14	.5557	.4443	.0557	0.39	.6517	.3483	.1517
0.15	.5596	.4404	.0596	0.40	.6554	.3446	.1554
0.16	.5636	.4364	.0636	0.41	.6591	.3409	.1591
0.17	.5675	.4325	.0675	0.42	.6628	.3372	.1628
0.18	.5714	.4286	.0714	0.43	.6664	.3336	.1664
0.19	.5753	.4247	.0753	0.44	.6700	.3300	.1700
0.20	.5793	.4207	.0793	0.45	.6736	.3264	.1736
0.21	.5832	.4168	.0832	0.46	.6772	.3228	.1772
0.22	.5871	.4129	.0871	0.47	.6808	.3192	.1808
0.23	.5910	.4090	.0910	0.48	.6844	.3156	.1844
0.24	.5948	.4052	.0948	0.49	.6879	.3121	.1879

Z-scores and Probability

- If I randomly select someone, what is the probability they scored ABOVE you if you had a z-score of .49?

□ $P(z > .49)$

■ .31

- What is the probability they scored ABOVE you if you had a z-score of .25?

□ $P(z > .25)$

■ .40

Notation: $P(x)$, pronounced P of x where x is some condition of interest.

TABLE B.1 The Unit Normal Table*

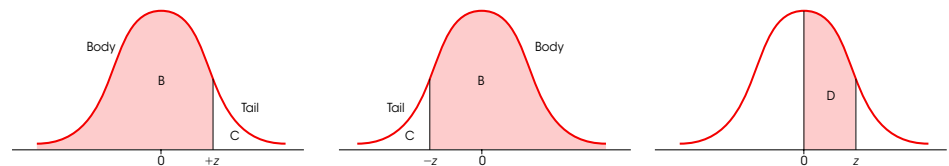
*Column A lists z-score values. A vertical line drawn through a normal distribution at a z-score location divides the distribution into two sections.

Column B identifies the proportion in the larger section, called the *body*.

Column C identifies the proportion in the smaller section, called the *tail*.

Column D identifies the proportion between the mean and the z-score.

Note: Because the normal distribution is symmetrical, the proportions for negative z-scores are the same as those for positive z-scores.



(A) z	(B) Proportion in Body	(C) Proportion in Tail	(D) Proportion Between Mean and z	(A) z	(B) Proportion in Body	(C) Proportion in Tail	(D) Proportion Between Mean and z
0.00	.5000	.5000	.0000	0.25	.5987	.4013	.0987
0.01	.5040	.4960	.0040	0.26	.6026	.3974	.1026
0.02	.5080	.4920	.0080	0.27	.6064	.3936	.1064
0.03	.5120	.4880	.0120	0.28	.6103	.3897	.1103
0.04	.5160	.4840	.0160	0.29	.6141	.3859	.1141
0.05	.5199	.4801	.0199	0.30	.6179	.3821	.1179
0.06	.5239	.4761	.0239	0.31	.6217	.3783	.1217
0.07	.5279	.4721	.0279	0.32	.6255	.3745	.1255
0.08	.5319	.4681	.0319	0.33	.6293	.3707	.1293
0.09	.5359	.4641	.0359	0.34	.6331	.3669	.1331
0.10	.5398	.4602	.0398	0.35	.6368	.3632	.1368
0.11	.5438	.4562	.0438	0.36	.6406	.3594	.1406
0.12	.5478	.4522	.0478	0.37	.6443	.3557	.1443
0.13	.5517	.4483	.0517	0.38	.6480	.3520	.1480
0.14	.5557	.4443	.0557	0.39	.6517	.3483	.1517
0.15	.5596	.4404	.0596	0.40	.6554	.3446	.1554
0.16	.5636	.4364	.0636	0.41	.6591	.3409	.1591
0.17	.5675	.4325	.0675	0.42	.6628	.3372	.1628
0.18	.5714	.4286	.0714	0.43	.6664	.3336	.1664
0.19	.5753	.4247	.0753	0.44	.6700	.3300	.1700
0.20	.5793	.4207	.0793	0.45	.6736	.3264	.1736
0.21	.5832	.4168	.0832	0.46	.6772	.3228	.1772
0.22	.5871	.4129	.0871	0.47	.6808	.3192	.1808
0.23	.5910	.4090	.0910	0.48	.6844	.3156	.1844
0.24	.5948	.4052	.0948	0.49	.6879	.3121	.1879

Two Students in Different Classes

There are two different instructors teaching Statistical Literacy. A student in Z's class had a score of 45 on the first exam and a student in D's class (assigned a different exam) had a score of 50.

Who did better?

What information do we need?

Two Students in Different Classes

There are two different instructors teaching Statistical Literacy. A student in Z's class had a score of 45 on the first exam and a student in D's class (assigned a different exam) had a score of 50.

$$z - score = \frac{x - \mu}{\sigma}$$

Z's Exam:

$$\mu = 40$$

$$\sigma = 5$$

$$x = 45$$

D's Exam:

$$\mu = 55$$

$$\sigma = 5$$

$$x = 50$$

Two Students in Different Classes

There are two different instructors teaching Statistical Literacy. A student in Z's class had a score of 45 on the first exam and a student in D's class (assigned a different exam) had a score of 50.

By transforming the data into a z-score, these scores are now comparable!

Z's Exam:

$$\mu = 40$$

$$\sigma = 5$$

$$x = 45$$

$$z = +1.0$$

D's Exam:

$$\mu = 55$$

$$\sigma = 5$$

$$x = 50$$

$$z = -1.0$$

$$z - score = \frac{x - \mu}{\sigma}$$

Two Students in Different Classes

There are two different instructors teaching Statistical Literacy. A student in Z's class had a score of 45 on the first exam and a student in D's class (assigned a different exam) had a score of 50.

By transforming the data into a z-score, these scores are now comparable!

Z's Exam:

$$\mu = 40$$

$$\sigma = 5$$

$$x = 45$$

$$z = +1.0$$

D's Exam:

$$\mu = 55$$

$$\sigma = 5$$

$$x = 50$$

$$z = -1.0$$

Two Students in Different Classes

There are two different instructors teaching Statistical Literacy. A student in Z's class had a score of 45 on the first exam and a student in D's class (assigned a different exam) had a score of 50. The Student in Z's class has a score that is one standard deviation above the mean, while the Student in D's class is one standard deviation below the mean.

Z's Exam:

$$\mu = 40$$

$$\sigma = 5$$

$$x = 45$$

$$z = +1.0$$

>

D's Exam:

$$\mu = 55$$

$$\sigma = 5$$

$$x = 50$$

$$z = -1.0$$

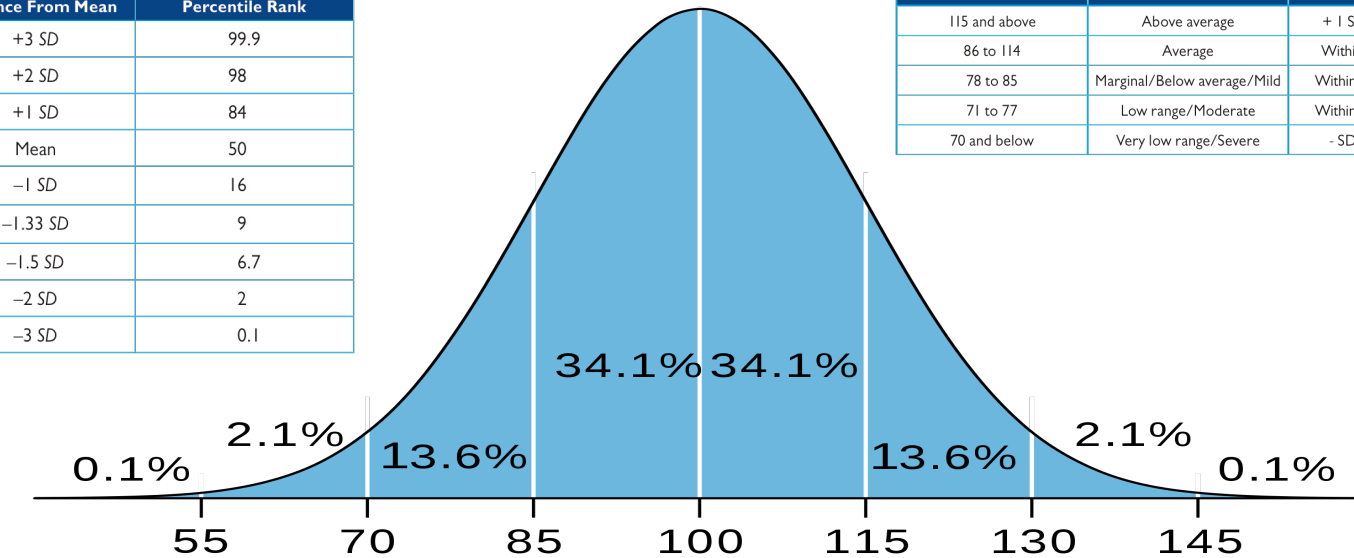
Real World Example Severity (CELF-5)

Table 4.3 Distances From the Mean of Selected Standard Scores

Standard Score	Distance From Mean	Percentile Rank
145	+3 SD	99.9
130	+2 SD	98
115	+1 SD	84
100	Mean	50
85	-1 SD	16
80	-1.33 SD	9
77.5	-1.5 SD	6.7
70	-2 SD	2
55	-3 SD	0.1

Guidelines for Describing the Severity of a Language Disorder Using Standard Scores

Core Language Score and Index Scores	Classification	Relationship to Mean	Percentile Ranks
115 and above	Above average	+ 1 SD and above	84 and above
86 to 114	Average	Within + or -1 SD	18 to 83
78 to 85	Marginal/Below average/Mild	Within -1 to -1.5 SD	7 to 17
71 to 77	Low range/Moderate	Within -1.5 to -2 SD	3 to 6
70 and below	Very low range/Severe	- SD and below	Less than 0.1 to 2



Very Low/Severe (≤ 70)

Low/Moderate (71-77)

Below Average/Mild (78-85)

Average (86-114)

Above Average (≥ 115)

Why Standardize?

- “Translating” raw data into a standardized scales allows us to interpret any one person’s score more deeply and relatively to other people.
 - ▣ Ex. A z-score of +2.0 would tell you a person did very well, and you don’t have to know anything about the raw score.
 - ▣ Ex. A z-score of -2.0 on a language test would tell you this person may need some speech therapy.
- The conversion of data to have a mean of zero with a standard deviation of one makes computations and interpretations easier. This type of “translating of data” is at the core of much of what we will be doing for the rest of the course.

Why are we doing this, again?

- We can take any scale and transform it into **simplified**, standard scale... the Z-scale.
- The Z-scale and the Z-scores associated with it allow us to make **comparison** across different scales, test, and situations.
- We can also use standardized Z-scores to help us figure out the **probability** of things. A core aspect of statistics.

Z

Up Next...

- We move beyond one person's score and start looking at what a standardized mean value you can tell us...

Probability of z-scores vs z-tests

Z-scores in R

Two Students in Different Classes

Z's Exam:

$$\mu = 40$$

$$\sigma = 5$$

$$x = 45$$

D's Exam:

$$\mu = 55$$

$$\sigma = 5$$

$$x = 50$$

```
#####  
##### z-scores #####  
#####  
  
#### Z's Exam z-score ####  
person_Z <- 45  
pop_mean_Z <- 40  
sd_Z <- 5  
  
z_score_Z <- (person_Z - pop_mean_Z)/sd_Z  
  
#### D's Exam z-score ####  
person_D <- 50  
pop_mean_D <- 55  
sd_D <- 5  
  
z_score_D <- (person_D - pop_mean_D)/sd_D  
  
z_score_Z  
z_score_D
```

SAT Z-Scores

$$\mu = 1000$$

$$\sigma = 210$$

$$x = 1555$$

The `pnorm()` functions will return the proportion (probability) rather than having to look up in a z-table.

```
#####  
### z-scores & Probability ###  
#####  
  
# Calculate a z-score  
  
SAT_mean <- 1000  
SAT_sd <- 210  
SAT_score <- 1555  
  
SAT_Z_score <- (SAT_score - SAT_mean)/SAT_sd  
  
# The function "pnorm()" can give you the proportion (i.e the probability,  
# the "p" in "pnorm") from a normal distribution (the "norm" in "pnorm").  
# You need to tell R the z-score (q), the mean, the sd, and the final  
# argument you need to give it is where or not to include the tail.  
#  
# You can tell lower.tail = FALSE, which is the equivalent of asking R the proportion of  
# people that score ABOVE that z-score, P (z < 2.64)  
#  
# If you have a 1555 on the SAT, less than 0.04% of people scored higher than you  
pnorm(q = 1555, mean=1000, sd=210, lower.tail=FALSE)  
pnorm(SAT_Z_score, mean = 0, sd = 1, lower.tail = FALSE)  
  
# You can tell lower.tail = TRUE, which is the equivalent of asking R the proportion of  
# people that score BELOW that z-score, this is like a percentile, P (z > 2.64)  
#  
# If you have a 1555 on the SAT, you are in the 99.6th percentile,  
# You scored better than 99.6% of people  
pnorm(q = 1555, mean=1000, sd=210, lower.tail=TRUE) #raw values  
pnorm(SAT_Z_score, mean = 0, sd = 1, lower.tail = TRUE) #standardized values  
  
# If you want to know how many people fall between two z-scores,  
# (symmetrical z-scores, meaning -1.96 to +1.96), P (-1.96 < z < 1.96)  
1-pnorm(q = 1.96, mean=0, sd=1, lower.tail=FALSE)*2
```