# “Simple Learning Pro” YouTube channel

Playlist for ***statistics***

There are many concepts (e.g., z-scores – videos 19 and 20) that we’ll learn this term that have videos you can watch.

https://www.youtube.com/playlist?list=PL0KQuRyPJoe6KjlUM6iNYgt8d0DwI-IGR

I also put this link on Moodle. Check it out and then remember to document your work on this week’s P3R survey!

# Tomorrow (Wednesday 2/21) – Harbert 301

- We'll be in the lab to work on Practice Lab #2, using JAMOVI
- Try to arrive 5-10 minutes early and when you arrive, find a computer that is turned on and connected to the internet
  - Please try to sit near the middle aisle and/or close to the front of room; this allows me to get to you more easily, in the event that you need help!
- Log in to Moodle, and open the Chapter 3 & 4 lecture slides as a reference
- Consider where you are going to save your practice lab work so that you'll have access to it later
  - E.g., USB, or desktop and then upload to *OneDrive*
- I'd recommend also having a pen/pencil and your notebook for taking notes on how to use JAMOVI

This YouTube channel has a lot of helpful tutorials about JAMOVI.
https://www.youtube.com/playlist?list=PLkk92zzyru5OAtc_ItUubaSSq6S_TGfRn

# Practice with measures of central tendency and measures of dispersion (spread, variability)

## DESCRIPTIVE STATISTICS TABLE

| VARIABLE | A | B | C |
|---|---|---|---|
| N | 150 | 150 | 150 |
| Mean | 6.28 | 13.94 | 3.81 |
| Median | 6.0 | 6.5 | 6.0 |
| Standard deviation | 2.75 | 6.51 | 4.50 |

For each Q below, select variable ***A, B, or C,*** whichever best answers the question.

1. Which variable above has scores that are *clustered most closely around the sample mean*?
2. Which variable above has scores that are *spread farthest from the sample mean*?
3. Which variable above has the most *symmetrical* distribution?
4. Which variable above is most likely to have a distribution with a *positive skew*?
5. Which variable above is most likely to have a distribution with a *negative skew*?

1. Which variable has scores that are clustered closely around the sample mean?
   - A (answer this question by comparing the standard deviations. Ignore everything else!)
2. Which variable has scores that are spread farthest from the sample mean?
   - B (answer this question by comparing the standard deviations. Ignore everything else!)
3. Which variable has the most symmetrical distribution?
   - A (answer this question by looking at the means and medians. Ignore everything else!)
4. Which variable is most likely to have a distribution with a positive skew?
   - B (answer this question by looking at the means and medians. Ignore everything else!)
5. Which variable is most likely to have a distribution with a negative skew?
   - C (answer this question by looking at the means and medians. Ignore everything else!)

| VARIABLE | A | B | C |
|---|---|---|---|
| N | 150 | 150 | 150 |
| Mean | 6.28 | 13.94 | 3.81 |
| Median | 6.0 | 6.5 | 6.0 |
| Standard deviation | 2.75 | 6.51 | 4.5 |

# CH. 4 – the Standard Normal Distribution and z-scores

PY 221 **Research Methods & Statistics I**

Dr. Valenti

# Outline for Ch. 4

1. Standard Normal Distribution
2. Defining and calculating z-scores (aka *standardized scores*)
3. z-scores and the area under the curve

# Histogram of Heights of Grad Students

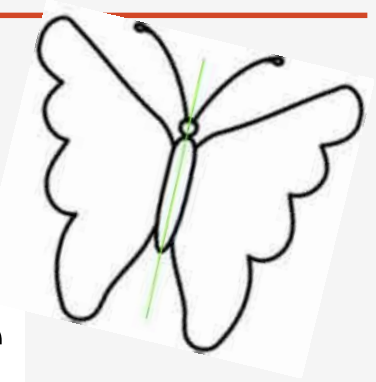
Frequency
(# Ps with each score)

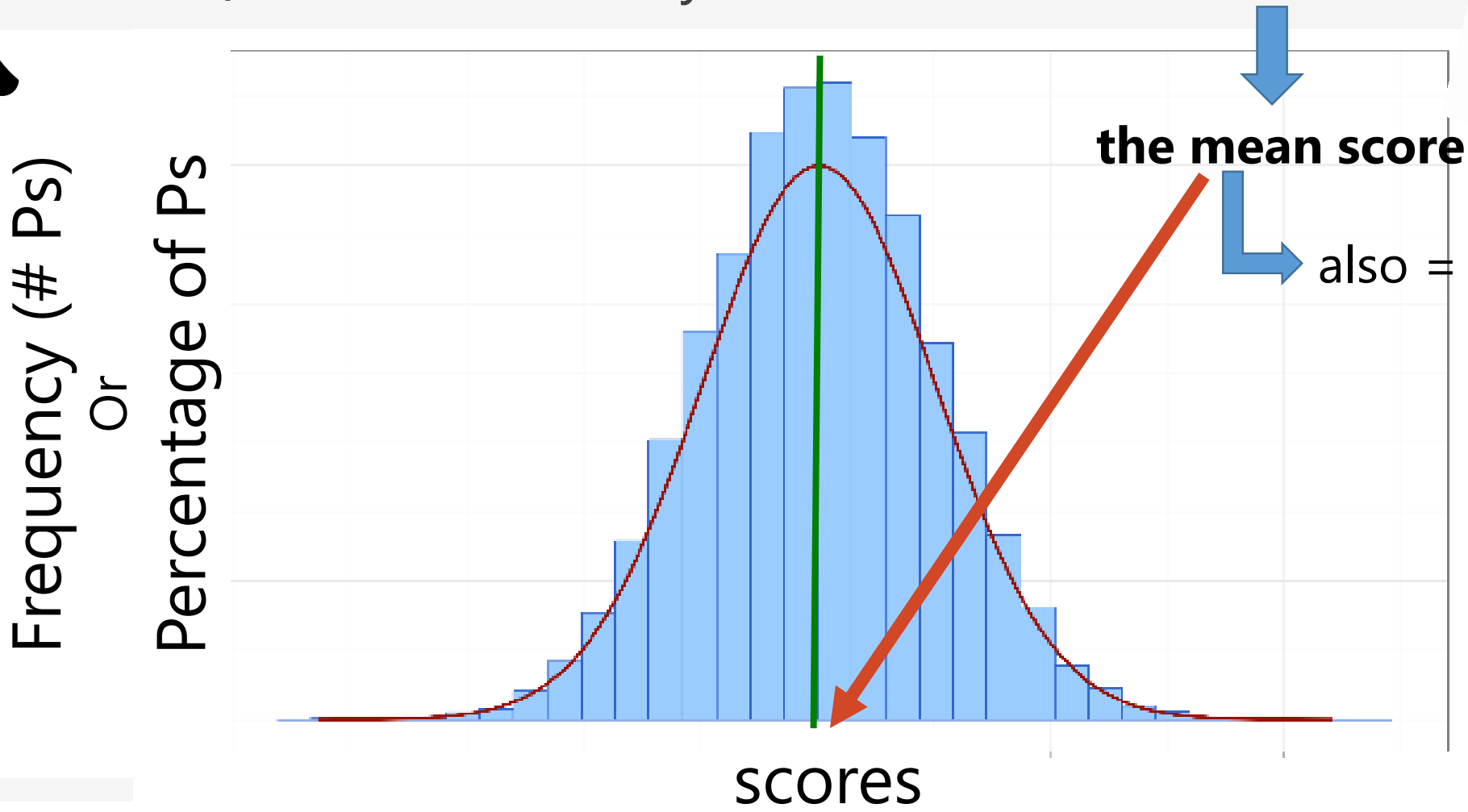

Score
(which is *height, in feet & inches*)

# Standard Normal Distribution

**What characteristics do data with this distribution have?**

- Bell shaped-*ish*
- Symmetrical around center

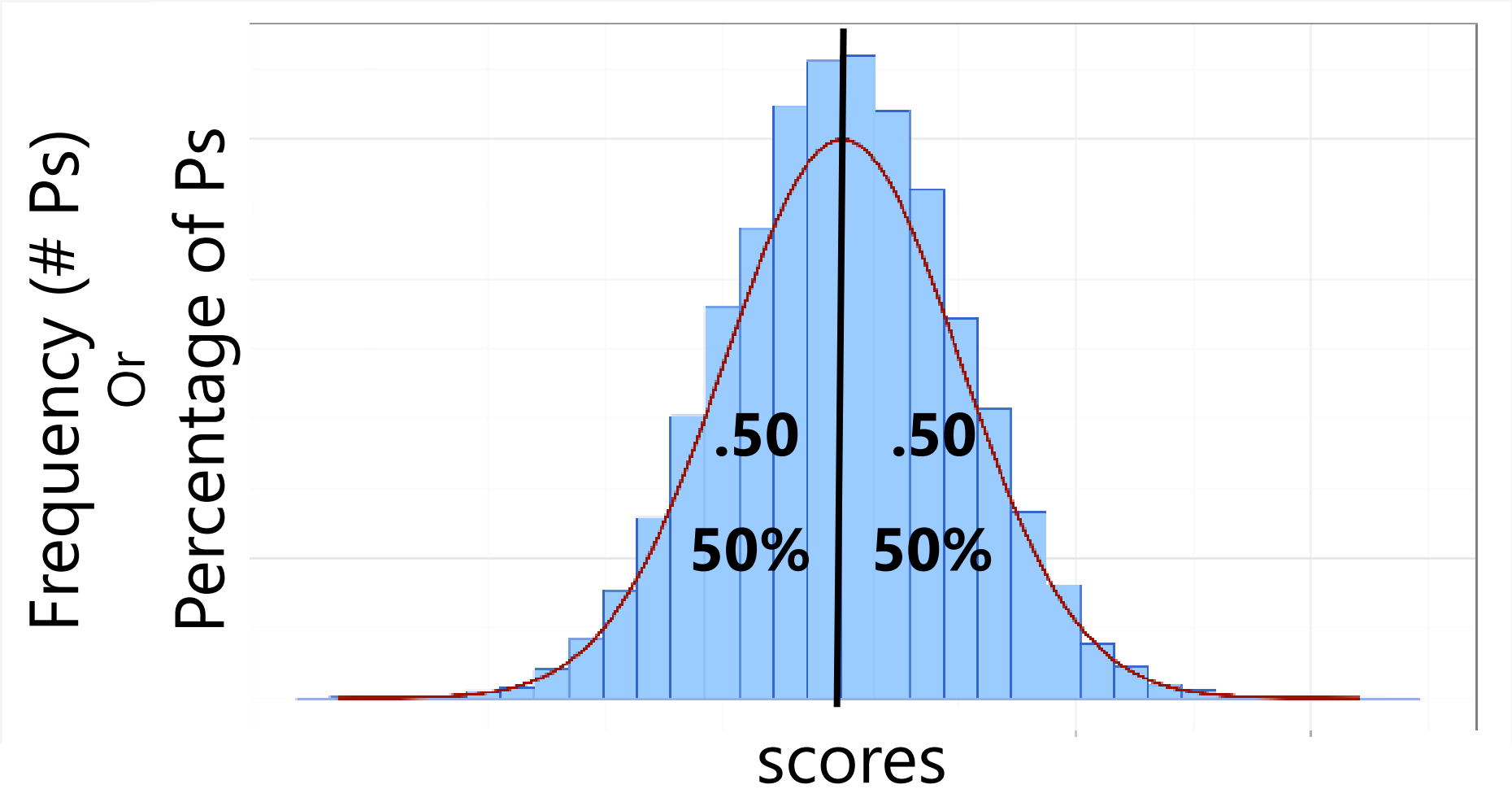




# Standard Normal Distribution

**What characteristics do data with this distribution have?**

- "area under the curve" – e.g., if the area = .50 (50%), it means 50% of Ps fall in that area under the curve.

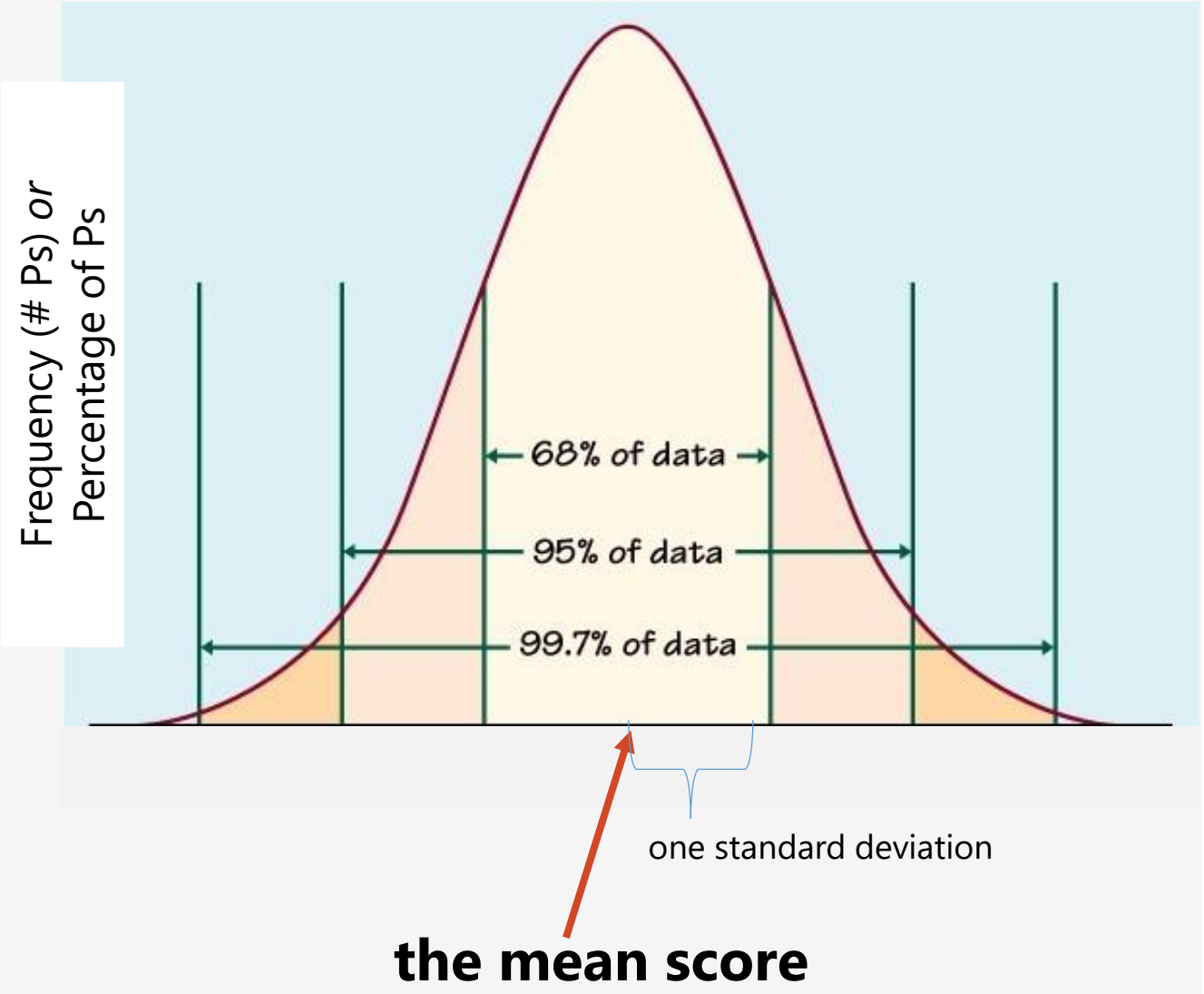


# Standard Normal Distribution

**What characteristics do data with this distribution have?**

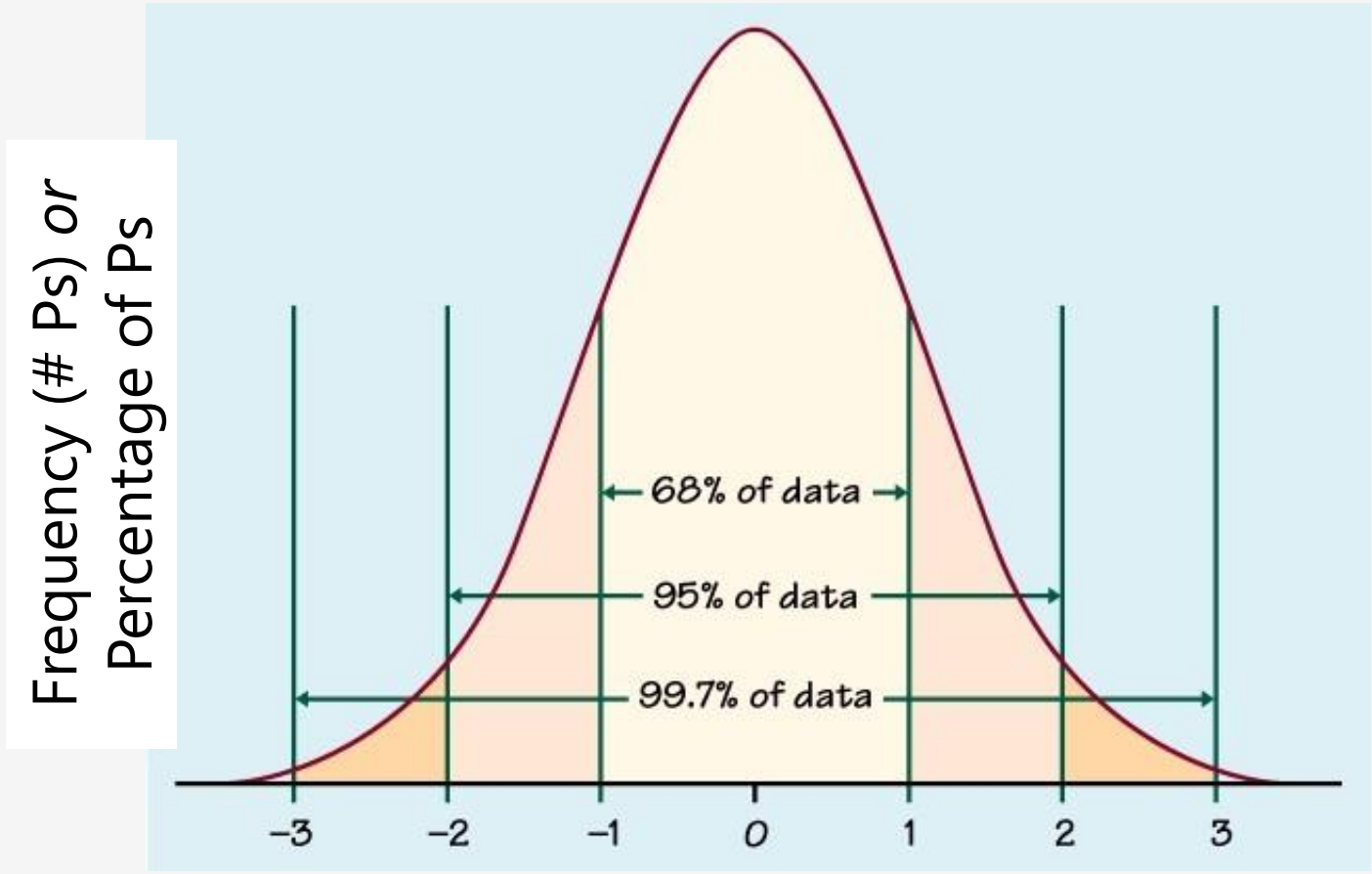
With standard normal distributions, it is always true that…

- **68%** of the area under the curve is within **1 standard deviation** of the mean
- (68% of people's scores are within 1 SD of the mean)
- **95%** of the area is (95% of people's scores are) within **2 standard deviations** of the mean
- **99.7%** of the area is (99.7% of people's scores are) within **3 standard deviations** of the mean



# Standard Normal Distribution

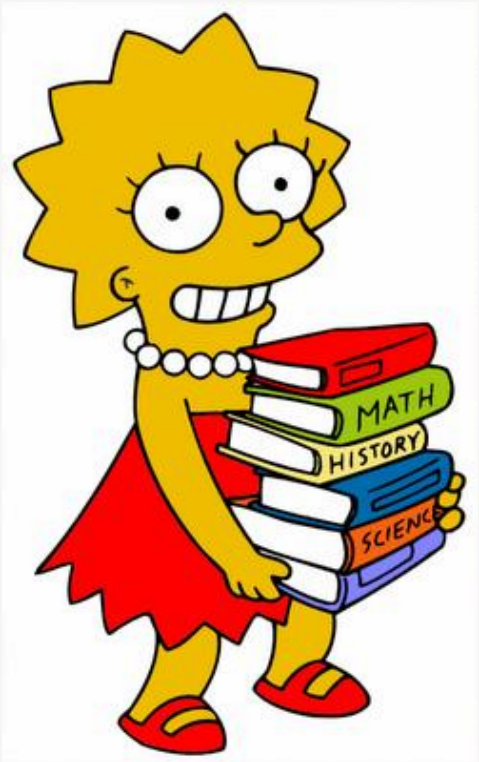
**What characteristics do data with this distribution have?**



# Outline for Ch. 4

1. Standard Normal Distribution
2. **Defining and calculating *z*-scores (aka *standardized scores*)**
3. z-scores and the area under the curve

# Suppose Lisa receives scores of 76% on both her PY 101 exam and her BI 110 exam . . .

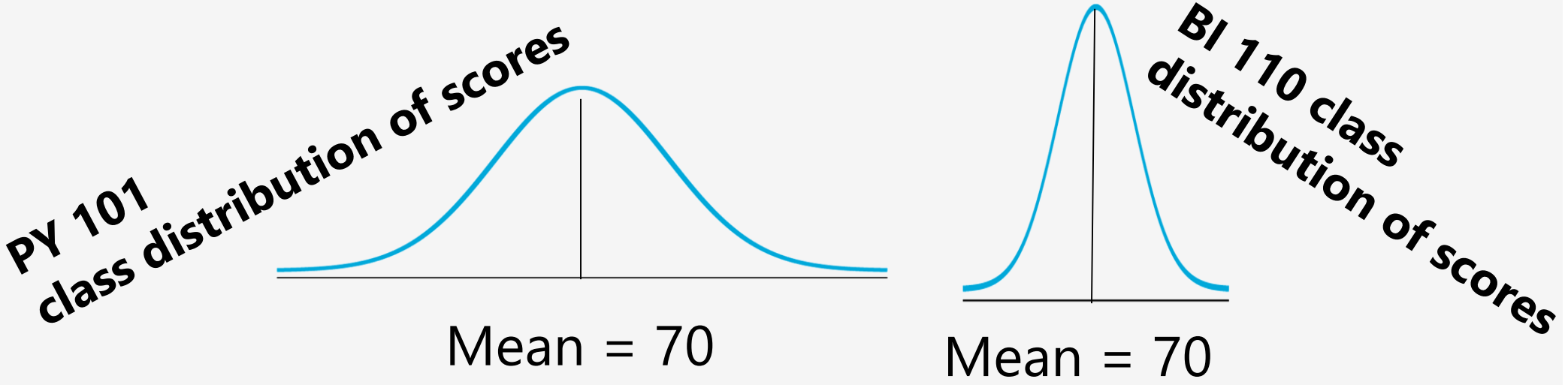



- how is Lisa performing in each class?
- is she equally talented at psychology and biology?

- On its own, Lisa's "*raw score*" of 76% doesn't provide all the information we need.
- **What additional information do we need?** (The pictures below are clues!)

# What additional information do we need?

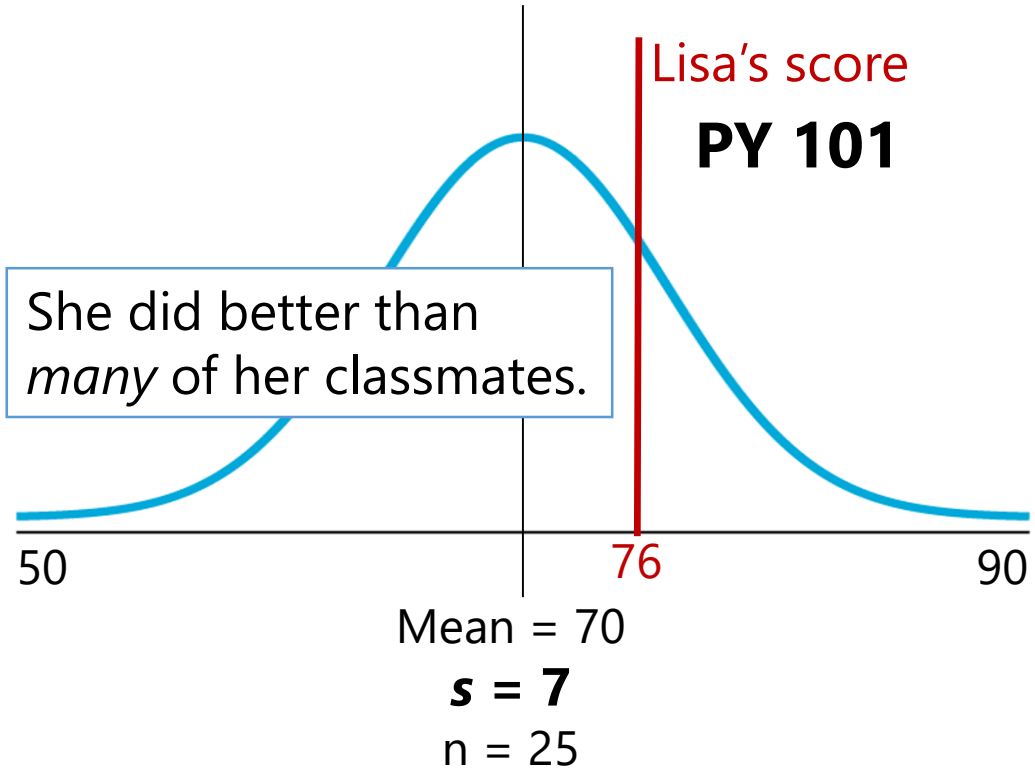
- Mean for each class → tells you whether she did better or worse than the average grade in that class
  - But what if mean for each class is the same? (70%) Is she performing equally well in PY & BI?
- Standard deviation for each class →
  - tells you whether the class' scores are spread out from mean or close to mean . . .

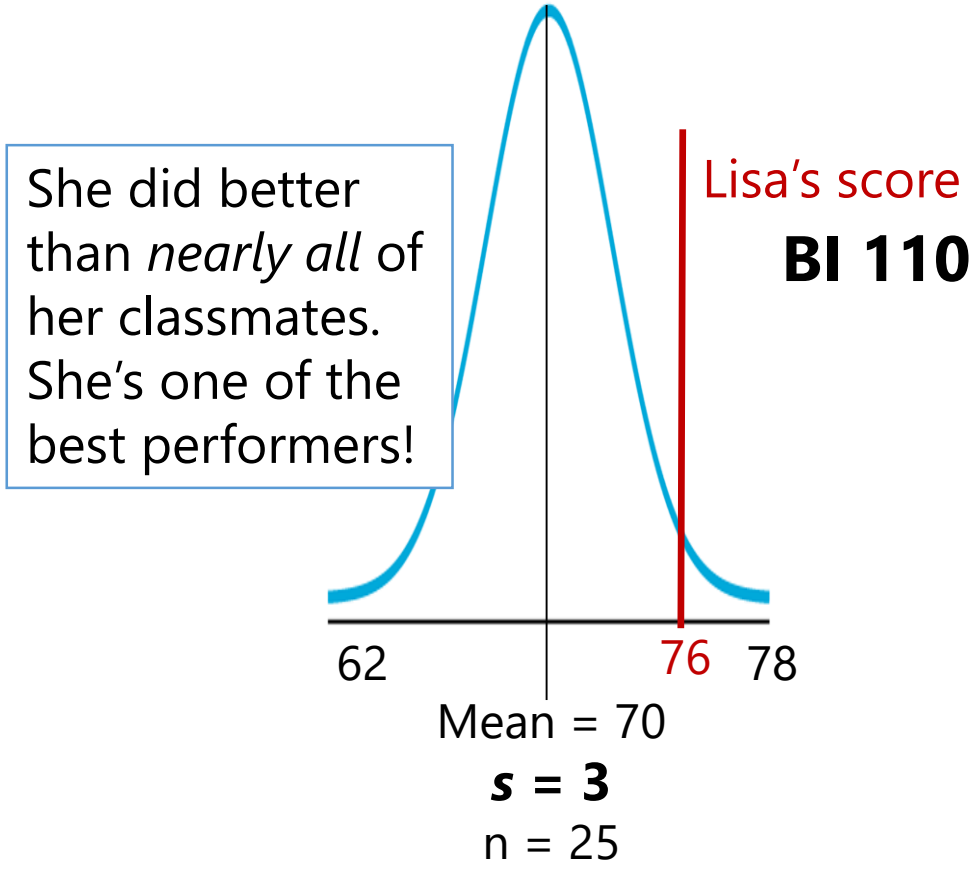



# It'd be helpful to know, for each class, the *location* of Lisa's score *within the distribution of the entire class' scores*

76 is Lisa's "raw score"

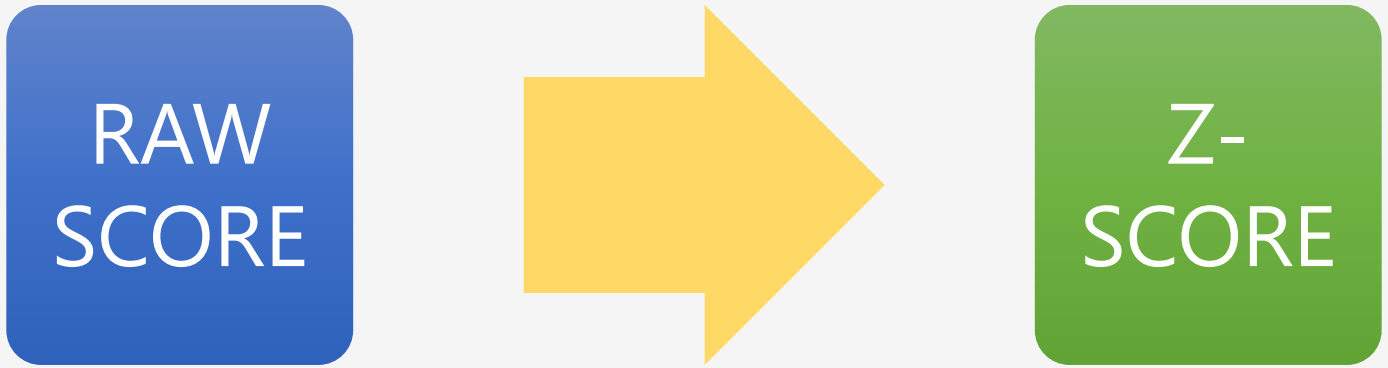
In which class is Lisa one of the best performers?
In which class is Lisa simply a good performer?



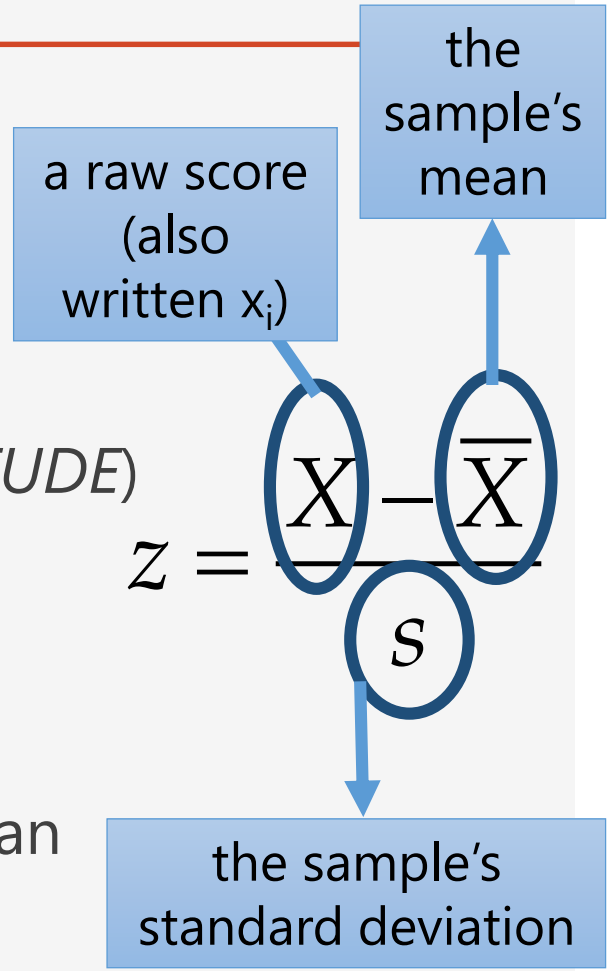
# What is a *z*-score?

- Each raw score can be converted into a *z*-score, as long as you know the mean (x-bar) and standard deviation (*s*) for the sample.
- Each *z*-score tells the *exact location* of the raw score *within the distribution*.
- In particular, on its own, a *z*-score tells us two things …
  - whether the raw score is *above* or *below* the mean (look at z-score's *SIGN, +/-* )
  - how many *standard deviations from the mean* the raw score is (look at *MAGNITUDE*)



# What is a *z*-score?

- Raw scores are converted into *z*-scores using a formula
  - *z*-scores can be positive or negative (*SIGN*)
    - when the raw score is *above (greater than)* the mean, *z* is *positive*
    - when the raw score is *below (less than)* the mean, *z* is *negative*
  - *z*-scores can be larger or smaller in *absolute value* (*MAGNITUDE*)
    - when the raw score is *far from* mean/center (i.e., raw score closer to tails), *z* will have *larger* absolute value
    - when the raw score is *close to* mean/center (i.e., raw score farther from tails), *z* will have *smaller* absolute value
  - a *z*-score of exactly 0 indicates the raw score is *equal to* mean

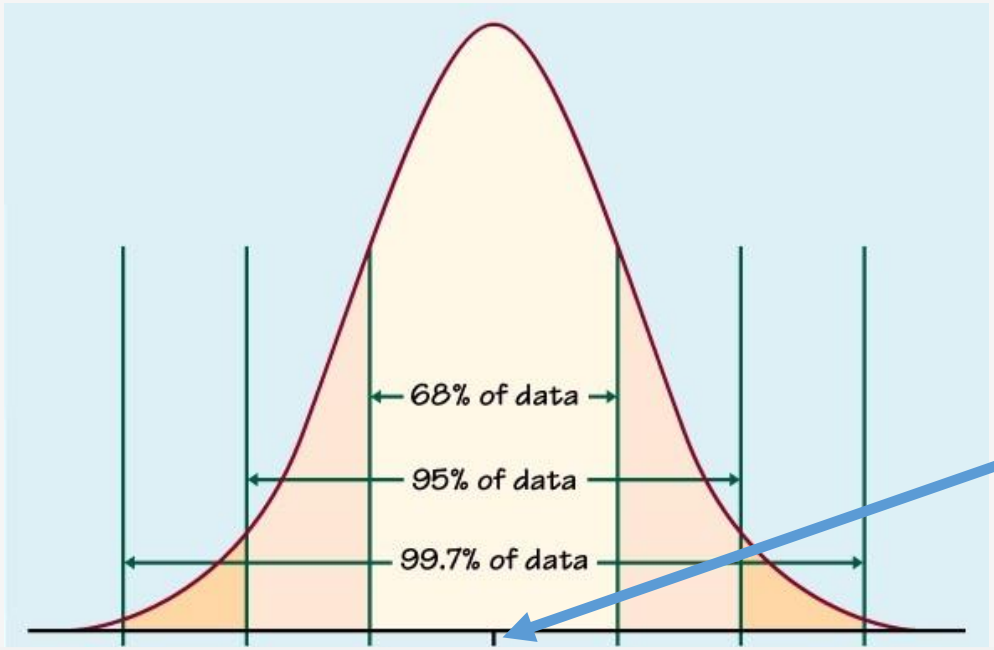



RAW SCORE → Z-SCORE

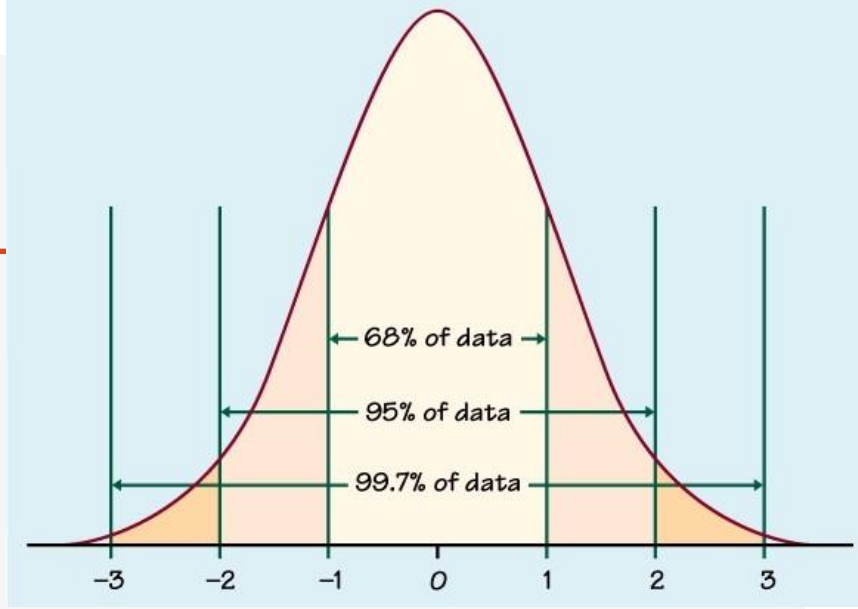
# What is a *z*-score?

You can convert raw scores into *z*-scores with a formula

$$z = \frac{X - \overline{X}}{s}$$



# What is a *z*-score? $z = \frac{X - \overline{X}}{s}$




You can convert raw scores into *z*-scores with a formula

- *z*-scores can be positive or negative (*SIGN*)
  - raw score is *above* the mean → *z* is *positive*
  - raw score is *below* the mean → *z* is *negative*
- *z*-scores can be larger or smaller in absolute value (*MAGNITUDE*)
  - when the raw score is *far from* mean/center (i.e., raw score closer to tails), *z* will have *larger* absolute value (closer to 3)
  - when the raw score is *close to* mean/center (i.e., raw score farther from tails), *z* will have *smaller* absolute value (closer to 0)
- a *z*-score of exactly zero indicates the raw score is *equal to* mean

# Let's convert Lisa's *raw score* into a *z*-score for each class

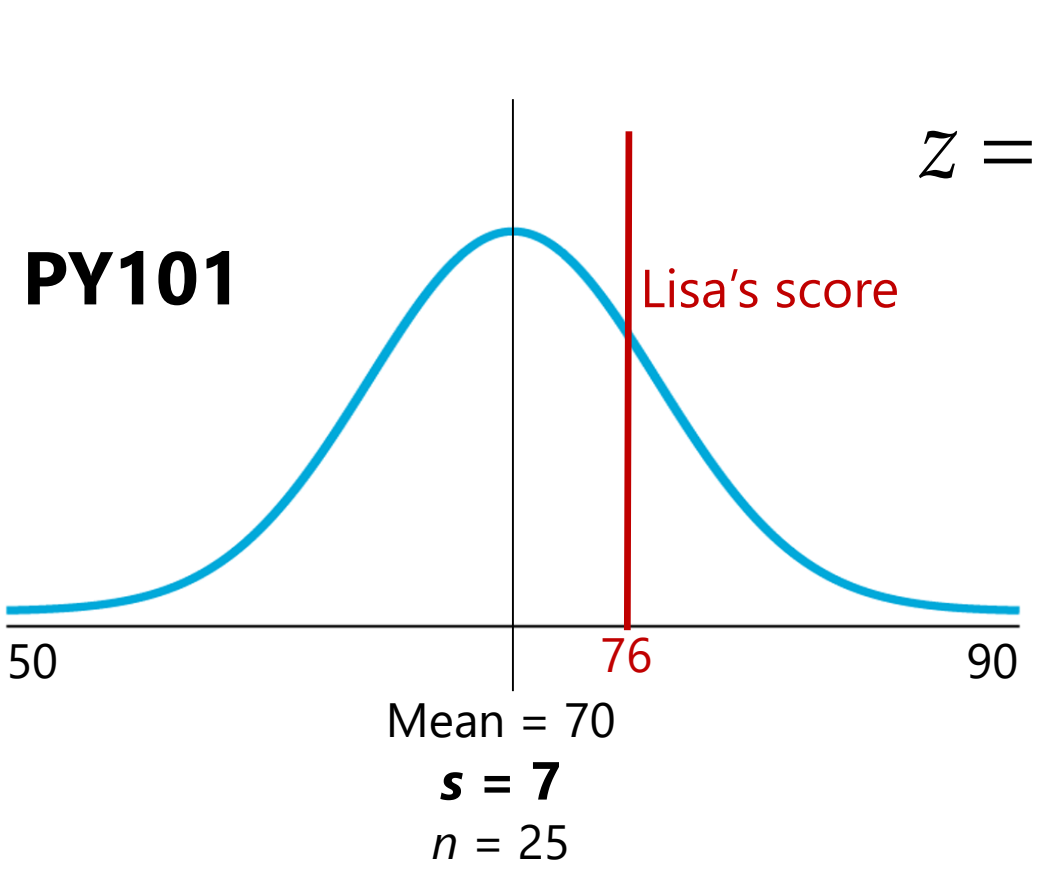
***Lisa's z-score for PY 101***

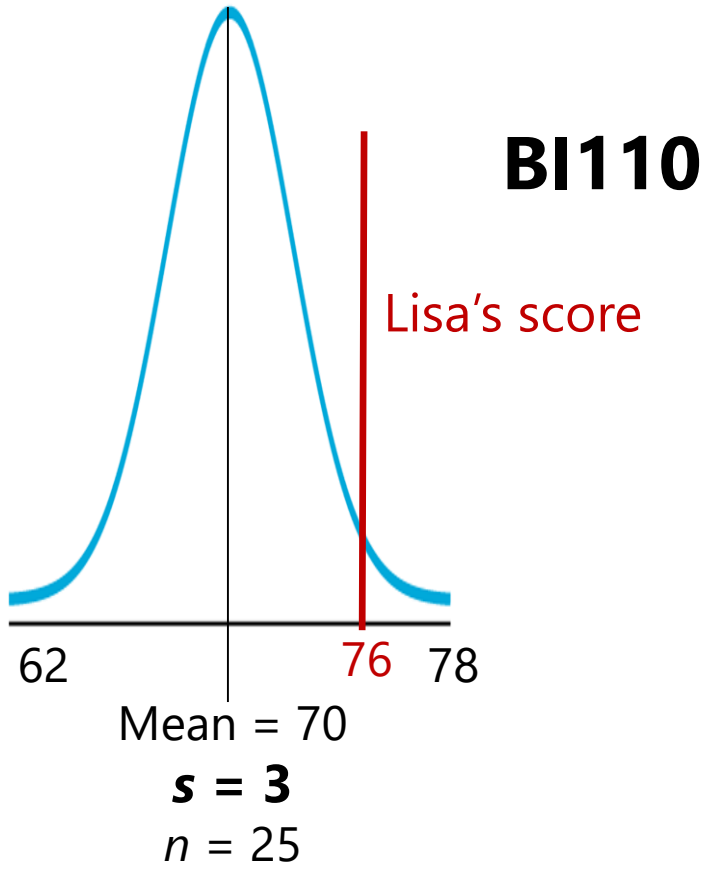
***z*** = (76-70) / 7 = 6/7 = **+0.86**

***Lisa's z-score for BI 110***

***z*** = (76-70) / 3 = 6/3 = **+2.00**

$$z = \frac{X - \overline{X}}{s}$$



Which class is Lisa doing better in, relative to the class' performance?

BI 110

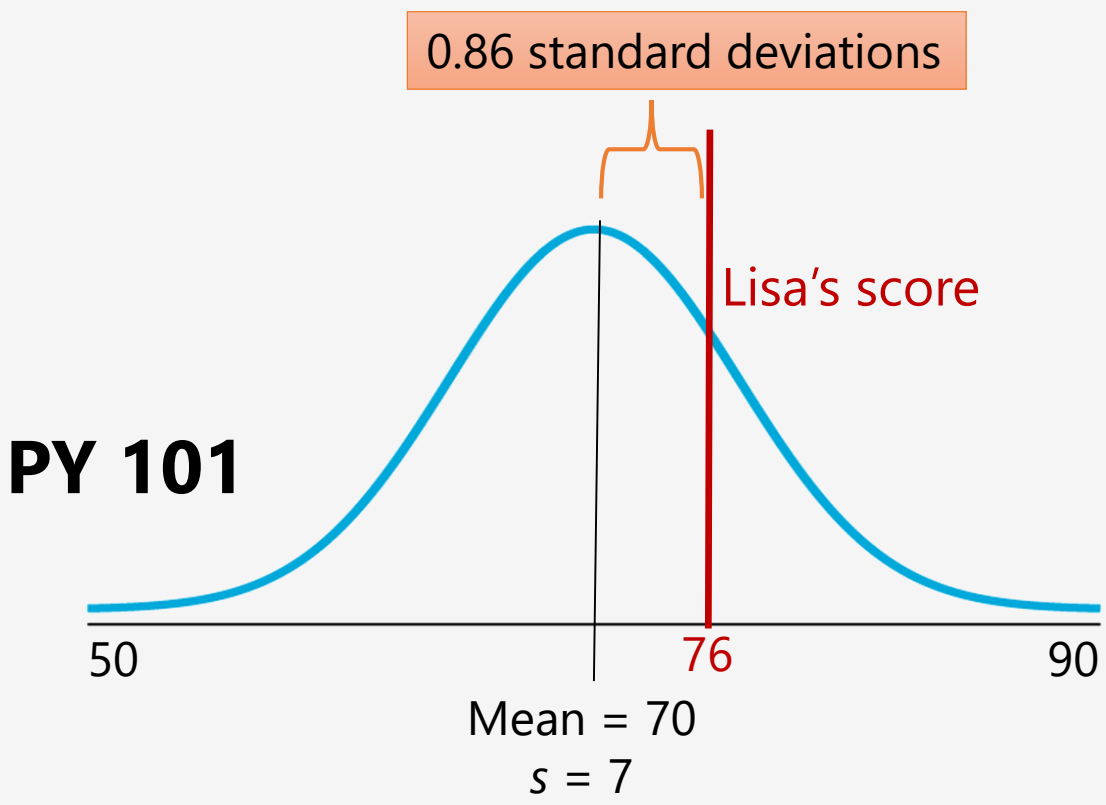
# How do you interpret *z*-scores? What do they mean, in words?

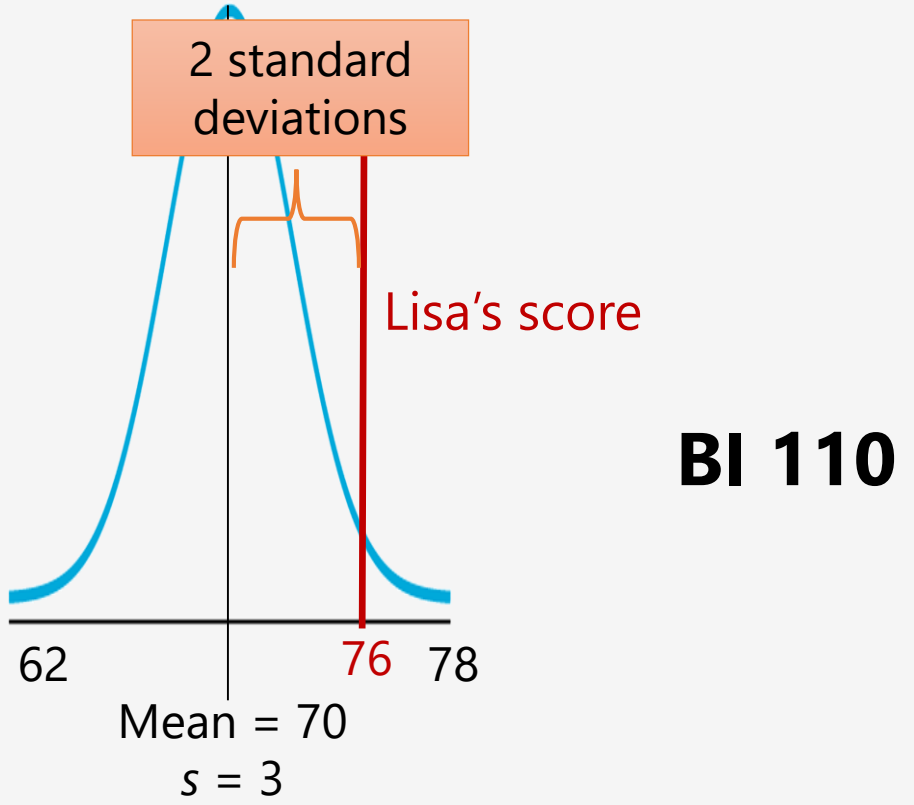
*From prior slide…for PY 101,* ***z* = +0.86**

- Lisa's raw score of 76 is **0.86 standard deviations above** the class mean in PY 101.

*From prior slide…for BI 110,* ***z* = +2.00**

- Lisa's raw score of 76 is **2.00 standard deviations above** the class mean in BI 110.



## Bart earned a 65% on both exams. What are his z-scores? *Calculate for each class distribution and try to write an interpretation of each z-score.*

*Reminder, for PY 101*

Mean = 70

Standard deviation = 7

- ***z* = -0.71**
- Bart's score of 65 is **0.71 standard deviations below** the mean PY 101 score.

*Reminder, for BI 110*

Mean = 70

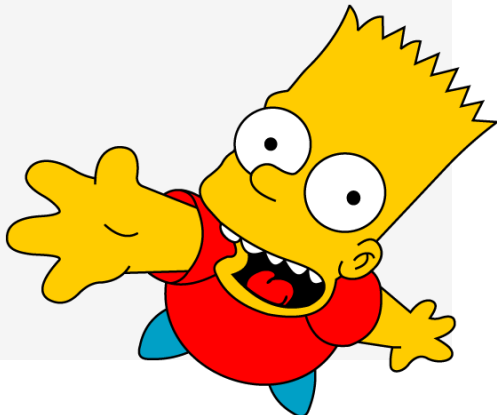
Standard deviation = 3

$$z = \frac{X - \overline{X}}{s}$$

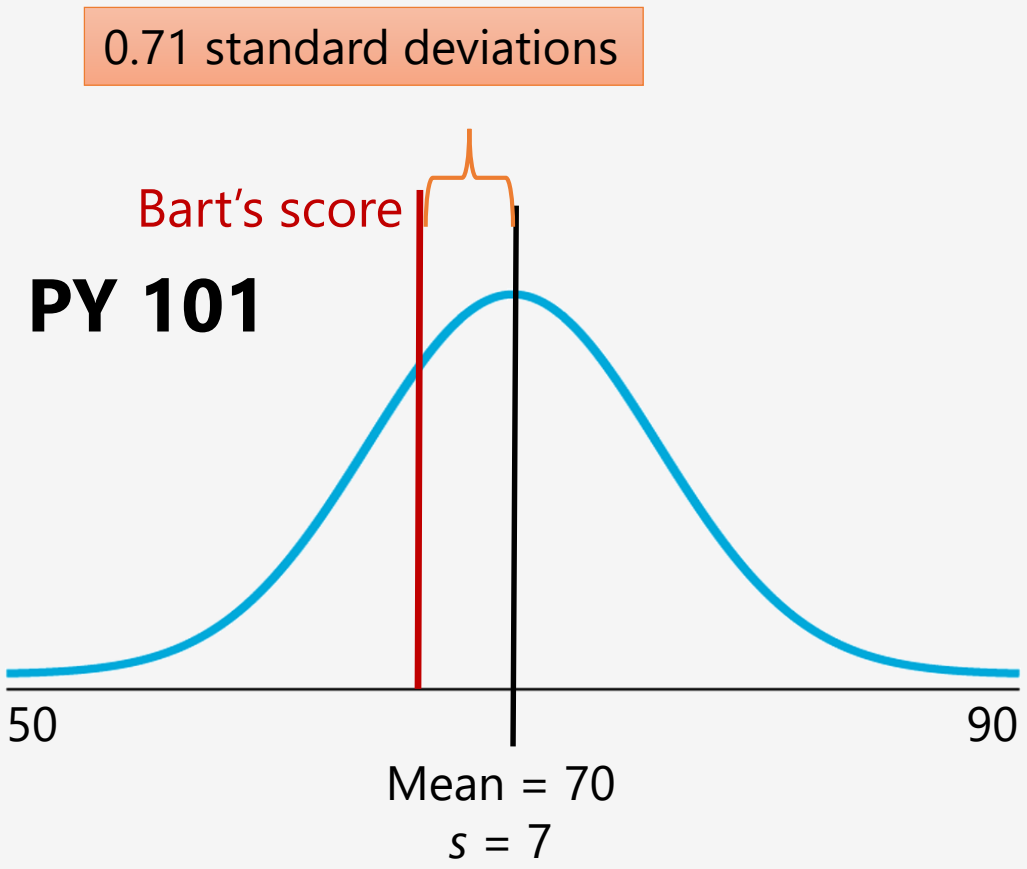
- ***z* = -1.67**
- Bart's score of 65 is **1.67 standard deviations below** the mean BI 110 score.

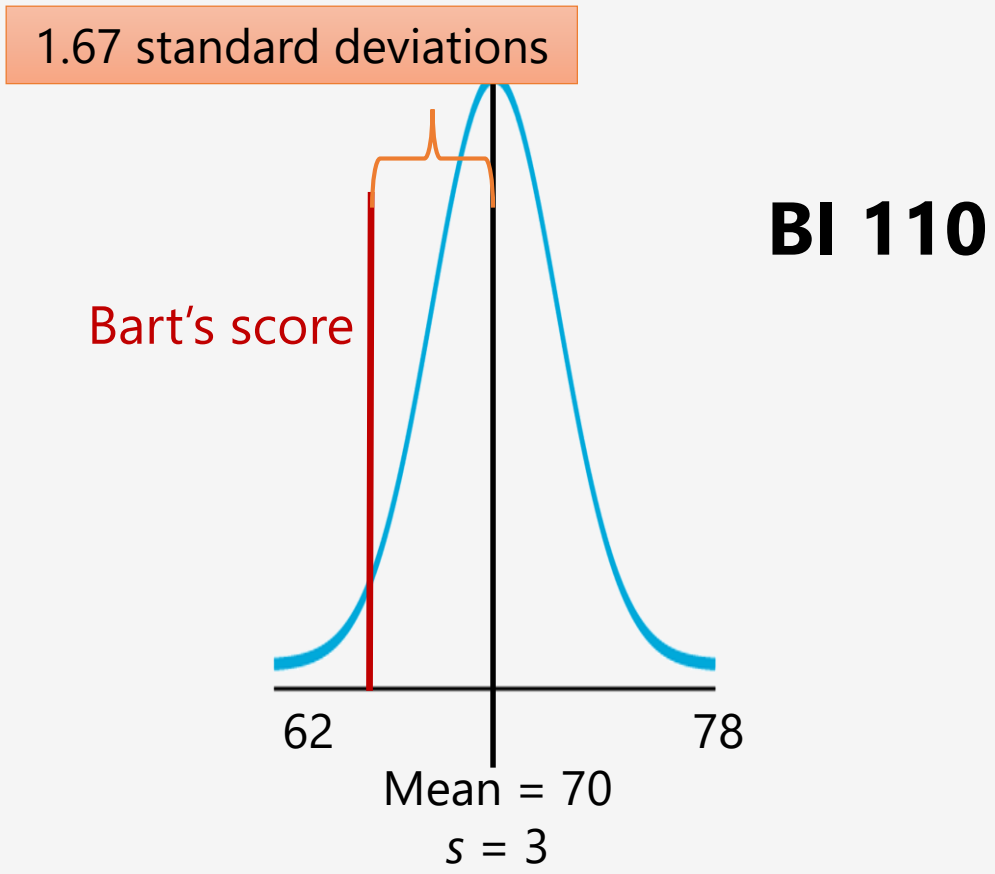
Which class is Bart doing *better* in, relative to the class' performance?

He's doing better in PY 101.
He is *only* 0.71 standard deviations below the mean score.



# Bart earned a 65% on both exams.



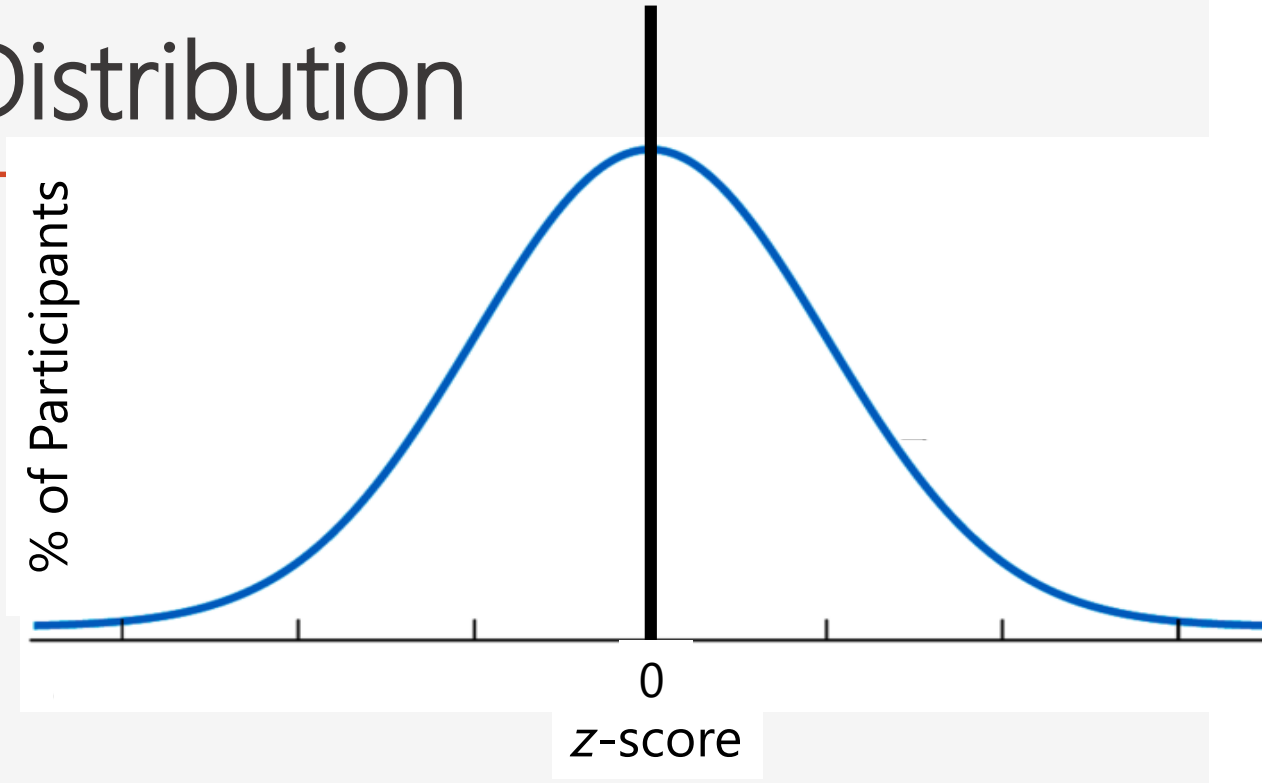
# What are two advantages of *z*-scores, over and above *raw scores*?

1. *Z*-scores make it possible to compare scores that come from different distributions.
   - Our original example: Lisa's *z*-score in **PY 101** & her *z*-score in **BI 110.**
   - Another example: Lisa's *z*-score in **MU 100 Section *A*** & Allison's *z*-score in **MU 100 Section *B*.**
2. *Z*-scores allow us to determine the **percentage** of raw scores that fall above or below a given raw score.
   - EX: what % of PY 101 classmates performed better than and what % performed worse than, Lisa, who scored a 76 on the exam?

*… the topic of Chapter 5*

# The Standard Normal (*z*) Distribution

- If you converted ALL raw scores in your normally-distributed sample to z-scores, & created a frequency distribution (histogram) from all of those z-scores, it would be called the *z* distribution, which always has a mean of zero.




If your z-score = 0, what % of scores in the sample are *above* your score?

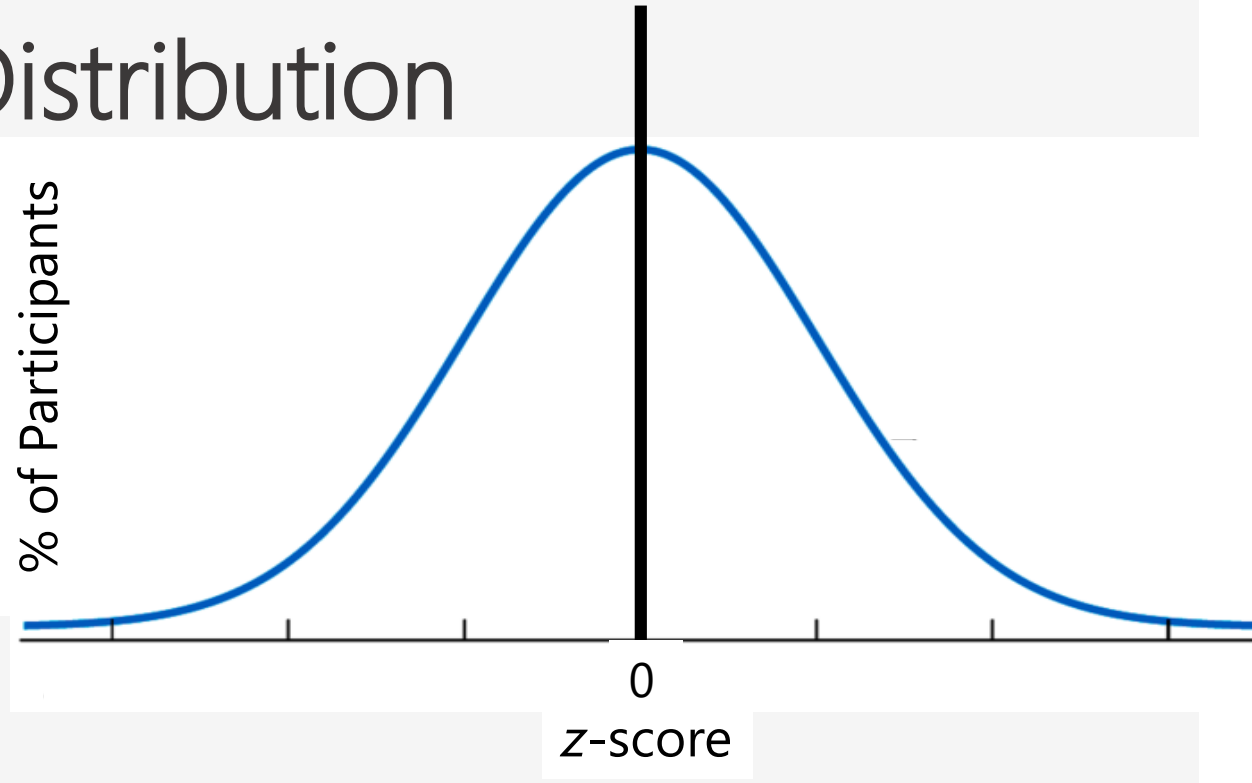
- **~50%**

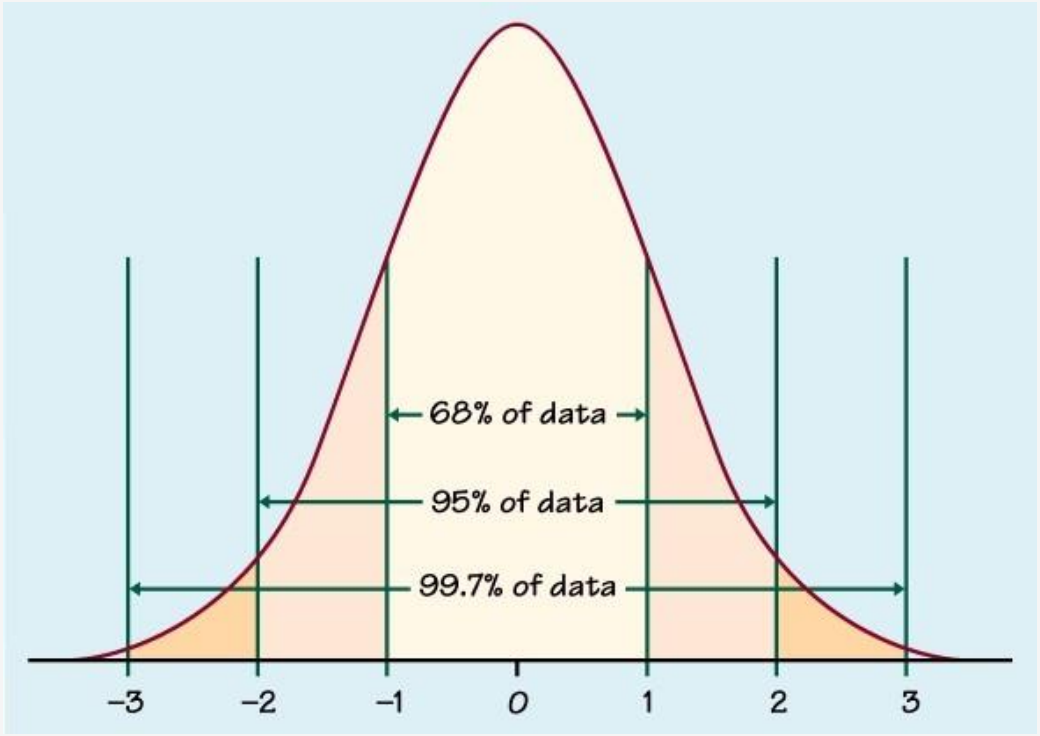
If your z-score = 0, what % of scores in the sample fall *below* your score?

- **~50%**

# The Standard Normal (*z*) Distribution

- If you converted ALL raw scores in your normally-distributed sample to z-scores, & created a frequency distribution (histogram) from all of those z-scores, it would be called the *z* distribution, which always has a mean of zero.



# Check your understanding of the z distribution

1. In any z-distribution, what % of all scores have a z-score of either >+2 or < –2?
2. If your z-score for an exam is +0.75, does your raw exam score fall *above* or *below* the class mean?
3. If your raw exam score equals the class mean, what is your z-score?
4. If your z-score is –2.98, are you closer to the center of the distribution or tails?
5. If you randomly select a z-score from all possible z-scores in our sample, what's the probability that the chosen score will be between -2 and +2?
6. In a given sample, Person A has a z-score of –1.75 and Person B has a z-score of 2.30. Whose raw score is closer to the mean of the sample?
7. In a given sample, Person J has a z-score of +1 and Person K has a z-score of -1. What % of participants in the sample have scores between Person J and Person K?
8. **If the mean height of all $5^{th}$ graders is 50 inches, with a standard deviation of 5 inches and assuming the distribution is normal, what % of $5^{th}$ graders have a height between 45 inches and 55 inches?

# Check your understanding of the z distribution

| | | |
|---|---|---|
| 5% | 1. | In any z-distribution, what % of all scores have a z-score of >+2 or < –2? |
| above | 2. | If your z-score for an exam is +.75, does your raw exam score fall *above* or *below* the mean? |
| $z = 0$ | 3. | If your raw exam score equals the class mean, what is your z-score? |
| tail | 4. | If your z-score is –2.98, are you closer to the center of the distribution or tails? |
| 95% | 5. | If you randomly select a z-score from all possible z-scores in our sample, what's the probability that the chosen score will be between -2 and +2? |
| A | 6. | In a given sample, Person A has a z-score of –1.75 and Person B has a z-score of 2.30. Whose raw score is closer to the mean of the sample? |
| 68% | 7. | In a given sample, Person J has a z-score of +1 and Person K has a z-score of -1. What % of participants in the sample have scores between Person J and Person K? |
| 68% | 8. | **If the mean height of all 5$^{th}$ graders is 50 inches, with a standard deviation of 5 inches and assuming the distribution is normal, what % of 5$^{th}$ graders have a height between 45 inches and 55 inches? |