
Last week:

Chapter 3:
Measures of Central tendency (mean, median, mean)
Measures of Dispersion
- variation ratio (v), the range and interquartile range
- standard deviation and the variance

Move onto Chapter 4: The Normal Curve

# Measures of Dispersion

# The Concept of Dispersion

- Dispersion = variety, diversity, amount of variation between scores
- Also referred to as "heterogeneity"
- The greater the dispersion of a variable, the greater the range of scores and the greater the differences between scores.

4-3

Which city has greater "variation" or "diversity" in terms of "ethnicity"


Toronto

Tokyo

# Variation Ratio ($v$)

- Variation Ratio ($v$) is one of only a few measures of dispersion for nominal-level variables.
  - $v$ provides a quick, easy way to quantify dispersion.
- Variation Ratio is simply the proportion of cases not in the modal category. That is:

$$v = 1 - \left(\frac{f_m}{n}\right)$$

where $f_m$ = the number of cases in the mode
$n$ = the total number of cases

- $v$ has a lower limit of 0.00 (no variation/all cases are in the mode) and increases to 1.00 as the proportion of cases in the mode decreases.
  - Thus, the larger the $v$, the more dispersion in a variable.

Country of Birth, Toronto and Tokyo

| | Toronto | Tokyo |
|---|---|---|
| | | |
| Born in Country of Residence | 3,172,620 | 26,072,629 |
| Born Abroad | 2,595,780 | 806,370 |
| | | |
| | 5,768,400 | 26,878,999 |

Which city has the greatest diversity?

$$v = 1 - \left(\frac{f_m}{n}\right)$$

where $f_m$ = the number of cases in the mode
$n$ = the total number of cases

For Toronto:

v = 1 - (3,172,620 / 5,768,400)

= 1 - .55 = .45

Toronto clearly has greater variety or ethnic diversity..

For Tokyo

v = 1 - (26,072,629 / 26,878,999)

= 1 - .97 = .03

# The Concept of Dispersion:

- V is not the preferred indicator for ordinal or interval/ratio variables

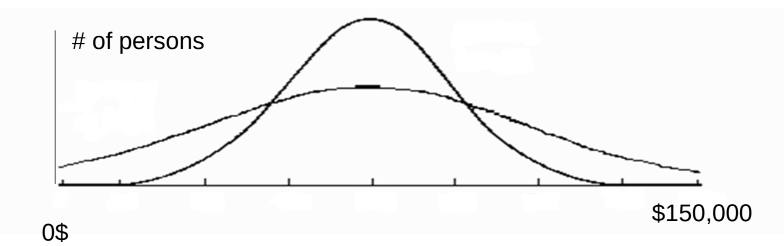
- Consider the income distribution of different societies?
- Interval ratio..
- We should work with other possibilities..,



# The Concept of Dispersion

*(continued)*

Consider the following distribution on income (interval ratio) variable in two hypothetical societies




- The taller curve has **less** dispersion.
- The flatter curve has **more** dispersion.

Range

**Range**
(*R* ) = High Score – Low Score

Consider the following income data:

If data are $130,000, 35,000, 30,000, 30,000, 30,000, 30,000, 25,000, 25,000
- then range = 130,000 – 25,000 = $105,000

- Quick and easy indication of variability.
- Can be used with ordinal or interval-ratio variables.
- Limitations because based on only two scores:
  1. Distorted by atypically high or low scores.
  2. No information about variation between high and low scores.

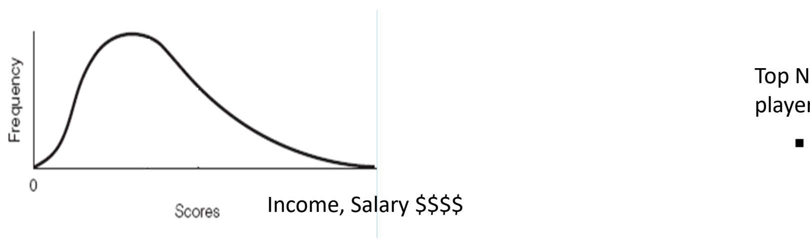
# The Range (*R*)

# The Range (*R*)



4-11

# The Range (*R*)





4-12

# The Range (*R*)


Income, Salary $$$$

Wider range,
but overwhelming
Majority fall within
A narrower
range

Top NBA
player


Income, Salary $$$$

4-13

# The Range (*R*)


Income, Salary $$$$

Wider range,
but overwhelming
Majority fall within
A narrower
range

Top NBA
player

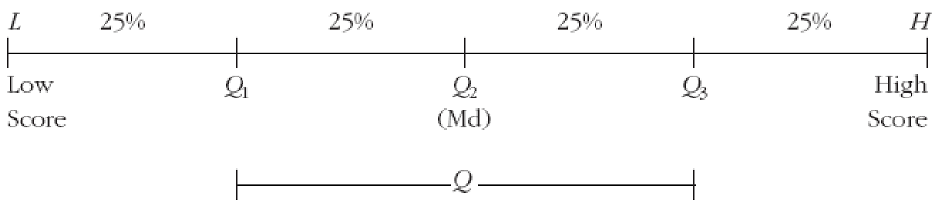
Range provides us
with little detail
on the distribution


Income, Salary $$$$

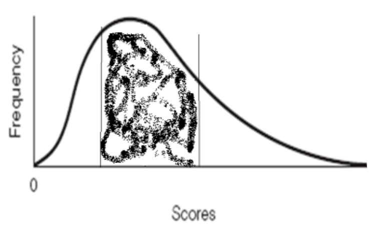
4-14

# Interquartile Range ($Q$)

- Avoids some problems of $R$ by focusing only on middle 50 percent of scores.
- Can isolate the scores between the third quartile ($Q_3$) to the first quartile ($Q_1$).



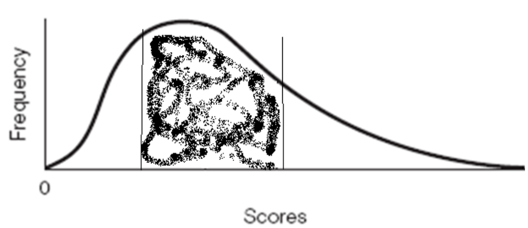

4-15




Wider range,
but overwhelming
Majority fall within
A narrower
range

Top NBA
player

Income, Salary $$$$




By considering
Interquartile range
better feel for the data

4-16

# Interquartile Range (*Q*)

- 1. Sort data from smallest to largest
-

| | Percentage using public transit across CMAs | |
|---|---|---|
| | | |
| 1 | St. John's | 3 |
| 2 | Halifax | 10 |
| 3 | Saint John | 4 |
| 4 | Chicoutimi | 2 |
| 5 | Quebec | 10 |
| 6 | Sherbrooke | 6 |
| 7 | Trois-Rivieres | 3 |
| 8 | Montreal | 22 |
| 9 | Ottawa Hull | 19 |
| 10 | Kingston | 4 |
| 11 | Oshawa | 7 |
| 12 | Toronto | 21 |
| 13 | Hamilton | 8 |
| 14 | St. Catherines-Niagara | 2 |
| 15 | Kitchener | 4 |
| 16 | London | 7 |
| 17 | Windsor | 3 |
| 18 | Greater Sudbury | 6 |
| 19 | Thunder Bay | 3 |
| 20 | Winnipeg | 13 |

# Interquartile Range (*Q*)

- 1. Sort data from smallest to largest
- 2. Locate case that corresponds to: Q1 = N*.25
- 3. Locate case that corresponds to: Q3 = N*.75

| | Percentage using public transit across CMAs | | | | |
|---|---|---|---|---|---|
| | | | | | |
| 1 | St. John's | 3 | | Chicoutimi | 2 |
| 2 | Halifax | 10 | | St. Catherines-Niaga | 2 |
| 3 | Saint John | 4 | | St. John's | 3 |
| 4 | Chicoutimi | 2 | | Trois-Rivieres | 3 |
| 5 | Quebec | 10 | | Windsor | 3 |
| 6 | Sherbrooke | 6 | | Thunder Bay | 3 |
| 7 | Trois-Rivieres | 3 | | Saint John | 4 |
| 8 | Montreal | 22 | | Kingston | 4 |
| 9 | Ottawa Hull | 19 | | Kitchener | 4 |
| 10 | Kingston | 4 | | Sherbrooke | 6 |
| 11 | Oshawa | 7 | | Greater Sudbury | 6 |
| 12 | Toronto | 21 | | Oshawa | 7 |
| 13 | Hamilton | 8 | | London | 7 |
| 14 | St. Catherines-Niagara | 2 | | Hamilton | 8 |
| 15 | Kitchener | 4 | | Halifax | 10 |
| 16 | London | 7 | | Quebec | 10 |
| 17 | Windsor | 3 | | Winnipeg | 13 |
| 18 | Greater Sudbury | 6 | | Ottawa Hull | 19 |
| 19 | Thunder Bay | 3 | | Toronto | 21 |
| 20 | Winnipeg | 13 | | Montreal | 22 |

# Interquartile Range (*Q*)

- 1. Sort data from smallest to largest
- 2. Locate case that corresponds to: Q1 = N*.25
- 3. Locate case that corresponds to: Q3 = N*.75

| | Percentage using public transit across CMAs | | | | |
|---|---|---|---|---|---|
| 1 | St. John's | 3 | Chicoutimi | 2 | Find Q1 = 20*.25 = 5 |
| 2 | Halifax | 10 | St. Catherines-Niaga | 2 | |
| 3 | Saint John | 4 | St. John's | 3 | Find Q3 = 20*.75 = 15 |
| 4 | Chicoutimi | 2 | Trois-Rivieres | 3 | |
| 5 | Quebec | 10 | Windsor | 3 | |
| 6 | Sherbrooke | 6 | Thunder Bay | 3 | |
| 7 | Trois-Rivieres | 3 | Saint John | 4 | |
| 8 | Montreal | 22 | Kingston | 4 | |
| 9 | Ottawa Hull | 19 | Kitchener | 4 | |
| 10 | Kingston | 4 | Sherbrooke | 6 | |
| 11 | Oshawa | 7 | Greater Sudbury | 6 | |
| 12 | Toronto | 21 | Oshawa | 7 | |
| 13 | Hamilton | 8 | London | 7 | |
| 14 | St. Catherines-Niagara | 2 | Hamilton | 8 | |
| 15 | Kitchener | 4 | Halifax | 10 | |
| 16 | London | 7 | Quebec | 10 | |
| 17 | Windsor | 3 | Winnipeg | 13 | |
| 18 | Greater Sudbury | 6 | Ottawa Hull | 19 | |
| 19 | Thunder Bay | 3 | Toronto | 21 | |
| 20 | Winnipeg | 13 | Montreal | 22 | |

**5th case**

**15th case**

# Interquartile Range (*Q*)

- 1. Sort data from smallest to largest
- 2. Locate case that corresponds to: Q1 = N*.25
- 3. Locate case that corresponds to: Q3 = N*.75
- 4. Find the corresponding scores associated with Q3 – Q1

| | Percentage using public transit across CMAs | | | | |
|---|---|---|---|---|---|
| 1 | St. John's | 3 | Chicoutimi | 2 | Find Q1 = 20*.25 = 5 |
| 2 | Halifax | 10 | St. Catherines-Niaga | 2 | |
| 3 | Saint John | 4 | St. John's | 3 | Find Q3 = 20*.75 = 15 |
| 4 | Chicoutimi | 2 | Trois-Rivieres | 3 | |
| 5 | Quebec | 10 | Windsor | 3 | |
| 6 | Sherbrooke | 6 | Thunder Bay | 3 | Interquartile range |
| 7 | Trois-Rivieres | 3 | Saint John | 4 | 10 - 3 = 7 |
| 8 | Montreal | 22 | Kingston | 4 | |
| 9 | Ottawa Hull | 19 | Kitchener | 4 | |
| 10 | Kingston | 4 | Sherbrooke | 6 | |
| 11 | Oshawa | 7 | Greater Sudbury | 6 | |
| 12 | Toronto | 21 | Oshawa | 7 | |
| 13 | Hamilton | 8 | London | 7 | |
| 14 | St. Catherines-Niagara | 2 | Hamilton | 8 | |
| 15 | Kitchener | 4 | Halifax | 10 | |
| 16 | London | 7 | Quebec | 10 | |
| 17 | Windsor | 3 | Winnipeg | 13 | |
| 18 | Greater Sudbury | 6 | Ottawa Hull | 19 | |
| 19 | Thunder Bay | 3 | Toronto | 21 | |
| 20 | Winnipeg | 13 | Montreal | 22 | |

**5th case** -> **score of 3**

**15th case** -> **score of 10**

**So, the interquartile range is:**

**10 – 3 or 7**

# Standard Deviation

- Very important statistic in sociology
- A measure of the degree of dispersion
-   -> most useful for interval ratio data
- Documents dispersion of data from the mean.
  - Specifically, the “typical” distance of each score from the mean.

- Symbolized as, *s,* for a sample and, *σ* (sigma), for a population.

- Can be used in combination with the mean to describe a "Normal” distribution (details to follow).

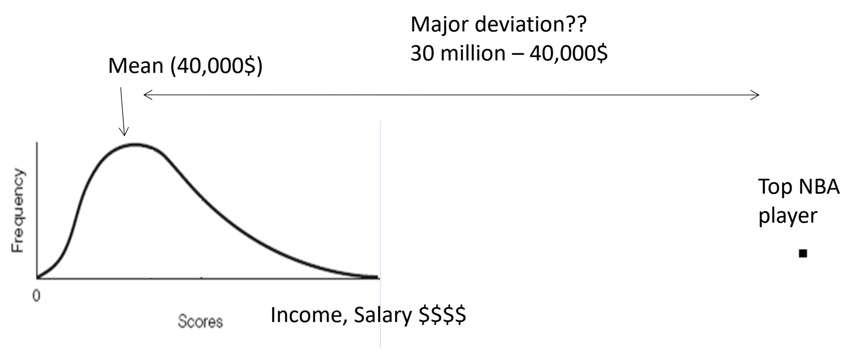


# Standard Deviation *(continued)*

- Meets criteria for good measure of dispersion:
  1. Use all scores in the distribution.
  2. Describe the average or typical deviation of the scores.
  3. Increase in value as the distribution of scores becomes more diverse.
- Should be used with interval-ratio variables, but is sometimes used with ordinal-level variables.

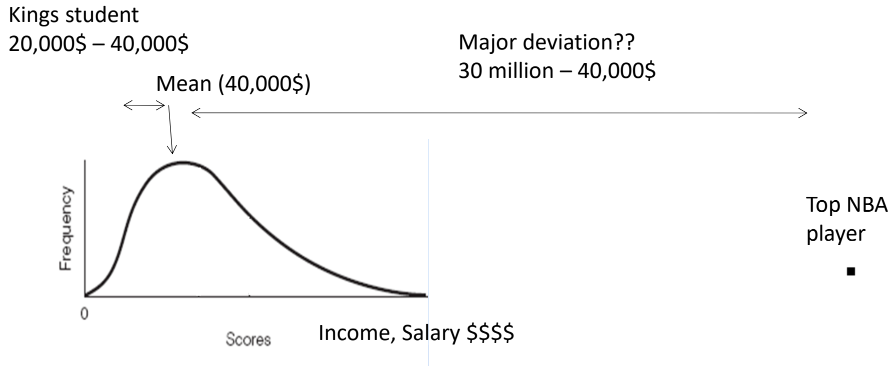

- What do I mean by “deviation of the scores from the mean”?

4-23




- What do I mean by “deviation of the scores from the mean”?

4-24




- What do I mean by "deviation of the scores from the mean"?
- Our NBA play deviates a lot further from the "mean salary" than does me or you.. Right?



## Standard Deviation *(continued)*

- The lowest value possible is 0 (no dispersion).
- The square of the standard deviation is the variance, $s^2$ (Variance is also commonly used!)
- Formula for standard deviation:

## Standard Deviation *(continued)*

- The lowest value possible is 0 (no dispersion).
- The square of the standard deviation is the variance, $s^2$ (Variance is also commonly used!)
- Formula for standard deviation:

$$s = \sqrt{\frac{\sum(X_i - \overline{X})^2}{N}}$$

4-27

## Computing Standard Deviation

- To solve:
  - First, most calculate the mean (if not available)
  - Subtract mean from each score.
  - Square the deviations.
  - Sum the squared deviations.
  - Divide the sum of the squared deviations by *N*.
  - Find the square root of the result.

4-28

## Computing Standard Deviation: An Example

**TABLE 4.4** COMPUTING THE STAN

| Scores ($X_i$) |
|---|
| 10 |
| 20 |
| 30 |
| 40 |
| 50 |
| $\Sigma(X_i) = 150$ |

## Computing Standard Deviation: An Example

**TABLE 4.4** COMPUTING THE STAN

| Scores ($X_i$) |
|---|
| 10 |
| 20 |
| 30 |
| 40 |
| 50 |
| $\Sigma(X_i) = 150$ |

Note: we need the mean

Across 5 cases,

Mean = 150/5 = 30

$$s = \sqrt{\frac{\Sigma(X_i - \overline{X})^2}{N}}$$

## Computing Standard Deviation: An Example

**TABLE 4.4** COMPUTING THE STANDARD DEVIATION

| Scores $(X_i)$ | Deviations $(X_i - \overline{X})$ |
|---|---|
| 10 | $(10 - 30) = -20$ |
| 20 | $(20 - 30) = -10$ |
| 30 | $(30 - 30) = 0$ |
| 40 | $(40 - 30) = 10$ |
| 50 | $(50 - 30) = 20$ |
| $\Sigma(X_i) = 150$ | $\Sigma(X_i - \overline{X}) = 0$ |

We use the mean in obtaining all our deviations

$$s = \sqrt{\frac{\sum(X_i - \overline{X})^2}{N}}$$

## Computing Standard Deviation: An Example

**TABLE 4.4** COMPUTING THE STANDARD DEVIATION

| Scores $(X_i)$ | Deviations $(X_i - \overline{X})$ | Deviations Squared $(X_i - \overline{X})^2$ |
|---|---|---|
| 10 | $(10 - 30) = -20$ | $(-20)^2 = 400$ |
| 20 | $(20 - 30) = -10$ | $(-10)^2 = 100$ |
| 30 | $(30 - 30) = 0$ | $(0)^2 = 0$ |
| 40 | $(40 - 30) = 10$ | $(10)^2 = 100$ |
| 50 | $(50 - 30) = 20$ | $(20)^2 = 400$ |
| $\Sigma(X_i) = 150$ | $\Sigma(X_i - \overline{X}) = 0$ | $\Sigma(X_i - \overline{X})^2 = 1000$ |

$$s = \sqrt{\frac{\sum(X_i - \overline{X})^2}{N}}$$

## Computing Standard Deviation: An Example

**TABLE 4.4** COMPUTING THE STANDARD DEVIATION

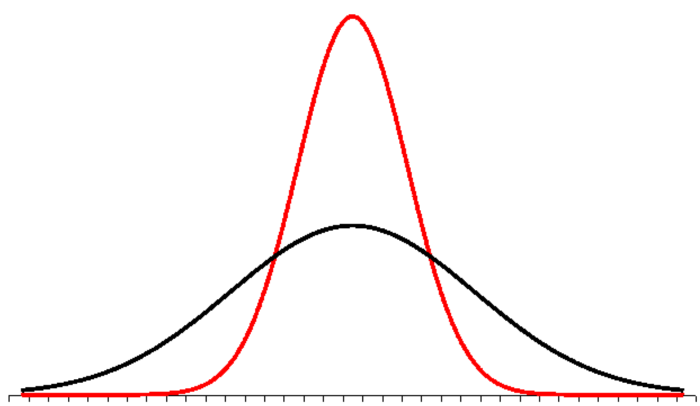
| Scores $(X_i)$ | Deviations $(X_i - \overline{X})$ | Deviations Squared $(X_i - \overline{X})^2$ |
|---|---|---|
| 10 | $(10 - 30) = -20$ | $(-20)^2 = 400$ |
| 20 | $(20 - 30) = -10$ | $(-10)^2 = 100$ |
| 30 | $(30 - 30) = 0$ | $(0)^2 = 0$ |
| 40 | $(40 - 30) = 10$ | $(10)^2 = 100$ |
| 50 | $(50 - 30) = 20$ | $(20)^2 = 400$ |
| $\Sigma(X_i) = 150$ | $\Sigma(X_i - \overline{X}) = 0$ | $\Sigma(X_i - \overline{X})^2 = 1000$ |

$$s = \sqrt{\frac{\sum(X_i - \overline{X})^2}{N}}$$

$$\begin{aligned} s &= \sqrt{\frac{\Sigma(X_i - \overline{X})^2}{N}} \\ &= \sqrt{\frac{1000}{5}} \\ &= \sqrt{200} \\ &= 14.14 \end{aligned}$$

4-33

## Standard Deviation: An Example



Two samples., both have the same means and ranges,..
Which has the larger standard deviation?
Obviously the dark line. The red line has less variance..

If you compared the GPA of PhD students with first year students, which would have the higher s?

4-34

# Standard Deviation: An Example

| | | | | | | |
|---|---|---|---|---|---|---|
| PHD students | | | | MA students | | |
| | | | | | | |
| Student # | | | | Student # | | |
| | | | | | | |
| 1 | 85 | | | 11 | 75 | |
| 2 | 81 | | | 12 | 65 | |
| 3 | 92 | | | 13 | 92 | |
| 4 | 88 | | | 14 | 66 | |
| 5 | 78 | | | 15 | 52 | |
| 6 | 82 | | | 16 | 40 | |
| 7 | 76 | | | 17 | 88 | |
| 8 | 84 | | | 18 | 92 | |
| 9 | 95 | | | 19 | 75 | |
| 10 | 82 | | | 20 | 74 | |
| | | | | | | |

4-35

| | |
|---|---|
| PHD students | |
| | |
| Student # | |
| | |
| 1 | 85 |
| 2 | 81 |
| 3 | 92 |
| 4 | 88 |
| 5 | 78 |
| 6 | 82 |
| 7 | 76 |
| 8 | 84 |
| 9 | 95 |
| 10 | 82 |
| | |

1. Obtain the mean, and then:
   2. Subtract mean from each score.
   3. Square the deviations.
   4. Sum the squared deviations.
   5. Divide the sum of the squared deviations by *N*
   6. Find the square root of the result.

$$s = \sqrt{\frac{\Sigma(X_i - \overline{X})^2}{N}}$$

| PHD students | |
|---|---|
| Student # | |
| 1 | 85 |
| 2 | 81 |
| 3 | 92 |
| 4 | 88 |
| 5 | 78 |
| 6 | 82 |
| 7 | 76 |
| 8 | 84 |
| 9 | 95 |
| 10 | 82 |

| 2. Deviations $(X_i - \overline{X})$ | 3. Deviations Squared $(X_i - \overline{X})^2$ |
|---|---|
| 0.70 | 0.49 |
| -3.30 | 10.89 |
| 7.70 | 59.29 |
| 3.70 | 13.69 |
| -6.30 | 39.69 |
| -2.30 | 5.29 |
| -8.30 | 68.89 |
| -0.30 | 0.09 |
| 10.70 | 114.49 |
| -2.30 | 5.29 |

1. $\Sigma(X_i) =$ 843

$\overline{X}$ = 84.30

4. $\Sigma(X_i - \overline{X})^2$ 318.1

5. $\Sigma(X_i - \overline{X})^2$ /N 31.81

6. $s = \sqrt{\dfrac{\Sigma(X_i - \overline{X})^2}{N}}$ → 5.64

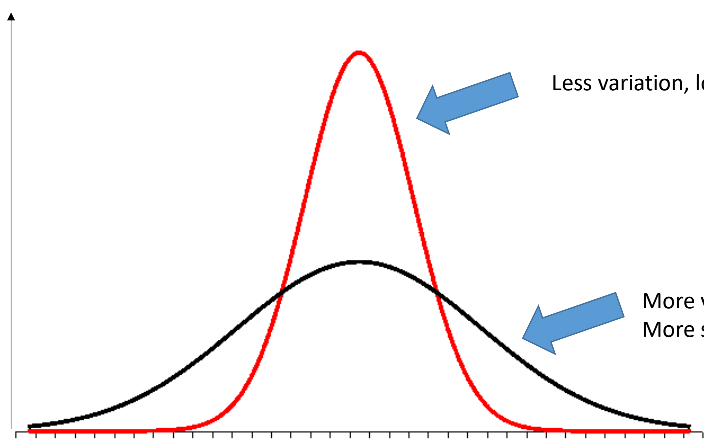
1. Obtain the mean, and then:
   2. Subtract mean from each score.
   3. Square the deviations.
   4. Sum the squared deviations.
   5. Divide the sum of the squared deviations by *N*
   6. Find the square root of the result.

$$s = \sqrt{\frac{\Sigma(X_i - \overline{X})^2}{N}}$$

Note: If you did the same with MA students, your standard deviation should be 16.08

## Standard Deviation: An Example

# of cases



Less variation, lower standard deviation

More variation, higher standard deviation,
More spread out!

Two samples., both have the same means and ranges,..
Which has the larger standard deviation?

Obviously the dark line. The red line has less variance..

4-38

## Summary: Relationship Between Level of Measurement and Measure of Central Tendency and Dispersion

THE RELATIONSHIP BETWEEN LEVEL OF MEASUREMENT AND MEASURE OF CENTRAL TENDENCY AND DISPERSION*

| Level of Measurement | Measure(s) of Central Tendency | Measure(s) of Dispersion |
|---|---|---|
| Nominal | *Mode* | *Variation Ratio* |
| Ordinal | Mode, *Median* | Variation Ratio, Range, *Interquartile Range* |
| Interval-ratio** | Mode, Median, *Mean* | Variation Ratio, Range, Interquartile Range, Variance, *Standard Deviation* |

*The most appropriate measure of central tendency and dispersion for each level of measurement is in italic.
**The median and interquartile range should be used for interval-ratio variables with highly skewed distributions.






# Chapter 4

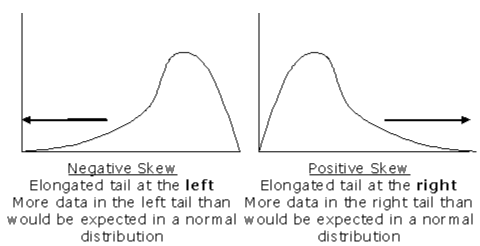
The Normal Curve

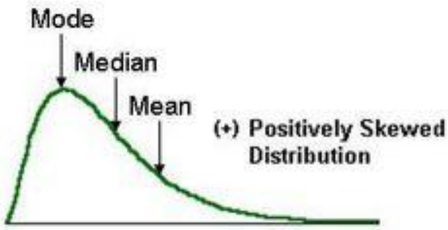
# In this presentation you will learn about:

- The Normal Curve
- *Z* scores (& how to use them!!)

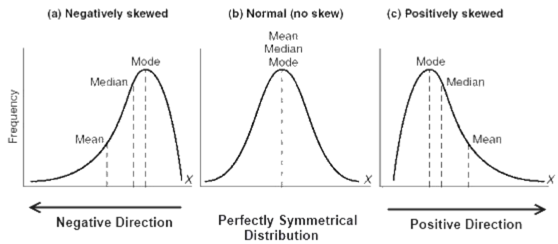
5-41

Recall last class, I mentioned other types of distributions:



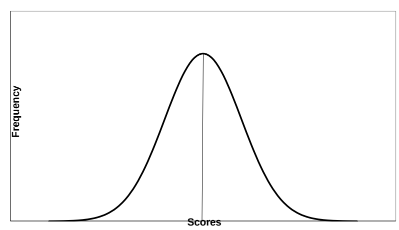


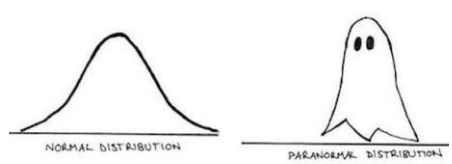

Eg. Income is positive skewed in Canada, right?

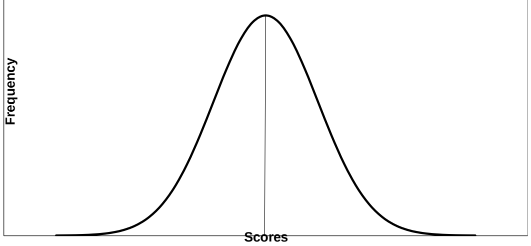

# The Normal Curve

- Also referred to as:
- “Gaussian distribution”
- “bell curve”
- “the perfect normal curve”
- In reality, many variables have a distribution that is similar to what is referred to as the “bell curve”..
- Note: in this distribution, the mean is equal to the median which is equal to the mode!

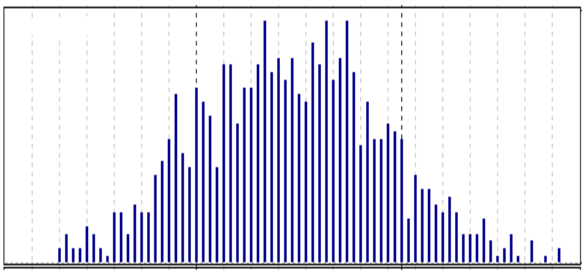


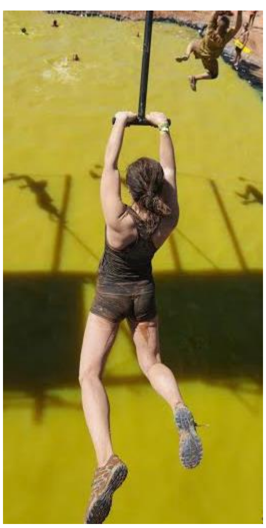



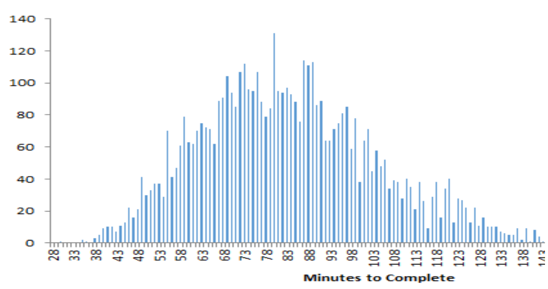
# The Normal Curve




In reality, many variables have a distribution that is similar to what is referred to as the "bell curve"..

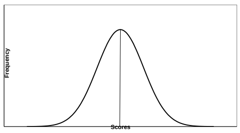







# The Normal Curve

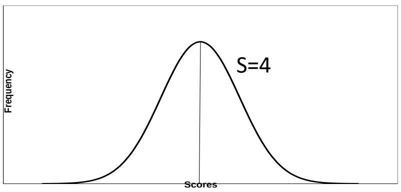
Frequency




S=2

Normal curves vary in terms of their dispersion, which we can summarize in the standard deviation of distributions

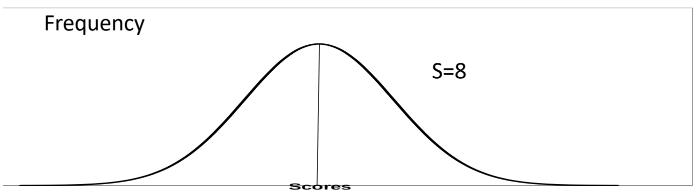
Frequency




The wider the distribution, the greater the standard deviation

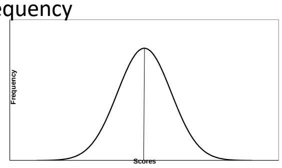
$$s = \sqrt{\frac{\sum (X_i - \overline{X})^2}{N}}$$

Frequency



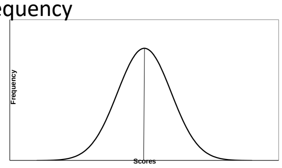

5-47

Frequency

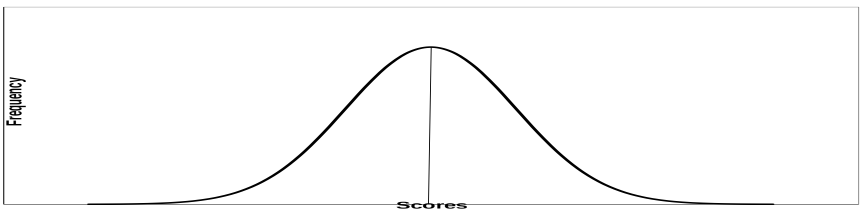



The standard deviation on times was much lower for the 2012 Olympic marathon than it was for the 2012 New York marathon

Frequency

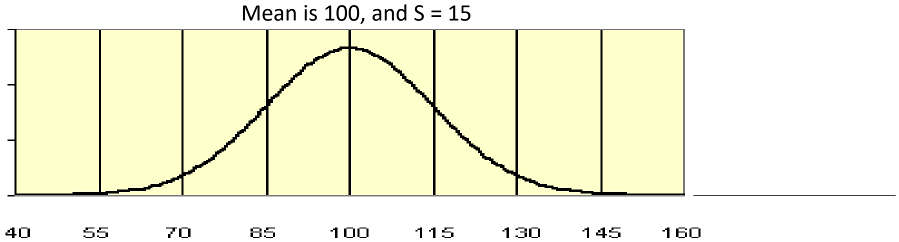


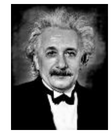

The standard deviation on times was much lower for the 2012 Olympic marathon than it was for the 2012 New York marathon




**IQ Normal Curve**

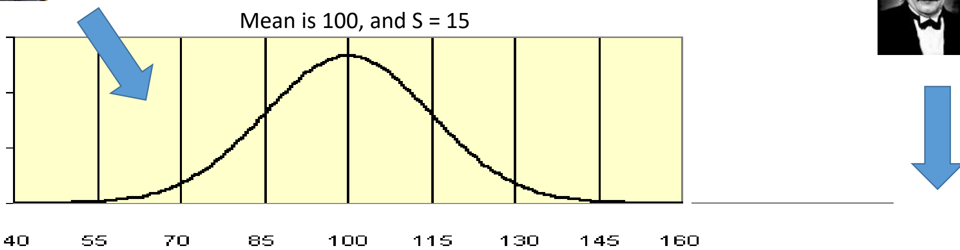
Mean is 100, and S = 15



**IQ Normal Curve**

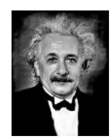
Mean is 100, and S = 15

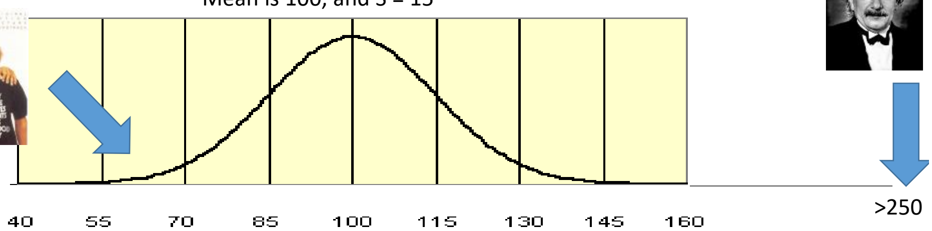



It is possible to think of a distribution in terms of actual scores on a variable (e.g. IQ)

**IQ Normal Curve**

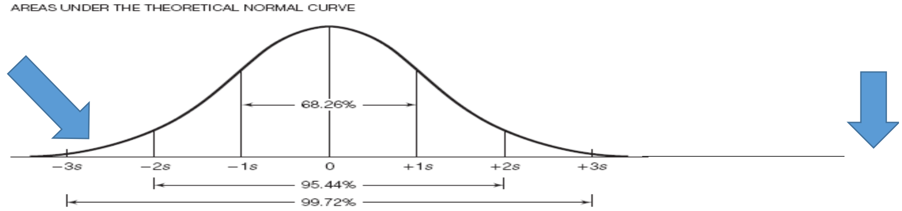
Mean is 100, and S = 15




.. Or in terms of a theoretical distribution (i.e. what we call the “theoretical normal curve”)

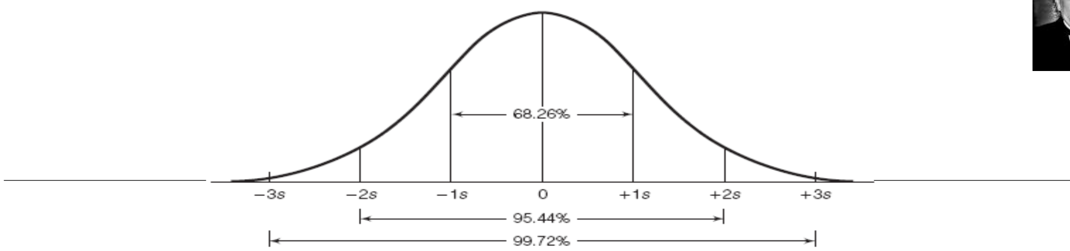
FIGURE 5.3 AREAS UNDER THE THEORETICAL NORMAL CURVE

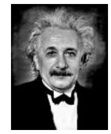


# Theoretical Normal Curve

- In Statistics we work with a “Theoretical distribution” meant to represent a normal distribution

FIGURE 5.3 AREAS UNDER THE THEORETICAL NORMAL CURVE



Z score of 10!!!

- The mean is assigned a value of 0
- The standard deviation is assigned a value of 1
- Describe this distribution in terms of Z scores (Standard score
- A Z score of 1 is 1 standard deviation above the mean,..
- A Z score of -1 is 1 standard deviation below the mean,.. etc.

| Between | Lies |
|---|---|
| ±1 standard deviation | 68.26% of the area |
| ±2 standard deviations | 95.44% of the area |
| ±3 standard deviations | 99.72% of the area |

5-53

- Z scores -> also called “standard scores”


$$Z = \frac{X_i - \overline{X}}{s}$$

- Are useful in determining the exact location of any value as observed $X_i$ in terms of this theoretical normal curve
- Tells us how many standard deviations a score is away from the mean…
- Correspondingly, can use Z scores to find the corresponding “proportions” of area under the curve associated with specific values
- Can use these proportions to determine what % higher or lower than a specific score

5-54

## Working with this theoretical distribution

- If we know the Z scores and corresponding location on the curves, we can answer all sorts of questions (using information on corresponding areas under the curve)
- Example: We know that the mean IQ in N.A. is 100, and the s = 15..



## Working with this theoretical distribution

- If we know the Z scores and corresponding location on the curves, we can answer all sorts of questions (using information on corresponding areas under the curve)
- Example: We know that the mean IQ in N.A. is 100, and the s = 15..
- 
- What proportion of the population is “dumb” as this guy? (e.g. IQ = 77.5)

## Working with this theoretical distribution

- If we know the Z scores and corresponding location on the curves, we can answer all sorts of questions (using information on corresponding areas under the curve)
- Example: We know that the mean IQ in N.A. is 100, and the s = 15..
- 
- What proportion of the population is "dumb" as this guy? (e.g. IQ = 77.5)
- Procedure:
- To find area, first compute the appropriate Z score (taking note of the appropriate sign) How may standard deviations away from the mean is he???
- Use the "Under the Normal curve Table" in the text book (Appendix A).
- This table allows us to find the areas above and below a specific Z score…

  (what proportion score lower, or higher than this fellow??)



# Using the Normal Curve: Appendix A

TABLE ONLY LISTS POSITIVE Z SCORES.. BUT THE AREAS ARE THE SAME REGARDLESS OF WHETHER THEY ARE POSITIVE OR NEGATIVE

AN ILLUSTRATION OF HOW TO FIND AREAS UNDER THE NORMAL CURVE USING APPENDIX A

| (a) *Z* | (b) Area between Mean and *Z* | (c) Area beyond *Z* |
|---|---|---|
| 0.00 | 0.0000 | 0.5000 |
| 0.01 | 0.0040 | 0.4960 |
| 0.02 | 0.0080 | 0.4920 |
| 0.03 | 0.0120 | 0.4880 |
| ⋮ | ⋮ | ⋮ |
| 1.00 | 0.3413 | 0.1587 |
| 1.01 | 0.3438 | 0.1562 |
| 1.02 | 0.3461 | 0.1539 |
| 1.03 | 0.3485 | 0.1515 |
| ⋮ | ⋮ | ⋮ |
| 1.50 | 0.4332 | 0.0668 |
| 1.51 | 0.4345 | 0.0655 |
| 1.52 | 0.4357 | 0.0643 |
| 1.53 | 0.4370 | 0.0630 |
| ⋮ | ⋮ | ⋮ |

# Working with Column B

AN ILLUSTRATION OF HOW TO FIND AREAS UNDER THE NORMAL CURVE USING APPENDIX A

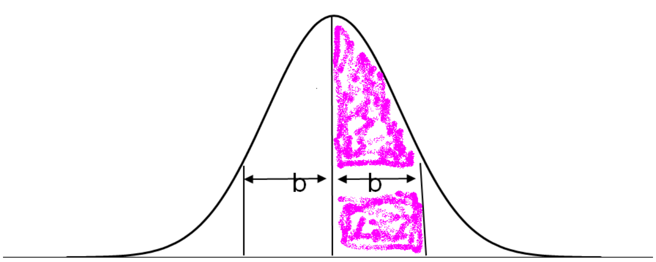
| (a) $Z$ | (b) Area between Mean and $Z$ | (c) Area beyond $Z$ |
|---|---|---|
| 0.00 | 0.0000 | 0.5000 |
| 0.01 | 0.0040 | 0.4960 |
| 0.02 | 0.0080 | 0.4920 |
| 0.03 | 0.0120 | 0.4880 |
| ⋮ | ⋮ | ⋮ |
| 1.00 | 0.3413 | 0.1587 |
| 1.01 | 0.3438 | 0.1562 |
| 1.02 | 0.3461 | 0.1539 |
| 1.03 | 0.3485 | 0.1515 |
| ⋮ | ⋮ | ⋮ |
| 1.50 | 0.4332 | 0.0668 |
| 1.51 | 0.4345 | 0.0655 |
| 1.52 | 0.4357 | 0.0643 |
| 1.53 | 0.4370 | 0.0630 |
| ⋮ | ⋮ | ⋮ |

5-59

# Column (b)

If $Z$ score is positive

column (b) = represents the area between the mean and the $+Z$ score

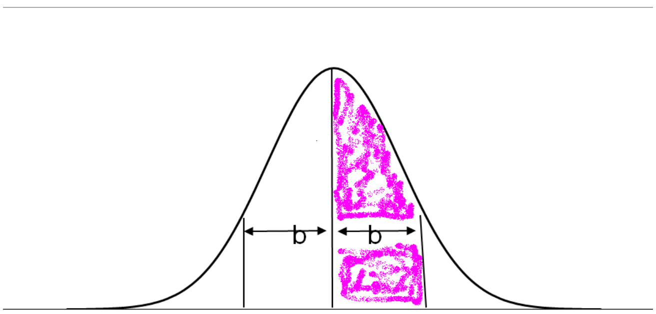


# Column (b)

If *Z* score is positive

column (b) = represents the area between the mean and the +*Z* score

Example: Average PhD student has an IQ of 122.5, and our population mean was 100, and S was 15




$$Z = \frac{X_i - \overline{X}}{s} = \frac{122.5 - 100}{15} = +1.5$$

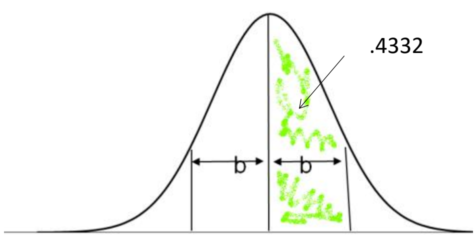
As an example Assume our Z score is + 1.5

i.e. 1.5 standard deviations above the mean

Column (b) gives us:

**TABLE 5.3** AN ILLUSTRATION OF HOW TO FIND AREAS UNDER THE NORMAL CURVE USING APPENDIX A

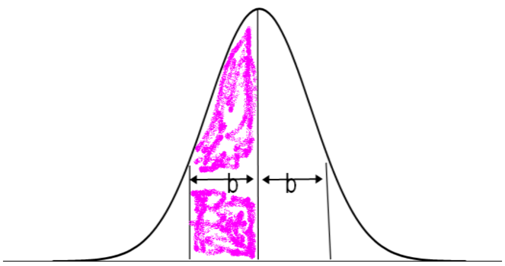
| (a) $Z$ | (b) Area between Mean and $Z$ | (c) Area beyond $Z$ |
|---|---|---|
| 0.00 | 0.0000 | 0.5000 |
| 0.01 | 0.0040 | 0.4960 |
| 0.02 | 0.0080 | 0.4920 |
| 0.03 | 0.0120 | 0.4880 |
| ⋮ | ⋮ | ⋮ |
| 1.00 | 0.3413 | 0.1587 |
| 1.01 | 0.3438 | 0.1562 |
| 1.02 | 0.3461 | 0.1539 |
| 1.03 | 0.3485 | 0.1515 |
| ⋮ | ⋮ | ⋮ |
| 1.50 → | 0.4332 | 0.0668 |
| 1.51 | 0.4345 | 0.0655 |
| 1.52 | 0.4357 | 0.0643 |
| 1.53 | 0.4370 | 0.0630 |
| ⋮ | ⋮ | ⋮ |




This tells us that the proportion of cases between the mean IQ and the IQ of interest (PhD students) is .4332

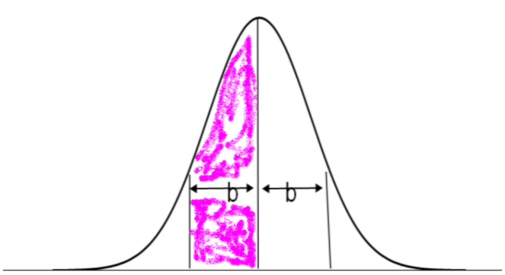
This translates into 43.32% of all cases

What if Z score is negative??
i.e. score on a variable is less than the mean
column (b) = also represents the area between the mean and the *-Z* score




5-63

What if Z score is negative??
i.e. score on a variable is less than the mean
column (b) = also represents the area between the mean and the *-Z* score




Example: Hab fan has an IQ of 77.5,
our population mean was 100, and S was 15

$$Z = \frac{X_i - \overline{X}}{s}$$

$$= \frac{77.5 - 100}{15} = -1.5$$

5-64

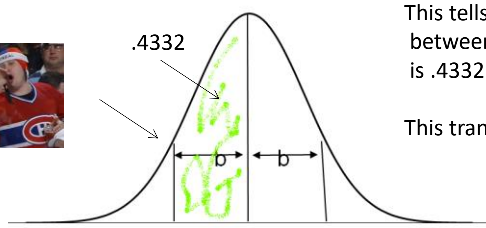
AN ILLUSTRATION OF HOW TO FIND AREAS UNDER THE NORMAL CURVE USING APPENDIX A

| (a) $Z$ | (b) Area between Mean and $Z$ | (c) Area beyond $Z$ |
|---|---|---|
| 0.00 | 0.0000 | 0.5000 |
| 0.01 | 0.0040 | 0.4960 |
| 0.02 | 0.0080 | 0.4920 |
| 0.03 | 0.0120 | 0.4880 |
| ⋮ | ⋮ | ⋮ |
| 1.00 | 0.3413 | 0.1587 |
| 1.01 | 0.3438 | 0.1562 |
| 1.02 | 0.3461 | 0.1539 |
| 1.03 | 0.3485 | 0.1515 |
| ⋮ | ⋮ | ⋮ |
| 1.50 → | 0.4332 | 0.0668 |
| 1.51 | 0.4345 | 0.0655 |
| 1.52 | 0.4357 | 0.0643 |
| 1.53 | 0.4370 | 0.0630 |
| ⋮ | ⋮ | ⋮ |

Z = - 1.5

i.e. 1.5 standard deviations below the mean

Note: areas are the same regardless of whether Z scores are positive or negative




This tells us that the proportion of cases between the average and Hab Fan's IQ is .4332

This translates into 43.32% of all cases

Hab fan's Z scrore was -1.5

5-65

# What about Column C?

AN ILLUSTRATION OF HOW TO FIND AREAS UNDER THE NORMAL CURVE USING APPENDIX A

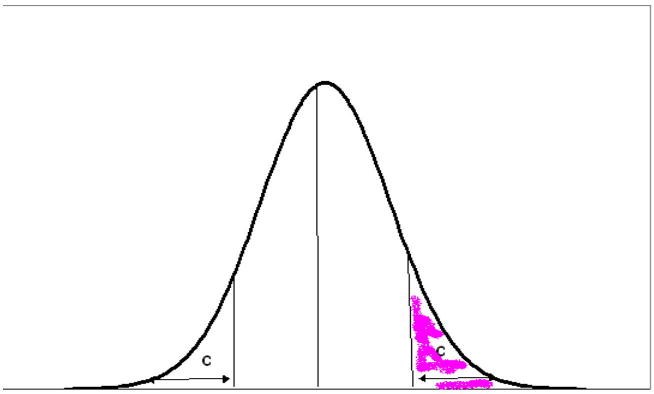
| (a) $Z$ | (b) Area between Mean and $Z$ | (c) Area beyond $Z$ |
|---|---|---|
| 0.00 | 0.0000 | 0.5000 |
| 0.01 | 0.0040 | 0.4960 |
| 0.02 | 0.0080 | 0.4920 |
| 0.03 | 0.0120 | 0.4880 |
| ⋮ | ⋮ | ⋮ |
| 1.00 | 0.3413 | 0.1587 |
| 1.01 | 0.3438 | 0.1562 |
| 1.02 | 0.3461 | 0.1539 |
| 1.03 | 0.3485 | 0.1515 |
| ⋮ | ⋮ | ⋮ |
| 1.50 | 0.4332 | 0.0668 |
| 1.51 | 0.4345 | 0.0655 |
| 1.52 | 0.4357 | 0.0643 |
| 1.53 | 0.4370 | 0.0630 |
| ⋮ | ⋮ | ⋮ |

5-66

# What about Column C in Appendix A

If *Z* score is positive

column (c) = represents the area "above" the +*Z* score



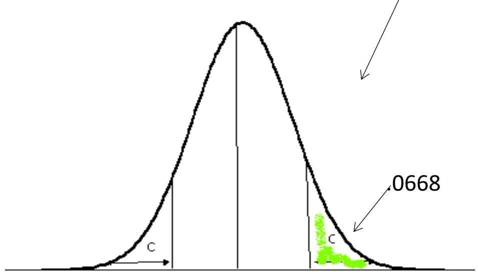



Z = 1.5

i.e. 1.5 standard deviations above the mean

AN ILLUSTRATION OF HOW TO FIND AREAS UNDER THE NORMAL CURVE USING APPENDIX A

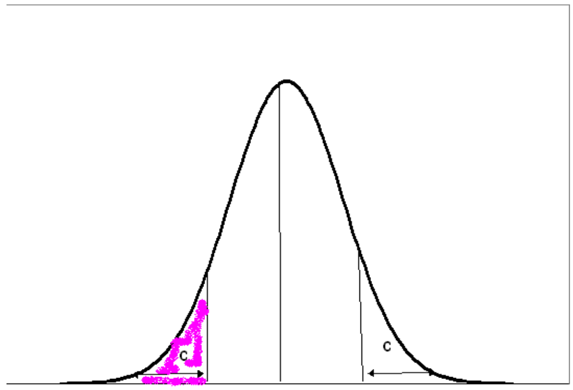
| (a) $Z$ | (b) Area between Mean and $Z$ | (c) Area beyond $Z$ |
|---|---|---|
| 0.00 | 0.0000 | 0.5000 |
| 0.01 | 0.0040 | 0.4960 |
| 0.02 | 0.0080 | 0.4920 |
| 0.03 | 0.0120 | 0.4880 |
| ⋮ | ⋮ | ⋮ |
| 1.00 | 0.3413 | 0.1587 |
| 1.01 | 0.3438 | 0.1562 |
| 1.02 | 0.3461 | 0.1539 |
| 1.03 | 0.3485 | 0.1515 |
| ⋮ | ⋮ | ⋮ |
| 1.50 | 0.4332 | 0.0668 |
| 1.51 | 0.4345 | 0.0655 |
| 1.52 | 0.4357 | 0.0643 |
| 1.53 | 0.4370 | 0.0630 |
| ⋮ | ⋮ | ⋮ |



# Using the Normal Curve: Appendix A *(continued)*

If *Z* score is negative

column (c) = represents the area “below” the - *Z* score



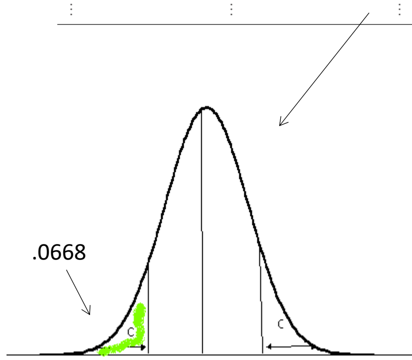



Z = - 1.5

i.e. 1.5 standard deviations below the mean

AN ILLUSTRATION OF HOW TO FIND AREAS UNDER THE NORMAL CURVE USING APPENDIX A

| (a) *Z* | (b) Area between Mean and *Z* | (c) Area beyond *Z* |
|---|---|---|
| 0.00 | 0.0000 | 0.5000 |
| 0.01 | 0.0040 | 0.4960 |
| 0.02 | 0.0080 | 0.4920 |
| 0.03 | 0.0120 | 0.4880 |
| ⋮ | ⋮ | ⋮ |
| 1.00 | 0.3413 | 0.1587 |
| 1.01 | 0.3438 | 0.1562 |
| 1.02 | 0.3461 | 0.1539 |
| 1.03 | 0.3485 | 0.1515 |
| ⋮ | ⋮ | ⋮ |
| 1.50 | 0.4332 | 0.0668 |
| 1.51 | 0.4345 | 0.0655 |
| 1.52 | 0.4357 | 0.0643 |
| 1.53 | 0.4370 | 0.0630 |
| ⋮ | ⋮ | ⋮ |

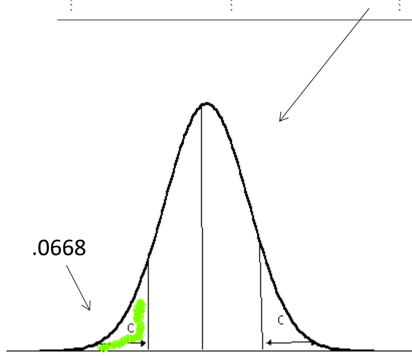


AN ILLUSTRATION OF HOW TO FIND AREAS UNDER THE NORMAL CURVE USING APPENDIX A

| (a) $Z$ | (b) Area between Mean and $Z$ | (c) Area beyond $Z$ |
|---|---|---|
| 0.00 | 0.0000 | 0.5000 |
| 0.01 | 0.0040 | 0.4960 |
| 0.02 | 0.0080 | 0.4920 |
| 0.03 | 0.0120 | 0.4880 |
| ⋮ | ⋮ | ⋮ |
| 1.00 | 0.3413 | 0.1587 |
| 1.01 | 0.3438 | 0.1562 |
| 1.02 | 0.3461 | 0.1539 |
| 1.03 | 0.3485 | 0.1515 |
| ⋮ | ⋮ | ⋮ |
| 1.50 | 0.4332 | 0.0668 |
| 1.51 | 0.4345 | 0.0655 |
| 1.52 | 0.4357 | 0.0643 |
| 1.53 | 0.4370 | 0.0630 |
| ⋮ | ⋮ | ⋮ |

Z = - 1.5

i.e. 1.5 standard deviations below the mean




Only 6.68 % of North Americans are dumber…

5-71


- Let us have an example
- I RECOMMEND FOUR BASIC STEPS:

- Canadian study on work/life balance

Among employed persons, the number of hours worked each week varies widely for many reasons. Using the 2010 data, we find that the mean number of hours worked last week was 38, with a standard deviation of 15 hours.

- What % of the population likely scored between 23 hours and the mean?

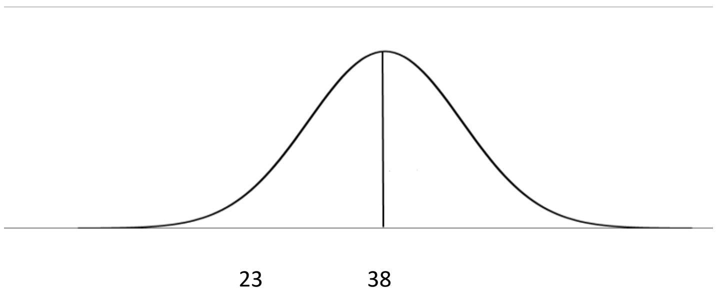
- Step 1. Get everything down: relevant formula and information provided

$$Z = \frac{X_i - \overline{X}}{s}$$

*mean is 38 and*
*s is 15*
*Xi is 23 hours*

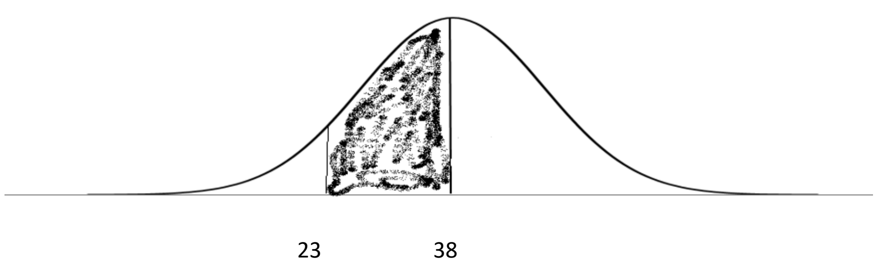
5-72

- *Step 2.* *Draw it : % between 23 and the mean 38*




5-73

- *Step 2.* *Draw it : % between 23 and the mean 38*




- *Step 3.*
- *Calculate your Z score*
- $$Z = \frac{X_i - \overline{X}}{s}$$ $= (23 – 38)/15 = -1$

5-74

# Using the Normal Curve: Appendix A *(continued)*

AN ILLUSTRATION OF HOW TO FIND AREAS UNDER THE NORMAL CURVE USING APPENDIX A

| (a) Z | (b) Area between Mean and Z | (c) Area beyond Z |
|---|---|---|
| 0.00 | 0.0000 | 0.5000 |
| 0.01 | 0.0040 | 0.4960 |
| 0.02 | 0.0080 | 0.4920 |
| 0.03 | 0.0120 | 0.4880 |
| ⋮ | ⋮ | ⋮ |
| 1.00 | 0.3413 | 0.1587 |
| 1.01 | 0.3438 | 0.1562 |
| 1.02 | 0.3461 | 0.1539 |
| 1.03 | 0.3485 | 0.1515 |
| ⋮ | ⋮ | ⋮ |
| 1.50 | 0.4332 | 0.0668 |
| 1.51 | 0.4345 | 0.0655 |
| 1.52 | 0.4357 | 0.0643 |
| 1.53 | 0.4370 | 0.0630 |
| ⋮ | ⋮ | ⋮ |

5-75

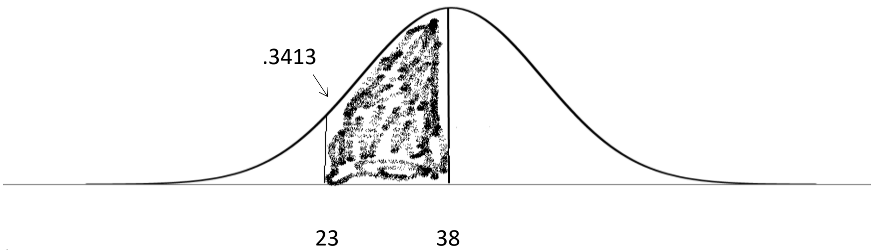
# Using the Normal Curve: Appendix A *(continued)*

- ( c) = areas beyond the *Z* score

AN ILLUSTRATION OF HOW TO FIND AREAS UNDER THE NORMAL CURVE USING APPENDIX A

| (a) Z | (b) Area between Mean and Z | (c) Area beyond Z |
|---|---|---|
| 0.00 | 0.0000 | 0.5000 |
| 0.01 | 0.0040 | 0.4960 |
| 0.02 | 0.0080 | 0.4920 |
| 0.03 | 0.0120 | 0.4880 |
| ⋮ | ⋮ | ⋮ |
| 1.00 | 0.3413 | 0.1587 |
| 1.01 | 0.3438 | 0.1562 |
| 1.02 | 0.3461 | 0.1539 |
| 1.03 | 0.3485 | 0.1515 |
| ⋮ | ⋮ | ⋮ |
| 1.50 | 0.4332 | 0.0668 |
| 1.51 | 0.4345 | 0.0655 |
| 1.52 | 0.4357 | 0.0643 |
| 1.53 | 0.4370 | 0.0630 |
| ⋮ | ⋮ | ⋮ |

5-76




- *Step 4.*
- *Interpret:*
- *% of population between 23 and 38 in terms of hours worked?*
- *.3413 is the proportion, and the % is: 34.13%*

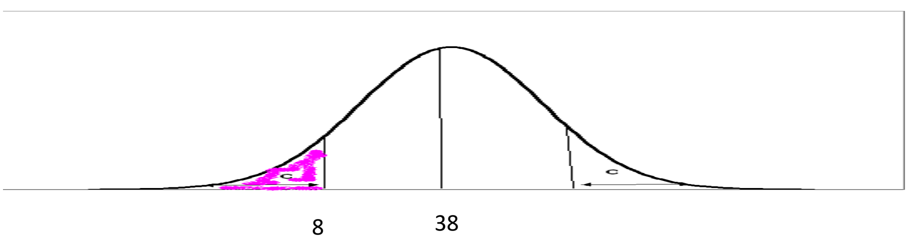
5-77

Another example: What if you were interested in documenting the percentage who worked fewer than 8 hours?

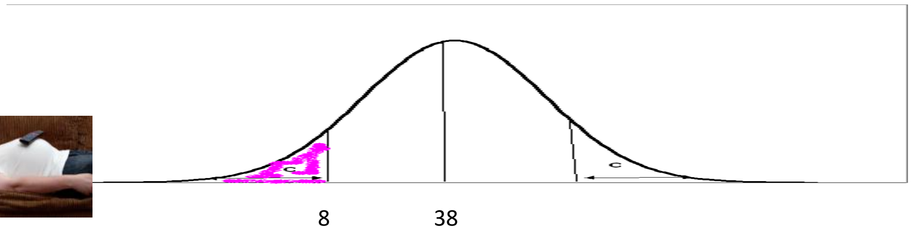
Another example: What if you were interested in documenting the percentage who worked fewer than 8 hours?

- Step 1. Get everything down: relevant formula and information provided
-
- *mean is 38 hrs and*
  *s is 15*
  *Xi is 8*

- *Step 2. Draw it : % worked fewer than 8 hours*




- *Step 2.   Draw it : % worked fewer than 8 hours*




- *Step 3.*
- *Calculate your Z score*

- $$Z = \frac{X_i - \overline{X}}{s}$$   = *(8 – 38)/15 = -2*



# Using the Normal Curve: Appendix A *(continued)*

- ( c) = areas beyond the *Z* score

AN ILLUSTRATION OF HOW TO FIND AREAS UNDER THE NORMAL CURVE USING APPENDIX A

| (a) *Z* | (b) Area between Mean and *Z* | (c) Area beyond *Z* |
|---|---|---|
| 0.00 | 0.0000 | 0.5000 |
| 0.01 | 0.0040 | 0.4960 |
| 0.02 | 0.0080 | 0.4920 |
| 0.03 | 0.0120 | 0.4880 |
| ⋮ | ⋮ | ⋮ |
| 1.00 | 0.3413 | 0.1587 |
| 1.01 | 0.3438 | 0.1562 |
| 1.02 | 0.3461 | 0.1539 |
| 1.03 | 0.3485 | 0.1515 |
| ⋮ | ⋮ | ⋮ |
| 1.50 | 0.4332 | 0.0668 |
| 1.51 | 0.4345 | 0.0655 |
| 1.52 | 0.4357 | 0.0643 |
| 1.53 | 0.4370 | 0.0630 |
| ⋮ | ⋮ | ⋮ |
| 2.00 | 0.4772 | 0.0228 |

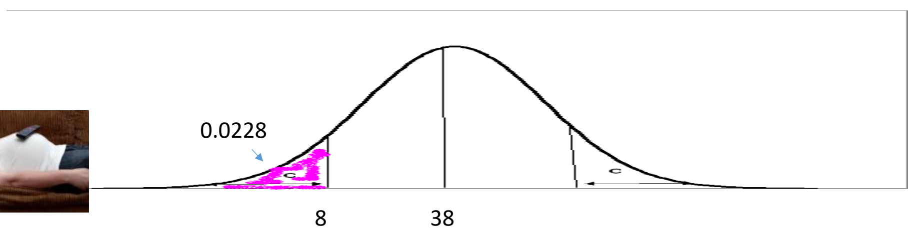
# Using the Normal Curve: Appendix A *(continued)*

- ( c) = areas beyond the *Z* score

AN ILLUSTRATION OF HOW TO FIND AREAS UNDER THE NORMAL CURVE USING APPENDIX A

| (a) *Z* | (b) Area between Mean and *Z* | (c) Area beyond *Z* |
|---|---|---|
| 0.00 | 0.0000 | 0.5000 |
| 0.01 | 0.0040 | 0.4960 |
| 0.02 | 0.0080 | 0.4920 |
| 0.03 | 0.0120 | 0.4880 |
| ⋮ | ⋮ | ⋮ |
| 1.00 | 0.3413 | 0.1587 |
| 1.01 | 0.3438 | 0.1562 |
| 1.02 | 0.3461 | 0.1539 |
| 1.03 | 0.3485 | 0.1515 |
| ⋮ | ⋮ | ⋮ |
| 1.50 | 0.4332 | 0.0668 |
| 1.51 | 0.4345 | 0.0655 |
| 1.52 | 0.4357 | 0.0643 |
| 1.53 | 0.4370 | 0.0630 |
| ⋮ | ⋮ | ⋮ |
| 2.00 | 0.4772 | 0.0228 |






- *Step 4.*
- *Interpret:*
- *% of population that worked 8 or fewer hours?*
- *0.0228 or 2.28%*

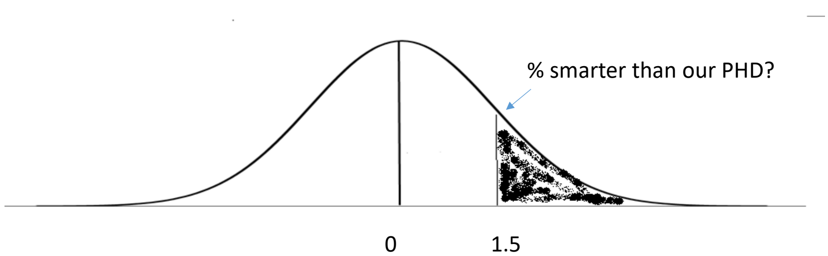
**We an use Appendix A to Describe Areas Under the Normal Curve**

-> area between a *Z* score and the mean.
(Section 4.3: I just gave you an example)
-> area either above or below a *Z* score (4.4)
-> area between two *Z* scores (4.5)
-> probability of randomly selected score (4.6)

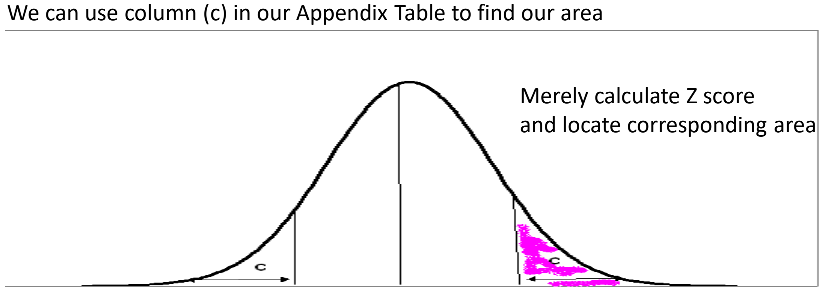
5-85

**Area above a given value?**

With a positive Z value (score above the mean)



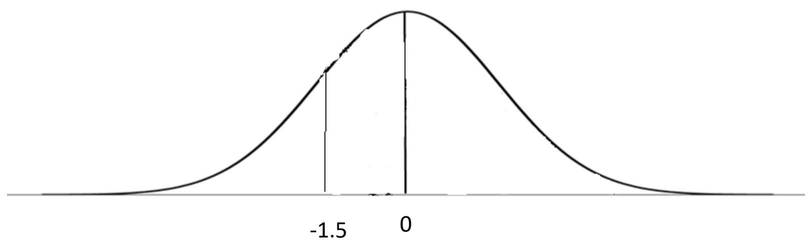

We can use column (c) in our Appendix Table to find our area




5-86

## Area above

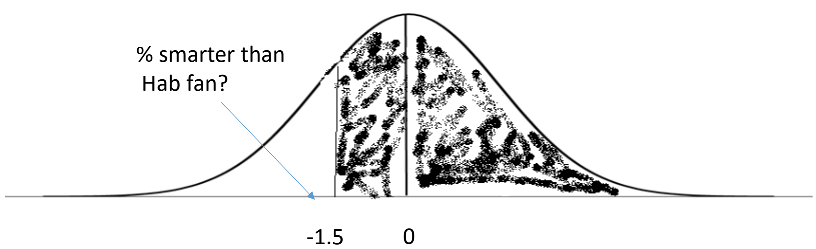
What about “above” a negative Z value (score below the mean)



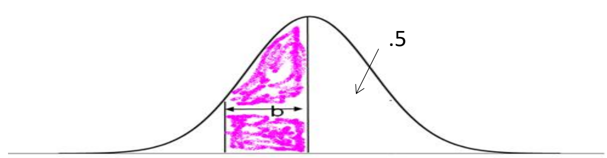



## Area above

What about “above” a negative Z value (score below the mean)

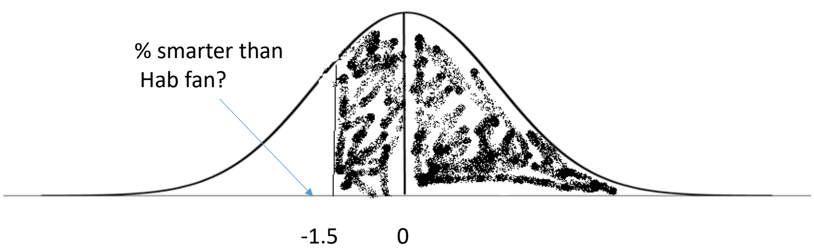



We can use column (b) in our Appendix Table to find our area **but:**

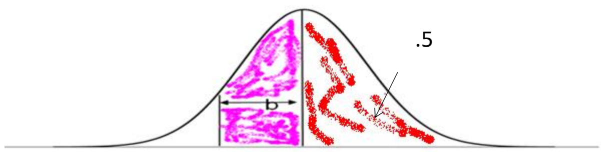


## Area above

What about "above" a negative Z value (score below the mean)



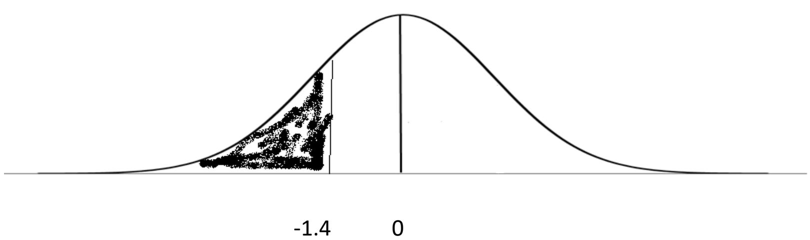

We can use column (b) in our Appendix Table to find our area **but:** then must then add .5 for the other half of the distribution



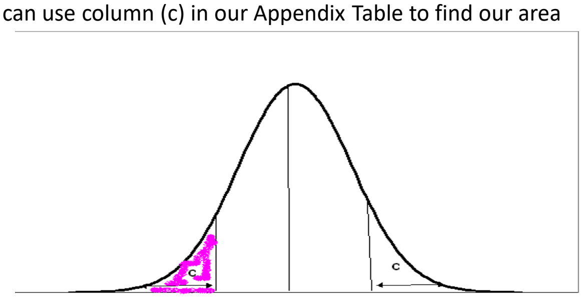



## What below area below?

With a negative Z value (score below the mean)

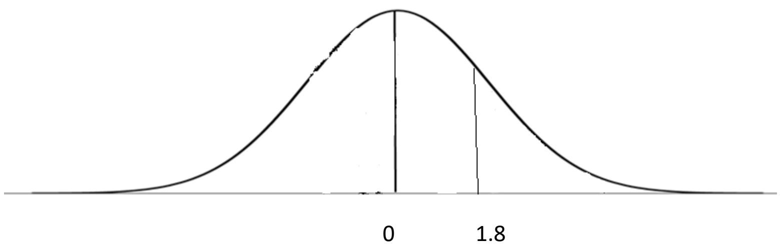



We can use column (c) in our Appendix Table to find our area



## below a given value

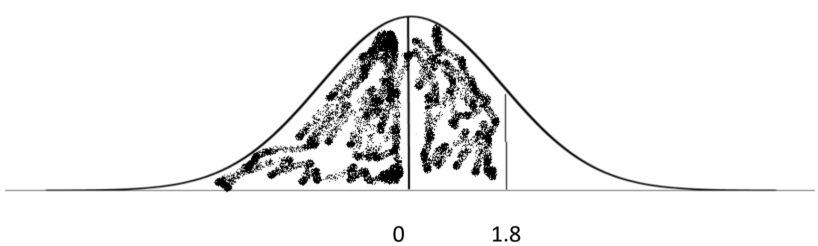
What about "below" a positive Z value?



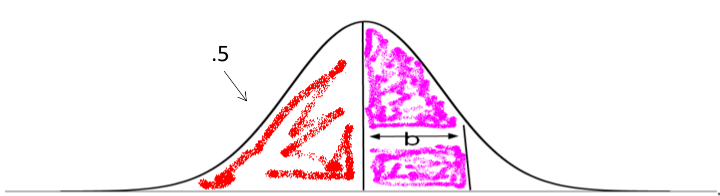



## below a given value

What about "below" a positive Z value



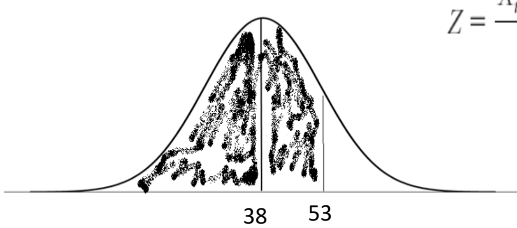

We can use column (b) in our Appendix Table to find our area and then add .5 for the other half of the distribution



### A few more examples

- The average full time Japanese worker is putting in 53 hour work weeks. What proportion of Canadian workers are putting in fewer hours (mean Canadian worker=38, standard deviation = 15)
- 1. Get everything down mean =38, s = 15 and X1= 53
- 2. Draw it:

$$Z = \frac{X_i - \bar{X}}{s}$$




- 3. The relevant score as a Z score is:

$$Z = \frac{53 - 38}{15} = +1.00$$



# Using the Normal Curve: Appendix A *(continued)*

AN ILLUSTRATION OF HOW TO FIND AREAS UNDER THE NORMAL CURVE USING APPENDIX A

| (a) $Z$ | (b) Area between Mean and $Z$ | (c) Area beyond $Z$ |
|---|---|---|
| 0.00 | 0.0000 | 0.5000 |
| 0.01 | 0.0040 | 0.4960 |
| 0.02 | 0.0080 | 0.4920 |
| 0.03 | 0.0120 | 0.4880 |
| ⋮ | ⋮ | ⋮ |
| 1.00 | 0.3413 | 0.1587 |
| 1.01 | 0.3438 | 0.1562 |
| 1.02 | 0.3461 | 0.1539 |
| 1.03 | 0.3485 | 0.1515 |
| ⋮ | ⋮ | ⋮ |
| 1.50 | 0.4332 | 0.0668 |
| 1.51 | 0.4345 | 0.0655 |
| 1.52 | 0.4357 | 0.0643 |
| 1.53 | 0.4370 | 0.0630 |
| ⋮ | ⋮ | ⋮ |

# Using the Normal Curve: Appendix A *(continued)*

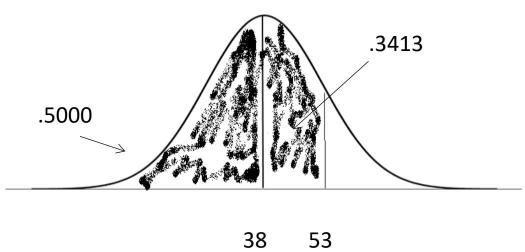
- ( c) = areas beyond the *Z* score

AN ILLUSTRATION OF HOW TO FIND AREAS UNDER THE NORMAL CURVE USING APPENDIX A

| (a) *Z* | (b) Area between Mean and *Z* | (c) Area beyond *Z* |
|---|---|---|
| 0.00 | 0.0000 | 0.5000 |
| 0.01 | 0.0040 | 0.4960 |
| 0.02 | 0.0080 | 0.4920 |
| 0.03 | 0.0120 | 0.4880 |
| ⋮ | ⋮ | ⋮ |
| 1.00 | 0.3413 | 0.1587 |
| 1.01 | 0.3438 | 0.1562 |
| 1.02 | 0.3461 | 0.1539 |
| 1.03 | 0.3485 | 0.1515 |
| ⋮ | ⋮ | ⋮ |
| 1.50 | 0.4332 | 0.0668 |
| 1.51 | 0.4345 | 0.0655 |
| 1.52 | 0.4357 | 0.0643 |
| 1.53 | 0.4370 | 0.0630 |
| ⋮ | ⋮ | ⋮ |



4. Interpret:




**Positive** *Z score*: 1.00
we consult Appendix A to find the area between the score and the mean (column b): 0.3413

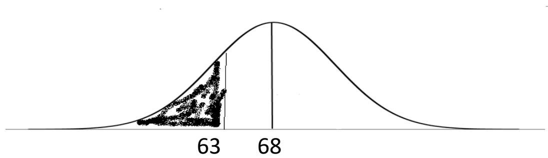
Then we add this area to the area below the mean: 0.5000
0.3413 + 0.5000 = 0.8413

Areas can be expressed as percentages: 84.13%.
The area below a *Z* score of +1.00 is 84.13%.
The average full time Japanese worker works longer hours than over 84 per cent of all Canadians

# Another example

- Average height in Canada is 68 with a standard deviation of 3, what if the height is below the mean? 63 inches? What proportion below?
- Mean = 68; s= 3; Xi = 63
- Draw it:

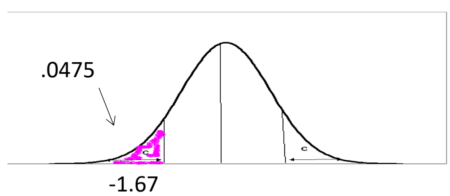
$$Z = \frac{X_i - \overline{X}}{s}$$




- 
- Find Z score: Z = (63-68)/3 = - 1.67



- The *Z* score = -1.67.
- To find the area **below** a **negative** score we use column c in Appendix A




- The area below a *Z* score of -1.67 is 0.0475
- Interpret :
- This person is taller than 4.75% of all persons.

# Summary: Finding an Area Above or Below a *Z* Score

FINDING AREAS ABOVE AND BELOW POSITIVE AND NEGATIVE *Z* SCORES

| | When the *Z* Score Is | |
|---|---|---|
| To Find Area: | Positive | Negative |
| Above *Z* | Look in column c | Add column b area to .5000 or 50.00% |
| Below *Z* | Add column b area to .5000 or 50.00% | Look in column c |

5-99