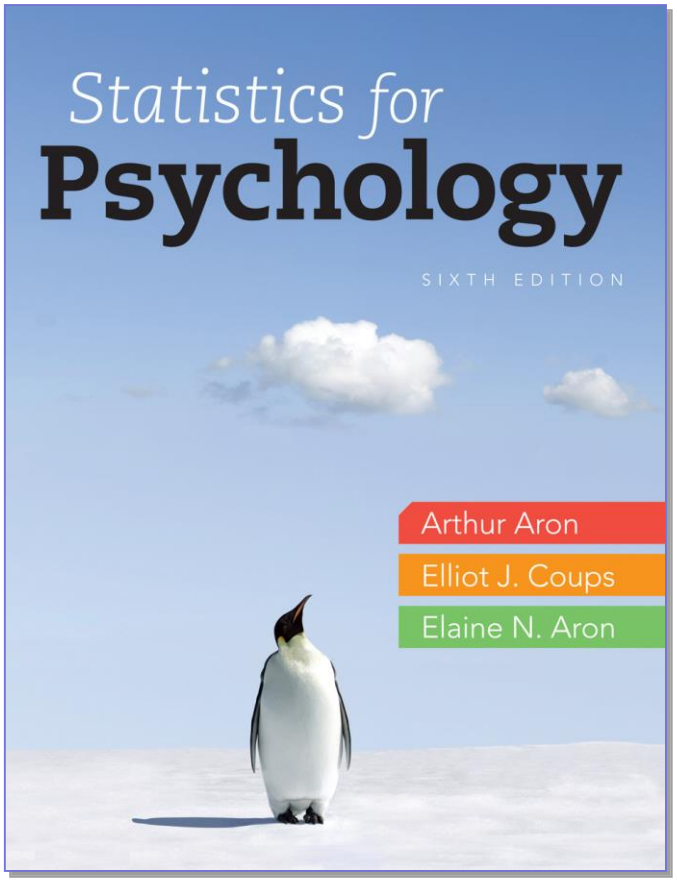
SIXTH EDITION



2

Central Tendency and Variability

Central Tendency

- Question: How might you make a group of scores more understandable?
  - Use tables and graphs (Chapter 1)
  - To use numbers to describe and summarize a group of scores

- You can describe a group of scores in terms of a representative (typical) value (eg. Average)
- A representative value gives the central tendency of a group of scores

Central Tendency

- The central tendency of a group of scores (a distribution) refers to the middle of the group of scores
- Central tendency: most "typical" or common score
- Three measures of central tendency:
  - Mean (most commonly used measure of central tendency)
  - Mode
  - Median

Mean

- Usually the best measure of central tendency is the ordinary average (mean)
- Mean is the sum of all the scores divided by the number of scores
- Symbol: M
- X means score
- N means number of scores

$$M = \frac{\sum X}{N}$$

Mean

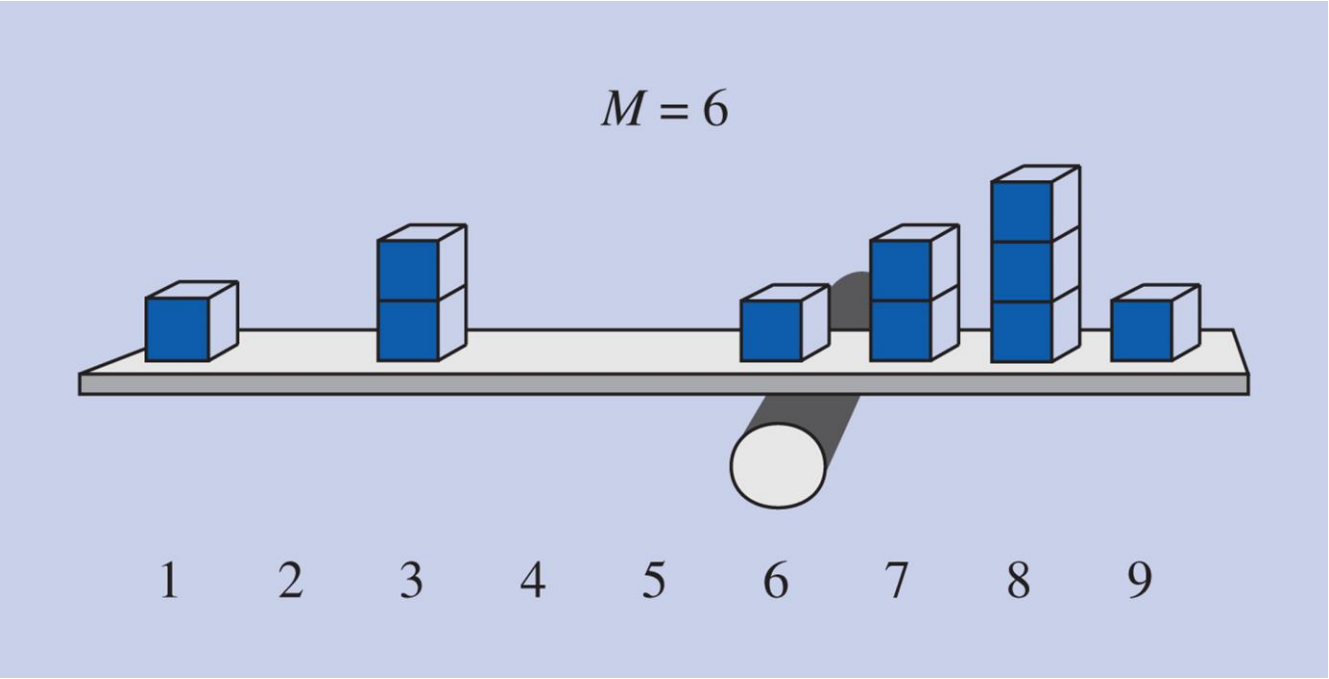
- An example:
- Let's calculate the mean of the distribution of the number of dreams during a week for 10 students
- 7,8,8,7,3,1,6,9,3,8
  - $\Sigma X = 7+8+8+7+3+1+6+9+3+8 = 60$
  - $N = 10$
  - Mean = 60/10 = 6

Mean

- Mathematically, you can think of the mean as **the point at which the total distance to all the scores above that point equals the total distance to all the scores below that point**
- Eg. 7,8,8,7,3,1,6,9,3,8

  M=6
- Scores above 6: 7,7,8,8,8,9
  - Difference 1+1+2+2+2+3= 11
- Scores below 6: 1,3,3,6,
  - Difference: 5+3+3+0 = 11

Figure 2-1 Mean of the distribution of the number of dreams during a week for 10 students, illustrated using blocks on a board balanced on a log.



The mean is the point on the board where the weight of the block on one side balances exactly w/ the weight on the other side

Mean

- Mean does not have to be a score actually in the distribution
- The mean can be a decimal number
  - Eg. the mean number of children per house is 2.50
- Lets calculate the mean of stress ratings of 30 students: 8,7,4,10,8,6,8,9,9,7,3,7,6,5,0,9,10,7,7,3,6,7,5,2,1, 6,7,10,8,8
- M = 6.43 which is clearly higher than the middle of the 0-10 scale
- On the histogram, M is between 6 and 7 scores on the horizontal line

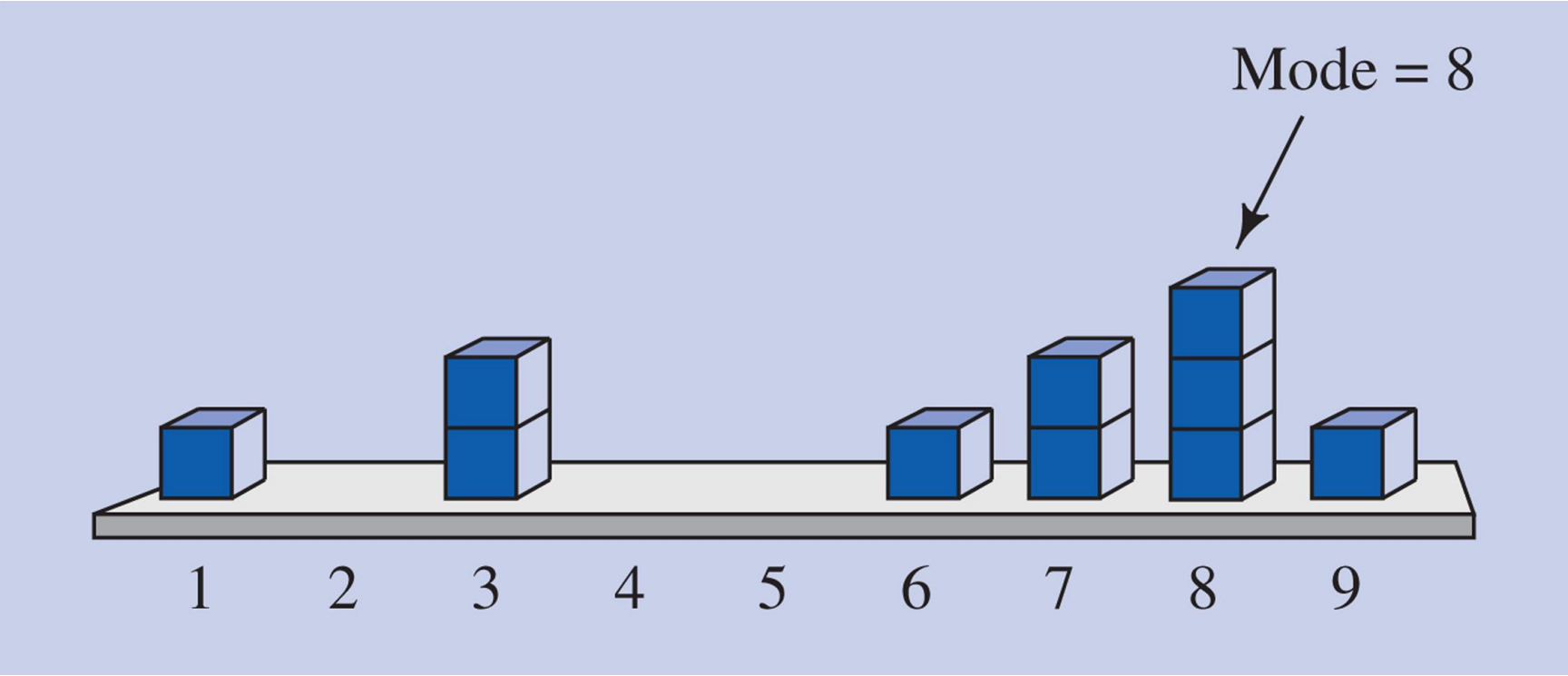
Mode

- Mode is another measure of central tendency
- Most common single value in a distribution
- Most frequently occurring number in a distribution
- A particular value w/ the largest frequency in a frequency table
- Eg. Lets calculate the mode of the below dataset 7,8,8,7,3,1,6,9,3,8
- Mode is 8 because its frequency is 3

Mode

- In a perfectly unimodal distribution, the mode is the same as the mean
- **What if mean and mode are not the same?**
  - **If so, the mode is usually not a very good way of describing the central tendency**
- Mode is the usual way of describing the central tendency for **nominal variables**
- Using the example of the number of dreams during a week for 10 students, the mode is the high point in the distribution's histogram

Figure 2-5 Mode as the high point in a distribution's histogram, using the example of the number of dreams during a week for 10 students.



Median

- Another alternative to the mean is median
- *If you line up the scores from lowest to highest*, the middle score is the median
- Median: the value at which 1/2 of the *ordered* scores fall above and 1/2 of the scores fall below

1 2 3 4 5

Median = 3

1 2 3 4

Median = 2.5

Median

- **The middle score when all scores are arranged from lowest to highest**
- Eg. Lets calculate the median of the below dataset
- 7,8,8,7,3,1,6,9,3,8
  - 1 3 3 6 **7 7** 8 8 8 9
  - The median is the average (mean) of the 5^{th} and 6^{th} scores, so the median is 7

Comparing the mean, mode and median

- **Sometimes, the median is better than mean (or mode) as a representative value for a group of scores**
- **This happens when a few extreme scores would strongly affect the mean but would not affect the median**
- Eg. Reaction times .74, .86, 2.32, .79 and .81
- Mean : 1.1040 (2.32 score affected the mean very much)
- Median : .81 (more representative for the dataset)
- **Using the median deemphasizes the one extreme score**

- The extreme score/case is called outlier
- Outlier can be extreme *either for the + side or the - side*

Measures of Central Tendency

- In a normal distribution, the mean, median and mode coincide
- The median is the usual way of describing the central tendency for a **rank-order variable**
- In skewed distributions, the mean is "pulled" toward the tail of the distribution
  - In a positively skewed distribution, the median is higher than the mean
  - In a negatively skewed distribution, the median is lower than the mean

Measures of Central Tendency

- In a positively skewed distribution, the median is lower than the mean
- In a negatively skewed distribution, the median is higher than the mean

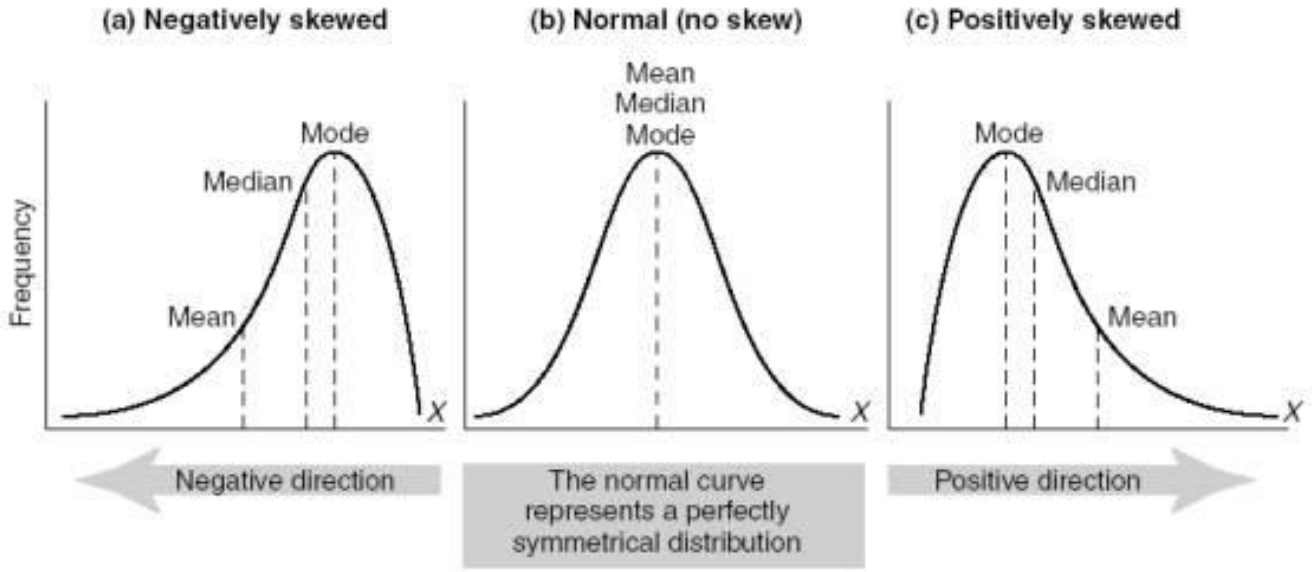


FIGURE 15.6 Examples of normal and skewed distributions

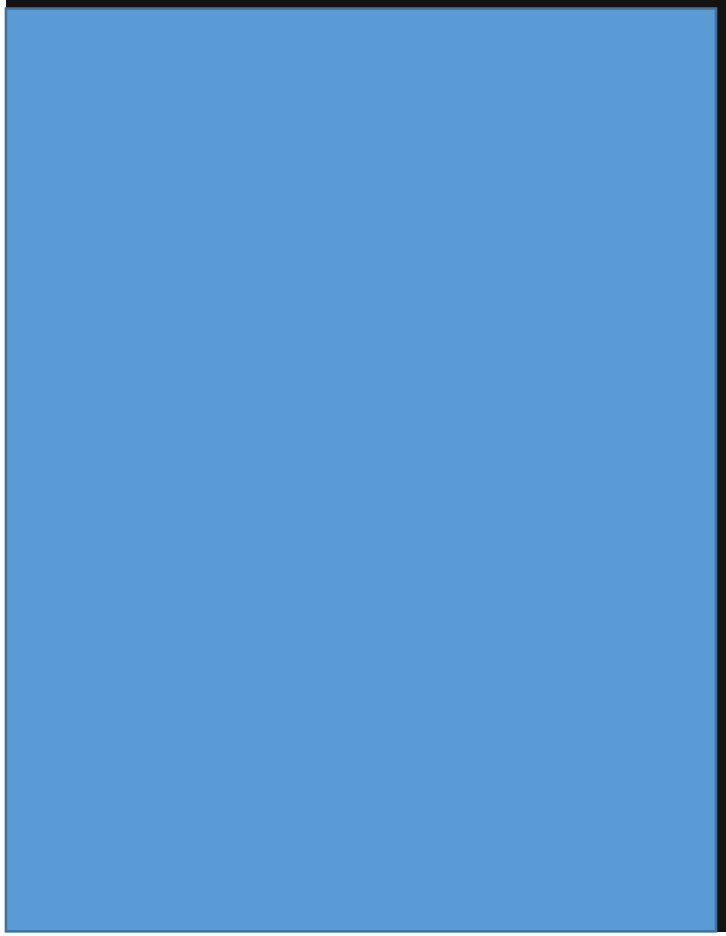
Advantages of the Mean

- The mean is the most stable measure of central tendency because all the scores in a distribution are included in its calculation (not true for mode or median)
  - Not as affected by addition or deletion of scores as mode and median
- The mean is used in many statistical procedures

The Mode

- It does not make sense to take the average in nominal data.
  - Gender: 67 males --- 1 → **mode**
    50 females ---- 2

14	14	13	15	11	15
13	10	12	13	14	13
14	15	17	14	14	15

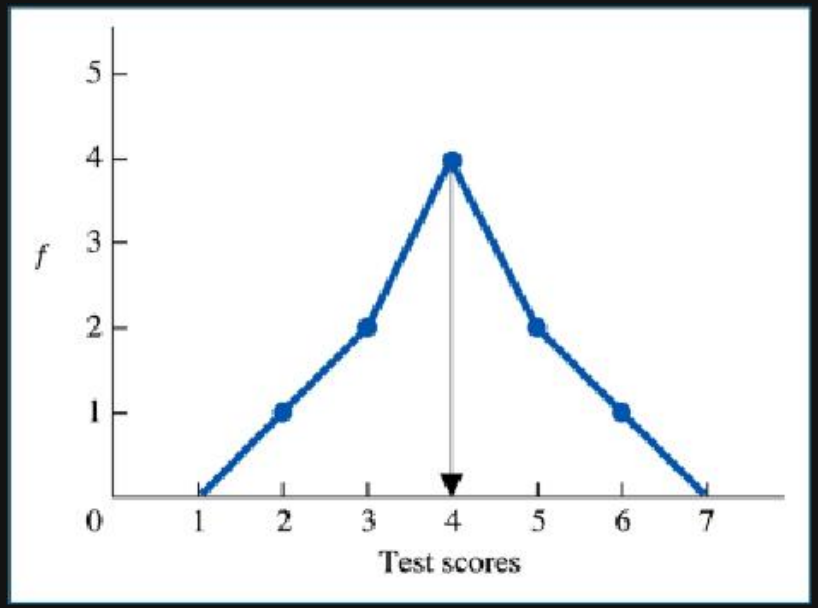


What is the mode?

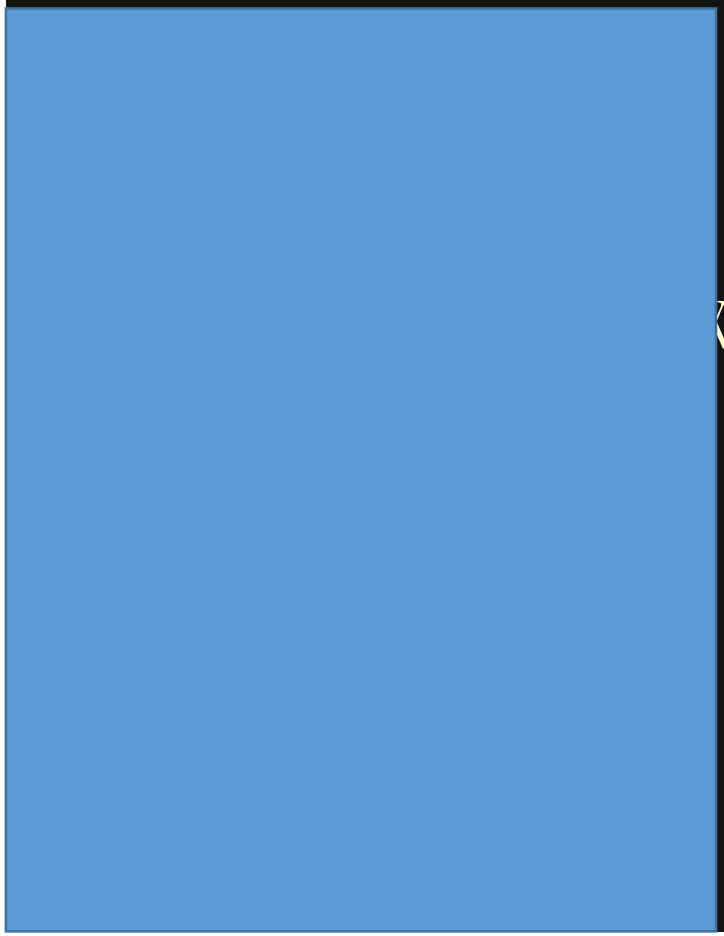
N=18

Unimodal Distributions

When a polygon has one hump (such as on the normal curve) the distribution is called **unimodal.**



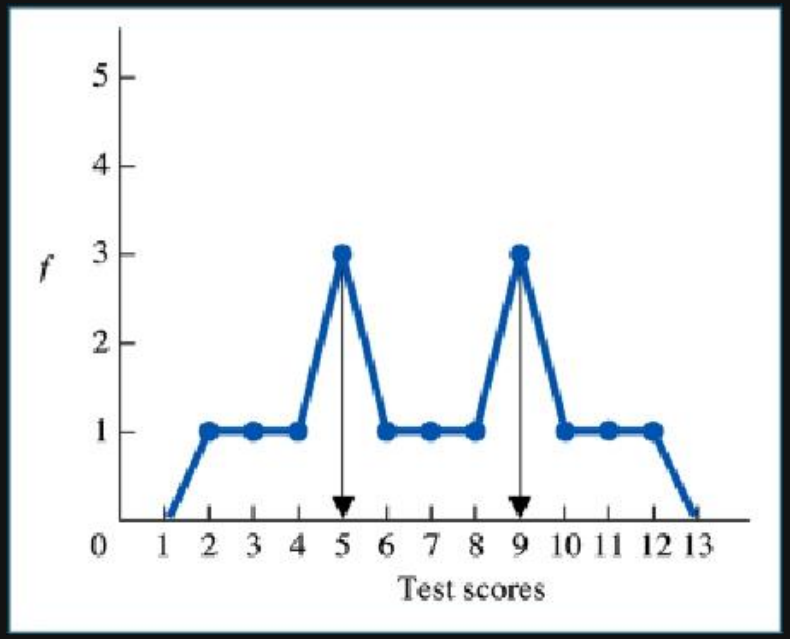
14	14	13	15	11	12
15	10	12	13	12	13
15	15	17	12	15	12



What is the mode?

Bimodal Distributions

When a distribution has two scores that are most frequently occurring, it is called **bimodal**.



Uses of The Mode

- In nominal data
  - Since we cannot use mean or median
- Also in ordinal, interval or ratio data, along with mean and median

Problems with The Mode

- Gives us limited information about a distribution
  - Might be misleading
  - EXP: 7 7 7 20 20 21 22 22 23 24
    - What is the mode here?

The Median (Mdn)

- The score at the 50th percentile, (in the middle)
- Used to summarize ordinal or highly skewed interval or ratio scores.

Determining the Median

- When data are normally distributed, the median is the same score as the mode.
- When data are not normally distributed, follow the following procedure:
  - Arrange the scores from highest to the lowest.
  - If there are an odd number of scores, the median is the score in the middle position.
  - If there are an even number of scores, the median is the average of the two scores in the middle.

The Median (Mdn)

- A better measure of central tendency than mode
  - Only one score can be the median
  - It will always be around where the most scores are.
- EXP: 1 2 3 3 4 7 9 10 11
- EXP: 1 2 3 3 4 6 7 9 10 11

14	14	13	15	11	15
13	10	12	13	14	13
14	15	17	14	14	15

What is the median?

17
15
15
15
15
14
14
14
14
14
13
13
13
13
12
11
10

N=18

The Mean

- The score located at the mathematical center of a distribution
- Used to summarize interval or ratio data in situations when the distribution is symmetrical and unimodal

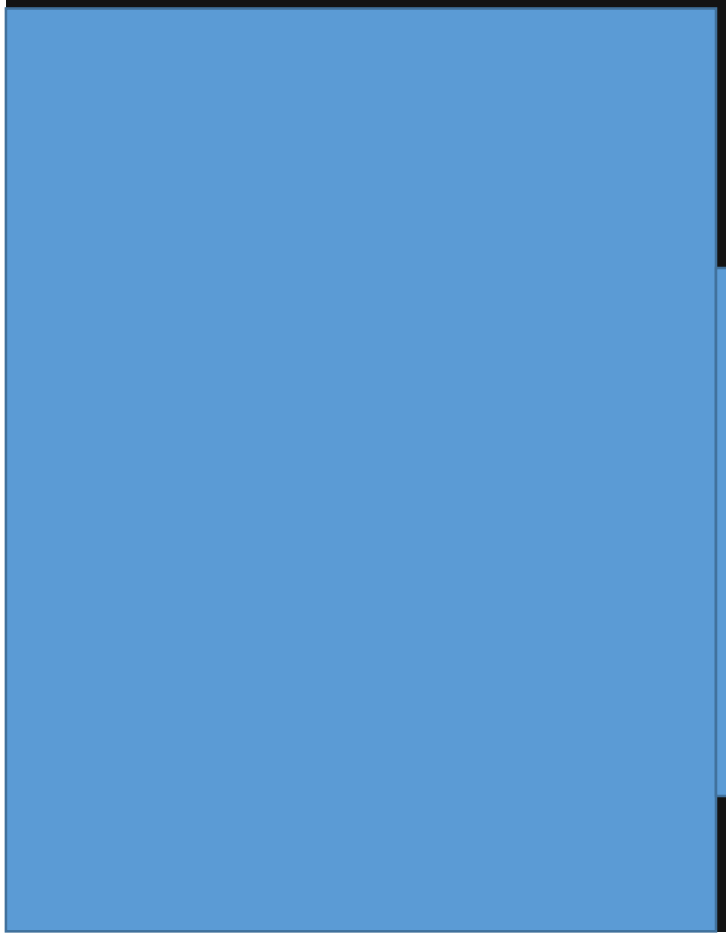
Determining the Mean

- The formula for the sample mean is

$$\overline{X} = \frac{\Sigma X}{N}$$

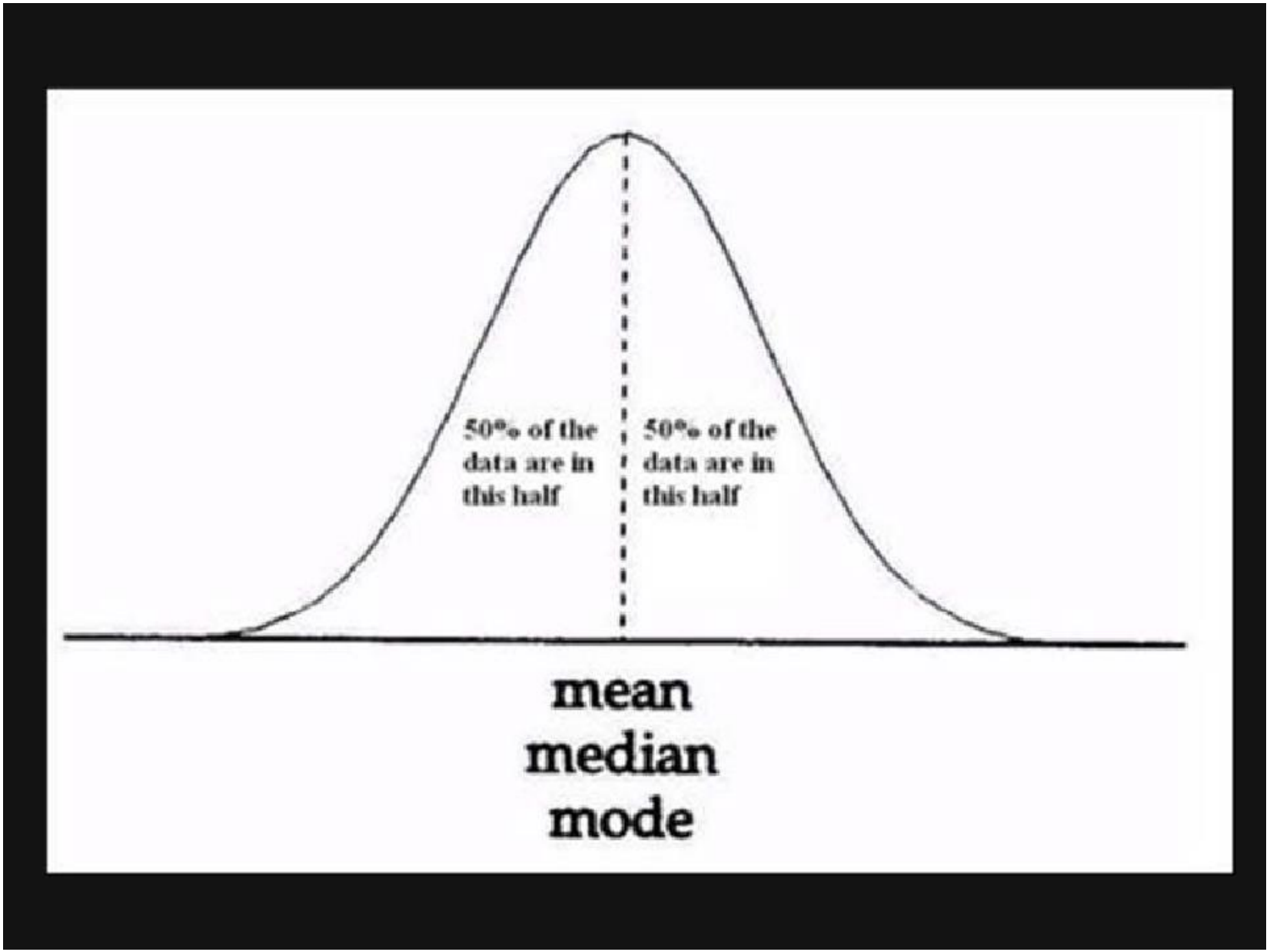
14	14	13	15	11	15
13	10	12	13	14	13
14	15	17	14	14	15

What is the mean?



Central Tendency and Normal Distributions

On a perfect normal distribution all three measures of central tendency are located at the same score.



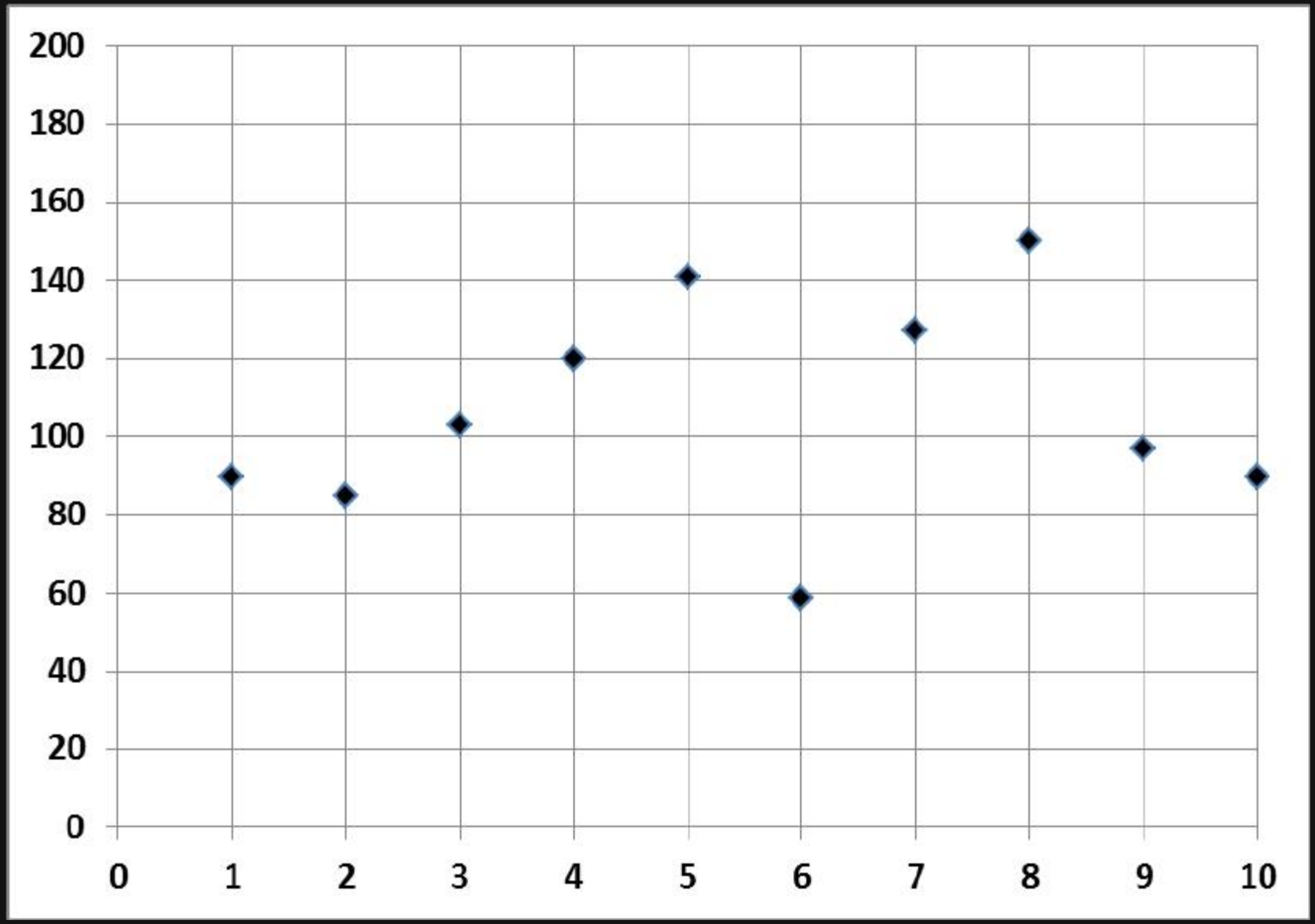
Central Tendency

- Measures of Central Tendency:
  - **Mean**
    - The sum of all scores divided by the number of scores.
  - **Median**
    - The score in the middle when the scores are ordered.
  - **Mode**
    - The most frequent score.

Measurement Scale	Measures you CANuse	BestMeasureofthe "Middle"
Nominal	Mode	Mode
Ordinal	Mode Median	Median
Interval	Mode Median Mean	Symmetricaldata:Mean Skeweddata:Median
Ratio	Mode Median Mean	Symmetricaldata:Mean Skeweddata:Median

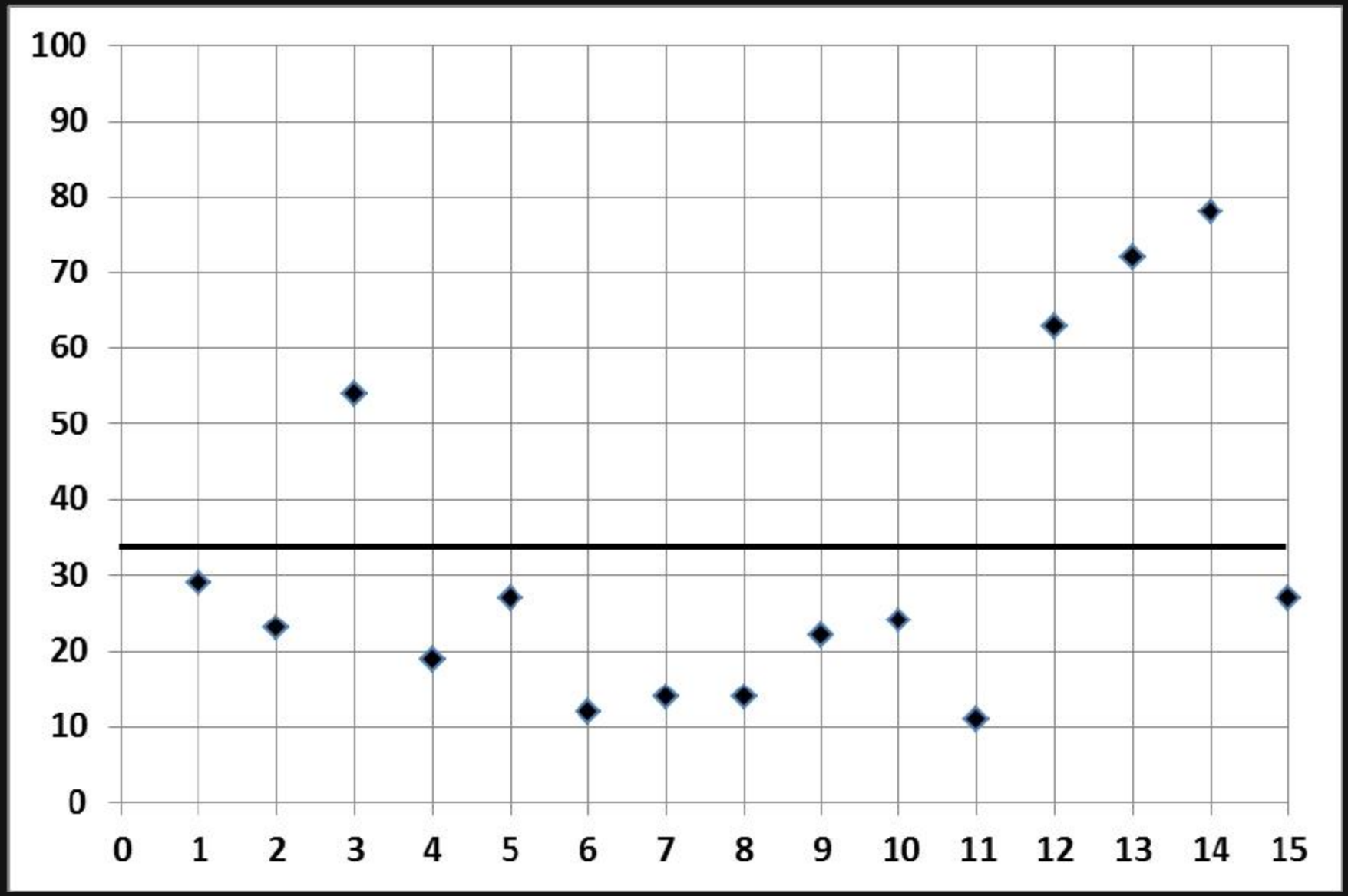
2.1

- The following data are representing intelligence in IQ points.
- 90, 85, 103, 120, 141, 59, 127, 150, 97, 90
- 1) Calculate mean, mode, median of IQ scores. 2) What kind of distribution of is this? 3) Draw scatter plot of this data below.


200
180
160
140
120
100
80
60
40
20
0
0
1
2
3
4
5
6
7
8
9
10

2.2

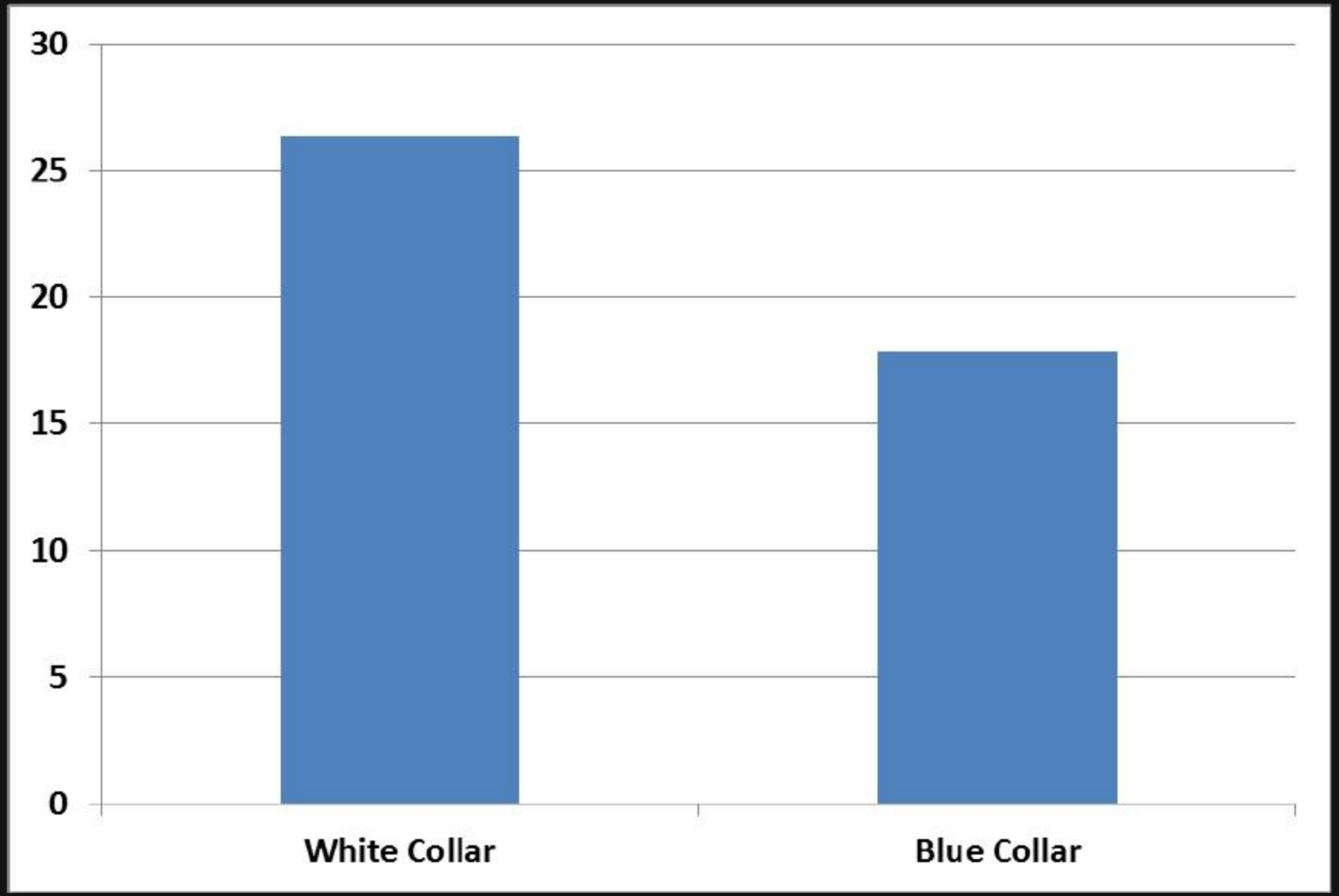
- The following data are representing anxiety score on psychological test shortly before exam.
- 29, 23, 54, 19, 27, 12, 14, 14, 22, 24, 11, 63, 72, 78, 27
- 1) Calculate mean, mode, median of anxiety scores. 2) What kind of distribution of is this? 3) Draw scatter plot of this data below. 4) Draw mean as a line on this scatter plot. 5) Calculate deviation from mean for each score 5) Draw deviation from mean as a perpendicular line between each score and the mean line on the graph.


100
90
80
70
60
50
40
30
20
10
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

2.3

- The following data are representing happiness scores on psychological test for white collar and blue collar workers.
- 1) Calculate mean, mode, median of anxiety scores. 2) What kind of distribution of is this? **3) Draw bar plot of this data using mean happiness scores of white and blue collar workers.**

Sno	WhiteCollar	Sno	BlueCollar
1	19	11	18
2	15	12	20
3	18	13	10
4	50	14	32
5	36	15	19
6	50	16	15
7	10	17	18
8	32	18	5
9	19	19	10
10	15	20	32


30
25
20
15
10
5
0
White Collar
Blue Collar

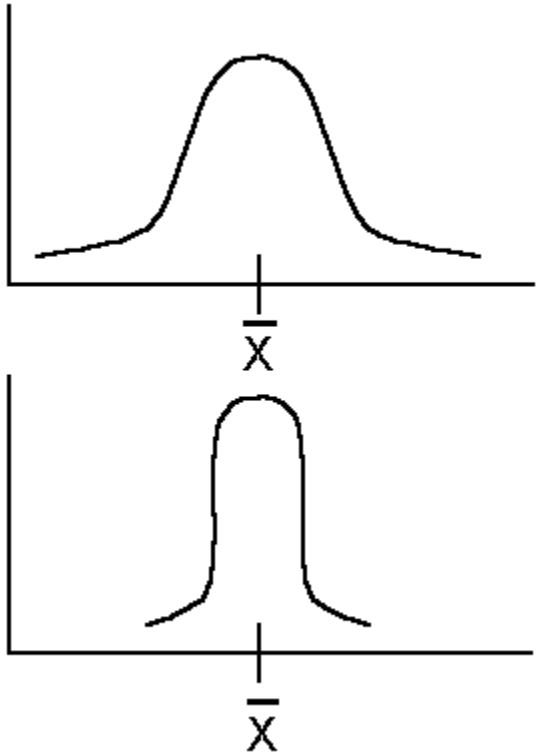
VARIABILITY:
Range
Variance
Standard Deviation

Variability

- Researchers also want to learn how spread out the scores are in a distribution
  - The amount of the variability in a distribution
- Eg. Mean age of the class is 29
  - Every student could be 29 yrs old.
    - variability: none
  - ½ of sts could be 19; the other ½ could be 39 yrs old
    - Variability: low but exists
- **Variability: the amount of the spread of the scores around the mean**
- If scores are mostly quite close to the mean; the distribution has less variability than if scores are further from the mean

Variability

- Distributions w/ the same mean can have very different amounts of spread around the mean
- Mean does not tell you a/ the variability of the scores
- Measures of variability: variance and standard deviation



Measures of Variability: The Variance

- The variance of a group of scores is one kind of number that tells you how spread out the scores are around the mean
- The average of each score's squared difference from the mean

Measures of Variability: The Variance

- Steps for computing the variance:
  1. Subtract the mean from each score
     - This gives each score's deviation score, which is how far away the score is from the mean
  2. Square each of these deviation scores
     - This gives each score's squared deviation score
  3. Add up the squared deviation scores
     - This total is called the sum of squared deviations
  4. Divide the sum of squared deviation by the number of scores
     - This gives the average (mean) of the squared deviations, called variance

Measures of Variability: The Variance

- The more spread-out distribution has a larger variance because being spread out makes the deviation scores bigger
- If deviation scores are bigger, squared deviation scores and the average of squared deviation scores are also bigger
- The variance is rarely used as a descriptive statistic
  - It is because variance is based on the squared deviation scores
  - It is not a easy-to-understand sense of how spread out the actual, nonsquared scores are

Measures of Variability: The Standard Deviation

- The most widely used number to describe the spread of a group of scores is the standard deviation
- The standard deviation is simply **the square root of the variance**
- Steps for computing the standard deviation:
    1. Figure the variance
    2. Take the square root

Measures of Variability: The Standard Deviation

- The variance is about the squared deviations from the mean; therefore, its square root (the standard deviation) is about direct, ordinary, not-squared deviations from the mean
- Roughly speaking; SD is the average amount that scores differ from the mean
- Eg. In a class, ages have a SD of 10
- The ages are spread out about 10 years in each direction from the mean

Formula for the Variance

- Definitional vs. computational formulas

$$SD^2 = \frac{\sum(X - M)^2}{N}$$

General symbol for variance

Sum of squared deviations **SS**

Formula for the Standard Deviation

$$SD = \sqrt{SD^2} = \sqrt{\frac{\Sigma(X - M)^2}{N}} = \sqrt{\frac{SS}{N}}$$

Example

Calculate the variance of the below data set

2,3,3,4

Deviation

Squared deviation

X	X-M	$(X-M)^2$
2	-1	1
3	0	0
3	0	0
4	1	1
M = 3		Total = 2

$SD^2 = 2/4 = 0.5$

Sum of squared deviations

Computational Formula

- In actual research situations, psychologists must often figure the variance and SD for distributions w/ many scores, often involving decimals or large numbers
- In days before computers, this could make the who process time consuming (even w/ a calculator)
- To deal w/ this problem, researchers developed various shortcuts to simplify the figuring
- The traditional computational formula for the variance

$$= \frac{\left(\sum_{i=1}^{N} X_i^2\right) - \frac{\left(\sum_{i=1}^{N} X_i\right)^2}{N}}{N}. \qquad (8)$$

Example 1

- Using the following data set, find
  - The range,
  - The sample variance and standard deviation,
  - The estimated population variance and standard deviation

14	14	13	15	11	15
13	10	12	13	14	13
14	15	17	14	14	15

Example1

- The range is the largest value minus the smallest value.

$$17 - 10 = 7$$

Example 1

$$S_X^2 = \frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N}$$

$$S_X^2 = \frac{3406 - \frac{(246)^2}{18}}{18} = \frac{3406 - 3362}{18} = 2.44$$

Example 1

$$S_X = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N}}$$

$$S_X = \sqrt{\frac{3406 - \frac{246_2}{18}}{18}} = \sqrt{2.44} = 1.56$$

Example 1

$$s_X^2 = \frac{\Sigma X_2 - \frac{(\Sigma X)_2}{N}}{N-1}$$

$$s_X^2 = \frac{3406 - \frac{(246)_2}{18}}{17} = \frac{3406 - 3362}{17} = 2.59$$

Example 1

$$s_X = \sqrt{\frac{\Sigma X_2 - \frac{(\Sigma X)_2}{N}}{N-1}}$$

$$s_X = \sqrt{\frac{3406 - \frac{246_2}{18}}{17}} = \sqrt{2.59} = 1.61$$

Example 2

- For the following sample data, compute the range, variance and standard deviation

  8 8 10 7 9 6 11 9 10 7

  11 11 7 9 11 10 11 8 10 7

Example 3

- For the data set below, calculate the mean, deviation, sum of squares, variance and standard deviation by creating a table.
- 15 12 13 15 16 17 13 16 11 18

Example 3 Solution

score	mean	deviation	sum of squares	variance	standard deviation
11	14,60	-3,60	12,96	4,64	2,15
12	14,60	-2,60	6,76		
13	14,60	-1,60	2,56		
13	14,60	-1,60	2,56		
15	14,60	0,40	0,16		
15	14,60	0,40	0,16		
16	14,60	1,40	1,96		
16	14,60	1,40	1,96		
17	14,60	2,40	5,76		
18	14,60	3,40	11,56		
			46,40		

Example 4

- For the data set below, calculate the mean, deviation, sum of squares, variance and standard deviation by creating a table.
- 1322243341

Example 4 Solution

score	mean	deviation	sum of squares	variance	standard deviation
1	2,50	-1,50	2,25		
1	2,50	-1,50	2,25		
2	2,50	-0,50	0,25		
2	2,50	-0,50	0,25		
2	2,50	-0,50	0,25		
3	2,50	0,50	0,25		
3	2,50	0,50	0,25		
3	2,50	0,50	0,25		
4	2,50	1,50	2,25		
4	2,50	1,50	2,25		
			10,50	1,05	1,02

Example 5

- For the data set below, calculate the mean, deviation, sum of squares, variance and standard deviation by creating a table.
- 1 3 30 12 15 20 5 13 2 4

Example 5 Solution

score	mean	deviation	sum of squares	variance	standard deviation
1	10,50	-9,50	90,25		
2	10,50	-8,50	72,25		
3	10,50	-7,50	56,25		
4	10,50	-6,50	42,25		
5	10,50	-5,50	30,25		
12	10,50	1,50	2,25		
13	10,50	2,50	6,25		
15	10,50	4,50	20,25		
20	10,50	9,50	90,25		
30	10,50	19,50	380,25		
			790,50	79,05	8,89

Variance as the sum of squared deviations divided by N-1

- Researchers often use a slightly different kind of variance
- Variance is the average of the sum of squared deviation; so, the sum of squared deviation scores are divided by the number of scores
- In chapter 7; you'll see that for many purposes it is better to define the variance as the sum of squared deviations divided by 1 less than the number of scores
  - For those purposes, the variance is SS/(N-1)
- The variances and SDs in research articles are usually figured using SS/(N-1)
- Calculators and computers give variance and SD automatically by using SS/(N-1) formula
- Don't worry; the things you learn in this chapter are correct ☺

Controversies and Limitations: The Tyranny of the Mean

- Behaviorism, B. F. Skinner was quite opposed to stats since he advocated that knowledge about the individual case is lost when taking averages
- 3 overeating mice: each have different learning curves for learning to press a bar for food
- If curves are merged statistically; the result would have represented no actual eating habits of any real mouse
- Clinical psychology and the study of personality also advocated in-depth study of one person instead of or as well as the averaging persons

Controversies and Limitations: The Tyranny of the Mean

- Today, the rebellion in psychology is led by qualitative research methods
- An approach that is more prominent in other behavioral and social sciences such as communication
  - Ex.: case studies, ethnography
- Carl Jung: statistical mood
  - Much like the impact of being on a jammed subway and observing the hundreds of blank faces and feeling diminished, «one of a crowd»
  - Important contributions to culture tend to come from people who feel unique and not ordinary
  - Jungian analyst Marie Louise von Franz (1979)
    - An act of loyalty is required t/wone's own feelings

The Mean and Standard Deviation in Research Articles

- Commonly reported in research articles