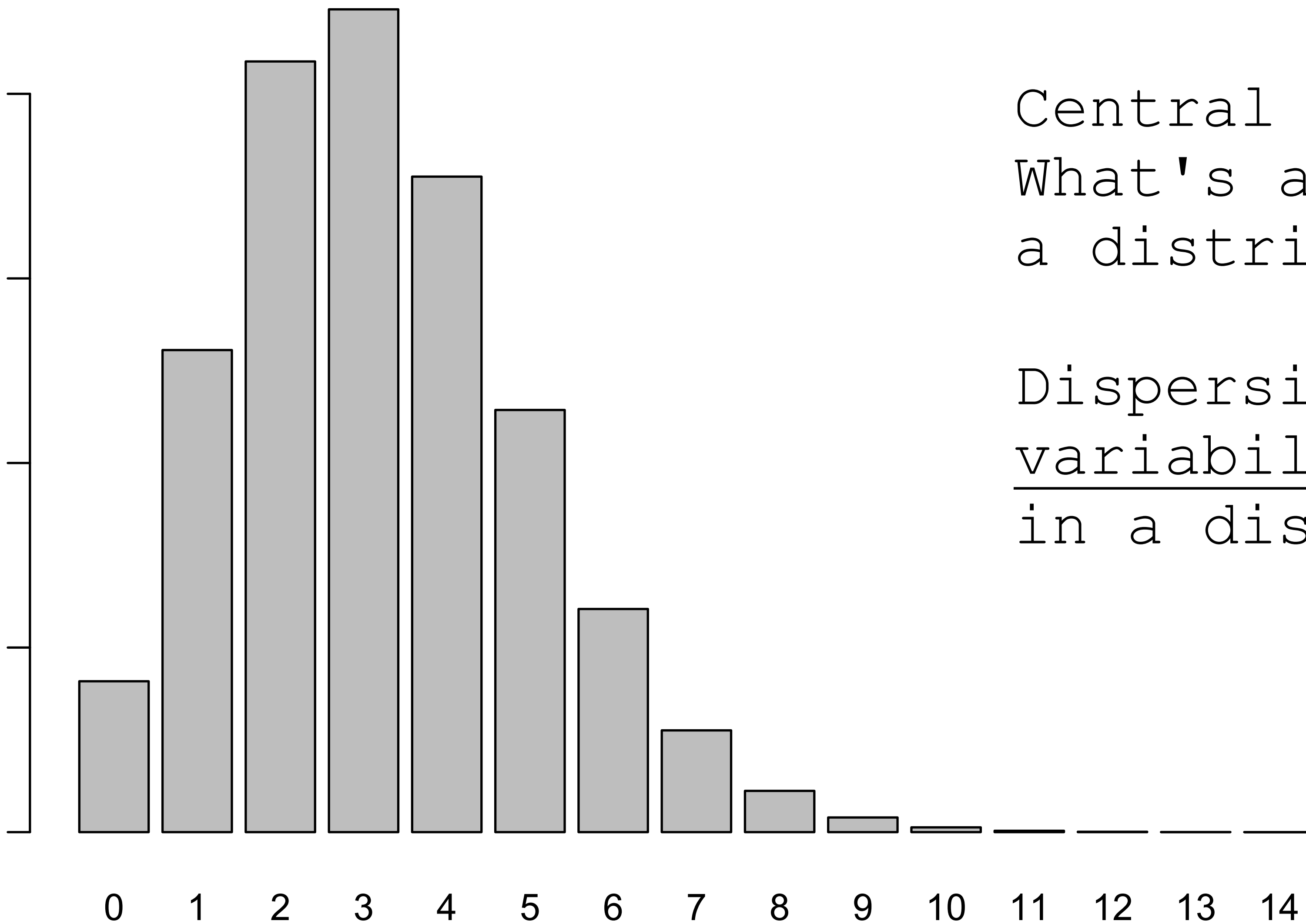
Lesson 11

Tuesday 3/5/24

Chapter 5: Measures of Dispersion



Central Tendency =
What's a Typical Value in a distribution

Dispersion = What is the variability or variation in a distribution

Example variable: people released from prison and followed for 5 years: # of rearrests during that time.

Variation Ratio (Freeman's Index)

$$\mathrm{VR} = 1 - \frac{f_m}{N}$$

A measure of dispersion used for nominal and ordinal scale (categorical or qualitative) variables; conveys the proportion of cases that are <u>not</u> in the modal category; another way to phrase this is to say it measures the variation about the mode.

Variation Ratio Example

$$\text{VR} = 1 - \frac{f_m}{N}$$

Consider a sample of people exiting prison. The distribution of "most serious" offenses for each of these people is listed below:

Value	N =	% of Total
Murder	14	1.0
Rape	8	0.6
Robbery	78	5.8
Assault	83	6.2
Burglary	247	18.3
Larceny	382	28.3
Drugs	350	25.9
Public Order	187	13.9
Total	1349	100.0

What is the Variation Ratio for this distribution?

Step 1: what is the mode? (larceny)

Step 2: What is the number of people in the modal category, f_m? (382)

Step 3: What is the total number of people in the sample? (N = 1349)

Step 4: What proportion of the sample is in the modal category? (382/1349 = 0.283)

Step 5: What is the *complement* of this proportion? (1-0.283 = 0.717)

Answer: The variation ratio for this distribution is 0.717.

Interpretation: probability that someone drawn at random from this sample is not in the modal category is 0.717.

Comparing 2 VR's

Time 1

Value	N =	% of Total
Murder	14	1.0
Rape	8	0.6
Robbery	78	5.8
Assault	83	6.2
Burglary	247	18.3
Larceny	382	28.3
Drugs	350	25.9
Public Order	187	13.9
Total	1349	100.0

VR = 1-382/1349 = 0.717

Time 2

Value	N =	% of Total
Murder	17	1.2
Rape	15	1.0
Robbery	84	5.7
Assault	78	5.3
Burglary	308	20.9
Larceny	503	34.1
Drugs	322	21.8
Public Order	150	10.2
Total	1477	100.0

VR = 1-503/1477 = 0.659

Sometimes, we would like to be able to compare the dispersion of 2 different distributions. In this case, we might conclude that the Time 1 distribution is more dispersed. This would be fine since the 2 distributions have the same number of categories.

A Problem...

Time 2

Value	N =	% of Total
Murder	17	1.2
Rape	15	1.0
Robbery	84	5.7
Assault	78	5.3
Burglary	308	20.9
Larceny	503	34.1
Drugs	322	21.8
Public Order	150	10.2
Total	1477	100.0

VR = 1-503/1477 = 0.659

Time 3

Value	N =	% of Total
Murder	9	0.7
Rape	5	0.4
Robbery	107	8.1
Assault	65	4.9
Burglary	291	22.1
Larceny	418	31.7
Drugs	301	22.8
Public Order	122	9.3
Arson	3	0.2
Total	1318	100.0

VR = 1-418/1318 = 0.683

Comparing the VR's is ambiguous in this case because different numbers of categories imply different maximum values of the VR...

Finding the Maximum VR

$$\max(\mathrm{VR}) = 1 - \frac{n/k}{n}$$

where n = the number of people in the sample and k = the number of categories in the distribution.

Time 2

$$\max(\mathrm{VR}) = 1 - \frac{1477/8}{1477} = 1 - \frac{184.625}{1477} = 0.875$$

Time 3

$$\max(\mathrm{VR}) = 1 - \frac{1318/9}{1318} = 1 - \frac{146.444}{1318} = 0.889$$

The Range (for Quantitative or Interval/Ratio Scales)

$$\text{Range}(y) = y_{\text{max}} - y_{\text{min}}$$

The range of a variable, *y*, is the maximum score minus the minimum score of the distribution.

Example

10 Largest American Cites' (FBI) Homicide Rates (per 100k population) in 2019

City	Homicide Rate
New York	3.4
Los Angeles	7.0
Chicago	18.3
Houston	11.5
Phoenix	9.6
Philadelphia	20.1
San Antonio	8.2
San Diego	2.5
Dallas	12.5
San Jose (CA)	3.1

What is the range of this distribution?

Step 1: what is the maximum value in the dataset? (y_{max} = 20.1)

Step 2: what is the minimum value in the dataset? (y_{min} = 2.5)

Step 3: calculate $y_{max} - y_{min}$ = 20.1 - 2.5 = 17.6

Solution: range = 17.6