

Panel 1

Stats: "A long, long time ago..."

Frequency distribution: # of items in categories categ.

Histogram: # of items in "bins" numeric

Mean: Avg $\frac{M}{n} = \frac{1}{n} \sum x$ numeric

Median: Middle numeric + ordinal, but not nominal

Mode: Most frequent categorical - nominal (numeric)

Which applies when?

1

Panel 2

Work sheet on Distributions and Histograms



Mean, Median, Mode:

~~1, 2, 3, 4, 5, 6, 7, 8, 9, 10~~

1, 1, 2, 3, 3, 4, 5, 7, 9, 10

$$\frac{45}{10} = 4.5 = \bar{x}$$

$$\text{median} = 3.5 \left(\frac{3+4}{2} \right)$$

$$\text{mode} = 1, 3$$

2

Panel 3

Quiz # 2

Name: _____

① Consider the following data values obtained from a random sample: 6, 4, 1, 2, 5, 6

Find

a) the mean

 $\bar{x} =$ _____

b) the median

median = _____

c) the mode(s)

mode(s): _____

3

Panel 4

② Match the measures of central tendency with the appropriate type of variable:

mean

numeric variable

median

nominal variable

mode

③ Create a histogram with 3 bins for the data 1, 5, 4, 3, 3, 6, 8, 3, 5, 10

min: 1

1-4

max: 10

4-7

bin width: $\frac{10-1}{3} = 3$

7-10

bin	count	%
1-4		
4-7		
7-10		

4

Panel 5

Median Value of a Histogram / Frequency Distribution

	count	%	cumulative %	
A		25%	25%	
B		10% 15%	35% 40%	
C		15%	50% 55%	
D		20%	70%	75%
E		20% 10%	100%	100%

Median in a Frequency Table or Histogram is the first category where cumulative percent is 50% or above.

Median category is C

Median category is still C

5

Panel 6

6

Panel 7

INCOME	Frequency	Relative Frequency	Percent
01 - Under \$1,000	22	0.018502943	1.8502944
02 - \$1,000 to \$2,999	39	0.032800674	3.2800672
03 - \$3,000 to \$3,999	27	0.022708159	2.2708158
04 - \$4,000 to \$4,999	13	0.0109335575	1.0933558
05 - \$5,000 to \$5,999	19	0.015979815	1.5979815
06 - \$6,000 to \$6,999	16	0.013456686	1.3456687
07 - \$7,000 to \$7,999	19	0.015979815	1.5979815
08 - \$8,000 to \$9,999	24	0.02018503	2.018503
09 - \$10,000 to \$12,499	48	0.04037006	4.037006
10 - \$12,500 to \$14,999	63	0.0529857	5.29857
11 - \$15,000 to \$17,499	40	0.033641715	3.3641715
12 - \$17,500 to \$19,999	42	0.035323802	3.53238
13 - \$20,000 to \$22,499	44	0.037005886	3.7005887
14 - \$22,500 to \$24,999	76	0.06391926	6.391926
15 - \$25,000 to \$29,999	70	0.058873	5.8873
16 - \$30,000 to \$34,999	94	0.07905803	7.905803
17 - \$35,000 to \$39,999	71	0.059714045	5.9714046
18 - \$40,000 to \$49,999	125	0.10513036	10.513036
19 - \$50,000 to \$59,999	85	0.07148865	7.1488647
20 - \$60,000 to \$74,999	87	0.07317073	7.3170733
21 - \$75,000 to \$89,999	61	0.051303618	5.1303616
22 - \$90,000 to \$109,999	38	0.03195963	3.195963
23 - \$110,000 to \$129,999	28	0.023549201	2.3549201
24 - \$130,000 to \$149,999	6	0.0050462573	0.50462574
25 - \$150,000 or over	32	0.026913373	2.6913373

cum. of
1.8
4
6.27
7.28
...

mean does not make much sense here.

first above 50% is the median category

FW

Panel 8

Shapes of Distributions

One shape shows up particularly often: it is bell-shaped distribution or normal distribution

normal overlay: mean = 47.708397, std. dev. = 17.35084

mean + median

Other distributions are skewed:

skewed left:
has small outliers
⇒ mean < median

skewed to right
has large outliers
⇒ mean > median

Panel 9

Mean vs. Median (and Mode)

Mean, median, mode all apply to unsorted vars.

Ex: 10,000, 20,000, 30,000 → income in \$, random sample of 3
 mean = 20,000 = median

mode is unclear

mean is easiest to find

median is the most robust measure of central tendency and is not impacted by extreme values (outliers)

Ex: 10,000, 20,000, 1,000,000 → mean: increased by extreme value 1,000,000
 median: 20,000

9

Panel 10

Skewed Distributions and Mean/Median



10

Panel 11

Measures of Variability (Dispersion)

variance, standard deviation, etc

next