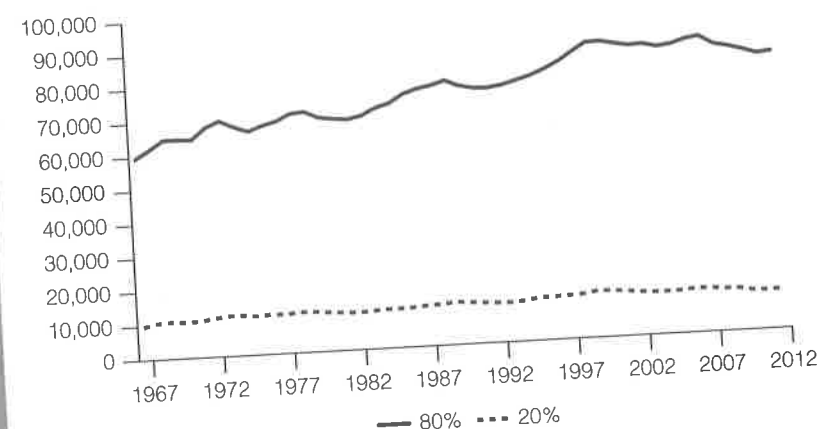


STATISTICS IN EVERYDAY LIFE (continued)

Mean Incomes for Lowest Fifth (20%) and Fourth Fifth (80%) of Households, 1967–2012 (In 2012 Dollars)



If we subtract the average income for the lowest fifth or 20% of the population from the average income for the 80% level, we have a measure of dispersion similar to the interquartile range (Q). The value of this measure increased throughout the time period. In 1967, it was about \$50,000 but the gap grew to over \$70,000 in 2012. (Had we measured from the 95% mark instead of the 80% mark, the gap would have grown from almost \$160,000 to a little over \$300,000.)

Thus, American incomes grew more variable and unequal over the time period. The increase in dispersion is solely a result of the rising incomes associated with the more affluent, not with any changes in the incomes associated with the first quintile (20%).

We saw in Chapter 3, that the income of the average American household (median) and the average income for all American households (mean) increased over this time period. Now we can see that the increase was not shared equally. It was virtually nonexistent for the less affluent (the first quintile or 20%) and quite robust for those with higher incomes. Thus, dispersion increased: Households with modest incomes continued to have modest incomes, and, consistent with the ancient folk wisdom, the rich got richer.

SUMMARY

- Measures of dispersion summarize information about the heterogeneity, or variety, in a distribution of scores. When combined with a measure of central tendency, these statistics convey a large volume of information in just a few numbers. While measures of central tendency locate the central points of the distribution, measures of dispersion indicate the amount of diversity in the distribution.
- The range (R) is the distance from the highest to the lowest score in the distribution. The interquartile range

(Q) is the distance from the third to the first quartile (the "range" of the middle 50% of the scores). These two ranges can be used with variables measured at either the ordinal or interval-ratio level.

- The standard deviation (s) is the most important measure of dispersion because of its central role in many more advanced statistical applications. The standard deviation has a minimum value of 0 (indicating no variation in the distribution) and increases in value as the

variability of the distribution increases. It is used most appropriately with variables measured at the interval-ratio level but is frequently computed for ordinal-level variables.

- The variance (s^2) is used primarily in inferential statistics and in the design of some measures of association.
- Boxplots provide a useful way of visualizing and analyzing dispersion.

SUMMARY OF FORMULAS

FORMULA 4.1	Range:	$R = \text{High score} - \text{Low score}$
FORMULA 4.2	Interquartile range:	$Q = Q_3 - Q_1$
FORMULA 4.3	Variance:	$s^2 = \frac{\sum (X_i - \bar{X})^2}{N}$
FORMULA 4.4	Standard deviation:	$s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N}}$

GLOSSARY

Box and whiskers plot. See **Boxplot**.

Boxplot. A graph that presents information about the central tendency and dispersion of a variable.

Deviation. The distance between the score and the mean.

Dispersion. The amount of variety, or heterogeneity, in a distribution of scores.

Interquartile range (Q). The distance from the third quartile to the first quartile.

Measures of dispersion. Statistics that indicate the amount of variety, or heterogeneity, in a distribution of scores.

Range (R). The highest score minus the lowest score.

Standard deviation. The statistic computed by summing the squared deviations of the scores around the mean, dividing by N , and, finally, taking the square root of the result. The most important and useful descriptive measure of dispersion; s represents the standard deviation of a sample; σ represents the standard deviation of a population.

Variance. The sum of the squared deviations of the scores around the mean, divided by N . A measure of dispersion used primarily in inferential statistics and also in correlation and regression techniques; s^2 represents the variance of a sample; σ^2 represents the variance of a population.

PROBLEMS

- Compute the range and standard deviation of the following 10 scores. (HINT: It will be helpful to organize your computations as in Table 4.4.)

10, 12, 15, 20, 25, 30, 32, 35, 40, 50

- Compute the range and standard deviation of the following 10 test scores.

66, 75, 69, 72, 84, 90, 96, 70, 55, 45

- SOC** In problem 3.1 at the end of Chapter 3, you calculated measures of central tendency for six variables for freshmen and seniors. Three of those variables are reproduced here. Calculate the mean (if necessary), the range, and the standard deviation for each variable. What information is added by the measures of dispersion? Write a paragraph summarizing the differences between freshmen and seniors.

Out-of-Pocket Expenses		Number of Movies		Rating of Cafeteria Food	
Freshmen	Seniors	Freshmen	Seniors	Freshmen	Senior
43	75	0	0	10	1
49	72	14	5	7	2
55	70	10	11	2	8
57	95	7	3	1	4
72	72	5	4	8	3
58	67	1	14	6	6
62	50	0	0	10	2
75	59	14	7	0	9
61	55	3	5	5	4
53	95	4	3	6	7
	88		5		4

- 4.4 **SOC** In problem 3.5 at the end of Chapter 3, you calculated measures of central tendency for four variables for 15 respondents. Two of those variables are reproduced here. Calculate the mean (if necessary), the range, and the standard deviation for each variable. What information is added by the measures of dispersion? Write a paragraph summarizing these statistics.

Respondent	Age	Attitude on Abortion (High score = Strong Opposition)
A	18	10
B	20	9
C	21	8
D	30	10
E	25	7
F	26	7
G	19	9
H	29	6
I	31	10
J	55	5
K	32	4
L	28	3
M	23	2
N	24	1
O	32	9

- 4.5 **SOC** In problem 3.6, you computed mean and median income for the 13 Canadian provinces and territories and for 13 U.S. states in two separate years. Now compute the standard deviation and range for each year, and, taking account of the two measures of

central tendency and the two measures of dispersion, write a paragraph summarizing the distributions. What do the measures of dispersion add to what you already knew about central tendency? Did the median income of the provinces and states become more or less variable over the period? The scores are reproduced here.

Median Income for Canadian Provinces and Territories, 2000 and 2011 (Canadian dollars)

Province or Territory	2000	2011
Newfoundland and Labrador	38,800	67,200
Prince Edward Island	44,200	66,500
Nova Scotia	44,500	66,300
New Brunswick	43,200	63,930
Quebec	47,700	68,170
Ontario	55,700	73,290
Manitoba	47,300	68,710
Saskatchewan	45,800	77,300
Alberta	55,200	89,930
British Columbia	49,100	69,150
Yukon	56,000	90,090
Northwest Territories	61,000	105,560
Nunavut	37,600	65,280
Mean =		
Median =		
Range =		
Standard Deviation =		

Source: Statistics Canada. Retrieved from <http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/famil108a-eng.htm>

Median Income for Thirteen States, 1999 and 2012 (U.S. dollars)

State	1999	2012
Alabama	36,213	43,464
Alaska	51,509	63,348
Arkansas	29,762	39,018
California	43,744	57,020
Connecticut	50,798	64,247
Illinois	46,392	51,738
Kansas	37,476	50,003
Maryland	52,310	71,836
Michigan	46,238	50,015
New York	40,058	47,680
Ohio	39,617	44,375
South Dakota	35,962	49,415
Texas	38,978	51,926
Mean =		
Median =		
Range =		
Standard Deviation =		

Sources: 1999: U.S. Bureau of the Census. *Statistical Abstract of the United States: 2001*, p. 436. Retrieved from <http://www.census.gov/prod/2002pubs/01statab/income.pdf>. 2012: U.S. Bureau of the Census. *American Community Survey, 2012*. Retrieved from <http://www.census.gov/hhes/www/income/data/statemedian/>

- 4.7 **SOC** Labor force participation rates (percent employed), percent high school graduates, and mean income for males and females in ten states are reported here. Calculate the mean and the standard

State	Labor Force Participation Rate		Percentage HS Graduates		Mean Income	
	Male	Female	Male	Female	Male	Female
A	65.8	54.3	81.0	81.9	55,623	50,012
B	76.7	63.0	88.4	89.7	52,345	51,556
C	71.8	57.2	82.4	84.6	55,789	48,231
D	76.1	66.6	89.5	90.9	48,907	46,289
E	75.1	63.1	86.9	88.7	62,023	58,034
F	69.9	61.1	86.3	86.4	55,000	53,897
G	73.6	59.6	87.1	87.6	49,145	47,148
H	70.5	60.3	87.0	87.6	51,897	50,659
I	66.3	55.2	81.7	84.1	51,238	45,289
J	74.5	67.1	89.1	91.6	60,746	56,489

- 4.8 **SOC** Data on several variables measuring overall health and well-being for 11 nations are reported here for 2010, with projections to 2020. Are nations becoming

- 4.6 Compute the standard deviation for the pretest and posttest scores that were used in problems 2.6 and 3.12. The scores are reproduced here. Taking into account all of the information you have on these variables, write a paragraph describing how the sample changed from test to test. What does the standard deviation add to the information you already had?

Case	Pretest	Posttest
A	8	12
B	7	13
C	10	12
D	15	19
E	10	8
F	10	17
G	3	12
H	10	11
I	5	7
J	15	12
K	13	20
L	4	5
M	10	15
N	8	11
O	12	20

deviation for both groups for each variable and describe the differences. Are males and females unequal on any of these variables? How great is the gender inequality?

more or less diverse on these variables? Calculate the mean, range, and standard deviation for each year for each variable. Summarize the results in a paragraph.

Nation	Life Expectancy (years)		Infant Mortality Rate*		Fertility Rate#	
	2010	2020	2010	2020	2010	2020
Canada	81	82	5.0	4.4	1.6	1.6
China	75	76	16.5	12.6	1.5	1.5
Egypt	72	75	26.2	17.9	3.0	2.7
Germany	79	81	4.0	3.6	1.4	1.5
Japan	82	83	2.8	2.7	1.2	1.3
Mali	52	57	113.7	91.9	6.5	5.5
Mexico	76	78	17.8	13.2	2.3	2.1
Peru	71	74	27.7	20.2	2.3	2.0
Ukraine	69	70	8.7	7.3	1.3	1.4
U.S.	78	80	6.1	5.4	2.1	2.1
Zambia	52	54	68.4	50.6	6.0	5.3

Source: U.S. Bureau of the Census, 2012. *Statistical Abstract of the United States*: 2012, p. 842.

Notes:

*Number of deaths of children under one year of age per 1000 live births.

#Average number of children per female.

- 4.9 **[SOC]** The price of a gallon of regular gas for 20 nations is reported below. Compute the mean, median, range, interquartile range, and standard deviation for this variable, and write a paragraph summarizing these statistics.

Nation	Cost of a Gallon of Gasoline (U.S. dollars) in 2013
1 Norway	10.08
2 Italy	8.61
3 France	8.13
4 Finland	8.05
5 Ireland	7.79
6 United Kingdom	7.75
7 South Korea	6.44
8 Japan	5.90
9 Argentina	5.38
10 South Africa	4.94
11 India	4.74
12 Canada	4.67
13 China	4.67
14 Colombia	4.52
15 United States	3.66
16 Mexico	3.43
17 Russia	3.39
18 Malaysia	2.31
19 Nigeria	2.27
20 Egypt	1.00

Source: <http://www.bloomberg.com/visual-data/gas-prices/2013:South%20Africa:USD:g>

- 4.10 **[SOC]** In problem 3.11, you computed measures of central tendency for the number of cars per 1000 population for 15 nations. The scores are reproduced here. Compute the standard deviation for this variable, and write a paragraph summarizing the mean, median, and standard deviation.

Nation	Number of Motor Vehicles per 1000 Population (2009)
United States	786
Germany	588
Canada	608
Japan	588
Australia	703
Israel	330
Russia	271
Mexico	278
Tunisia	130
Bolivia	87
China	69
Niger	8
Kenya	25
Thailand	172
India	18
Mean =	
Median =	
Standard Deviation =	

Source: World Bank, 2014. Retrieved from <http://data.worldbank.org/indicator/IS.VEH.NVEH.P3>

- 4.11 **[PA]** Per capita expenditures for public transportation for 20 cities are reported here for 2000 and 2010.

Compute the mean and standard deviation for each year, and describe the differences in expenditures for the ten-year period.

City	2000	2010
A	52	197
B	87	124
C	101	131
D	52	197
E	115	119
F	88	87
G	100	150
H	101	209
I	95	110
J	78	140
K	103	178
L	107	94
M	85	125
N	117	200
O	167	225
P	180	210
Q	120	201
R	78	141
S	55	248
T	92	131

- 4.12 a. **[PA]** Compute the range, interquartile range, and standard deviation for the data presented in problem 3.4. The data are reproduced here. (NOTE: To simplify computations for Q , round off the case numbers that mark the first and third quartiles to the nearest whole number.)
- b. What would happen to the value of the standard deviation if you removed Chicago (the city with the highest score) from this distribution and recalculated? Why?

City	Annual Person-Hours Lost to Traffic Congestion
Baltimore	33
Boston	28
Buffalo	11
Chicago	44
Cleveland	13
Dallas	32
Detroit	23
Houston	37
Kansas City	14
Los Angeles	40
Miami	26
Minneapolis	27

(continued next column)

(continued)

City	Annual Person-Hours Lost to Traffic Congestion
New Orleans	20
New York	24
Philadelphia	26
Phoenix	23
Pittsburgh	23
San Antonio	19
San Diego	23
San Francisco	30
Seattle	27
Washington, D.C.	41

Source: U.S. Bureau of the Census, 2012. *Statistical Abstract of the United States*: 2012, Table 1099. Available at <http://www.census.gov/prod/2011pubs/12statab/trans.pdf>

- 4.13 **[SOC]** Listed here are the rates of abortion per 100,000 women for 20 states in 1973 and 1975. Compute the median, mean, standard deviation, range, and interquartile range for both years. Describe what happened to these distributions over the two-year period. Did the average rate increase or decrease? What happened to the dispersion of this distribution? Did the measures of dispersion increase or decrease? What happened between 1973 and 1975 that might explain these changes in central tendency and dispersion? (HINT: It was a Supreme Court decision.)

State	1973	1975
1 Mississippi	0.2	0.6
2 Arkansas	2.9	6.3
3 Montana	3.1	9.9
4 Maine	3.5	9.5
5 South Carolina	3.8	10.3
6 Tennessee	4.2	19.2
7 Texas	6.8	19.1
8 Arizona	6.9	15.8
9 Ohio	7.3	17.9
10 Nebraska	7.3	14.3
11 Virginia	7.8	18.0
12 Iowa	8.8	14.7
13 Massachusetts	10.0	25.7
14 Pennsylvania	12.1	18.5
15 Colorado	14.4	24.6
16 Florida	15.8	30.5
17 Michigan	18.7	20.3
18 Hawaii	26.3	31.6
19 California	30.8	33.6
20 New York	53.5	40.7

Source: U.S. Bureau of the Census, 1977. *Statistical Abstract of the United States*: 1977. Washington, DC: Government Printing Office, 1977.