Measures of Central Tendency

When given a set of numerical data, it is a good idea first to organize the data set in ascending order. This is accomplished by hand as well as in calculator and computer spreadsheet software with a "sort".

Another very useful skill is to summarize the numerical data set with one number, or statistic. There are two general categories of numerical statistics: measures of central tendency and measures of spread or variation. The most common measures of central tendency are the mean, the median, the mode, and the midrange.

Mean: The arithmetic mean is commonly called "the average". The mean is rarely a data value, but it often gives a good idea of the "center" or what is typical in the quantitative data.

In general, finding the arithmetic mean involves three steps:

(1) Calculate the sum of all the data values, symbolized by ΣX.
(2) Count the number of data values. (We use n for the sample size and N for the population size.)
(3) Divide the sum of all the data values by the number of data values. We typically round to one more decimal place than the values in the data set.

When dealing with complete data sets or entire populations, the notation commonly given is

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

where μ stands for the population mean, ΣX stands for the sum of the data values, and N stands for the number of data values in the population.

When dealing with incomplete data sets or samples from populations, the notation commonly given is

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

where $\overline{X}$ stands for the sample mean, ΣX stands for the sum of the data values, and n stands for the number of data values in the sample.

Another common average is the **weighted mean or average**. This is used when the individual data entries have varying "weights" or "values". The formula for computing a weighted mean is:

$$\overline{X} = \frac{\Sigma(w \cdot x)}{\Sigma w}$$

where x stands for the individual data value and w stands for its weight.

One example would be when an instructor figures your overall average in a course. An instructor may choose to assign a different weight to portfolio, projects, quizzes and tests in a course. Two other instances involving weighted averages include (1) the computation of a baseball player's slugging average, computed by dividing a player's total bases (rather than just hits) by their official "at bats", and (2) the computation of a student's grade point average (GPA), computed by dividing their total quality points by their total credits.

Before we discuss another statistic, we need to deal with a problem that can arise with data sets. An **outlier** is an extreme value (on the high or low end). Often there's a traceable reason for the outlier. One such example of this involves rain gauge readings. It seems that a certain town was giving readings to the weather station that stood out as much higher than other readings in the immediate area. Upon further investigation, it was found that the particular rain gauge in question was "filling up" with water from an automatic sprinkler system. In this case, it would be reasonable to throw out the outlier as a false reading for the region.

Another example of an outlier involves a well-known baseball player named Ted Williams. (We'll see his home run statistics later.) He played an outstanding and long career. He set numerous batting records despite missing nearly 5 seasons due to military service and two major injuries. He had a 0.406 average in 1941, earned 2 Triple Crowns, 2 MVPs, 6 batting championships, 18 All-Star games, and a lifetime 0.344 batting average. The "absent years" are given in baseball records, but shouldn't be included in many of his career statistical averages because those were atypical years for his baseball career.

In general, because each data value is involved in the calculation of the mean, outliers affect the mean more than any other measure of central tendency.

<u>Median</u>:

When there is an outlier in a data set, the median is very often a better measure of the "center" of a data set than the mean.

The median is in the middle of a quantitative data set. It is actually a data value in the case where there are an odd number of values, and it's between two data values when there's an even number of values.

There are 3 major steps for finding the median:

(1) Sort the data set (in ascending or descending order).

(2) Use the formula $\frac{n+1}{2}$ or $\frac{1}{2}(n+1)$ to calculate the position of the middle number.

(3) If the result of this calculation is whole, the data value in that position (from either end) is the median. If the result is midway between two whole numbers, use the data values in those two positions and average them to find the median.

A similar way to compute the median is to use the formula $c = \frac{n}{2}$ to calculate the position of the middle number. Here, if the result is whole, use the value halfway between the numbers in the c and c + 1 positions. If the result is not whole, round c to the nearest whole number and find the value in that position. Either method gives the same result for median. To summarize, sort first, then use one of the two tricks to find the position of the middle number. The median is either a value in the data set (odd n) or between two data values (even n).

Mode:

The mode of a set of data is the value that occurs most often.

There are two unique things about the mode:

- A set of data could have no mode, one mode, two modes, three modes, or several modes; and
- The mode can be found for both quantitative and qualitative data. For example, if a poll is conducted in which students are asked, "What is your academic major?" or "What is your favorite sport?", the most frequently occurring response is a particular major or a sport, and this major (like "math!") or sport (like "baseball") is the mode.

Midrange:

The midrange of a quantitative data set is midway between the highest (Max) and lowest (Min) data values. The formula is: $\text{Midrange} = \dfrac{\text{Min} + \text{Max}}{2}$.

Technology Notes:

In a TI graphing calculator, after the data values have been entered into a list, you can calculate 1-variable statistics using the STAT key and CALC menu, and the TI calculator uses this symbol " $\overline{X}$ " for both the population and sample mean, "Med" for the median, and you're on your own for noticing the mode.

The computer spreadsheet software, Microsoft Excel, uses the following commands:
"=AVERAGE(first cell:last cell)" for the population or sample mean,
"=MEDIAN(first cell:last cell)" for the median,
"=MODE(first cell:last cell)" for the mode,
"=(MAX(first cell:last cell)+MIN(first cell:last cell))/2" for the midrange.

Examples:

(1) The number of consecutive hours a light bulb will last before it burns out is tested. The number of hours for 20 bulbs is shown below. Compute the mean, mode, and median.

402	405	409	389	456	423	432	441	425	436
421	421	408	411	434	448	392	400	439	430

Solutions:

Mean: Divide the total by 20. $\overline{X} = \dfrac{\Sigma X}{n} = \dfrac{8{,}422}{20} = 421.1$ h

Mode: 421 h (the only value occurring twice)

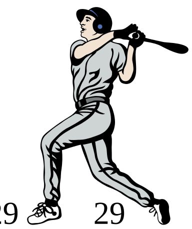
Median (Don't forget to sort the data first!):
389 392 400 402 405 408 409 411 421 421 • 423 425 430 432 434 436 439 441 448 456

With 20 data values arranged in order, the middle value is the average of the 10^{th} and 11^{th} values. The average of 421 and 423 is clearly 422 hours.

(2) Here are Ted Williams' home run totals by season for 1939-1960. Calculate the mean, median, mode, and midrange (excluding outliers).

31	23	37	36	0*	0*	0*	38	32	25	43	28
30	1**	13**	29	28	24	38	26	10	29		

Note: It's reasonable to toss out the five service year data as outliers. *Active service, WWII **Active service, Korean War



Solutions:

Mean: $\mu = \dfrac{\Sigma X}{N} = \dfrac{507}{17}$ B 29.8

Median:

10	23	24	25	26	28	28	29	29
30	31	32	36	37	38	38	43	

With 17 values arranged in order, the 9^{th} value (from either end) is the middle, or median, value; the median is 29 home runs.

Mode: There are 3 modes for this data set (28, 29, and 38), each occurring twice.

Midrange: Using the min (10) and the max (43), we have $\dfrac{10+43}{2} = 26.5$.

(3) The computation of a grade point average (GPA) is a very important example of a weighted average. The formula for calculating GPA is:

$$\text{GPA} = \frac{\Sigma\left(\text{credits} \cdot \text{quality points}\right)}{\Sigma \text{ credits}}$$

where an A in a course means 4 quality points, B = 3, C = 2, D = 1, and F or WF = 0. The grade of W in a course does not count toward a student's GPA both in the credits and quality points areas of the formula (but will affect financial aid!).

Jennifer makes the following grades one semester. Compute her GPA.

Course	Credits	Grade	Course	Credits	Grade
MATH 1101	3	A	PHED 1100C	2	C
ENGL 1101	3	W	ABAC 1000	1	A
HIST 2112	3	B	MUSC 1100	3	D

Solution: Ignore the W grade in ENGL 1101. The sum of the earned credits for the remaining 5 courses is 12.

$$\text{GPA} = \frac{3 \cdot 4 + 3 \cdot 3 + 2 \cdot 2 + 1 \cdot 4 + 3 \cdot 1}{12} = \frac{32}{12} \text{ B } 2.67 \text{ or } 2.7$$

Exercises [In general, round the mean to one more decimal place than the given data values.]:

1. Refer to the following cattle weights:

748 lb	485 lb	807 lb	1,023 lb	761 lb
765 lb	934 lb	579 lb	865 lb	1,064 lb

Compute the mean, mode, midrange, and median.

2. On a particularly busy hospital evening, 10 babies are born to 9 happily married couples. (One husband and wife had twins.) The babies' weights are recorded below.

8 lb, 9 oz	7 lb, 4 oz	4 lb, 5 oz	5 lb, 3 oz	10 lb, 2 oz
4 lb, 1 oz	7 lb, 13 oz	7 lb, 3 oz	9 lb, 10 oz	8 lb, 1 oz

(a) Find the mean, mode, and median.

(b) Decide which of these statistics would give the best prediction for the weight of the next baby born.

3. A policeman records the speeds to the nearest mile per hour of vehicles as they observe 30 minutes of rural highway traffic. P.S. The speed limit is 45 mph.

44	54	50	46	45	49	55	44	42	55	51	52
54	44	60	44	59	41	44	47	51	42	43	

(a) Calculate the mean, mode, and median of speeds for the drivers.

(b) If you lived in the area and wanted to promote more enforcement of the speed limit, which statistic would you use to argue your case?

4. In an experiment on the effect of a drug on reaction time, a patient is asked to depress a button whenever a light flashes. The reaction time (in milliseconds) for 15 trials is:

102	135	123	143	134	130	122	123	103	98	93	112
117	105	128									

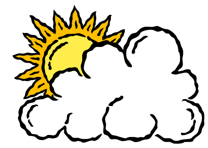
(a) Compute the mean, median, and mode for these reaction times.

(b) Which of these statistics is most "central" for the data set?

5. (a) Compute the mean, median, and mode for the following new home prices in Tift County.

\$42,000	\$82,000	\$100,000	\$76,500	\$95,000
\$67,000	\$85,000	\$102,500	\$115,000	\$310,000

(b) Which of these statistics is most “central” for the data set?



6. The values below are record high temperatures in °F for each of the United States (as of August, 2006, Source: *U.S. National Climatic Data Center*).

State	**Record High (°F)**
Alabama	112
Alaska	100
Arizona	128
Arkansas	120
California	134
Colorado	118
Connecticut	106
Delaware	110
Florida	109
Georgia	112
Hawaii	100
Idaho	118
Illinois	117
Indiana	116
Iowa	118
Kansas	121
Kentucky	114
Louisiana	114
Maine	105
Maryland	109
Massachusetts	107
Michigan	112
Minnesota	114
Mississippi	115
Missouri	118
Montana	117

State	**Record High (°F)**
Nebraska	118
Nevada	125
New Hampshire	106
New Jersey	110
New Mexico	122
New York	108
North Carolina	110
North Dakota	121
Ohio	113
Oklahoma	120
Oregon	119
Pennsylvania	111
Rhode Island	104
South Carolina	111
South Dakota	120
Tennessee	113
Texas	120
Utah	117
Vermont	105
Virginia	110
Washington	118
Wisconsin	112
West Virginia	114
Wyoming	114

Compute the median, mode, mean, and midrange for these temperatures.

7. Listed below are measured amounts of lead (in micrograms per cubic centimeter or $\mu g/cm^3$) in the air. The Environmental Protection Agency has established an air quality standard of 1.5 $\mu g/cm^3$. The measurements were recorded at Building 5 of the World Trade Center site on different days following the destruction caused by the terrorist attacks on September 11, 2001. After the collapse of the 2 World Trade Center buildings, there was considerable concern about air quality. Find the mean and median for this sample of measured levels of lead in the air.

5.40 1.10 0.42 0.73 0.48 1.10

8. Distinguish which measure of central tendency (mean, median, or mode) would best characterize the given scenario. Explain your reasoning, and compute the statistic.

(a) Employee salaries at A & B Co.

$24,000	$17,500	$21,000	$32,000	$40,000
$27,900	$30,850	$28,400	$26,500	$85,250

(b) Mickey Mantle's home run output (by season)

13	23	21	27	37	52	34	42	31	40	54
30	15	35	19	23	22	18				

(c) Salesperson ordering ladies athletic shoes based on sales figures for the month

$$6,\ 6\frac{1}{2},\ 5,\ 7,\ 7\frac{1}{2},\ 8\frac{1}{2},\ 8,\ 5\frac{1}{2},\ 6,\ 6,\ 7\frac{1}{2},\ 8,\ 7,\ 5\frac{1}{2},\ 6,\ 5\frac{1}{2},\ 5,\ 7\frac{1}{2},\ 7,\ 9,\ 8\frac{1}{2},\ 7$$

9. The following words were selected at random from a college catalog. The pages were selected using a random integer function, and the 10^{th} word was selected from each randomly selected page. Compute the mean word length.

of	are	with	be	Schools	law
for	Financial	a	Turf	built	or

10. Consider the following test scores: 86, 83, 88, 95, ??? .

(a) What is the average of the first four test scores?
(b) What would the fifth test score need to be for the student to earn an A? [Hint: An 89.5 average rounds to a 90, for an A.]
(c) If the fifth test score is a 70, compute the median before and after this fifth test score.
(d) Which statistic is affected more by this new fifth test score, the mean or the median?

11. A dietician obtains the amounts of sugar (in grams) in a sample of 1 gram from each of 16 different cereals, including Cheerios, Corn Flakes, Fruit Loops, Trix, and 12 other brands. Find the mean of these values. Is this mean likely to be a good estimate of the mean amount of sugar in each gram of cereal consumed by the population of all Americans who eat cereal? Explain.

0.03	0.24	0.30	0.47	0.43	0.07	0.47	0.13
0.44	0.39	0.48	0.17	0.13	0.09	0.45	0.43

12. Find the mean, mode, median, and midrange for the following reaction times. Each measurement is rounded to the nearest hundredth of a second. Notice the consistency.

19	20	17	21	21	21	19	18	19	19
17	17	15	17	18	17	18	18	18	17

13. When investigating times required for drive-through service, the following results (in seconds) are obtained (based on data from QSR Drive-Thru Time Study).

McDonald's:	287	128	92	267	176	240	192	118	153
Jack in the Box:	190	229	74	377	300	481	428	255	328

Which of the two fast food giants appears to be faster in their customer service? Does the difference appear to be significant?

14. The Aaron/Ruth Problem:

Refer to the career statistics for each player on the next page.

(a) Calculate the batting averages for Hank Aaron and Babe Ruth by season. Round each average to the nearest thousandths place.

$$\text{AVG} = \frac{\text{H}}{\text{AB}}$$

(b) Calculate each player's career category averages. (For example, the average number of doubles Hank Aaron hit over his entire career or the average number of home runs Babe Ruth hit over his entire career, etc.)

(c) Based on your calculations, who do you think was the better player, and why?

(d) Compute the slugging averages for Aaron and/or Ruth for any season. Slugging average is an example of a weighted average. Round to the nearest thousandths place.

KEY to abbreviations:

AB: official times at bat	H: number of hits, including 1B, 2B, 3B, HR
1B: singles	2B: doubles
3B: triples	HR: home runs
RBI: runs batted in	AVG: batting average
SA: slugging average	BB: bases on balls (walks)
SO: strikeouts	SB: stolen bases
FA: fielding average	R: runs

Henry Louis Aaron, Height: 6'0", Weight: 180 lbs.,
Bats: right, Throws: right,
Born: 2/5/34 , Mobile, AL, Home: Atlanta, GA, "The Hammer"

YEAR	CLUB	AB	H	1B	2B	3B	HR	RBI	AVG	SA	BB	SO	GAMES	SB	FA	R
1954	Milwaukee	468	131	85	27	6	13	69			28	39	122	2	0.970	58
1955	Milwaukee	602	189	116	37	9	27	106			49	61	153	3	0.967	105
1956	Milwaukee	609	200	126	34	14	26	92			37	54	153	2	0.962	106
1957	Milwaukee	615	198	121	27	6	44	132			57	58	151	1	0.983	118
1958	Milwaukee	601	196	128	34	4	30	95			59	49	153	4	0.984	109
1959	Milwaukee	629	223	131	46	7	39	123			51	54	154	8	0.983	116
1960	Milwaukee	590	172	101	20	11	40	126			60	63	153	16	0.982	102
1961	Milwaukee	603	197	114	39	10	34	120			56	64	155	21	0.983	115
1962	Milwaukee	592	191	112	28	6	45	128			66	73	156	15	0.981	127
1963	Milwaukee	631	201	124	29	4	44	130			78	94	161	31	0.979	121
1964	Milwaukee	570	187	131	30	2	24	95			62	46	145	22	0.981	103
1965	Milwaukee	570	181	108	40	1	32	89			60	81	150	24	0.987	109
1966	Atlanta	603	168	100	23	1	44	127			76	96	158	21	0.988	117
1967	Atlanta	600	184	105	37	3	39	109			63	97	155	17	0.979	113
1968	Atlanta	606	174	108	33	4	29	86			64	62	160	28	0.989	84
1969	Atlanta	547	164	87	30	3	44	97			87	47	147	9	0.984	100
1970	Atlanta	516	154	89	26	1	38	118			74	63	150	9	0.979	103
1971	Atlanta	495	162	90	22	3	47	118			71	58	139	1	0.994	95
1972	Atlanta	449	119	75	10	0	34	77			92	55	129	4	0.984	75
1973	Atlanta	392	118	65	12	1	40	96			68	51	120	1	0.977	84
1974	Atlanta	340	91	55	16	0	20	69			39	29	112	1	0.986	47
1975	Milwaukee	465	109	79	16	2	12	60			70	51	137	0	1	45
1976	Milwaukee	271	62	44	8	0	10	35			35	38	85	0	1	22
T0TALS	Career Totals															
AVG	Average															

George Herman Ruth, Height: 6'2", Weight: 215 lbs, Bats: left, Throws: left,
"The Sultan of Swat", "The Bambino"

YEAR	CLUB	AB	H	1B	2B	3B	HR	RBI	AVG	SA	BB	SO	GAMES	SB	FA	R
1914	Boston	10	2	1	1	0	0	0			0	4	5	0	1	1
1915	Boston	92	29	14	10	1	4	21			9	23	42	0	0.976	16
1916	Boston	136	37	26	5	3	3	16			10	23	67	0	0.973	18
1917	Boston	123	40	29	6	3	2	12			12	18	52	0	0.984	14
1918	Boston	317	95	47	26	11	11	66			57	58	95	6	0.95	50
1919	Boston	432	139	64	34	12	29	114			101	58	130	7	0.988	103
1920	New York	458	172	73	36	9	54	137			148	80	142	14	0.936	158
1921	New York	540	204	85	44	16	59	171			144	81	152	17	0.967	177
1922	New York	406	128	61	24	8	35	99			84	80	110	2	0.964	94
1923	New York	522	205	106	45	13	41	131			170	93	152	17	0.973	151
1924	New York	529	200	108	39	7	46	121			142	81	153	9	0.962	143
1925	New York	359	104	65	12	2	25	66			59	68	98	2	0.974	61
1926	New York	495	184	102	30	5	47	145			144	76	152	11	0.979	139
1927	New York	540	192	95	29	8	60	164			138	89	151	7	0.963	158
1928	New York	536	173	82	29	8	54	142			135	87	154	4	0.975	163
1929	New York	499	172	94	26	6	46	154			72	60	135	5	0.984	121
1930	New York	518	186	100	28	9	49	153			136	61	145	10	0.966	150
1931	New York	534	199	119	31	3	46	163			128	51	145	5	0.972	149
1932	New York	457	156	97	13	5	41	137			130	62	133	2	0.961	120
1933	New York	459	138	80	21	3	34	103			114	90	137	4	0.967	97
1934	New York	365	105	62	17	4	22	84			103	63	125	1	0.962	78
1935	Boston	72	13	7	0	0	6	12			20	24	28	0	0.952	13
TOTALS	Career Totals															
AVG	Average															