TOPIC 2.1

Frequency Distributions and Their Graphs

OBJECTIVES Build a frequency distribution with limits, midpoints, relative frequencies, and cumulative frequencies. Draw a frequency histogram, a frequency polygon, and an ogive.

GETTING STARTED Babe Ruth hit these home run totals for the Yankees from 1920 to 1934 (Source: Baseball-Reference):

54, 59, 35, 41, 46, 25, 47, 60, 54, 46, 49, 46, 41, 34, 22

a.) Put the data in order from least to greatest.

b.) In a typical season, about how many home runs did Ruth hit? Why is that hard to answer from the unsorted list?

Frequency Distributions

DEFINITION **Frequency Distribution**

A **frequency distribution** is a table that shows **classes** (intervals) of data entries with a count of the entries in each class. The count is the **frequency**, f.

Each class has a **lower class limit** (least number that can belong to the class) and an **upper class limit** (greatest number).

The **class width** is the distance between the _____ of consecutive classes. The **range** is the maximum entry minus the _____ entry.

PROCEDURE **Building a Frequency Distribution**

1. Choose the number of classes, usually 5 to 20.
2. Class width = $\frac{\text{range}}{\text{number of classes}}$, rounded **up** to the next convenient number.
3. The minimum entry is the first lower limit. Add the class width to get each next lower limit.
4. Each upper limit is one less than the next lower limit.
5. Tally the entries and count the frequency of each class.

DEFINITION **More Columns**

Midpoint = $\frac{\text{lower limit+upper limit}}{2}$

Relative frequency = $\frac{f}{n}$, where n is the _____ number of entries.

Cumulative frequency of a class is the sum of the frequency of that class and all _____ classes.

Example 1 Build a frequency distribution of Ruth's home runs with 5 classes.

a.) Find the range. _____

b.) Find the class width. _____

c.) Complete the table.

Class	Frequency, f	Midpoint	Relative frequency	Cumulative frequency

d.) In what percent of these seasons did Ruth hit 54 or more home runs? _____

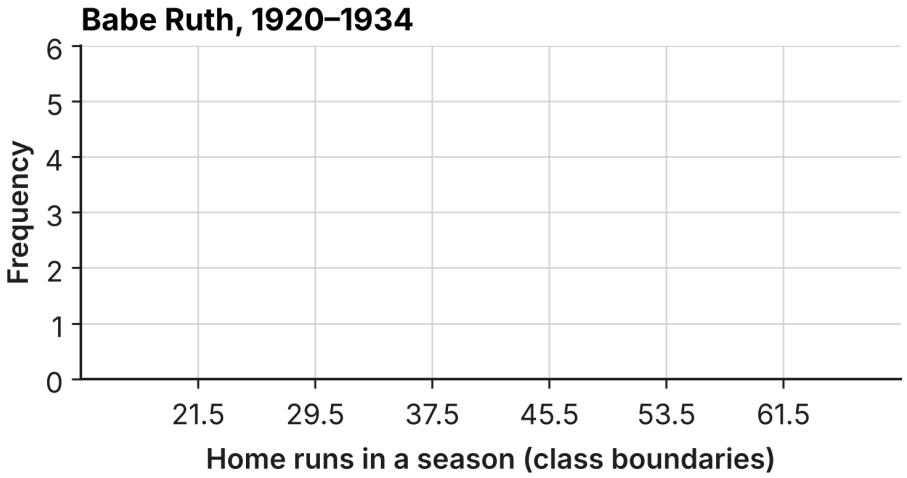
Graphs of Frequency Distributions

DEFINITION **Frequency Histogram**

A **frequency histogram** is a bar graph whose horizontal scale is quantitative. Its bars _____.

The bars run between **class boundaries**: the numbers halfway between the upper limit of one class and the lower limit of the next. For example, the boundary between 22–29 and 30–37 is _____.

Example 2 Draw a frequency histogram of Ruth's home runs.

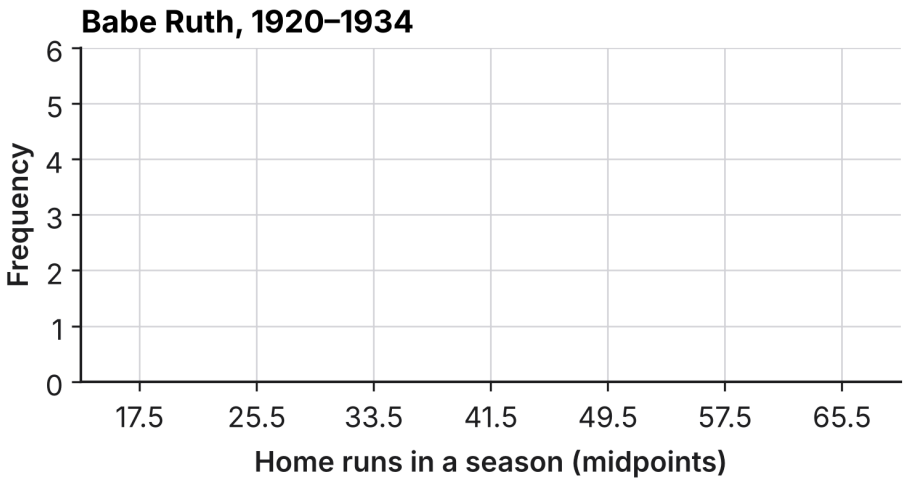


DEFINITION **Frequency Polygon and Relative Frequency Histogram**

A **frequency polygon** plots each class _____ against its frequency and connects the points. It starts and ends on the horizontal axis, one class width beyond the first and last midpoints.

A **relative frequency histogram** has the same shape as the frequency histogram, but its vertical scale shows relative frequencies.

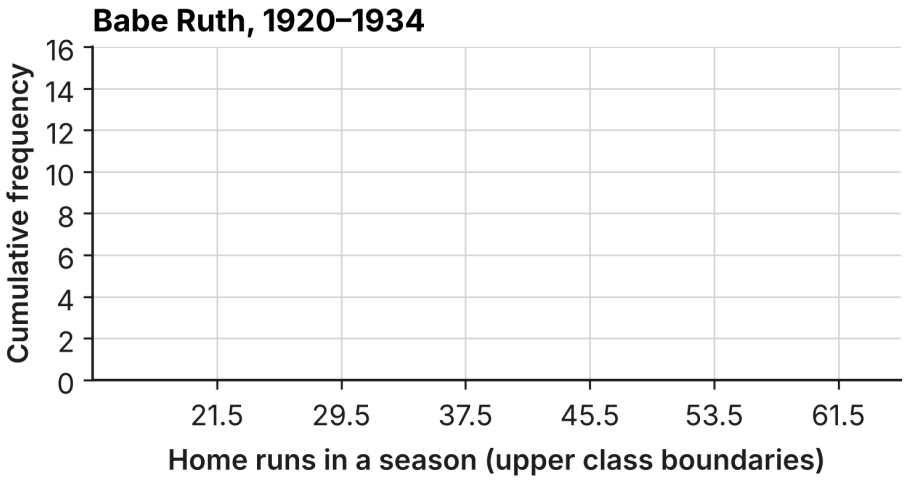
Example 3 Draw a frequency polygon of Ruth's home runs.



DEFINITION **Ogive**

An **ogive** (oh-jive) is a cumulative frequency graph. It plots the _____ class boundaries against the cumulative frequencies, starting at the first lower boundary with a frequency of 0.

Example 4 Draw an ogive of Ruth's home runs. Then answer the question.



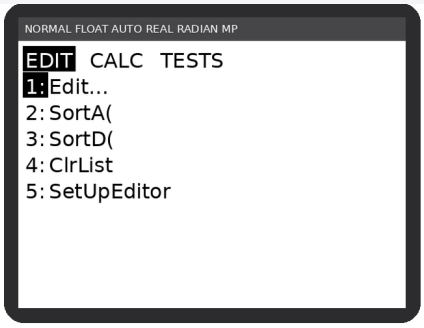
In how many of these seasons did Ruth hit 45 or fewer home runs? _____

TIP TI-84: Histogram

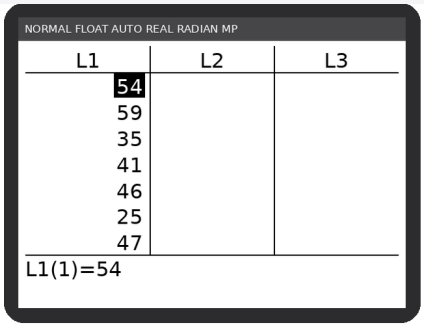
Enter the data: **STAT** ▸ **1:Edit**, type the data in **L1**.

Turn on the plot: **2nd** [STAT PLOT] ▸ **1:Plot1** ▸ **On**, Type: histogram (third picture), Xlist: **L1**, Freq: **1**.

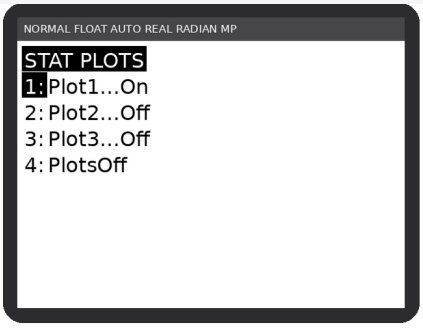
Set the classes: **WINDOW** ▸ Xmin = 21.5, Xmax = 61.5, Xscl = 8 (the class width), Ymin = 0, Ymax = 6. Press **GRAPH**, then **TRACE** to see each class and its count.



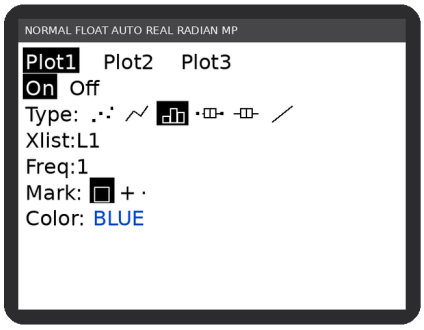
1 STAT ▸ 1:Edit



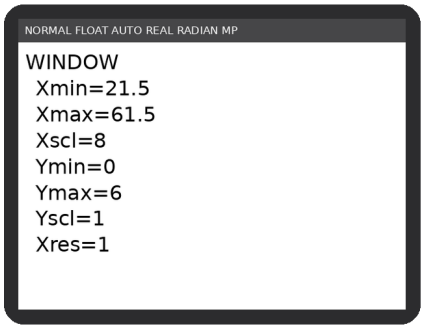
2 Type the data in L1



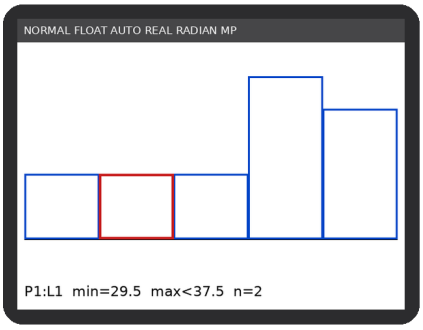
3 2nd Y= (STAT PLOT) ▸ 1:Plot1



4 On, histogram (third picture), Xlist: L1, Freq: 1



5 WINDOW: class boundaries, Xscl = class width



6 GRAPH, then TRACE: each bar's class and count

CHECK YOUR UNDERSTANDING

A frequency distribution has classes 10–19, 20–29, 30–39, and 40–49. What is the class width?

(A) 9
(B) 10
(C) 19
(D) 40

TOPIC 2.2

More Graphs and Displays

OBJECTIVES Graph quantitative data with stem-and-leaf plots and dot plots. Graph qualitative data with pie charts and Pareto charts. Graph paired data with scatter plots and time series charts.

GETTING STARTED Roger Maris broke Babe Ruth's record with 61 home runs in 1961. Here are Maris's home run totals for 1957 to 1968 (Source: Baseball-Reference):

14, 28, 16, 39, 61, 33, 23, 26, 8, 13, 9, 5

Do you think 61 was a typical season for Maris? Explain.

Graphing Quantitative Data

DEFINITION **Stem-and-Leaf Plot**

In a **stem-and-leaf plot**, each number is split into a **stem** (the leftmost digits) and a **leaf** (the rightmost digit). It shows the shape of the data and _____ every original data entry.

Always include a **key**, such as 6 | 1 = 61. Leaves are written in order, least to greatest.

Example 1 Make a stem-and-leaf plot of Maris's home runs.

Stem	Leaves
0	
1	
2	
3	
4	
5	
6	

Key: 6 | 1 = 61 home runs

DEFINITION Back-to-Back Stem-and-Leaf Plot

A **back-to-back** stem-and-leaf plot compares two data sets that share the same _____. Leaves for one set go to the left, read outward from the stem.

Example 2 The back-to-back plot compares Ruth (1920–1934) and Maris (1957–1968).

Ruth	Stem	Maris
	0	5 8 9
	1	3 4 6
5 2	2	3 6 8
5 4	3	3 9
9 7 6 6 6 1 1	4	
9 4 4	5	
0	6	1

Key: 5 | 4 | = 45 for Ruth, | 1 | 3 = 13 for Maris

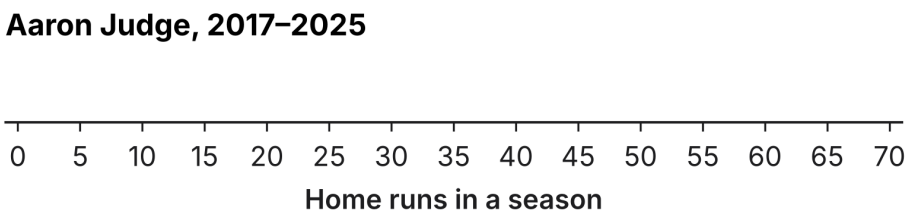
Watch out: check the stems carefully. In this plot the stems are tens of home runs.

a.) Compare the two players. Who usually hit more home runs? _____

DEFINITION Dot Plot

In a **dot plot**, each entry is plotted as a point above a _____. Repeated values stack up.

Example 3 Make a dot plot of Aaron Judge's home runs for 2017 to 2025: 52, 27, 27, 9, 39, 62, 37, 58, 53. (Source: Baseball-Reference)



What might explain the lone 9? _____

Graphing Qualitative Data

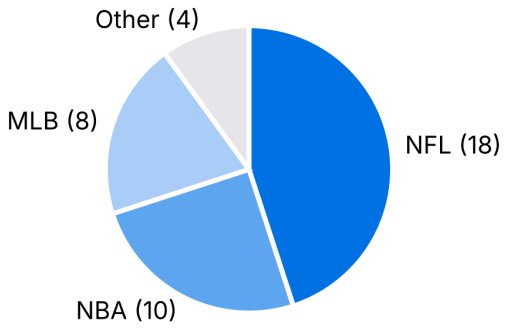
DEFINITION **Pie Chart and Pareto Chart**

A **pie chart** is a circle divided into sectors that represent categories. Each sector's angle is the relative frequency times _____.

A **Pareto chart** is a vertical bar graph with the bars in order of decreasing _____. It shows the most common categories first.

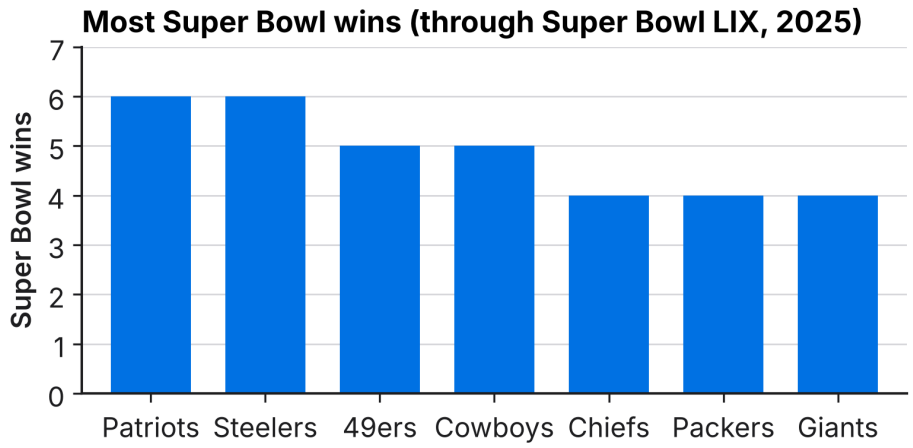
Example 4 A class of 40 students named their favorite pro league.

Favorite league: a class of 40 students



League	Frequency	Relative frequency	Angle
NFL	18		
NBA	10		
MLB	8		
Other	4		

Example 5 The Pareto chart shows the NFL teams with the most Super Bowl wins.



a.) Which teams are tied for the most wins? _____

b.) How many Super Bowls did these seven teams win in all? _____

Graphing Paired Data

DEFINITION **Scatter Plot and Time Series Chart**

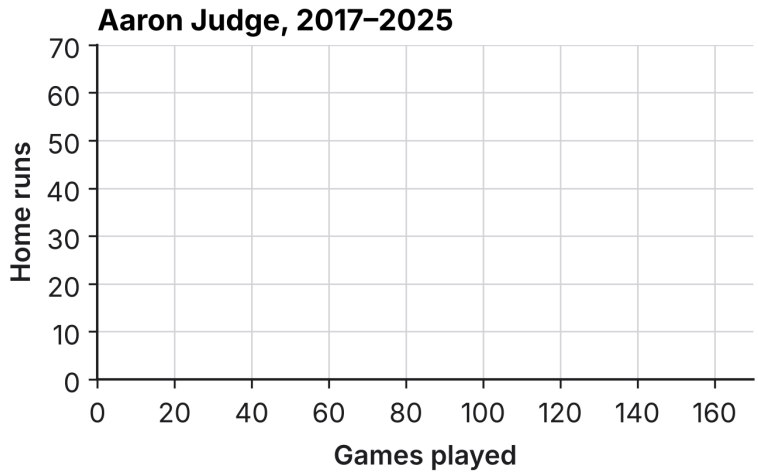
Paired data match each entry in one data set with an entry in a second data set, like (games, home runs).

A **scatter plot** graphs paired data as ordered pairs (x, y). It can show a _____ between two variables.

A **time series chart** graphs data entries that are measured over _____, with time on the horizontal axis.

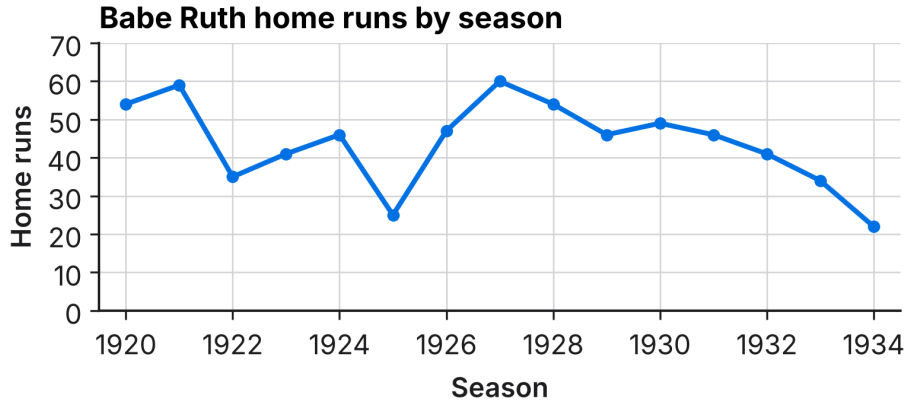
Example 6 Make a scatter plot of Judge's games played and home runs. Then describe the relationship.

Season	2017	2018	2019	2020	2021	2022	2023	2024	2025
Games	155	112	102	28	148	157	106	158	152
HR	52	27	27	9	39	62	37	58	53



Relationship: _____

Example 7 The time series chart shows Ruth's home runs by season.

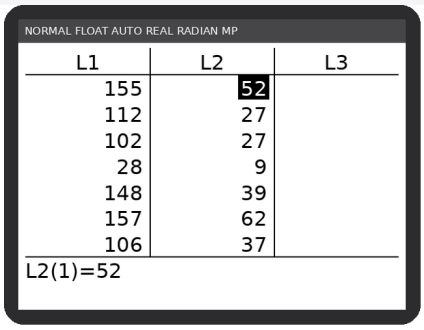


a.) In which season did Ruth hit the most home runs? _____

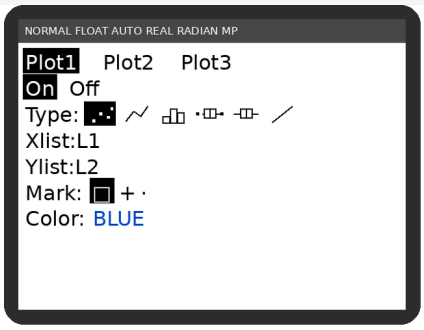
b.) Describe the trend after 1930. _____

TIP TI-84: Scatter Plot

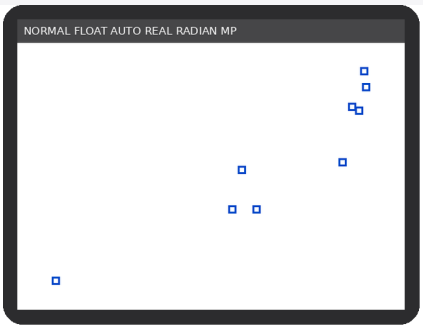
Enter x values in **L1** and y values in **L2**. **2nd** [STAT PLOT] ▸ **Plot1** ▸ **On**, Type: scatter (first picture), Xlist: **L1**, Ylist: **L2**. Press **ZOOM** ▸ **9:ZoomStat**.



❶ Games in L1, home runs in L2



❷ Plot1: On, scatter (first picture), Xlist: L1, Ylist: L2



❸ ZOOM ▸ 9:ZoomStat

CHECK YOUR UNDERSTANDING

Which graph is best for showing how one player's batting average changed from season to season?

(A) pie chart
(B) Pareto chart
(C) time series chart
(D) stem-and-leaf plot

TOPIC 2.3

Measures of Central Tendency

OBJECTIVES Find the mean, median, and mode of a data set. Find a weighted mean and the mean of a frequency distribution. Describe the shape of a distribution.

GETTING STARTED Maris hit 61 home runs in 1961, but under 40 in every other season. If you describe a "typical" Maris season with one number, should the 61 count? Explain.

Mean, Median, and Mode

DEFINITION Measures of Central Tendency

A **measure of central tendency** is a value that represents a typical, or central, entry of a data set.

Mean: the sum of the entries divided by the number of entries. Population mean $\mu = \frac{\Sigma x}{N}$.

Sample mean $\bar{x} = \frac{\Sigma x}{n}$.

Median: the middle entry when the data are in _____. With an even number of entries, it is the mean of the two middle entries.

Mode: the entry that occurs most often. A data set can have one mode, more than one mode, or _____ mode. It is the only measure that works for qualitative data.

Example 1 Find the mean, median, and mode of Ruth's home runs, 1920–1934.

22, 25, 34, 35, 41, 41, 46, 46, 46, 47, 49, 54, 54, 59, 60

a.) Mean _____

b.) Median _____

c.) Mode _____

Example 2 Find the mean, median, and mode of Maris's home runs, 1957–1968.

5, 8, 9, 13, 14, 16, 23, 26, 28, 33, 39, 61

a.) Mean _____

b.) Median _____

c.) Mode _____

Outliers

DEFINITION **Outlier**

An **outlier** is a data entry that is far removed from the other entries. An outlier can strongly affect the _____, but it has little effect on the _____.

Example 3 Suppose Maris had hit 31 home runs in 1961 instead of 61.

a.) Find the new mean. _____

b.) Find the new median. _____

c.) Which measure better describes a typical Maris season? Explain.

TIP Choosing a Measure

Use the **mean** when the data have no outliers: it uses every entry. Use the **median** when the data have outliers or are skewed. Use the **mode** for qualitative data, such as a favorite team.

Weighted Mean and Grouped Data

DEFINITION Weighted Mean

A **weighted mean** is the mean of entries that have different weights w: $\bar{x} = \frac{\Sigma(x \cdot w)}{\Sigma w}$.

Example 4 A course grade is 50% tests, 20% homework, and 30% final exam. A student has a test average of 86, a homework average of 95, and a final exam score of 78. Find the course grade. _____

DEFINITION Mean of a Frequency Distribution

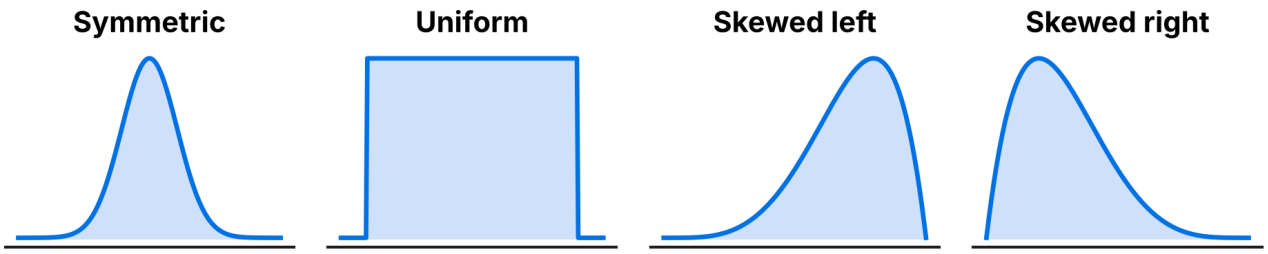
For grouped data, use each class _____ x and frequency f: $\bar{x} = \frac{\Sigma(x \cdot f)}{n}$, where $n = \Sigma f$.

Example 5 Use the frequency distribution of Ruth's home runs from 2.1 to estimate his mean.

Midpoint, x	Frequency, f	x · f
25.5	2	
33.5	2	
41.5	2	
49.5	5	
57.5	4	
	n = 15	

Estimated mean _____

Shapes of Distributions



DEFINITION **Shape and Center**

Symmetric: the left and right halves are mirror images. Mean = median.

Uniform: all entries or classes have about equal frequencies.

Skewed left: the tail extends to the left. The mean is usually _____ than the median.

Skewed right: the tail extends to the right. The mean is usually _____ than the median.

TIP **Remember**

The mean gets pulled toward the **tail**. The tail tells the direction of the skew.

Example 6 Use the means and medians from Examples 1 and 2 to describe each shape.

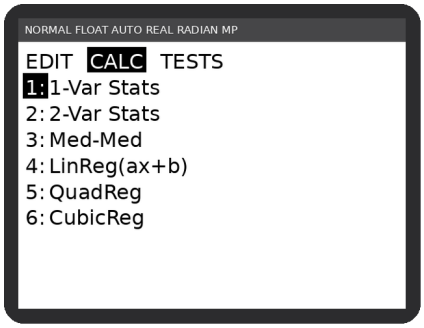
a.) Ruth: mean 43.9, median 46 _____

b.) Maris: mean 22.9, median 19.5 _____

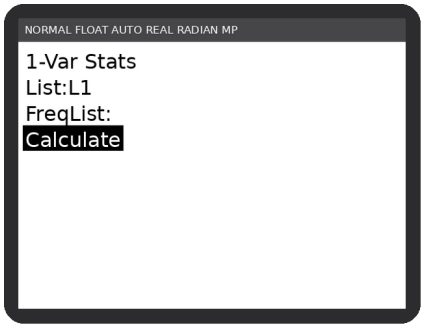
TIP **TI-84: 1-Var Stats**

Enter the data in **L1**. Press **STAT** ▸ **CALC** ▸ **1:1-Var Stats**, List: **L1**, then **Calculate**. The screen shows $\bar{x}$, Σx, n, and (scroll down) Med, minX, and maxX.

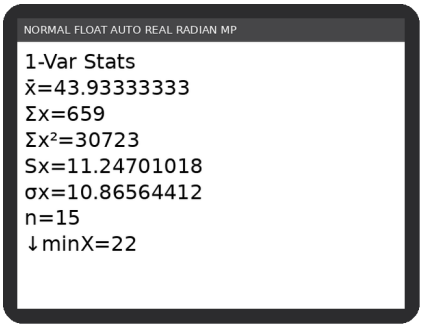
For a frequency distribution, put midpoints in **L1** and frequencies in **L2**, and set FreqList: **L2**.



❶ STAT ▸ CALC ▸ 1:1-Var Stats



❷ List: L1, leave FreqList blank, Calculate



❸ x̄, Σx, n; scroll down for Med, minX, maxX

CHECK YOUR UNDERSTANDING

A data set has a mean of 52 and a median of 45. What is the most likely shape of its distribution?

(A) symmetric
(B) uniform
(C) skewed left
(D) skewed right

TOPIC 2.4

Measures of Variation

OBJECTIVES Find the range, variance, and standard deviation of a data set. Interpret standard deviation. Use the Empirical Rule and Chebyshev's Theorem. Find the coefficient of variation.

GETTING STARTED Two hitters each average 30 home runs over three seasons. Hitter A hits 29, 30, and 31. Hitter B hits 10, 30, and 50. The means are the same. How are the hitters different?

Range

DEFINITION Range

The **range** of a quantitative data set is the difference between the maximum and minimum entries: Range = (Maximum entry) – (_____).

Example 1 Find the range of each data set.

a.) Maris, 1960–1963: 39, 61, 33, 23 _____

b.) Ruth, 1920–1934 (from 2.3): min 22, max 60 _____

c.) The ranges are equal. Does that mean the two data sets have the same spread? Explain.

Variance and Standard Deviation

DEFINITION **Deviation, Variance, and Standard Deviation**

The **deviation** of an entry x is its distance from the mean: $x - \mu$ for a population, $x - \bar{x}$ for a sample. The sum of the deviations is always _____.

Population: variance $\sigma^2 = \frac{\Sigma(x-\mu)^2}{N}$ and standard deviation $\sigma = \sqrt{\sigma^2}$.

Sample: variance $s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$ and standard deviation $s = \sqrt{s^2}$.

The **standard deviation** measures the typical distance of an entry from the mean. The more spread out the data, the _____ the standard deviation.

1. **PROCEDURE** **Finding a Standard Deviation by Hand**

1. Find the mean.
2. Subtract the mean from each entry to get the deviations.
3. Square each deviation.
4. Add the squares to get the sum of squares, SS.
5. Divide by N (population) or $n - 1$ (sample) to get the variance.
6. Take the square root.

Example 2 Roger Maris hit 39, 61, 33, and 23 home runs from 1960 to 1963. Treat these 4 seasons as a population.

x	x – μ	$(x - \mu)^2$
39		
61		
33		
23		
Σx = 156		

a.) Find the mean. _____

b.) Find the population variance. _____

c.) Find the population standard deviation. _____

Example 3 Now treat the same 4 seasons as a sample of Maris's career.

a.) Find the sample variance. _____

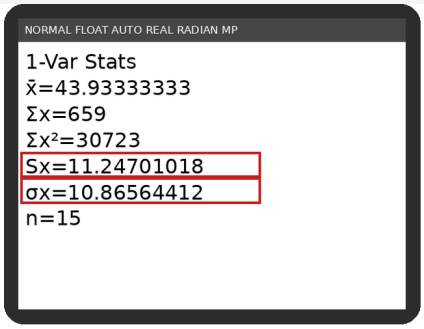
b.) Find the sample standard deviation. _____

c.) Why is s larger than σ?

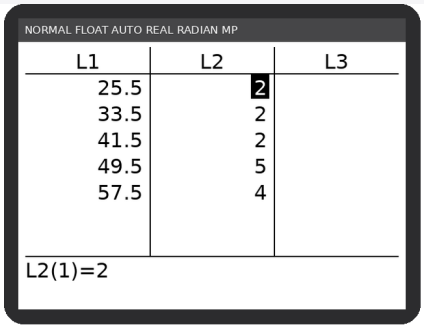
TIP **TI-84: Standard Deviation**

Enter the data in **L1**. Press **STAT** ▸ **CALC** ▸ **1:1-Var Stats**, then **Calculate**. **Sx** is the sample standard deviation s. **σx** is the population standard deviation σ. Square either one to get the variance.

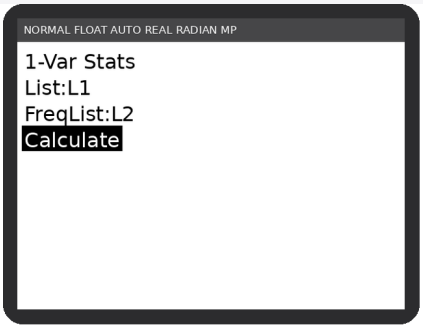
For a frequency distribution, put midpoints in **L1**, frequencies in **L2**, and set FreqList: **L2**.



1 Sx = sample s, σx = population σ



2 Frequency distribution: midpoints in L1, frequencies in L2



3 Set FreqList: L2, then Calculate

Example 4 Use a TI-84 and 1-Var Stats on the career data from 2.3. Treat each as a sample.

a.) Ruth, 1920–1934: $\bar{x} \approx 43.9$, $s \approx$ _____

b.) Maris, 1957–1968: $\bar{x} \approx 22.9$, $s \approx$ _____

c.) Which player's home run totals were more spread out? Explain.

Interpreting Standard Deviation

DEFINITION **Coefficient of Variation**

The **coefficient of variation** (CV) compares the standard deviation to the mean:

$CV = \frac{s}{\bar{x}} \cdot 100\%$

Use it to compare the variation of data sets with very different means. A larger CV means _____ relative variation.

Example 5 Use Example 4 to find each player's coefficient of variation.

a.) Ruth _____

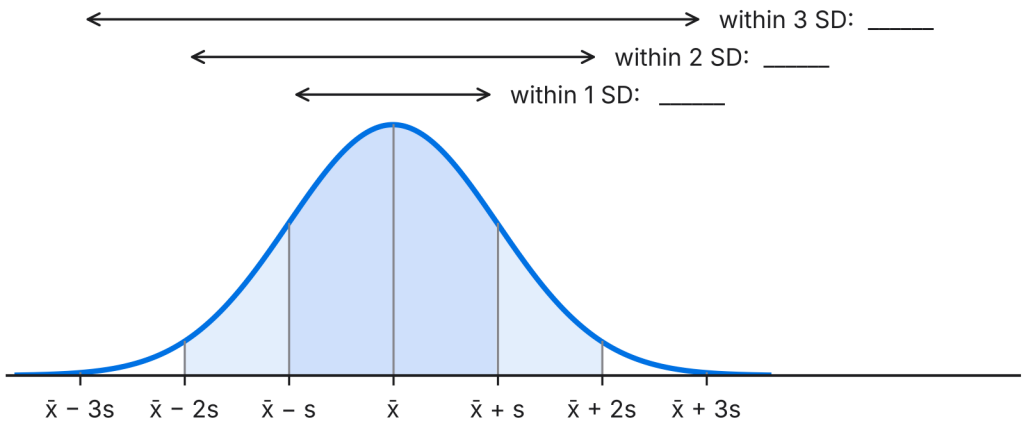
b.) Maris _____

The Empirical Rule

DEFINITION **Empirical Rule (68–95–99.7 Rule)**

For data with a symmetric, **bell-shaped** distribution:

About _____ of the data lie within 1 standard deviation of the mean. About _____ lie within 2 standard deviations. About _____ lie within 3 standard deviations.



Example 6 Suppose the heights of a group of basketball players are bell-shaped with a mean of 79 inches and a standard deviation of 3.5 inches. Estimate the percent of players with each height.

a.) between 75.5 and 82.5 inches _____

b.) between 72 and 86 inches _____

c.) between 79 and 86 inches _____

d.) taller than 82.5 inches _____

Chebyshev's Theorem

DEFINITION **Chebyshev's Theorem**

For **any** data set, no matter its shape, the portion of the data within k standard deviations of the mean ($k > 1$) is **at least** $1 - \frac{1}{k^2}$.

$k = 2$: at least _____. $k = 3$: at least _____.

Example 7 Maris's 1957–1968 seasons are skewed right, so the Empirical Rule does not apply. Use $\bar{x} \approx 22.9$ and $s \approx 16.0$.

a.) Find the interval within 2 standard deviations of the mean. _____

b.) By Chebyshev's Theorem, at least what percent of his seasons lie in this interval? _____

c.) What percent actually do? _____

CHECK YOUR UNDERSTANDING

A data set is bell-shaped with a mean of 50 and a standard deviation of 5. About what percent of the data lie between 40 and 60?

(A) 68%
(B) 75%
(C) 95%
(D) 99.7%

TOPIC 2.5

Measures of Position

OBJECTIVES Find the quartiles and interquartile range of a data set. Draw a box-and-whisker plot. Find percentiles. Find and interpret z-scores.

GETTING STARTED Babe Ruth hit 60 home runs in 1927. Roger Maris hit 61 in 1961. Which season stood out more compared with the rest of that player's own career? Make a guess.

Quartiles

DEFINITION **Quartiles**

The three **quartiles** Q_1, Q_2, and Q_3 divide an ordered data set into four parts with about the same number of entries.

Q_2 is the _____. Q_1 is the median of the entries below Q_2. Q_3 is the median of the entries above Q_2.

The **interquartile range** is IQR = $Q_3 - Q_1$. It measures the spread of the middle _____ of the data.

Example 1 Find the quartiles and IQR of Ruth's home runs, 1920–1934.

22 25 34 35 41 41 46 46 46 47 49 54 54 59 60

a.) Q_2 _____ Q_1 _____ Q_3 _____

b.) IQR _____

Example 2 Find the quartiles and IQR of Maris's home runs, 1957–1968.

5 8 9 13 14 16 23 26 28 33 39 61

a.) Q_2 _____ Q_1 _____ Q_3 _____

b.) IQR _____

DEFINITION **Identifying Outliers with the IQR**

An entry is an **outlier** if it is less than $Q_1 - 1.5(\text{IQR})$ or greater than $Q_3 + 1.5(\text{IQR})$. These two values are called _____.

Example 3 Is Maris's 61 an outlier? Use the fences.

Box-and-Whisker Plots

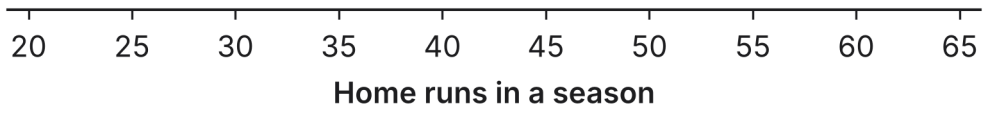
DEFINITION **Five-Number Summary and Box Plot**

The **five-number summary** is the minimum, Q_1, Q_2, Q_3, and maximum. A **box-and-whisker plot** (box plot) displays it: a box from Q_1 to Q_3, a vertical line at the _____, and whiskers out to the minimum and maximum.

In a **modified** box plot, outliers are marked with a star, and the whisker stops at the most extreme entry that is not an outlier.

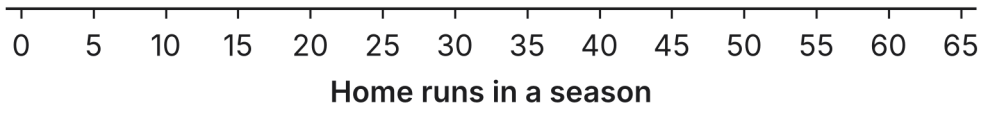
Example 4 Write the five-number summary for Ruth: _____. Then draw the box plot.

Babe Ruth, 1920–1934



Example 5 Draw a modified box plot for Maris. Mark the outlier with a star.

Roger Maris, 1957–1968

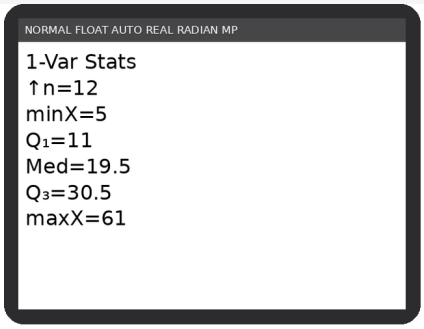


Compare the two plots. What do they tell you?

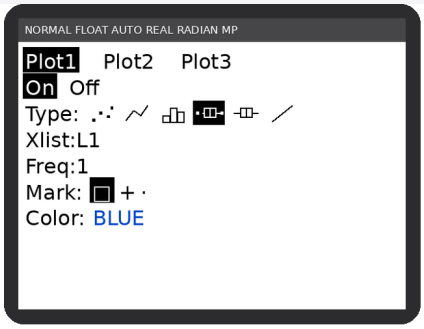
TIP TI-84: Five-Number Summary and Box Plot

Enter the data in **L1**. **1-Var Stats** shows minX, Q_1, Med, Q_3, and maxX (scroll down).

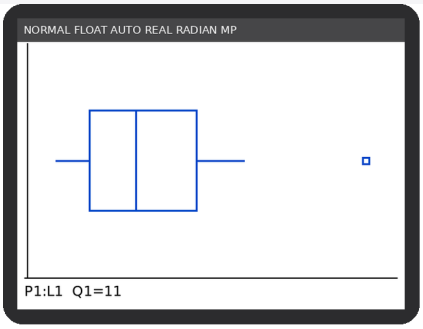
Press **2nd Y=** (STAT PLOT) ▸ **1:** Plot1 ▸ **On**. Choose the first box plot icon (modified, shows outliers). Xlist: **L1**. Press **ZOOM 9** (ZoomStat).



❶ 1-Var Stats, scroll down: minX, Q₁, Med, Q₃, maxX



❷ Plot1: On, first box plot icon (modified), Xlist: L1



❸ ZOOM 9 (ZoomStat). The outlier shows as a small square.

Percentiles

DEFINITION Percentiles

The **percentiles** divide an ordered data set into 100 parts. If an entry is at the 80th percentile, then about _____ of the entries are at or below it.

Percentile of $x = \dfrac{\text{number of entries less than } x}{\text{total number of entries}} \cdot 100$, rounded to the nearest whole number.

Q_1, Q_2, and Q_3 are about the 25th, 50th, and 75th percentiles.

Example 6 Ruth hit 54 home runs in 1920 and in 1928. Find the percentile of 54 in his 1920–1934 data. _____

The Standard Score (z-Score)

DEFINITION z-Score

The **standard score**, or **z-score**, tells how many standard deviations an entry x is from the mean: $z = \frac{x-\mu}{\sigma}$ (or use $\bar{x}$ and s for a sample).

A negative z-score means the entry is _____ the mean. A z-score less than –2 or greater than 2 is **unusual**. Less than –3 or greater than 3 is **very unusual**.

Example 7 Compare Ruth's 60 home runs in 1927 with Maris's 61 in 1961, each against his own career. Use the sample statistics.

a.) Ruth: $\bar{x} \approx 43.93$, $s \approx 11.25$ _____

b.) Maris: $\bar{x} \approx 22.92$, $s \approx 15.98$ _____

c.) Which season was more unusual for that player? Explain.

CHECK YOUR UNDERSTANDING

A test has a mean of 72 and a standard deviation of 8. What is the z-score of a test score of 60?

(A) –1.5
(B) –0.67

(C) 1.5
(D) 12