AP Stats Ch. 2   Modeling Distributions of Data

2.1 Describing Location in a Distribution

MEASURE position using percentiles
INTERPRET cumulative relative frequency graphs
MEASURE position using *z*-scores
TRANSFORM data
DEFINE and DESCRIBE density curves

2.2 Modeling Distributions of Data

DESCRIBE and APPLY the 68-95-99.7 Rule
DESCRIBE the standard Normal Distribution
PERFORM Normal distribution calculations
ASSESS Normality

2.1 Describing Location in a Distribution

Percentile:

Ex. 1

Jenny earned a score of 86 on her test. How did she perform relative to the rest of the class?

6	7
7	2334
7	5777899
8	00123334
8	5**6**9
9	03

Key: 8|6 represents a score of 86 on the test.

Ex. 2

The stemplot below shows the number of wins for each of the 30 Major League Baseball teams in 2009.

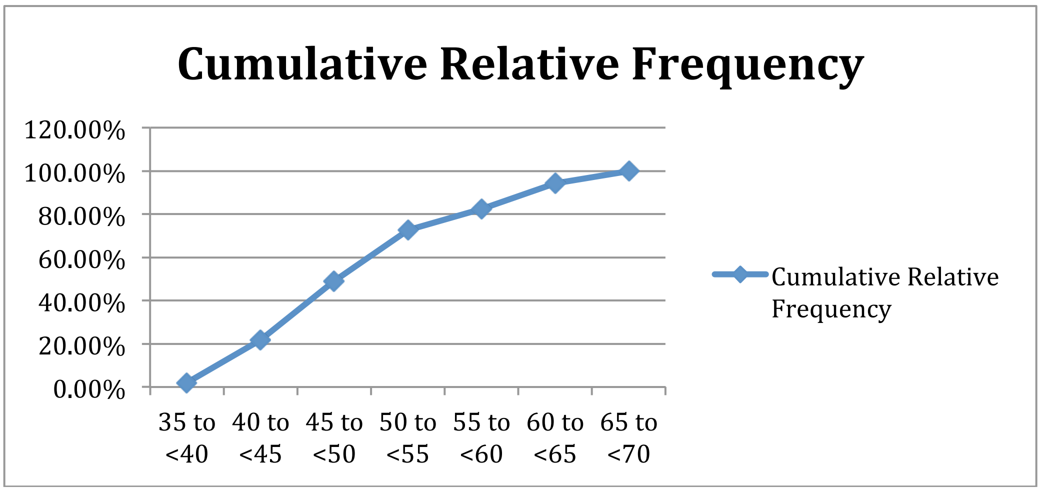
5	9
6	2455
7	00455589
8	0345667778
9	123557
10	3

Key: 5|9 represents a team with 59 wins.

Problem: Find the percentiles for the following teams:
(a) The Colorado Rockies, who won 92 games.
(b) The New York Yankees, who won 103 games.
(c) The Kansas City Royals and Cleveland Indians, who both won 65 games.

Cumulative Relative Frequency (Ogive) Graph:

Median Income ($1000s)	Frequency	Relative Frequency	Cumulative Frequency	Cumulative Relative Frequency
35 to < 40	1	1/51 = 0.020	1	1/51 = 0.020
40 to < 45	10	10/51 = 0.196	11	11/51 = 0.216
45 to < 50	14	14/51 = 0.275	25	25/51 = 0.490
50 to < 55	12	12/51 = 0.236	37	37/51 = 0.725
55 to < 60	5	5/51 = 0.098	42	42/51 = 0.824
60 to < 65	6	6/51 = 0.118	48	48/51 = 0.941
65 to < 70	3	3/51 = 0.059	51	51/51 = 1.000



Example:

a) At what percentile is California with a median income of $57,445?

b) Estimate and interpret the first quartile of this distribution.

<u>Z-scores:</u>

A *z*-score tells us how ______________________________ from the __________an observation falls, and in what direction.

Definition of z-score:

Example: The single-season home run record for major league baseball has been set just three times since Babe Ruth hit 60 home runs in 1927. Roger Maris hit 61 in 1961, Mark McGwire hit 70 in 1998 and Barry Bonds hit 73 in 2001. In an absolute sense, Barry Bonds had the best performance of these four players, since he hit the most home runs in a single season. However, in a relative sense this may not be true. Baseball historians suggest that hitting a home run has been easier in some eras than others. This is due to many factors, including quality of batters, quality of pitchers, hardness of the baseball, dimensions of ballparks, and possible use of performance-enhancing drugs. To make a fair comparison, we should see how these performances rate relative to others hitters during the same year.

Problem: Compute the standardized scores for each performance.
Which player had the most outstanding performance relative to his peers?

Year	Player	HR	Mean	SD	Z-score
1927	Babe Ruth	60	7.2	9.7	
1961	Roger Maris	61	18.8	13.4	
1998	Mark McGwire	70	20.7	12.7	
2001	Barry Bonds	73	21.4	13.2	

Conclusion:

Transforming Data: Transformations can effect the ________________,

________________, and ______________ of a distribution.

Adding (or Subtracting) a constant to each observation:

-

-

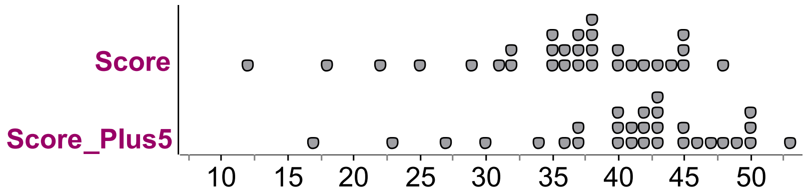
Multiplying (or Dividing) each observation by a constant :

-

-

-

Example: Here are a graph and table of summary statistics for a sample of 30 test scores. The maximum possible score on the test was 50 points. Suppose that the teacher was nice and added 5 points to each test score (who is this teacher?). How would this change the shape, center, and spread of the distribution? Here are graphs and summary statistics for the original scores and the +5 scores:



	n	$\bar{x}$	s_x	Min	$Q1$	M	$Q3$	Max	IQR	Range
Score	30	35.8	8.17	12	32	37	41	48	9	36
New Score (+5)										

DENSITY CURVE:

1. Always plot your data: make a graph.

2. Look for the overall pattern (shape, center, and spread) and for striking departures such as outliers.

3. Calculate a numerical summary to briefly describe center and spread.

4.

Area Under a density curve

A **<u>Density Curve</u>** is a curve that:

-

-

The ________________________________, within a given interval, gives the

____________________________ that fall within that interval.

Uniform Density

2.1 Assignment:

2.2 Modeling Distributions of Data

<u>Definition:</u>

A **Normal distribution** is described by a Normal density curve. Any particular Normal distribution is completely specified by two numbers: ____________________ and ____________________

•The __________________of a Normal distribution is the ___________________of the symmetric **Normal curve**.

•The __________________________________ is the distance from the center to the change-of-curvature points on either side.

•We abbreviate the Normal distribution with mean μ and standard deviation σ as $N(\mu,\sigma)$.

Normal Distributions are:

•

•

Note:

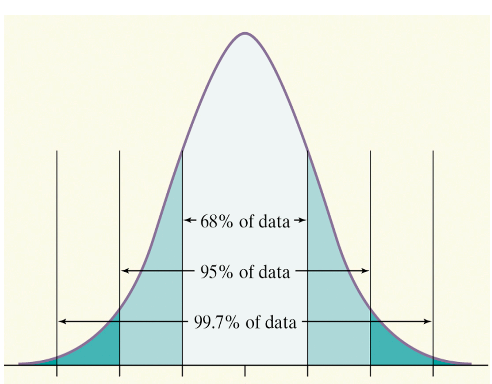
<u>Definition:</u> **The 68-95-99.7 Rule ("The Empirical Rule")**

In the Normal distribution with mean μ and standard deviation σ:

•Approximately ___________of the observations fall within __________ of ___________

•Approximately ___________ of the observations fall within __________ of ___________

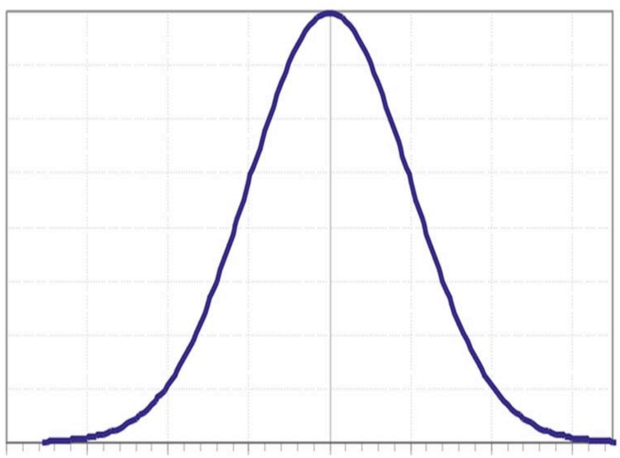
•Approximately ___________ of the observations fall within __________ of ___________



Example: Batting averages for Major League Baseball players in 2009, the mean of the 432 batting averages was 0.261 with a standard deviation of 0.034. Suppose that the distribution is exactly Normal with = 0.261 and = 0.034.

Problem:
(a)Sketch a Normal density curve for this distribution of batting averages.
(b)Label the points that are 1, 2, and 3 standard deviations from the mean.
(c) What percent of the batting averages are above 0.329? Show your work.

(d) What percent of the batting averages are between 0.193 and .295?



Standard Normal Distribution:

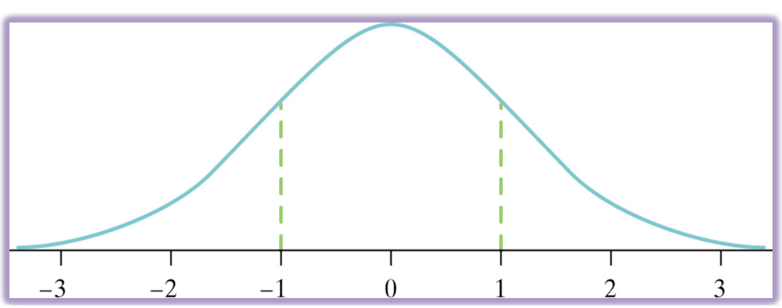
• The ______________________Normal distribution is the Normal distribution with mean

________and standard deviation __________

• If a variable x has any Normal distribution $N(\mu,\sigma)$ with mean μ and standard deviation σ, then the standardized variable

z =   has the standard normal distribution N(0, 1)

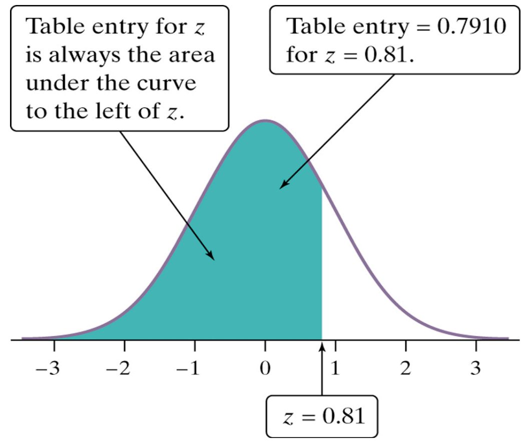
Standard Normal Distribution:



<u>Definition:</u> **The Standard Normal Table**
Table A :

Z	.00	.01	.02
0.7	.7580	.7611	.7642
0.8	.7881	.7910	.7939
0.9	.8159	.8186	.8212

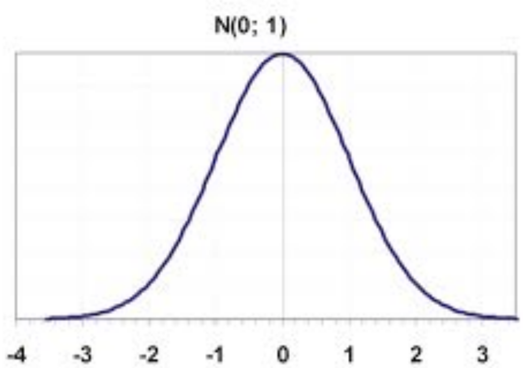
$P(z < 0.81) =$ ___________



How to solve a problem using Normal Distributions:

-
-
-
-

Example: Find the proportion of observations from the standard Normal distribution that are between -0.58 and 1.79.



State:

Plan:

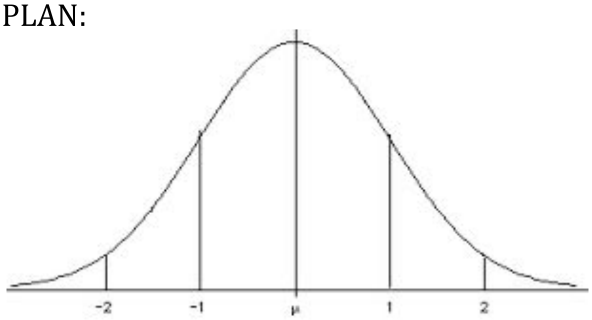
Do:

Conclude:

Example: In the 2008 Wimbledon tennis tournament, Rafael Nadal averaged 115 miles per hour (mph) on his first serves[1]. Assume that the distribution of his first serve speeds is Normal with a mean of 115 mph and a standard deviation of 6 mph. About what proportion of his first serves would you expect to exceed 120 mph?

STATE:

PLAN:



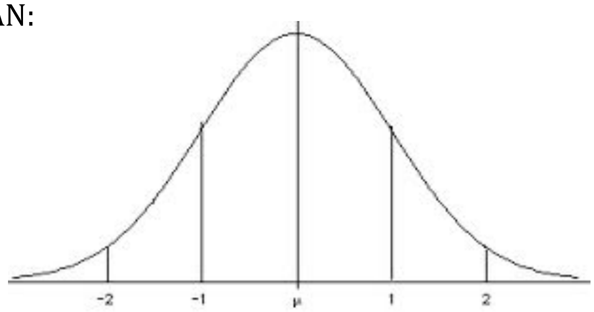
DO:

CONCLUDE:

Example: According to http://www.cdc.gov/growthcharts/, the heights of 3 year old females are approximately Normally distributed with a mean of 94.5 cm and a standard deviation of 4 cm. What is the third quartile of this distribution?

STATE:

PLAN:



DO:

CONCLUDE:

2.2 Assignment (day 1):

[1] http://sports.espn.go.com/sports/tennis/wimbledon08/columns/story?columnist=garber_greg&id=3472238

Assessing Normality:

- ______________________________

  Make a dotplot, stemplot, or histogram and see if the graph is approximately symmetric and bell-shaped.

- ______________________________

  Count how many observations fall within one, two, and three standard deviations of the mean and check to see if these percents are close to the 68%, 95%, and 99.7% targets for a Normal distribution.

Example: The measurements listed below describe the useable capacity (in cubic feet) of a sample of 36 side-by-side refrigerators. <source: Consumer Reports, May 2010> Are the data close to Normal?

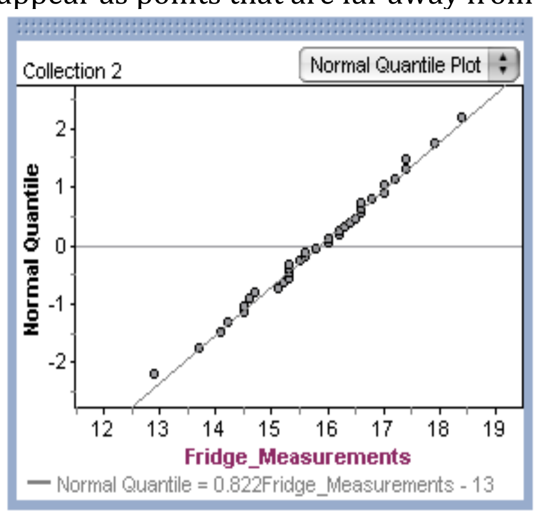
12.9 13.7 14.1 14.2 14.5 14.5 14.6 14.7 15.1 15.2 15.3 15.3
15.3 15.3 15.5 15.6 15.6 15.8 16.0 16.0 16.2 16.2 16.3 16.4
16.5 16.6 16.6 16.6 16.8 17.0 17.0 17.2 17.4 17.4 17.9 18.4

a) Use your graphing calculator to create a histogram of the data. Does the data appear Normal?

b) Use your graphing calculator to calculate the mean and standard deviation. Calculate what percent of data fall between +/- $1s_x$, +/- $2s_x$, +/- $3s_x$

Normal Probability Plot: If the points on a Normal probability plot lie close to a straight line, the plot indicates that the data are Normal. Systematic deviations from a straight line indicate a non-Normal distribution. Outliers appear as points that are far away from the overall pattern of the plot.

Refrigerator Example (from above):



2.2 Assignment (day 2):