A. South Carolina Standards

1. Investigate how a change in one variable relates to a change in a second variable.
2. Create charts and graphs.
3. Represent situations with number tables, graphs, and verbal descriptions.
4. Associate tables, graphs, and stories of the same event.
5. Understand measurable attributes of objects and their units, systems, and processes of measurement.
6. Identify equivalent relationships among fractions, decimals, and percents.

B. Identification of the Concepts

Statistics, Statistical Distributions, Components of Statistical Distributions, Normal Distributions, z-score, percentile, Skewed Distributions, Creating and Interpreting Graphs

C. Background Information for the Teacher

<u>What is Statistics?</u>

The field of statistics is concerned with the collection, description, and interpretation of data. (data are numbers obtained through measurement) In the field of statistics, the term “statistic” denotes a measurement taken on a sample (as opposed to a population). In general conversation, “statistics” also refers to facts and figures.

<u>What is a Statistical Distribution?</u>

A statistical distribution describes the numbers of times each possible outcome occurs in a sample. If you have 10 test scores with 5 possible outcomes of A, B, C, D, or F, a statistical distribution describes the relative number of times an A,B,C,D or F occurs. For example, 2 A’s, 4 B’s, 4 C’s, 0 D’s, 0 F’s.

<u>What are the Components of A Distribution?</u>

<u>Measures of Central Tendency</u>

(Suppose we have a sample with 4 observations: 4, 1, 4, 3)

Mean - the sum of a set of numbers divided by the number of observations.

Median - the middle point of a set of numbers (for odd numbered samples).
the mean of the middle two points (for even samples).

Mode - the most frequently occurring number. Mode=4 (4 occurs most).

$$\text{Mean} = \frac{4+1+4+3}{4} = \frac{12}{4} = 3$$

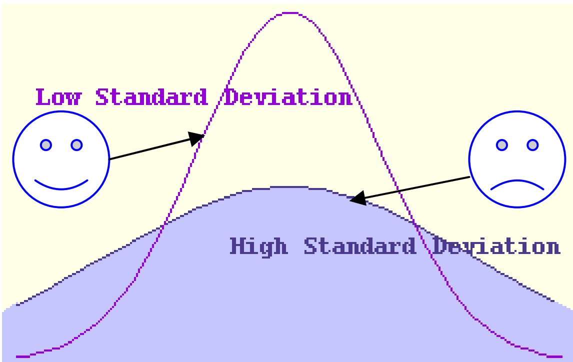
$$\text{Median}=1,\underline{3},\underline{4},4 \text{ or } \frac{3+4}{2} = \frac{7}{2} = 3.5$$

Measures of Variation

Range - the maximum value minus the minimum value in a set of numbers. Range = 4-1 = 3.

Standard Deviation - the average distance a data point is away from the mean.

$$\text{standard deviation} = \frac{|4-3|+|1-3|+|4-3|+|3-3|}{4} = \frac{1+2+1+0}{4} = \frac{4}{4} = 1$$



Hints for calculations:

1. Standard deviation: compute the difference between each data point and the mean. Take the absolute value of each difference. Sum the absolute values. Divide this sum by the number of data points. Median: first arrange data points in increasing order.

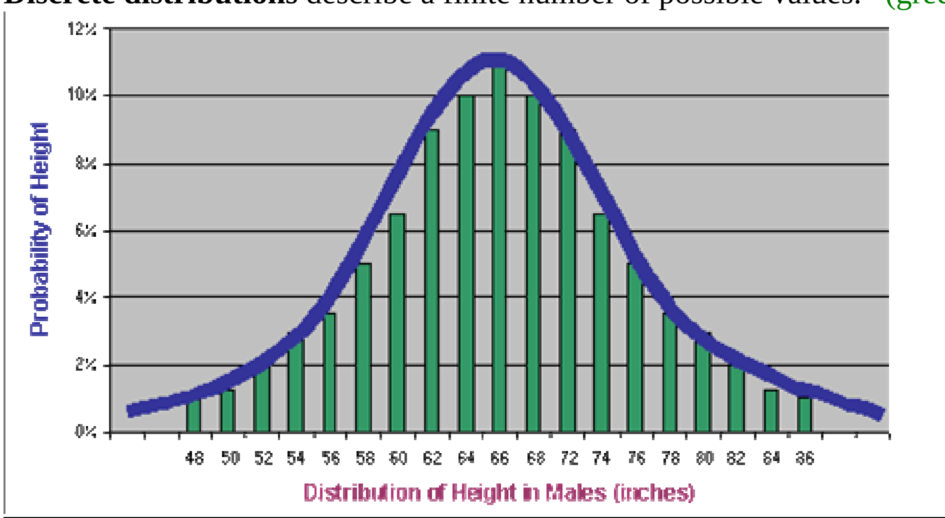
Mean, Median, Mode, Range, and Standard Deviations are measurements in a sample (statistics) and can also be used to make inferences on a population.

What is the Best Way to Show Graphically How Data Are Distributed?

1. Stem-and-Leaf Plots are useful for showing gaps, clusters, and outliers in distributions.
2. Bar graphs use bars to compare frequencies of possible data values.
3. Double bar graphs use two sets of bars to compare frequencies of data values between two levels of data (e.g. boys and girls)
4. Histograms use bars to show how frequently data occur within equal spaces within an intervals.
5. Stick Plots use sticks or skinny bars to compare frequencies among many data values.

What is the Difference between a Continuous and a Discrete Distribution?

Continuous distributions describe an infinite number of possible data values. (blue curve)
Discrete distributions describe a finite number of possible values. (green bars)



What is a Normal Distribution?

A normal distribution is a continuous distribution that is “bell-shaped”.   Data are often assumed to be normal. Normal distributions can estimate probabilities over a continuous interval of data values.

Which Data Values Are Most Likely to be Observed in a Normal Distribution?

In a normal distribution, data are most likely to be at the mean. Data are less likely to farther away from the mean. Are the people around more likely to be short, tall, or average in height?

What is a Standard Normal Distribution?

All normal distributions can be converted into a standard normal distribution**. A standard normal distribution** is a normal distribution with a mean=0 and standard deviation = 1.

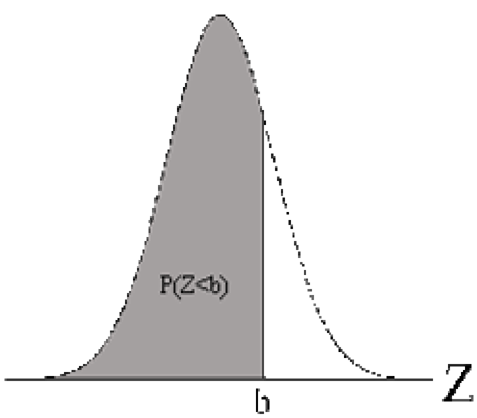
Why Convert to a Standard Normal Distribution?
The values for points in a standard normal distribution are **z-scores**. We can use a standard normal table to find the probability of getting at or below a z-score. (a percentile).

How do You Convert a Normal Distribution to a Standard Normal Distribution?

1. Subtract the mean from each observation in your normal distribution, the new mean=0.
2. Divide each observation by the standard deviation, the new standard deviation=1.

What is a Percentile?

A percentile (or **cumulative probability**) is the proportion of data in a distribution less than or equal to a data point. If you scored a 90 on a math test and 80% of the class had scores of 90 or lower; your percentile is 80. In the figure below, b=90 and $P(Z<b)=80$.



How do you compute probabilities in Discrete Distributions?

Suppose for your 10 tests you received 5 As, 2 Bs, 2 Cs, 1 D and want to find the probability of receiving an A or a B. Sum the frequencies for A and B and divide by the sample size. The probability of receiving an A or a B is (5+2)/10 = .7 (a 70% chance).

What are Skewed Distributions?

Skewed distributions are not symmetric. Imagine stretching out one side of a bell-curve. The stretched out side is considered skewed. Is the distribution of wealth skewed?

Activity I. Collecting Data and Computing Statistics

1. Commit to An Outcome

How well can you conceptualize height distributions? Remember standard deviation is how far away, on average, each height will be from the mean. I predict:

Name(1) ___________	Name(2) ___________	Name(3) ___________	Name(4) __________
My height = ____ in.	My height = ___ in.	My height = ___ in.	My height = ____ in.
The mean = ____ in.	The mean = ____ in.	The mean = ____in.	The mean = ____in.
Range = [__ , __] in.	Range = [__ , __] in.	Range = [__ , __] in.	Range = [__ , __] in.
Std. Dev. = ____ in.	Std. Dev. = ____ in.	Std. Dev. = ___ in.	Std. Dev.= ___ in.
Tallest ÷ Shortest =___	Tallest ÷ Shortest =___	Tallest ÷ Shortest =___	Tallest ÷Shortest=__

2. Expose Beliefs

Write down your predictions and your explanations and discuss them with your group.
Be prepared to discuss your final predictions and explanations.

3. Confront Beliefs

Form a group with either 4 boys or 4 girls. Let each group member measure your height. You should have 3 measurements of your height. Measurements should be to the nearest ¼ of an each.   Your mean estimate should be rounded to nearest inch.

Name(1) _________	Name(2) _________	Name(3) _________	Name(4) _________
Trial 1 = ____ in.	Trial 1 = ____ in	Trial 1 = ____ in.	Trial 1 = ____ in
Trial 2 = ____ in.	Trial 2 = ____ in.	Trial 2 = ____ in.	Trial 2 = ____ in
Trial 3 = ____ in.	Trial 3 = ____ in.	Trial 3 = ____ in.	Trial 3 = ____ in.
Total = _____ in.	Total = _____in.	Total = _____in.	Total = _____in.
Total÷3 = ____ in.	Total÷3 = ____in.	Total÷3 = ____in.	Total ÷ 3 = ____ in.
Mean = ______ in.	**Mean =___ in.**	**Mean=___ in.**	**Mean =___ in.**

Record your mean estimate and your classmates mean estimate on your student sheet. Compute the statistics listed on your student sheet. How do your calculations compare to your predictions?

4.   Accommodate and Extend the Concept

What do the computed statistics tell you about the distribution of data you collected? Are the median, mean, and mode close together in value? What does that tell us? Are measurements more likely to be near the mean, median, or mode.   Can you think of examples in your every day life in which computed statistics like these can be applied?

Activity II Find the Frequency Distribution of Heights

1. Commit to an Outcome

What will the distribution of your data look like? Do you expect any outliers or gaps?

If 6 equal intervals the range, predict how many heights fall within those intervals.

Which type of graph will show the distribution of heights best? Why?

2. Expose Beliefs

Write down your predictions and your explanations and discuss them with your group.
Be prepared to discuss your final predictions and explanations.

3. Confront Beliefs

Construct a stem-and-leaf plot. Is the stem-and-leaf plot appropriate for the data?

Count the number heights that fall in your defined intervals. Construct a histogram.

Construct a “stick plot” (a skinny bar chart) to represent the frequency of each distinct numerical outcome value.

Which plots characterize the data distribution best? Which did you predict?

4. Accommodate the Concept

How is the distribution of the data shaped? Does the distribution resemble a bell-curve.

Was the stem-and-leaf plot appropriate for the data? Why or why not?
Is it reasonable to split the stems in half in order to have more stems?

Could a line plot have been used to show the distribution of data? Why or why not?

5. Extend the Concept

What is the difference between a stick plot and a bar chart?

Can you think of a situation in which you may want to use a stem-and-leaf plot, a bar chart, a histogram, or a stick plot in your everyday life?

Activity III   Exact and Cumulative Probability Distributions

1. Commit to an Outcome

Name(1)____________________

Predict the following exact and cumulative probabilities. Cumulative probability is the probability of being at least a particular height.

The probability of being exactly 52 inches (4 ft. 2 in.) or (4'2") is _____ or ____%.
The probability of being exactly 55 inches (4'5") is _____ or ____%.
The probability of being exactly 58 inches (4'8") is _____ or ____%.

The probability of being at least 52 inches (4 ft. 2 in.) or (4'2") is _____ or ____%.
The probability of being at least 55 inches (4'5") is _____ or ____%.
The probability of being at least 58 inches (4'8") is _____ or ____%.

Name(2)____________________

The probability of being exactly 52 inches (4 ft. 2 in.) or (4'2") is _____ or ____%.
The probability of being exactly 55 inches (4'5") is _____ or ____%.
The probability of being exactly 58 inches (4'8") is _____ or ____%.

The probability of being at least 52 inches (4 ft. 2 in.) or (4'2") is _____ or ____%.
The probability of being at least 55 inches (4'5") is _____ or ____%.
The probability of being at least 58 inches (4'8") is _____ or ____%.

Name(3)____________________

The probability of being exactly 52 inches (4 ft. 2 in.) or (4'2") is _____ or ____%.
The probability of being exactly 55 inches (4'5") is _____ or ____%.
The probability of being exactly 58 inches (4'8") is _____ or ____%.

The probability of being at least 52 inches (4 ft. 2 in.) or (4'2") is _____ or ____%.
The probability of being at least 55 inches (4'5") is _____ or ____%.
The probability of being at least 58 inches (4'8") is _____ or ____%.

Name(4)____________________

The probability of being exactly 52 inches (4 ft. 2 in.) or (4'2") is _____ or ____%.
The probability of being exactly 55 inches (4'5") is _____ or ____%.
The probability of being exactly 58 inches (4'8") is _____ or ____%.

The probability of being at least 52 inches (4 ft. 2 in.) or (4'2") is _____ or ____%.
The probability of being at least 55 inches (4'5") is _____ or ____%.
The probability of being at least 58 inches (4'8") is _____ or ____%.

How can you answer these questions with your data.

Which type of plot can be used to graph exact probabilities for <u>all</u> measured heights? What are your dependent (y) and independent variables (x)?

Which type of plot can be used to graph exact probabilities for <u>all</u> measured heights? What are your dependent (y) and independent variables (x)?

Is the cumulative always larger than the exact probability? Why or why not?

2. Expose Beliefs

Write down your predictions and your explanations and discuss them with your group.
Be prepared to discuss your final predictions and explanations.

3. Confront Beliefs

On your student sheet, divide the frequency of each measurement by the number of students to get an exact probability of each possible height measurement.

Plot the exact probabilities of all possible measurements using a stick plot.

Find the cumulative probability of each measurement. This is the probability of being equal to or less than the measurement value. To find cumulative probability of each measurement: sum each exact probability with all the exact probabilities that precede it.

Plot the cumulative probabilities of all possible measurements using a stick plot.

How do your predictions compare to these plots?

4. Accommodate and Extend the Concept

Should the exact probabilities add up to 1? Should the largest cumulative probability be 1? Why?

How does a probability distribution compare to a frequency distribution?
Which do you prefer? Why?

Can cumulative probabilities ever decrease as measurements increase?

Suppose we rounded each measurement to a tenth of an inch (e.g. 50.1 inches instead of 50 inches). Would the probabilities for each measurement change?

How can you graph the probabilities of all possible measurements (e.g. all measurements with one or more decimal points)? How can you find the probability over a specific interval of these possible values?

Activity IV. Estimating Percentiles Using a Standard Normal Distribution

1. Commit to an Outcome

The percentiles computed in Activity III were computed from a discrete probability distribution. In this activity, we will estimate percentiles using a standard normal (continuous) distribution.

Will previously computed percentiles change?
If so, which percentiles do you expect to change the most?

Which sample statistics will be needed to estimate a standard normal distribution?

2. Expose Beliefs

Write down your predictions and your explanations and discuss them with your group.
Be prepared to discuss your final predictions and explanations.

3. Confront Beliefs

Compute the standard deviation. Compute the z-score for each height measurement on your student sheet. The z-score is the (Height measurement – Mean) ÷ standard deviation.
Use the attached table to compute the percentile for each z-score.

How do these estimated percentiles compare with the class percentiles previously computed?
Which sample statistics did we use?

4. Accommodate the Concept

A standard normal distribution is a normal distribution with mean =0 and standard deviation = 1.
Do the z-score values have a standard normal distribution?

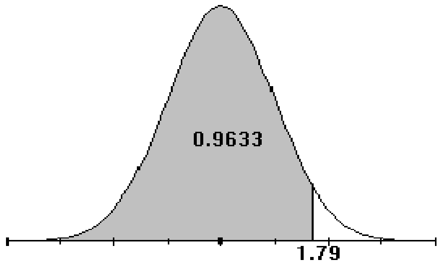
Height measurements were rounded. Was this necessary for computing the z-scores and estimated percentiles? Was this necessary for computing cumulative probabilities in Activity 3?

What assumption about the data did we make in computing the estimated percentiles? Could we make that assumption for weight measurements? (See attached charts)

5. Extend the Concept

We did not adjust for your age or gender and used rounded measurements? Did any of these 3 factors effect your percentile? The standard deviation? (See attached charts for help). What other factors would make our percentiles differ from the attached charts).

Table of Cumulative Probability of a Standard Normal Distribution



Total Area =1

Suppose my z-score is 1.79. I go to the 1.7 row and the .09 column to get 1.7 + .09 =1.79. I get a cumulative probability of .9633 or 96th percentile. Suppose my z-score is -1.79. My cumulative probability goes up to the same point, but on the opposite side of the distribution. My cumulative probability is 1-.9633= .0367 or 4th percentile.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5190	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7157	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7969	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8513	.8554	.8577	.8529	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9215	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9492	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	**.9633**
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952

2 to 20 years: Boys
Stature-for-age and Weight-for-age percentiles

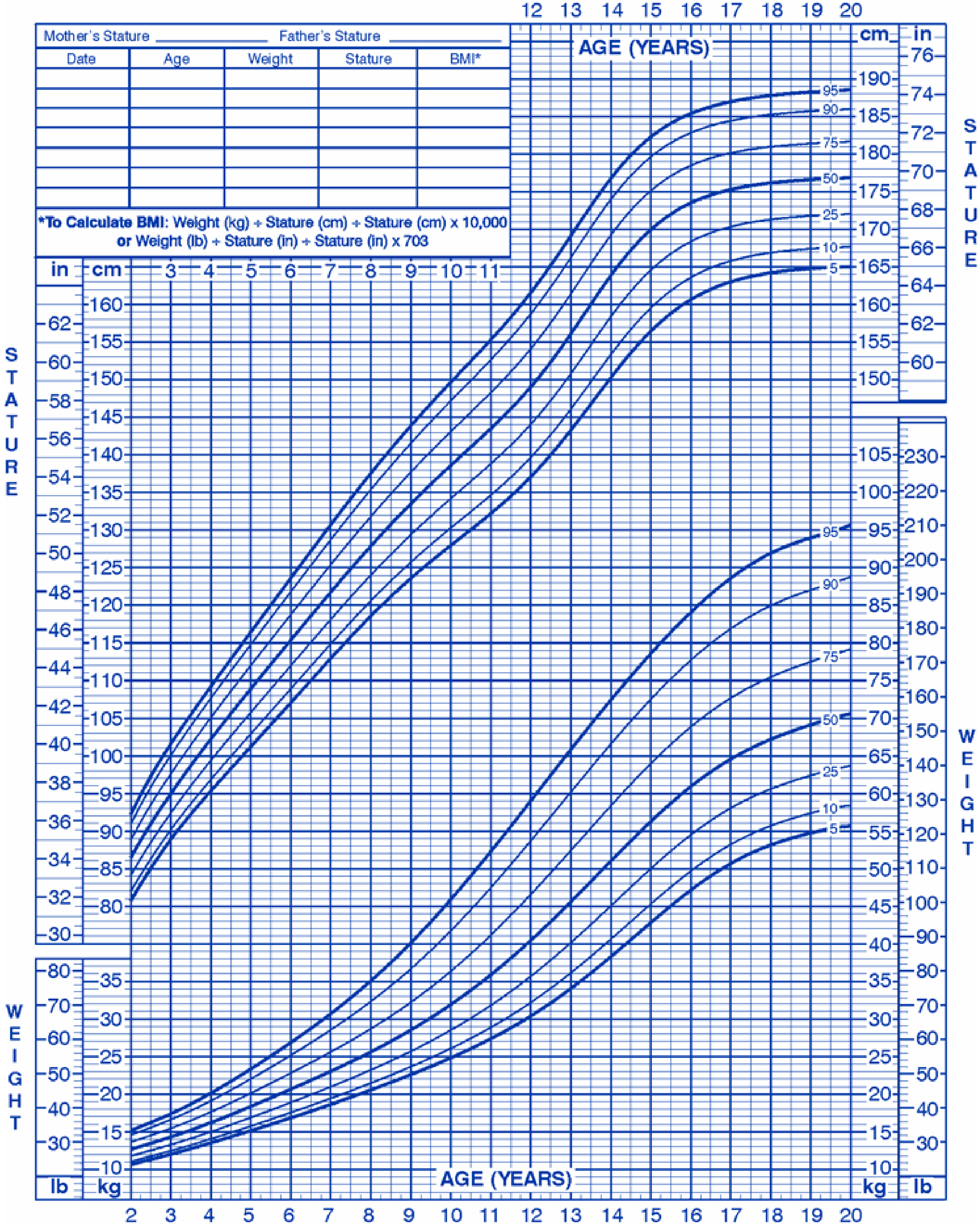
NAME ______________________

RECORD # ______________

Mother's Stature ____________ Father's Stature ____________

Date	Age	Weight	Stature	BMI*

***To Calculate BMI**: Weight (kg) ÷ Stature (cm) ÷ Stature (cm) x 10,000
or Weight (lb) ÷ Stature (in) ÷ Stature (in) x 703



Published May 30, 2000 (modified 11/21/00).
SOURCE: Developed by the National Center for Health Statistics in collaboration with the National Center for Chronic Disease Prevention and Health Promotion (2000).
http://www.cdc.gov/growthcharts

CDC
SAFER • HEALTHIER • PEOPLE™

2 to 20 years: Girls
Stature-for-age and Weight-for-age percentiles

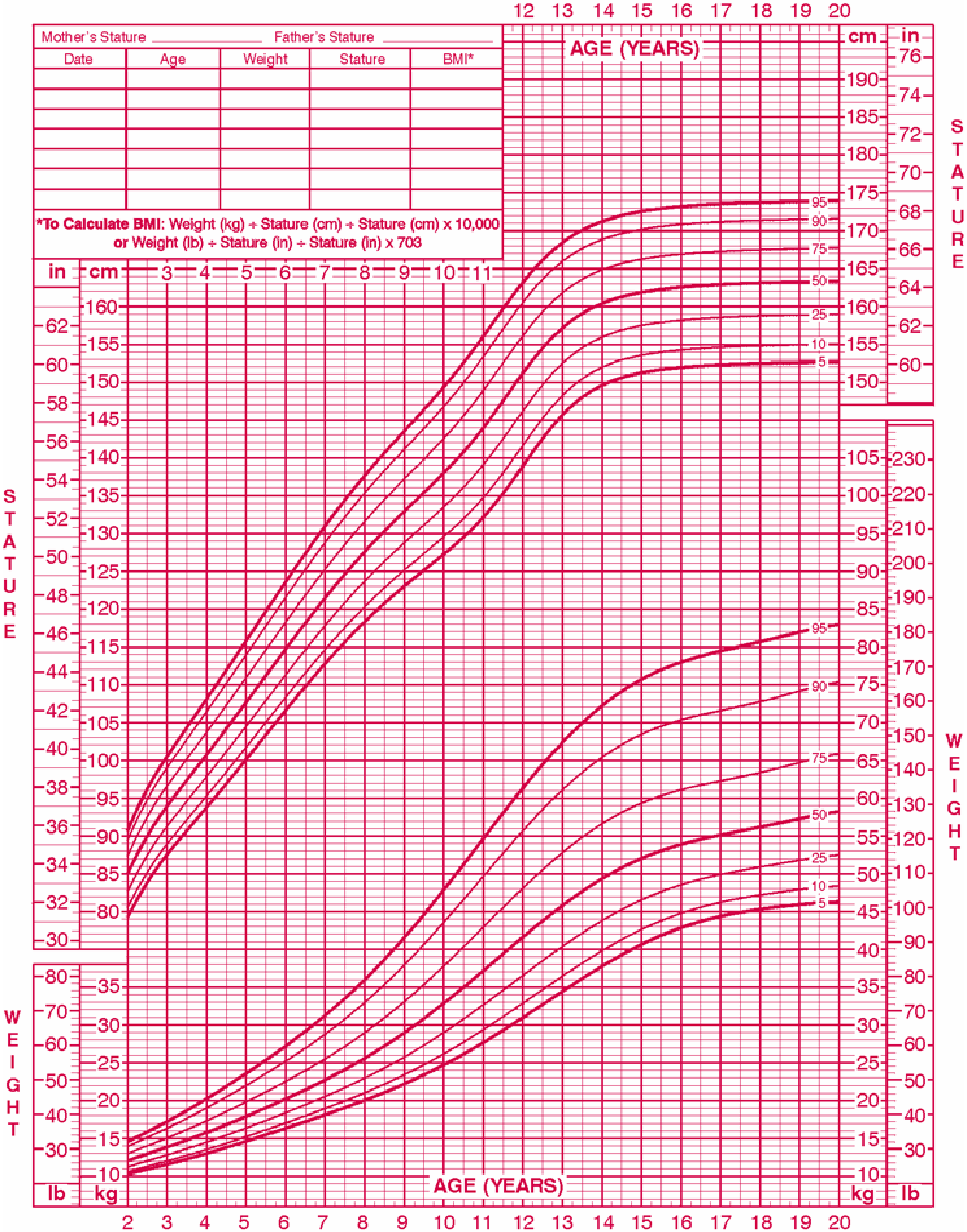
NAME ____________________

RECORD # ____________

Mother's Stature ____________ Father's Stature ____________

Date	Age	Weight	Stature	BMI*

***To Calculate BMI:** Weight (kg) ÷ Stature (cm) ÷ Stature (cm) x 10,000
or Weight (lb) ÷ Stature (in) ÷ Stature (in) x 703



Published May 30, 2000 (modified 11/21/00).
SOURCE: Developed by the National Center for Health Statistics in collaboration with
the National Center for Chronic Disease Prevention and Health Promotion (2000).
http://www.cdc.gov/growthcharts

SAFER • HEALTHIER • PEOPLE™