Chapter 5: The standard deviation as a ruler and the normal model p131

Which is the better exam score?

- 67 on an exam with mean 50 and SD 10
- 62 on an exam with mean 40 and SD 12?

Is it fair to say:

- 67 is better because 67 > 62?

- 62 is better because it is 22 marks above the mean and 67 is only 17 marks above the mean?

Key: *z-scores*.

Effect of a Linear Transformation on summary statistics

- Multiplying each observation in a data set by a number *b* multiplies the mean, median, by *b* and the measures of spread (standard deviation, *IQR*) by *abs(b)* .
- Adding the same number *a* to each observation in a data set adds *a* to measures of center, quartiles, percentiles but does not change the measures of spread.

Standardizing and *z*-scores p132

If *x* is an observation from a distribution that has mean μ and std. dev. σ , the standardized value of *x* is

$z = \frac{x - \mu}{\sigma}$ A standardized value is often called a *z*-score.

A *z*-score tells us how many standard deviations the original observation falls away from the mean.

What are the mean and SD **of z**?

No matter what mean and SD x has, z has mean 0, SD 1.

- Calculating a z-score sometimes called "standardizing". Above says why.
- Gives a basis for comparison for things with different means and sds.

Those exam scores above:

Which is the better exam score?

- 67 on an exam with mean 50 and SD 10
- 62 on an exam with mean 40 and SD 12?

Turn them into z-scores:

- 67 becomes (67-50)/10=1.70
- 62 becomes (62-40)/12=1.83

so the 62 is a (slightly) better performance, relative to the mean and SD.

Density Curve and the Normal Model (p137)

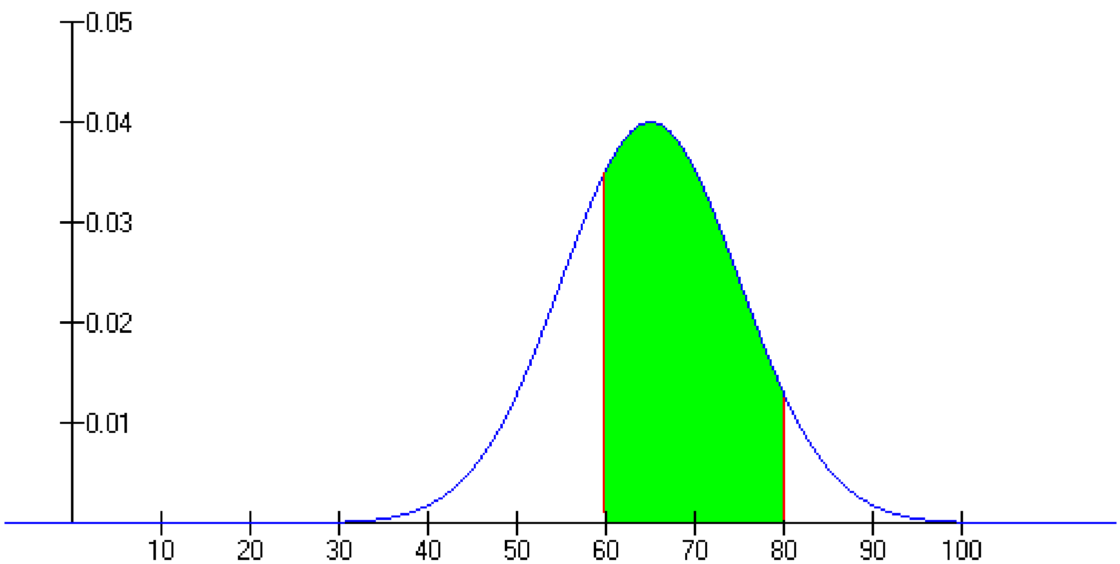
Density curve is a curve that

- is always on or above the horizontal axis.

- has area exactly 1 underneath it.

- A density curve describes the overall pattern of a distribution.

Example: The curve below shows the density curve for scores in an exam and the area of the shaded region is the proportion of students who scored between 60 and 80.



Normal distributions p139

- An important class of density curves are the symmetric unimodal bell-shaped curves known as **normal curves**. They describe **normal distributions**.

- All normal distributions have the same overall shape.

- The density curve for a particular normal distribution is specified by giving the mean μ and the standard deviation σ.

- The mean is located at the center of the symmetric curve and is the same as the median (and mode).

The standard deviation σ controls the spread of a normal curve.

•There are other symmetric bell- shaped density curves that are not normal.

•The normal density curves are specified by a particular equation. The height of a normal density curve at any point *x* is given by

$$y=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Roma tomatoes have weights that have a normal distribution with mean 74 grams and SD 2.5 grams. What proportion of these tomatoes will weigh less than 70 grams?

Ans z=(70-74)/2.5=-1.60;
look up -1.60 in table Z to get 0.0548.

Roma tomatoes again (mean 74, SD 2.5): What proportion less than 77.4 grams?

And z=(77.4-74)/2.5=1.36

0.9131

What proportion of the Roma tomatoes in the previous question will weigh more than 80 grams? (Mean 74, SD 2.5.)

Ans $z=(80-74)/2.5=2.40$

Table: 0.9918 *less*

so $1-0.9918=0.0082$

What proportion of the Roma tomatoes of the previous two questions will weigh between 70 and 80 grams? (There are two ways to do this, both of which use the previous work.)

Ans : 70 as z-score is -1.60, table gives 0.0548.

80 as z-score is 2.40, table gives 0.9918.

Subtract: 0.9918-0.0548=0.9370

Getting values from proportions

- Use Table Z backwards to get z that goes with proportion less
- Turn z back into original scale by using x=mean + SD * z.
- Why? z=(x-mean)/SD, solve for x

Newborn babies in Canada have weights that follow a normal distribution, with mean 3500 grams and SD 500 grams. (The mean is a little less than 8 pounds.)

A baby is defined as being "high birth weight" if it is in the top 2% of birth weights. What weight would make a baby "high birth weight"?

Ans 2% more = 98% less = 0.9800 less

z=2.05 (closest)

weight = 3500 + 500*2.05 = 4525 **grams** (or more)

A baby is defined as being "very low birth weight" if it is in the bottom 0.1% of birth weights. What weight would make a baby "very low birth weight"?

0.1% less = 0.0010 less

z = -3.09 (I picked the middle one)

weight = 3500 + 500*(-3.09) = 1955 **grams** (or less)

Normal quantile plots p148

- A histogram or stem plot can reveal distinctly nonnormal features of a distribution.

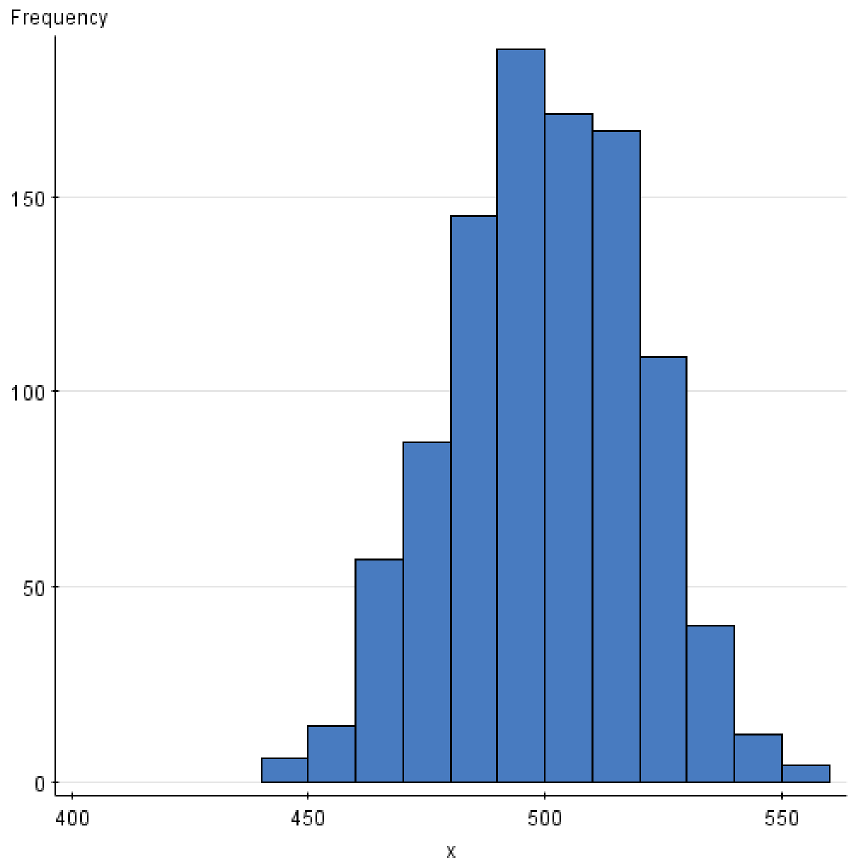
- If the stemplot or histogram appears roughly symmetric and unimodal, we use another graph, the normal quantile plot as a better way of judging the adequacy of a normal model

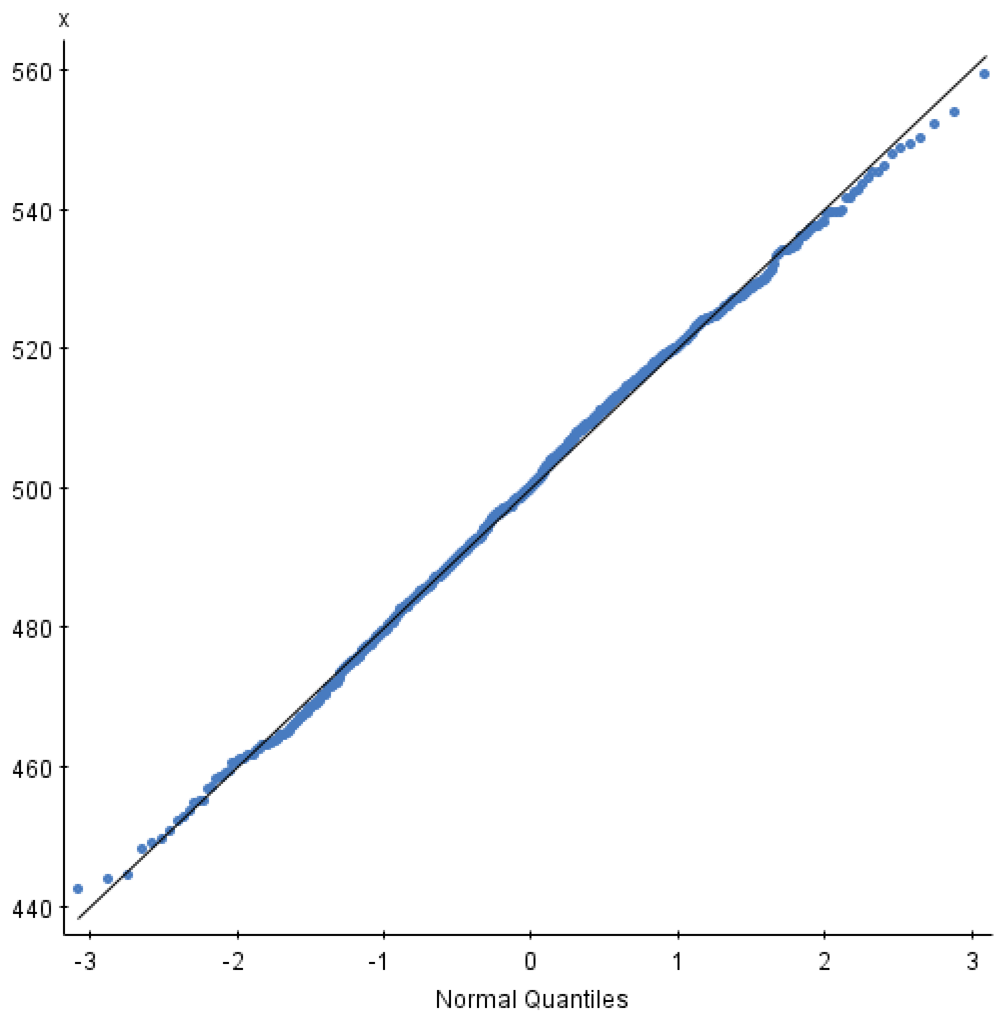
- **Use of normal quantile plots.**

If the points on a normal lie close to a straight line, the plot indicates that the data are normal.

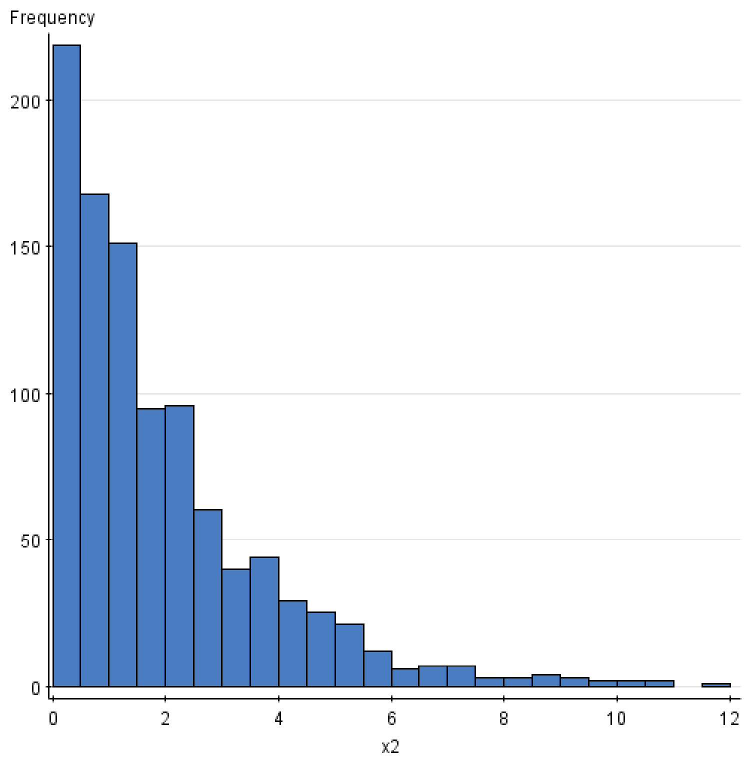
Outliers appear as points that are far away from the overall pattern of the plot.

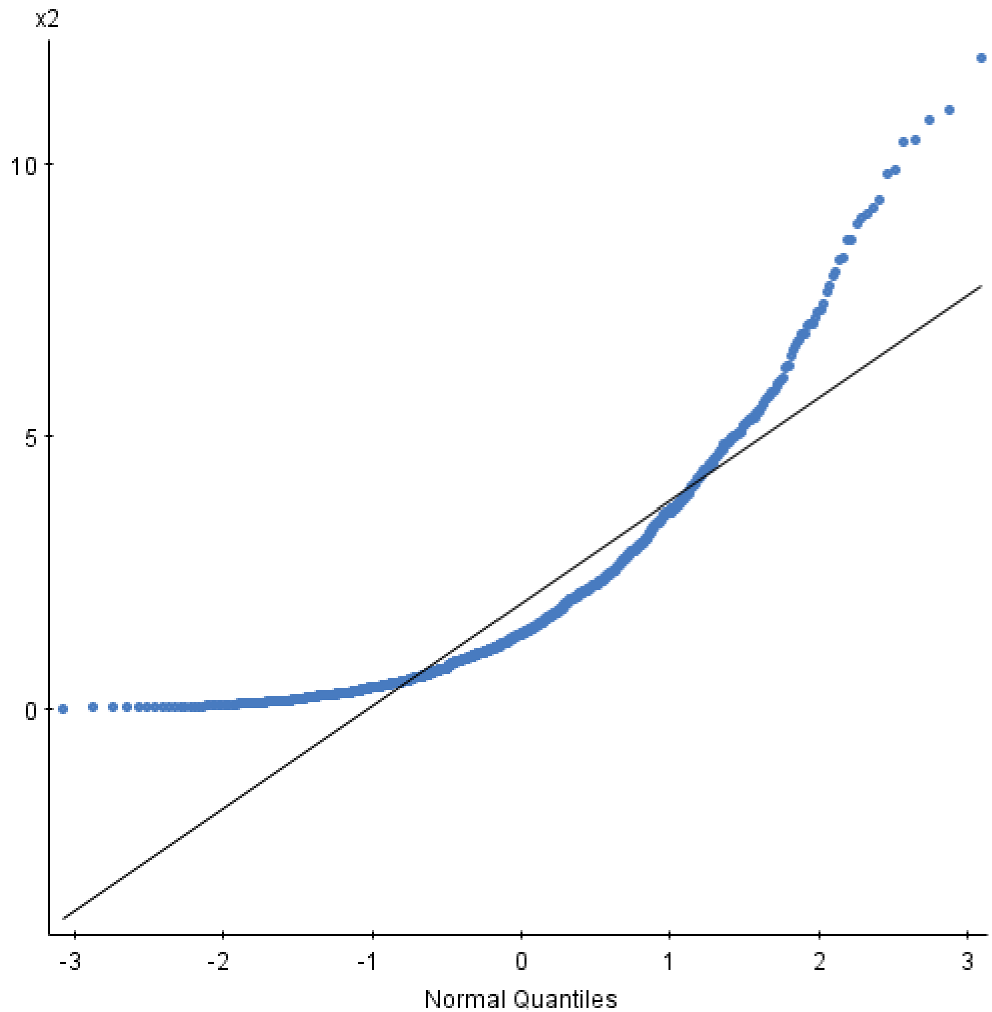
Histogram and the nscores plot for data generated from a normal distribution (*N*(500, 20)) (for 1000 observations)



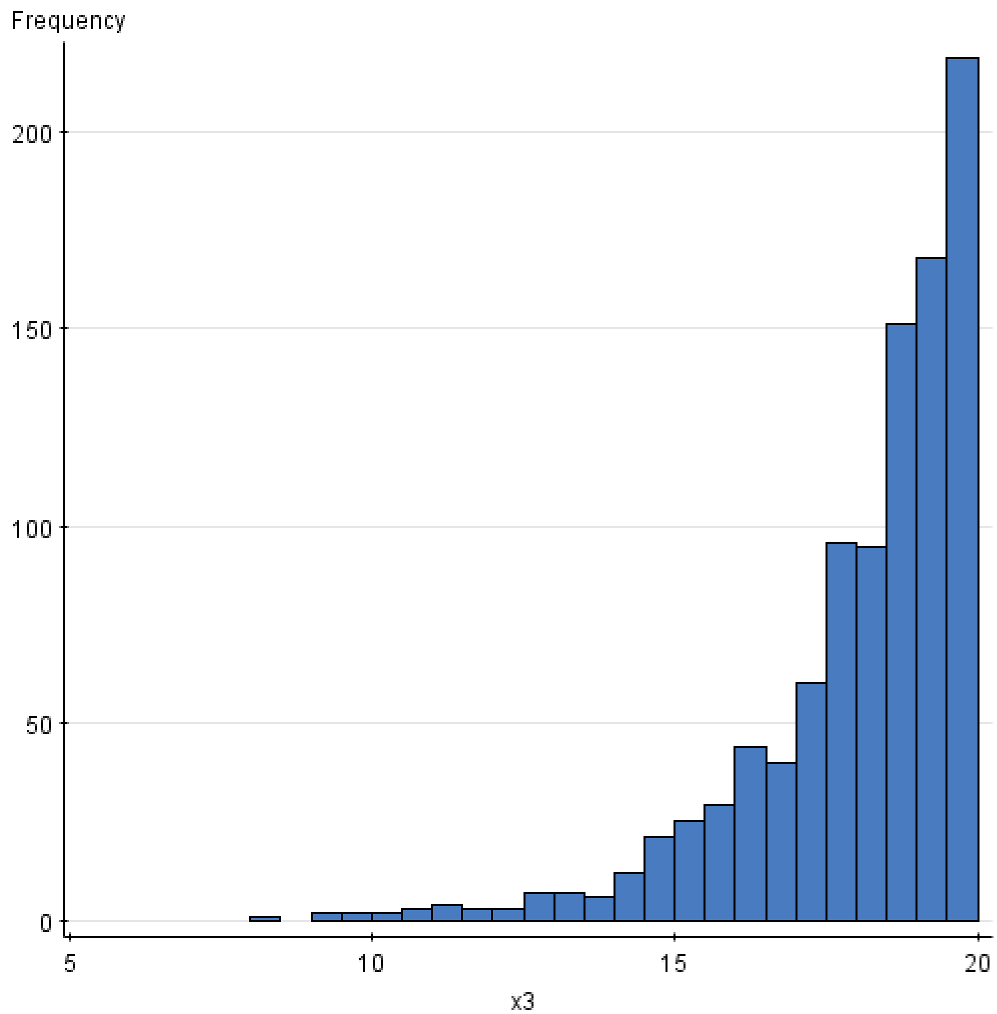

x
560
540
520
500
480
460
440
-3
-2
-1
0
1
2
3
Normal Quantiles

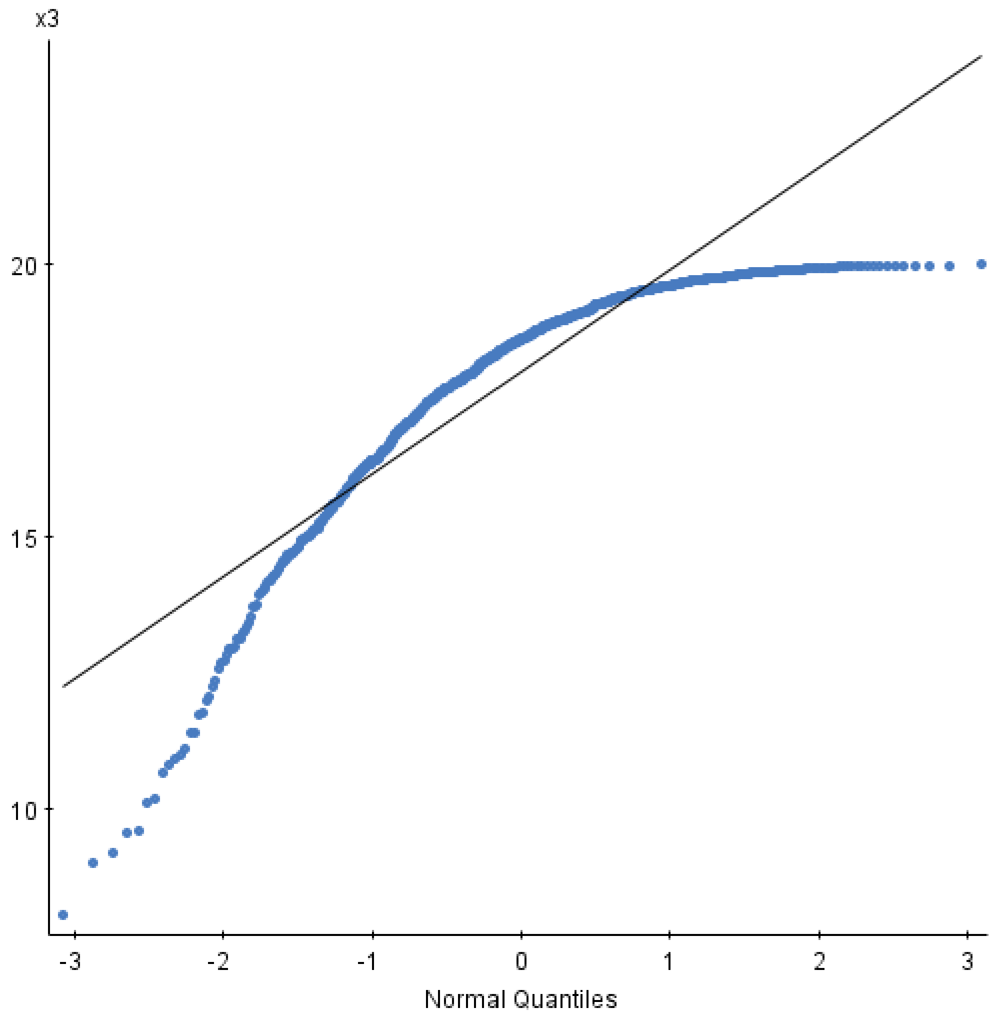
Histogram and the nscores plot for data generated from a right skewed distribution




x2
10
5
0
-3
-2
-1
0
1
2
3
Normal Quantiles

Histogram and the nscores plot for data generated from a left skewed distribution




x3
20
15
10
-3
-2
-1
0
1
2
3
Normal Quantiles

- **The 68-95-99.7 rule p141**

In the normal distribution with mean m and std. deviation s,

- Approx. 68% of the observations fall within one standard deviation of the mean.
- Approx. 95% of the observations fall within two standard deviations of the mean .
- Approx. 99.7% of the observations fall within 3 standard deviations of the mean.

- Example

The distribution of heights of women aged 18-24 is approx. normal with mean μ = 64.5 inches and std. dev.

σ = 2.5 inches.

2σ = 5 inches. The 68-95-99.7 rule says that the middle 95% (approx.) of women are between 64.5-5 to 64.5+5 inches tall.

The other 5% have heights outside the range from 64.5-5 to 64.5+5 inches .

2.5% of the women are taller than 64.5+5 .

Ex. 1) The middle 68% (approx.) of women are between ____ to ___ inches tall.

2) ___% of the women are taller than 67

3) ___% of the women are taller than 72