Lecture Notes & Examples 2.2 Part A

Section 2.2 –Normal Distributions

Learning Targets

- Use the 68–95–99.7 rule to estimate (i) the proportion of values in a specified interval, or (ii) the value that corresponds to a given percentile in a Normal distribution.
- Find the proportion of values in a specified interval in a Normal distribution using Table A or technology.
- Find the value that corresponds to a given percentile in a Normal distribution using Table A or technology.

Probably the most famous of all *density curves* are **Normal curves**. The distributions they describe are called **Normal distributions.** They play a very large part in statistics.

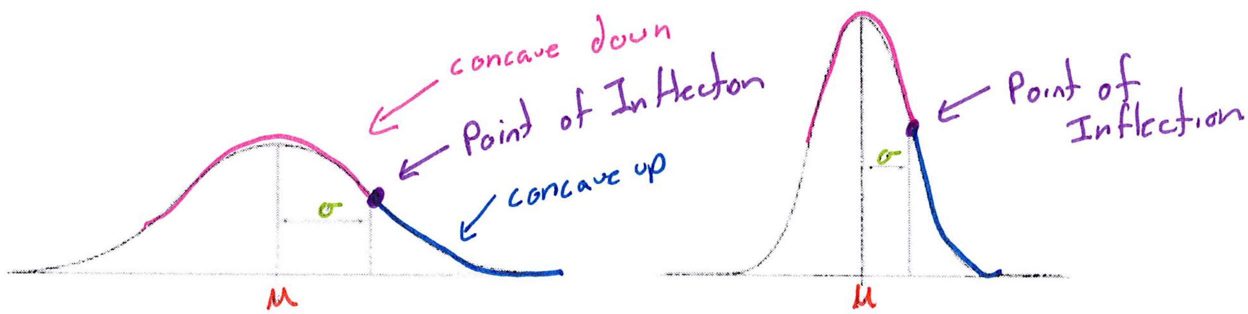


FIGURE 2.11 Two Normal curves, showing the mean μ and standard deviation σ.

Normal curves have several properties:

- All Normal curves have the same overall shape: symmetric, single-peaked, bell-shaped.
- Any specific Normal curve is completely described by its mean μ and standard deviation σ.
- The mean is located at the center and is equal to the median. Changing μ without changing σ moves the Normal curve along the horizontal axis without changing its shape.
- The standard deviation σ controls the spread of a Normal curve. Normal curves with larger standard deviations are more spread out.

* The points at which the Normal curve changes from *concave down* to *concave up* occurs one standard deviation from the mean. Because of this, the standard deviation can be estimated by the graph.

(Inflection Point)

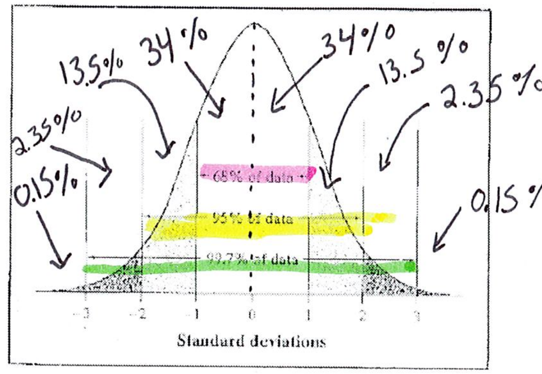
Definition: A **Normal distribution** is described by a Normal density curve. Any particular Normal distribution is completely specified by its mean μ and standard deviation σ. The mean of a Normal distribution is at the center of the symmetric **Normal curve** and equals the median. The standard deviation is the distance from the center to the inflection points (where concavity changes) on either side.

Notation: We abbreviate the Normal distribution with mean μ and standard deviation σ as ***N*(μ, σ)**.

The 68-95-99.7 Rule

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

- Approximately 68% of the observations fall within 1 σ of the mean μ.
- Approximately 95% of the observations fall within 2 σ's of the mean μ.
- Approximately 99.7% of the observations fall within 3 σ's of the mean μ.

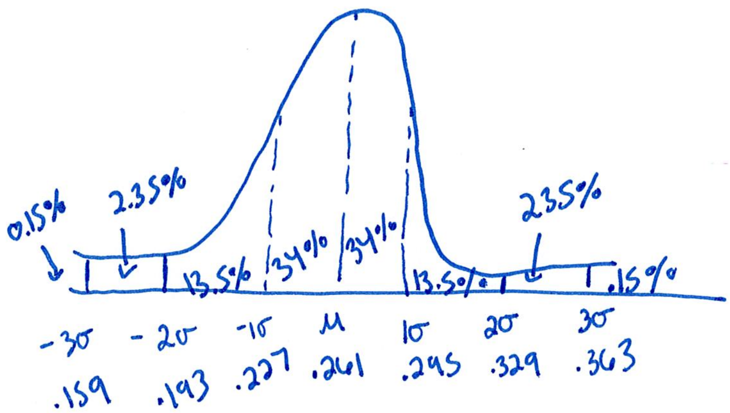


** (Note: this rule does not apply to any distribution – only the Normal. Common error on AP Exam.) **

Example: The mean batting average for the 432 Major League Baseball players in 2009 was 0.261 with a standard deviation of 0.034. Suppose the distribution is exactly Normal with $\mu = 0.261$ and $\sigma = 0.034$.

$N(0.261, 0.034)$

a. Sketch a Normal density curve for this distribution. Label the points that are 1, 2, and 3 standard deviations from the mean.



b. What percent of batting averages are above 0.329?

2.35% + .15% = 2.5%

(You could also think of this problem as 95% within 2σ, so 100 - 95% = 5% so 5% past 2σ, and $5\%/2 = 2.5\%$)

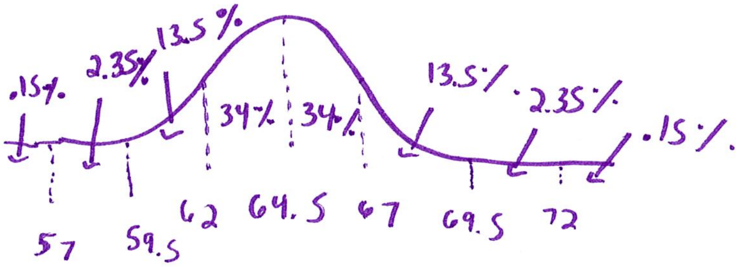
c. What percent of batting averages are between 0.193 and 0.295?

13.5% + 68% = 81.5%

CHECK YOUR UNDERSTANDING

The distribution of heights of young women aged 18 to 24 is approximately N(64.5, 2.5).

1. Sketch a Normal density curve for the distribution of young women's heights. Label the points one, two, and three standard deviations from the mean.



2. What percent of young women have heights greater than 67 inches? Show your work.

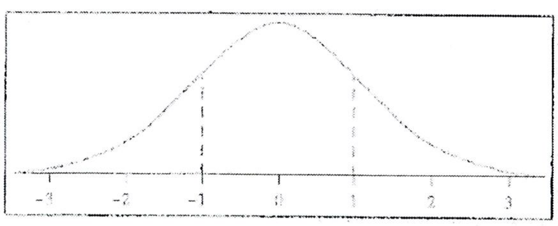
$P(\text{heights} > 67) = \frac{100-68}{2} = \frac{32}{2} = 16\%$ or $P(\text{heights} > 67) = 13.5 + 2.35 + .15 = 16\%$

3. What percent of young women have heights between 62 and 72 inches? Show your work.

$P(62 < h < 72) = 68\% + 13.5\% + 2.35\% = 83.85\%$

The Standard Normal Distribution (Distribution of z-scores)

Definition: The **standard Normal distribution** is the Normal distribution with mean 0 and standard deviation 1. If a variable x has any Normal distribution $N(\mu, \sigma)$ with mean μ and standard deviation σ, then the standardized variable $z = \frac{x-\mu}{\sigma}$ has the standard Normal distribution.



68-95-99.7 Rule: For the standard Normal distribution

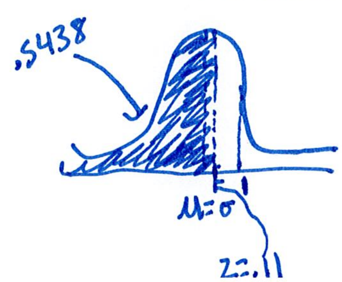
68 % between $\pm 1\sigma$
95 % between $\pm 2\sigma$
99.7 % between $\pm 3\sigma$

The **standard Normal table** is contained in Table A. It is a table of areas under the Normal curve. The table entry for each value z is the area under the curve to left of z. This is also known as the *lower tail*.

$P(z < .11) = .5438$

Table A (Continued)

z	.00	.01	.02
0.0	.5000	.5040	.5080
0.1	.5398	.5438	.5478
0.2	.5793	.5832	.5871
0.3	.6179	.6217	.6255



Draw Pictures!

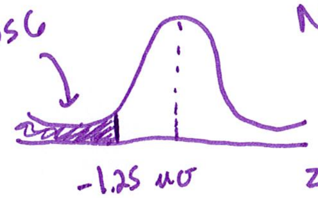
Example: Finding areas under the standard Normal curve.
Use *Table A* to find the proportion of observations from the standard Normal distribution given the following z-values. Draw a diagram for each.

Remember in Standard Normal Distribution

mean: $\mu = 0$

stand. dev. $\sigma = 1$

a. Less than z = -1.25



N(0,1)

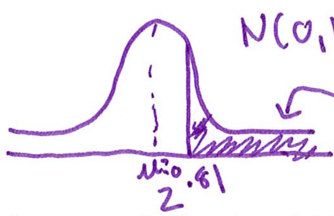
$= \boxed{.1056}$

b. Less than z = 0.81

N(0,1)

$\boxed{=.7910}$

c. Greater than z = 0.81



N(0,1)

$1 - .7910 = \boxed{.2090}$

d. Between z = -1.25 and z = 0.81

N(0,1)

$z < 0.81 = .7910$

$z < -1.25 = .1056$

$.7910 - .1056 = .6854$

Example: Repeat the previous example using *technology*.

normalcdf = normal cumulative density function

TI-84

a. Less than z = -1.25

normalcdf(-1000, -1.25, 0, 1) ≈ .1056

$\boxed{2nd}$ $\boxed{Vars}$ (Distr)

b. Less than z = 0.81

#2: normalcdf

Lower Bound

Upper Bound

μ (mean)

σ (stand deviation)

normalcdf(-1000, .81, 0, 1) ≈ .7910

c. Greater than z = 0.81

normalcdf(.81, 1000, 0, 1) ≈ .20897

d. Between z = -1.25 and z = 0.81

normalcdf(-1.25, .81, 0, 1) ≈ .6854

lower bound: -1E99, or -1000

upper bound: 1E99 or 1000

to get E type $\boxed{2nd}$ $\boxed{,}$ (EE)

TI-Inspire

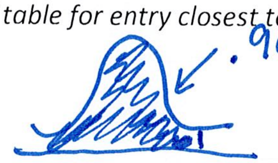
Menu, 6, 5, 2

* when using calculator you must draw pictures *

Example: Working backwards.....
Find the 90th percentile of standard Normal distribution

a) Using *Table A (Look in the body of table for entry closest to .90.)*

This is the entry corresponding to



N(0,1)

z ≈ 1.28

b. Using *technology*

invNorm(.9, 0, 1) = 1.282

z ≈ 1.282

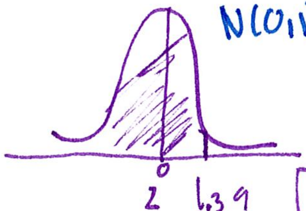
* when finding a corresponding data value to a give proportion, use InvNorm(%, μ, σ)

$\boxed{2nd}$ $\boxed{Vars}$ #3 InvNorm

CHECK YOUR UNDERSTANDING

Use the z table (Table A in the back of the book) to find the proportion of observations from a standard Normal distribution that fall in each of the following regions. In each case, sketch a standard Normal curve and shade the area representing the region.

1. $z < 1.39$

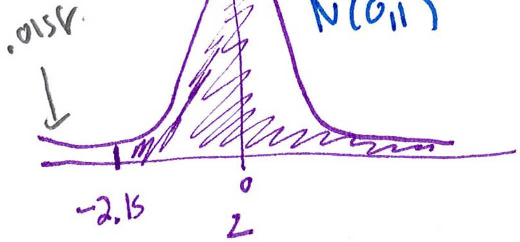


N(0,1)

$\boxed{.9177}$

2. $z > -2.15$

so $z < -2.15$ is .0158

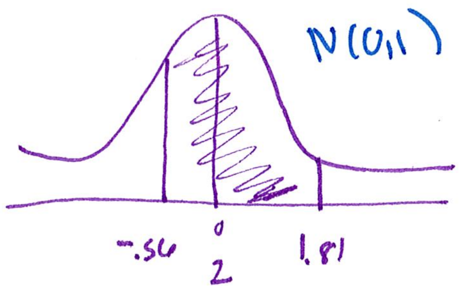


N(0,1)

$\boxed{\text{so } z > -2.15 \text{ is } 1 - .0158 = .9824}$

3. $-0.56 < z < 1.81$

$z < 1.81 = .9649$



N(0,1)

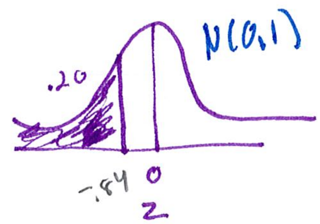
$z < -.56 = .2877$

$\boxed{.9649 - .2877 = .6772}$

Work Backwards

4. The 20th percentile (Look in body of table for entry closest to .20)

$\boxed{\text{z-score for 20th percentile} \approx -.84}$



N(0,1)

5. 45% of all observations are greater than z

This is the z score at the 55th percentile

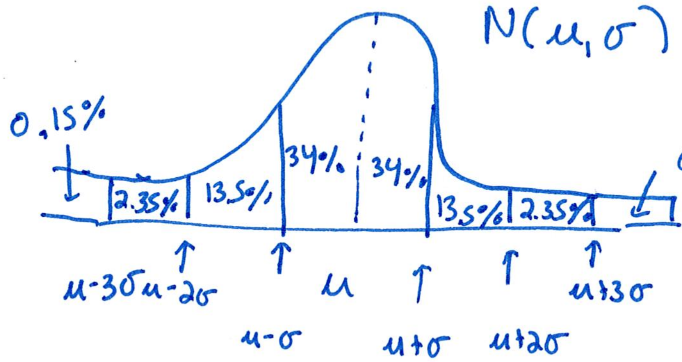
N(0,1)

$\boxed{\text{z score} \approx 0.13}$

Summarize Big Ideas:

Learning Target #1:

In a Normal Distribution the 68-95-99.7 rule applies to estimate proportions.



Learning Target #2:

Finding Proportion of values in a specified Interval Using Table A
- Convert data value to z-score (standardize data value)
- Draw curve, Find z-score on table and find corresponding percentage
- Table gives proportion to the left of z-score

Finding Proportion of values in a specified Interval using Technology
- normalcdf (lower bound, upper bound, mean μ, standard deviation σ)
  must always draw picture when using technology

Learning Target #3:

Find value that corresponds to given percentile in Normal Distribution
- Using Table A, look in the body of Table and find proportion that is the closest to the one given; then match up to corresponding z-score, then unstandardize if necessary
- Use Technology: InvNorm(proportion) = corresponding data value

Lecture Notes & Examples 2.2 Part B

Section 2.2 Normal Distributions (continued)

******Normal Distribution Calculations******

We will use the previous procedures to answer questions about observations in *any* Normal distribution by *standardizing* and then using the standard Normal table.

4-Step Process

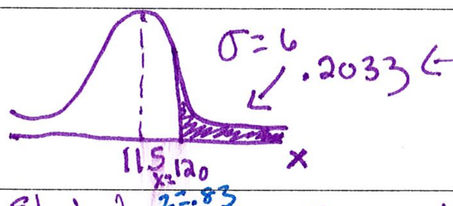
1. *State*: Express the problem in terms of the observed variable x.
2. *Plan*: Draw a picture of Distribution and shade area of interest
3. *Do*: Perform Calculations
   - Standardize x to restate problem in terms of z
   - Use Table A and the fact that total area under curve is 1 to find required area of interest.
4. *Conclude*:

   Write conclusion in context of problem

Example: In the 2008 Wimbledon tennis tournament, Rafael Nadal averaged 115 miles per hour on his first serves. Assume that the distribution of his first serves 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?

1. State: Let x = speed of 1st serve. X has N(115, 6). We want the proportion of serves ≥ 120 mph.

2. Plan:



3. Do: Standardize 120: $z_{120} = \frac{120-115}{6} = \frac{5}{6} = .83$

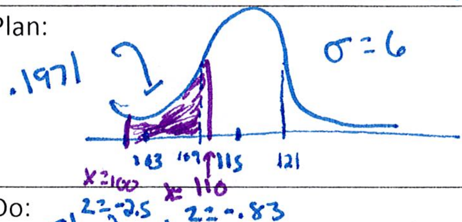
Table A $z < .83 = .7967$

so $z > .83 = 1 - .797 = .2033$

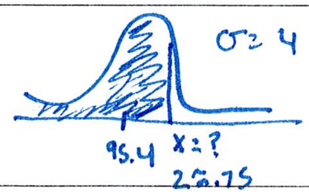
4. Conclude: About 20 % of Nadal's 1st serves exceed 120mph

* proportion of observations with $x \geq 120$ is the same as $x > 120$ is true when using any density curve.

Example continued: What percent of Rafael Nadal's first serves are between 100 and 110 mph?

1. State:	Let x = speed of 1st serve. X has $N(115, 6)$. We want proportion of 1st serves with $100 < x < 110$ mph.
2. Plan:	
3. Do:	Standardize: $z_{100} = \frac{100-115}{6} = \frac{-15}{6} = -2.5$ so for $100 < x < 110$ is ⇒ $z_{110} = \frac{110-115}{6} = \frac{-5}{6} = -.83$ ↗ $-2.5 < z < -.83$ Table A $z < -2.5 = .0062$ $z < -.83 = .2033$ so for $-2.5 < z < -.83$ = $.2033 - .0062 = .1971$
4. Conclude:	About 19.71 % of Nadals serves are between 100 and 110 mph.

Example: According to the Centers for Disease Control (CDC), the heights of three-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?

1. State:	Let X = height of randomly selected 3 year old female X has $N(95.4, 4)$. What is the 3rd quartile of this distribution.
2. Plan:	
3. Do:	Using Table A, the closest value to 0.75 is .7486. This corresponds to z-score 0.67. Now unstandardize to find x. $0.67 = \frac{x - 94.5}{4}$ $x = 97.18$ cm
4. Conclude:	The 3rd quartile of 3 y.o. females' heights is 97.18 cm

Pg 124 AP Tip About Calculator Speak

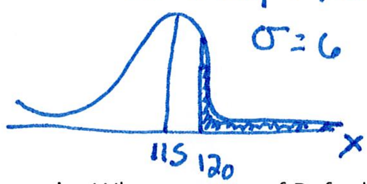
need to choose an upper bound many σ from mean

****Normal Distribution Calculations with Technology****

must sketch curve if using calc to show work!

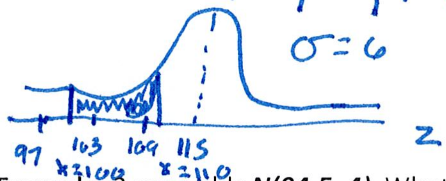
Example: Nadal *N(115, 6)*. Find the proportion of first serves we expect to exceed 120 mph.

Normal CDF (120, 10,000, 115, 6) = .202328 => 20 % of the time

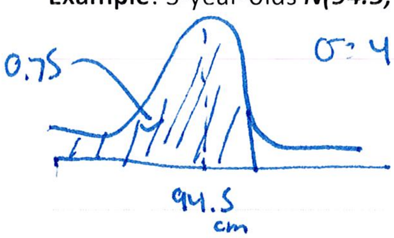


Example: What percent of Rafael Nadal's first serves are between 100 and 110 mph?

Normal CDF (100, 110, 115, 6) = .1961186447 => 19.6 % of the time



Example: 3-year-olds *N(94.5, 4)*. What is the third quartile of this distribution?



InvNorm (Prob, μ, σ)

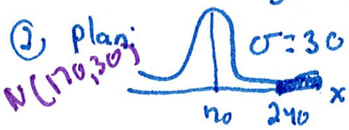
InvNorm (.75, 94.5, 4) = 97.19795 cm

Check Your Understanding. Use the 4-Step Process for each of these. Include a properly labeled diagram.

1. Cholesterol levels in 14-year-old boys is approximately Normally distributed with a mean of 170 mg/dl of blood and standard deviation 30 mg/dl. What percent of 14-year-old boys have more than 240 mg/dl of cholesterol?

① State: Let x = cholesterol levels in 14 y.o. boys. X has N (170, 30). what % 14 y.o. boys have > 240 mg/dl?

② Plan:



③ Do: Normal CDF (240, 10000, 170, 30) From calc, area = .009815

④ Conclude: About .98 % of 14 y.o. boys have cholesterol > 240 mg/dl.

2. What percent of 14-year-old boys have blood cholesterol between 200 and 240 mg/dl?

① Same as above, % between 200 and 240

② Plan:

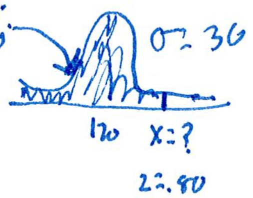
③ Do: Normal CDF (200, 240, 170, 30)
From calculator area = .1488

④ Conclude: About 14.9 % of 14 y.o. boys have cholesterol between 200 and 240.

3. What level of cholesterol would represent the 80th percentile?

① Same as above, what level is 80th percentile?

② plan:



③ InvNorm (.80, 170, 30)
From calculator, x = 195.249

④ 80th percentile is ≈ 195 mg/dl
80 % of 14 yo. boys have levels below 195 mg/dl.

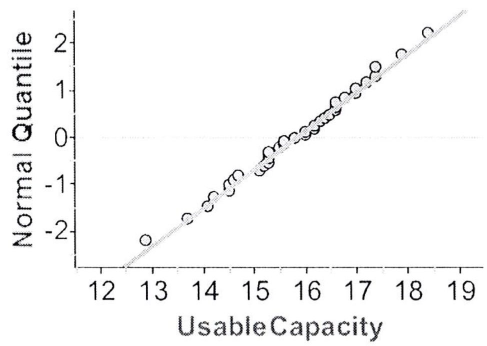
Normal Probability Plot

If a graph of the data is clearly skewed, has multiple peaks, or isn't bell-shaped, that's evidence that the distribution is *not* Normal. However, just because a plot of the data *looks* Normal, we can't say that the distribution *is* Normal. The 68–95–99.7 rule can give additional evidence in favor of or against Normality. A **Normal probability plot** also provides a good assessment of whether a data set follows a Normal distribution.

Normal Probability Plot: a graph that pairs an observation with its expected z-score to decide whether a distribution is approximately Normal

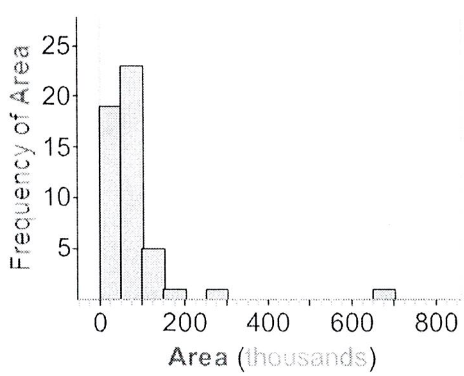
If the points on a Normal Probability Plot lies close to a straight line, the plot indicates that the data are normal. Systematic deviations from a straight line indicate a non-Normal distribution. Outliers appear as points that are far away from the overall pattern of the plot.

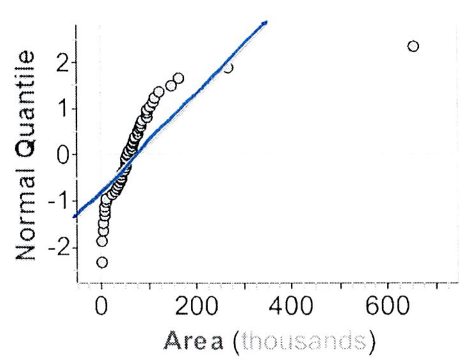
Example: Here is a Normal probability plot (also called a Normal quantile plot) of the refrigerator data from the previous page. It is quite linear, supporting our earlier decision that the distribution is close to Normal.



Example: State land areas

Problem: The histogram and Normal probability plot below display the land areas for the 50 states. Is this distribution approximately Normal?





Both histogram and Normal probability plot indicate distribution is strongly skewed right. In particular, there is 1 state whose area is much larger than we would expect if the distribution was approximately Normal.

Assessing Normality

The Normal distributions provide good models for some distributions of real data. In the latter part of this course, we will use various statistical inference procedures to try to answer questions important to us. These tests involve sampling individuals and analyzing data to gain insights about populations. Many of these procedures are based on the assumption that the population is *approximately Normally distributed*. Because of this we need to develop a strategy for assessing Normality.

Procedure.

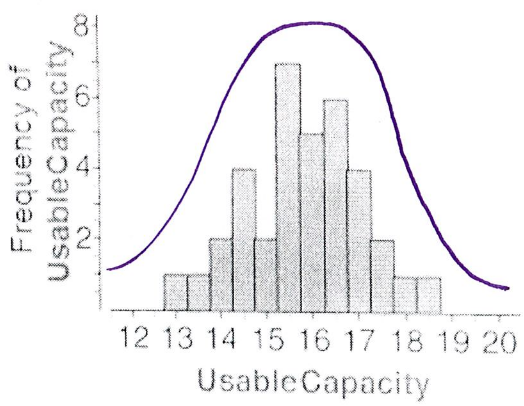
Step 1: *Plot the data* – make a dotplot, stemplot, or histogram. See if the graph is approximately symmetric and bell-shaped. Is the mean close to the median?

Step 2: *Check whether the data follow the 68-95-99.7 rule*. Find the mean and standard deviation. Then count the number of observations within one, two, and three standard deviations from the mean and compute these to percents.

Example. The measurements listed below describe the usable capacity (in cubic feet) of 36 side-by-side refrigerators. 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

The mean and standard deviation of these data are 15.825 and 1.217 cubic feet. The histogram is shown below.



$\bar{x} \pm 1s_x = (14.608, 17.042) = 24/36 = 66.7\%$

$x \pm 2s_x = (13.391, 18.259) = 34/36 = 94.4\%$

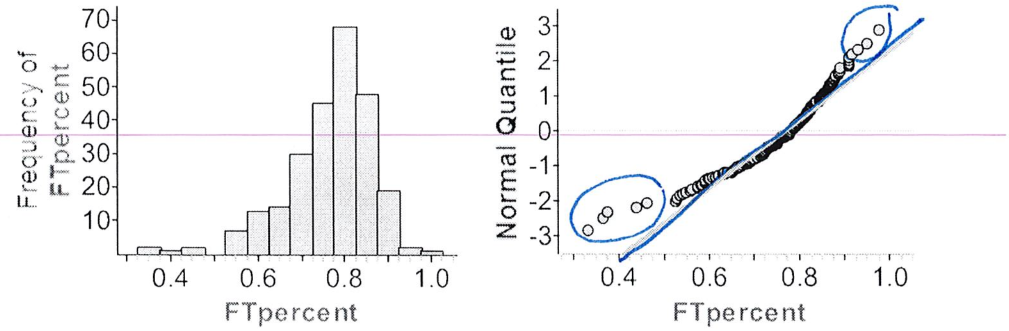
$x \pm 3s_x = (12.174, 19.476) = 36/36 = 100\%$

Graph: Roughly Symmetric
%'s follow 68-95-99.7 Rule Roughly
∴ Good evidence that this Distribution is close to Normal

* Normal Probability Plot * Page 128

Example: NBA free throw percentage

This is an example of a distribution that is skewed to the left. Notice that the lowest free throw percentages are too the left of what we would expect and the highest free throw percentages are not as far to the right as we would expect.



Technology Corner (Page 128)