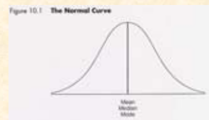
The Normal Distribution

- Properties of the Normal Distribution
- Shapes of Normal Distributions
- Standard (Z) Scores
- The Standard Normal Distribution

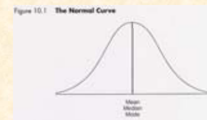
Normal Distributions

- **Normal Distribution**
  - Used with **linear** variables
  - A bell-shaped and symmetrical theoretical distribution,
  - with the mean, the median, and the mode all coinciding at its peak and
  - with frequencies gradually decreasing at both ends of the curve.



Normal Distributions

- **Normal Distribution**
- is a theoretical ideal distribution. Real-life empirical distributions never match this model perfectly.
- However, many things in life do approximate the normal distribution, and are said to be "normally distributed."



Scores "Normally Distributed?"

Table 10.1 Final Grades in Social Statistics of 1,200 Students (1983-1993)

Midpoint Score	Frequency Bar Chart	Freq.	Cum. Freq. (below)	%	Cum % (below)
40	*	4	4	0.33	0.33
50	*******	78	82	6.5	6.83
60	***************	275	357	22.92	29.75
70	***********************	483	840	40.25	70
80	***************	274	1114	22.83	92.83
90	*******	81	1195	6.75	99.58
100	*	5	1200	0.42	100

- Is this distribution normal?
- There are two things to initially examine: (1) look at the shape illustrated by the bar chart, and (2) calculate the mean, median, and mode.

Scores "Normally Distributed?"

Table 10.1 Final Grades in Social Statistics of 1,200 Students (1983-1993)

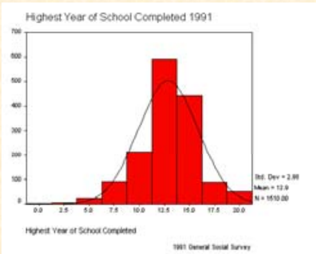
Midpoint Score	Frequency Bar Chart	Freq.	Cum. Freq. (below)	%	Cum % (below)
40	*	4	4	0.33	0.33
50	*******	78	82	6.5	6.83
60	***************	275	357	22.92	29.75
70	***********************	483	840	40.25	70
80	***************	274	1114	22.83	92.83
90	*******	81	1195	6.75	99.58
100	*	5	1200	0.42	100

Mean = (40 x 4) + (50 x 78) +
(60 x 275) + (70 x 483) + (80 x 274)
+ (90 x 81) + (100 x 5) / total number of cases

Scores Normally Distributed!

- The **Mean** = 70.07
- The **Median** = 70
- The **Mode** = 70
- Since all three are essentially equal, and the bar graph appears to be normally distributed, we can conclude these data are ***normally distributed***.
- Also, since the median is approximately equal to the mean, we know that the distribution is **symmetrical** (equal on both sides of the mid point, that is, mirror images on each half).

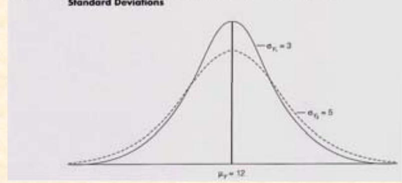
The Shape of a Normal Distribution



Notice the shape of the normal curve in this graph. Some normal distributions are tall and thin, while others are short and wide. All normal distributions, though, are wider in the middle and symmetrical.

Different Shapes of the Normal Distribution

Figure 10.2 Two Normal Distributions with Equal Means but Different Standard Deviations



Notice that the **standard deviation** changes the relative width of the distribution; the larger the standard deviation, the wider the curve.

Since the Standard Deviation reflects to width of the curve, lets review what the standard deviation is:

A measure that reflects how far the values range from the mean value on average.

A measure of variation for interval-ratio variables; it is equal to:

$$\sqrt{\frac{\sum(Y-\overline{Y})^2}{N-1}}$$

Percentage Change in the Nursing Home Population, 1980-1990

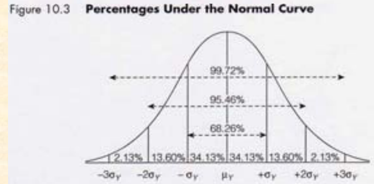
Nine Regions of U.S.	Percentage	$Y - \overline{Y}$	$(Y - \overline{Y})^2$ (squared deviations)
Pacific	15.7	15.7 - 31.5 = -15.8	249.64
West North Central	16.2	16.2 - 31.5 = -15.3	234.09
New England	17.6	17.6 - 31.5 = -13.9	193.21
East North Central	23.2	23.2 - 31.5 = - 8.3	68.89
West South Central	24.3	24.3 - 31.5 = - 7.2	51.84
Middle Atlantic	28.5	28.5 - 31.5 = - 3.0	9.00
East South Central	38.0	38.0 - 31.5 = 6.5	42.25
Mountain	47.9	47.9 - 31.5 = 16.4	268.96
South Atlantic	71.7	71.7 - 31.5 = 40.2	1616.04
(mean = 283.1/9 = 31.5)	283.1		$\sum (Y - \overline{Y})^2 =$ 2733.92

The standard deviation is the square root of the variance (variance = 2733.92/9-1 = 349.99) or 18.7. This is the average distance from the mean value (31.5).

A Leap of Faith:

We can use the **normal curve** and **standard deviation** to determine (or estimate) the percentage of cases that are close to the mean as well as the percentage that are far from the mean. Consider our example of a mean grade of 70.07 and SD of 10.27.

Figure 10.3 Percentages Under the Normal Curve



The **normal curve** allows us to predict what percentage of the cases fall between any two points.

For example, the percentage of students who scored between 70 and 80 on the statistics tests. Or, the percent that scored above 95.

The **normal curve** also allows us to predict what percentage of the cases fall above or below a specific value (or in this case grade).

To predict percentages, we must:

1. Convert the number (or in this case grade) we are interested in into a **Z score**, and
2. Look up the Z score in the "**Standard Normal Table**" (found at the end of most statistics books). This score will show us the percentage above and below our value of interest.

Standard (Z) Scores

- A ***standard score*** (also called ***Z score***) is the number of standard deviations that a given **raw score** (such as a grade of 95) is above or below the mean.
- It is calculated by using the following formula (the number, minus the mean, divided by the standard deviation):

$$Z = \frac{Y - \overline{Y}}{S_y}$$

Standard (Z) Scores

For example, what percentage of students scored above 95? First we must determine the Z score for a score of 95:

$$Z = \frac{95 - 70.0}{10.27} = \frac{24.3}{10.27} = 2.37$$

$$Z = \frac{Y - \overline{Y}}{S_y}$$

The standard normal table provides two pieces of information:

1. The percent of cases that exist between the mean and the Z score (**column B**)
2. The percent of cases that exist beyond the Z score (**column C**)

(see Standard Normal Table)

In our example, look up the Z score of 2.37 (grade of 95) and find the area (or percent of cases) beyond this score.

The number found is .0089. This means that less than 1 percent (.89 percent) of the students scored greater than a 95 on their test and

roughly 99 percent of the students scored less than 95 (those below the mean or 50% + those between the mean and the 2.37 Z score or .4911 = .9911.

Finding the Area Between the Mean and a Positive Z Score

- Using the data presented in Table 10.1, find the percentage of students whose scores range from the mean (70.07) to 85 when the standard deviation is 10.27.

Table 10.1 Final Grades in Social Statistics of 1,200 Students (1983-1993)

Midpoint Score	Frequency Bar Chart	Freq.	Cum. Freq. (below)	%	Cum % (below)
40	*	4	4	0.33	0.33
50	*******	78	82	6.5	6.83
60	***************	275	357	22.92	29.75
70	**********************	483	840	40.25	70
80	***************	274	1114	22.83	92.83
90	*******	81	1195	6.75	99.58
100	*	5	1200	0.42	100

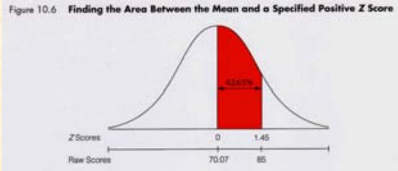
Finding the Area Between the Mean and a Positive Z Score

(1) Convert 85 to a Z score:

Z = (85-70.07)/10.27 = 1.45

(2) Look up the Z score (1.45) in column A, finding the proportion (.4265). (The Z scores are found in Appendix B of your book.)

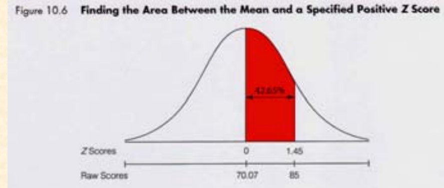
Figure 10.6 Finding the Area Between the Mean and a Specified Positive Z Score



Finding the Area Between the Mean and a Positive Z Score

(3) Convert the proportion (.4265) to a percentage (42.65%); this is the percentage of students scoring between the mean and 85 in the course.

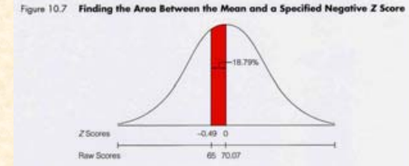
Figure 10.6 Finding the Area Between the Mean and a Specified Positive Z Score



Finding the Area Between the Mean and a Negative Z Score

Using the data presented in Table 10.7, find the percentage of students scoring between 65 and the mean (70.07)

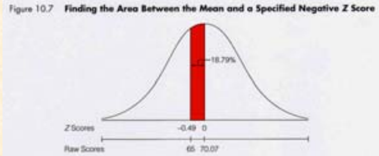
Figure 10.7 Finding the Area Between the Mean and a Specified Negative Z Score



Finding the Area Between the Mean and a Negative Z Score

- (1) Convert 65 to a Z score:
  Z = (65-70.07)/10.27 = -.49
- (2) Since the curve is symmetrical, use .49 to find the area in the standard normal table (Appendix B of book): .1879 or 18.79% of the students scored between 65 and 70.07.

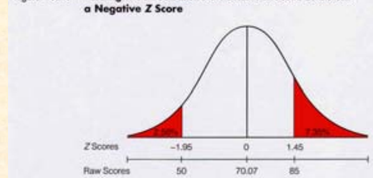
Figure 10.7 Finding the Area Between the Mean and a Specified Negative Z Score



Finding Area Above a Positive Z Score or Below a Negative Z Score

- Find the percentage of students who did (a) very well, scoring above 85, and (b) those students who did poorly, scoring below 50.

Figure 10.10 Finding the Area Above a Positive Z Score or Below a Negative Z Score



Finding Area Above a Positive Z Score or Below a Negative Z Score

- (a) Convert 85 to a Z score, then look up the value in column C of the standard normal table:
  Z = (85-70.07)/10.27 = 1.45 → 7.35%
- (b) Convert 50 to a Z score, then look up the value (look for a positive Z score!) in column C:
  Z = (50-70.07)/10.27 = -1.95 → 2.56%

Finding Area Above a Positive Z Score or Below a Negative Z Score

Figure 10.10 **Finding the Area Above a Positive *Z* Score or Below a Negative *Z* Score**