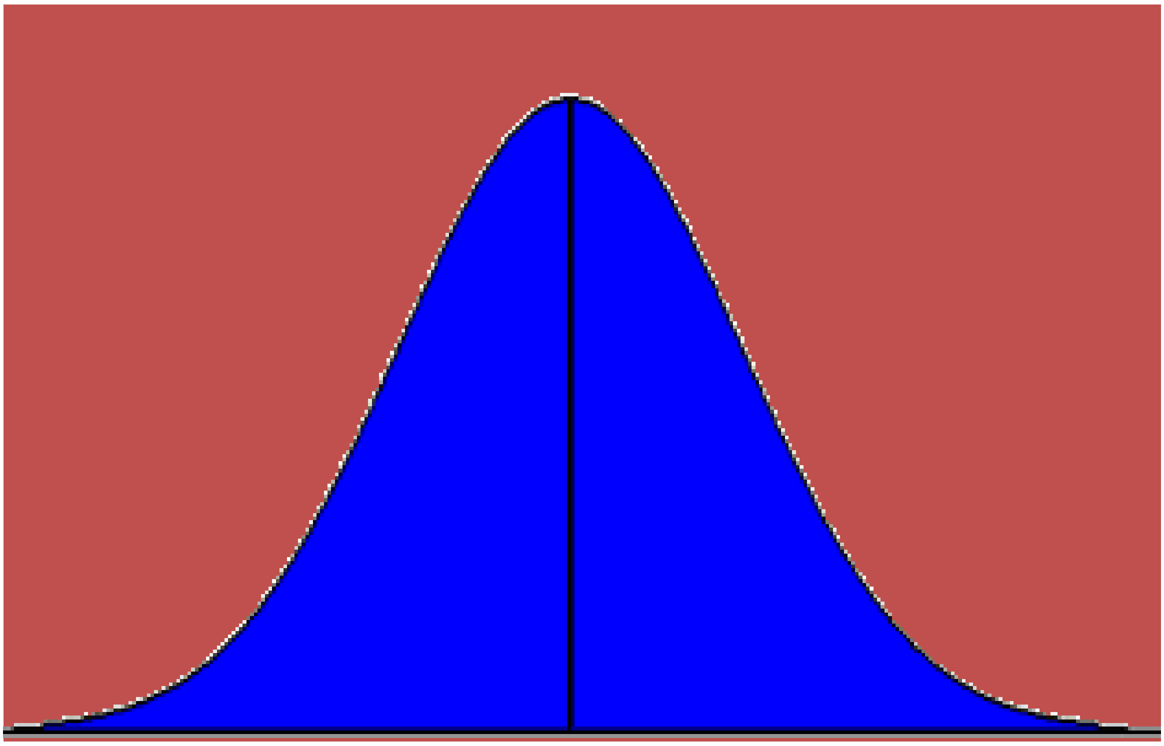
Standard Scores and the Normal Distribution



Standard Scores and the Normal Distribution

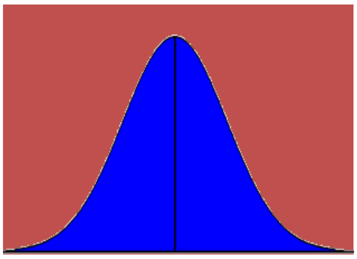
- ***The normal probability distribution is a mathematical model which is generated by a specific mathematical function.***

- ***This mathematical model is useful because many variables observed in nature closely approximate a normal distribution.***

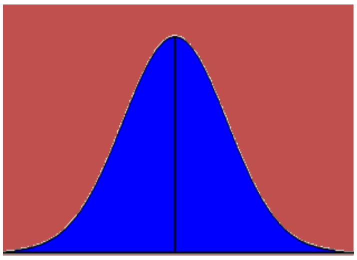
- ***Normal distributions are symmetric and the tails are asymptotic, that is, the tails approach the baseline but never quite reach it no matter how far away from the mean you are.***

Continuous distributions

- *Continuous distributions are formed because everything in the world that can be measured varies to some degree.*
- *Measurements are like fingerprints, no two are exactly alike. The degree of variation will depend on the precision of the measuring instrument used.*
- *The more precise the instrument, the more variation will be detected.*
- *A distribution, when displayed graphically, shows the variation with respect to a central value.*



Everything that can be measured forms some type of distribution that contains the following characteristics



Measures of central tendency

- ❑ **Arithmetic mean**
- ❑ **Median**
- ❑ **Mode**

- **Measures of spread or dispersion**

- ➢ **Range**
- ➢ **Variance**
- ➢ **Standard deviation**

Shapes of distributions

- Symmetrical - normal
- Symmetrical - not normal
- Skewed right or left
- More than one peak

Frequency Distributions

- ***A frequency distribution is a model that indicates how the entire population is distributed based on sample data.***

- ***Since the entire population is rarely considered, sample data and frequency distributions are used to estimate the shape of the actual distribution.***

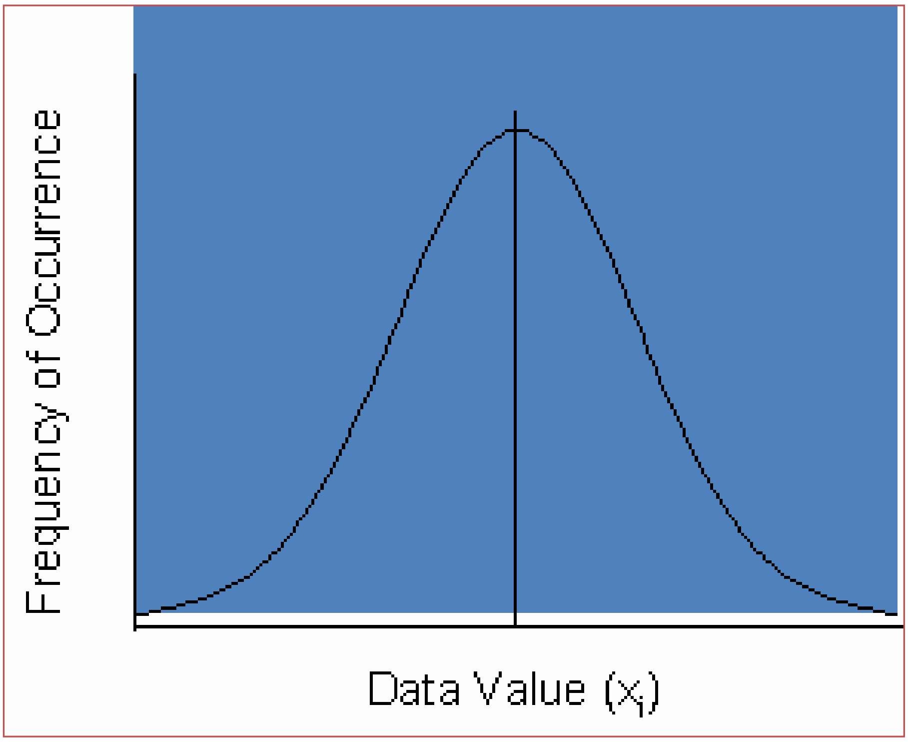
- ***This estimate allows inferences to be made about the population from which the sample data were obtained.***

- ***It is a representation of how data points are distributed.***

- ***It shows whether the data are located in a central location, scattered randomly or located uniformly over the whole range.***

The graph of the frequency distribution will display the general variability and the symmetry of the data.

.



The frequency distribution may be represented in the form of an equation and as a graph

Frequency Distribution

- ***When using a frequency distribution, the interest is rarely in the particular set of data being investigated.***
- ***In virtually all cases, the data are samples from a larger set or population.***
- ***Sometimes, it is wrongfully assumed that data follow the pattern of a known distribution such as the normal.***
- ***The data should be tested to determine if this is true.***
- ***Goodness of Fit tests are used to compare sample data with known distributions.***
- ***The inferences made from a frequency distribution apply to the entire population.***

$s/\sqrt{n}$

$s/\sqrt{n}$

Central limit theory

- *statisticians deal with distributions formed from individual measurements as well as distributions formed by sets of averages.*

- *If the data are taken from the same population, there is a relationship between the distribution of individual measurements and the distribution of averages.*

- *The means will be equal to*

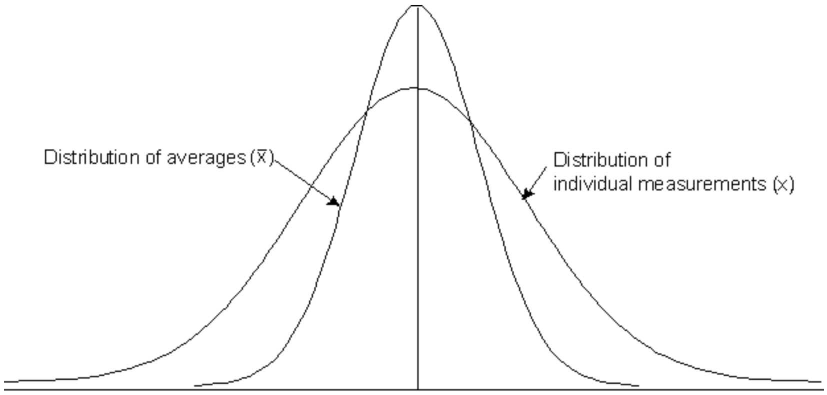
$$\bar{X} = \bar{\bar{X}}$$

- *If the standard deviation for individual measurements is s, then the standard error for the distribution of averages is* .

$$s/\sqrt{n}$$

- *If a sample of 100 parts is divided into 20 subsets of 5 parts each, then n is 100 when calculating the variance and standard deviation of individual measurements and n is 5 when calculating the standard error using .*

Distribution of individual measurements versus averages



Pattern of distribution of data

- ***Some distributions have more than one point of concentration and are called multimodal.***

- ***When multimodal distributions occur, it is likely that portions of the output were produced under different conditions.***

- ***A distribution with a single point of concentration is called unimodal.***

- ***A distribution is symmetrical if the mean, median and mode are at the same location.***

Pattern of distribution of data

- ***The symmetry of variation is indicated by skewness.***
- ***If a distribution is asymmetrical it is considered to be skewed.***
- ***The tail of a distribution indicates the type of skewness.***
- ***If the tail goes to the right, the distribution is skewed to the right and is positively skewed.***
- ***If the tail goes to the left, the distribution is skewed to the left and is negatively skewed.***
- ***A symmetrical distribution has no skewness***

Pattern of distribution of data

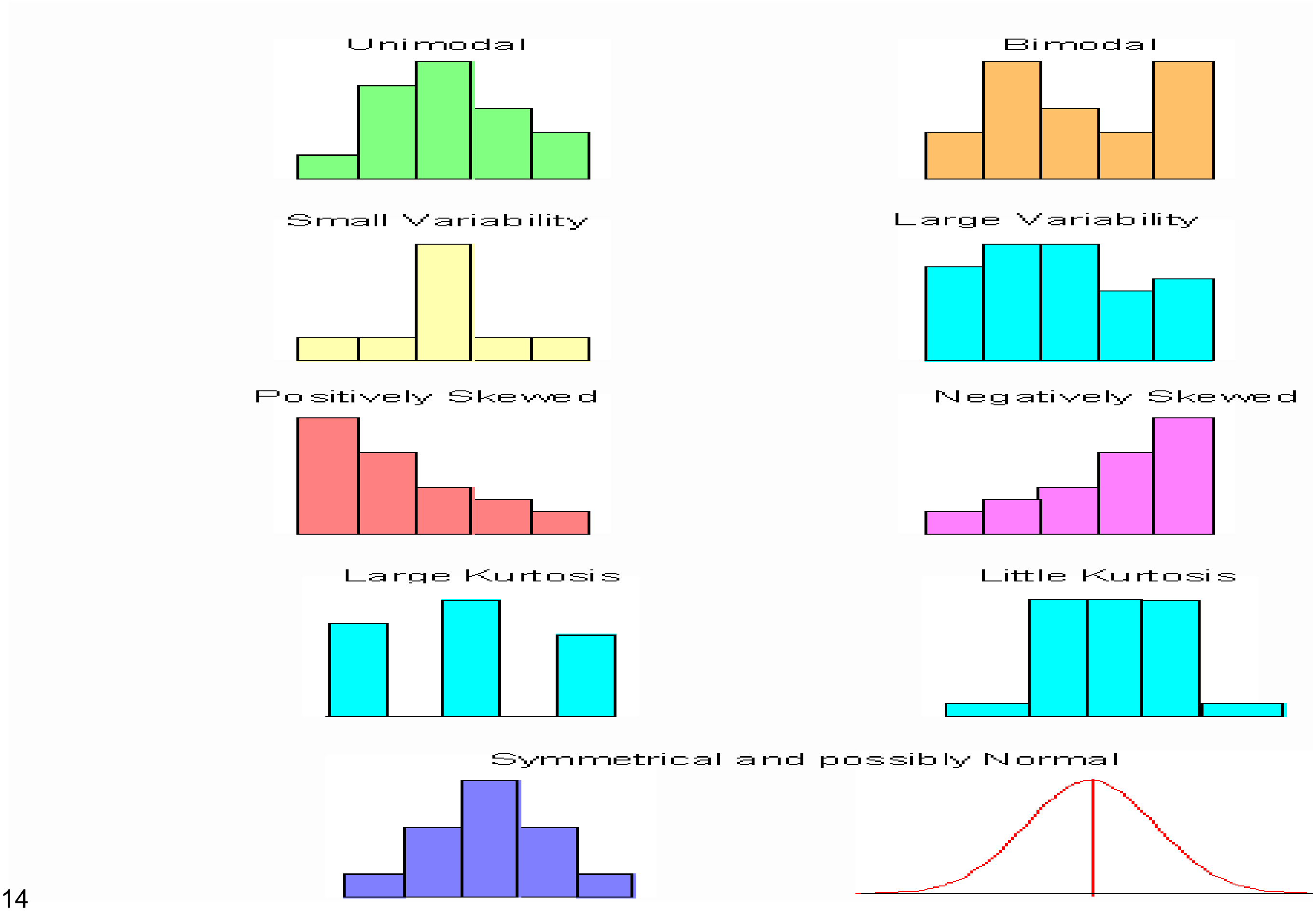
- ***Kurtosis* is defined as the *state or quality of flatness or peakedness of a distribution*.**

- ***If a distribution has a relatively high concentration of data in the middle and out on the tails, but little in between, it has large kurtosis.***

- ***If it is relatively flat in the middle and has thin tails, it has little kurtosis.***

- ***If the frequencies of occurrence of a frequency distribution are cumulated from the lower end to the higher end of a scale, a cumulative frequency distribution is formed.***

SHAPES OF DISTRIBUTIONS



THE NORMAL CURVE

- ***The normal curve is one of the most frequently occurring distributions in statistics.***
- ***The pattern that most distributions form tend to approach the normal curve.***
- ***It is sometimes referred to as the Gaussian curve named after Karl Friedrich Gauss (1777-1855) a German mathematician and astronomer.***
- ***The normal curve is symmetrical about the average, but not all symmetrical curves are normal.***
- ***For a distribution or curve to be normal, a certain proportion of the entire area must occur between specific values of the standard deviation.***

THE NORMAL CURVE

- ***There are two ways that the normal curve may be represented: The actual normal curve and the standard normal curve.***

- ***[1] Actual Normal***

The curve represents the distribution of actual data. The actual data points (xi) are represented on the abscissa

(x-scale) and the number of occurrences are indicated on the ordinate (y-scale).

- ***[2] Standard Normal***

The sample average and standard deviation are transformed to standard values with a mean of zero and a standard deviation of one. The area under the curve represents the probability of being between various values of the standard deviation.

THE NORMAL CURVE

- ***By transforming the actual measurements to standard values, one table is used for all measurement scales.***

- ***The abscissa on the actual normal curve is denoted by x and the abscissa on the standard normal curve is denoted by Z.***

- The relationship between x and Z:

$$Z = \frac{(X_i - \bar{X})}{s}$$

- This is known as the transformation formula. It transforms the x value to its corresponding Z value.

- A distribution of averages may also be represented with the normal curve.

Normal curve

- The abscissa on the actual normal curve for a distribution of averages is denoted $\bar{X}$

- The center is denoted by $\bar{\bar{X}}$

  (the average of averages.)

Normal curve

- The relationship between $\bar{x}$ and Z:

$s/\sqrt{n}$

$$Z = \frac{(\bar{x}_i - \bar{\bar{x}})}{s/\sqrt{n}}$$

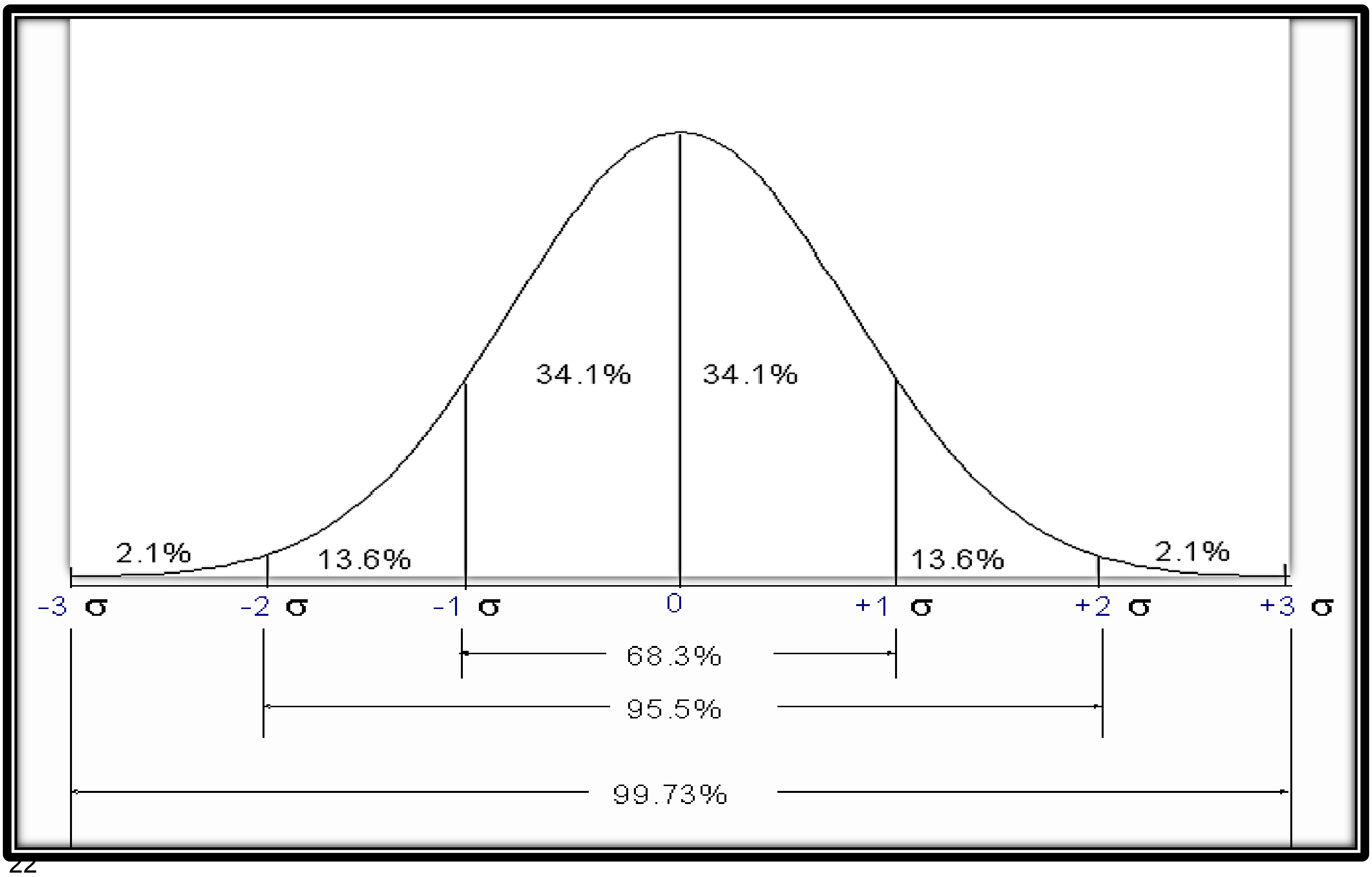
- The statistic $s/\sqrt{n}$ is the standard error or the standard deviation for a set of averages.

Normal curve

- *The standard normal curve areas are used to make certain forecasts and predictions about the population from which the data were taken.*

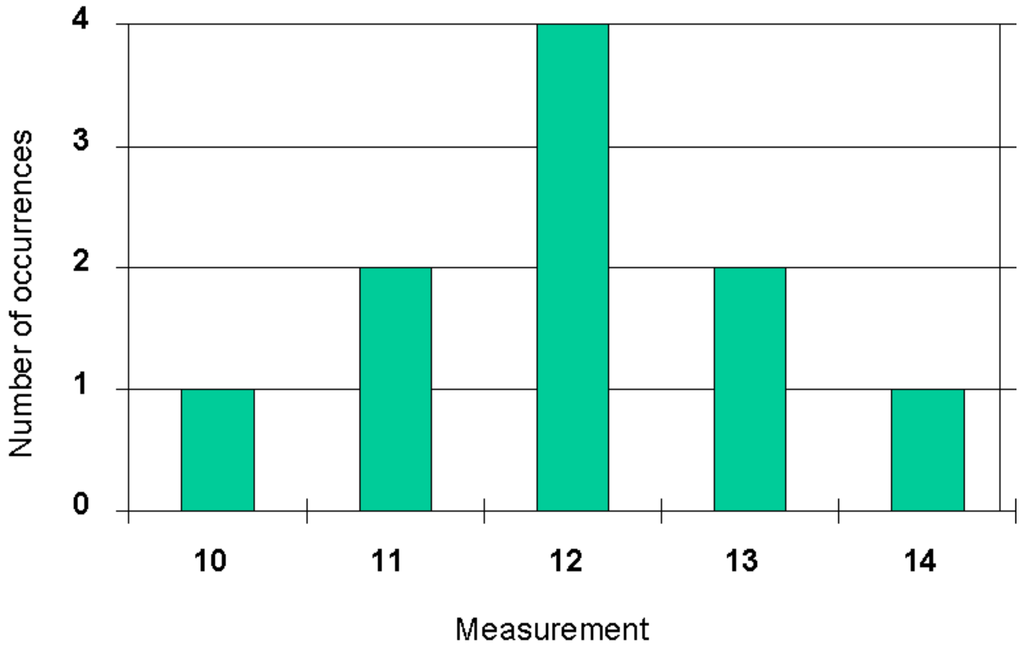
- *The standard normal curve areas are probability numbers. The area indicates the probability of being between two values on the Z scale.*

Areas Under the Standard Normal Curve



- **Example :**
- The following data represent ten measurements (weights of under five children in Kgs) attending a PHC center. This is a sample taken from a population attending the center during a month.
- 10, 11, 11, 12, 12, 12, 12, 13, 13, 14

A histogram is drawn to get a general idea of the shape of the distribution.



- The mean and standard deviation are calculated (12, and 1.15 respectively).

- The normal curve areas are used to make predictions about the weights.

- To use the standard normal tables the x values must be converted to their equivalent Z values.

- Using $$Z = \frac{(x_i - \overline{x})}{s}$$

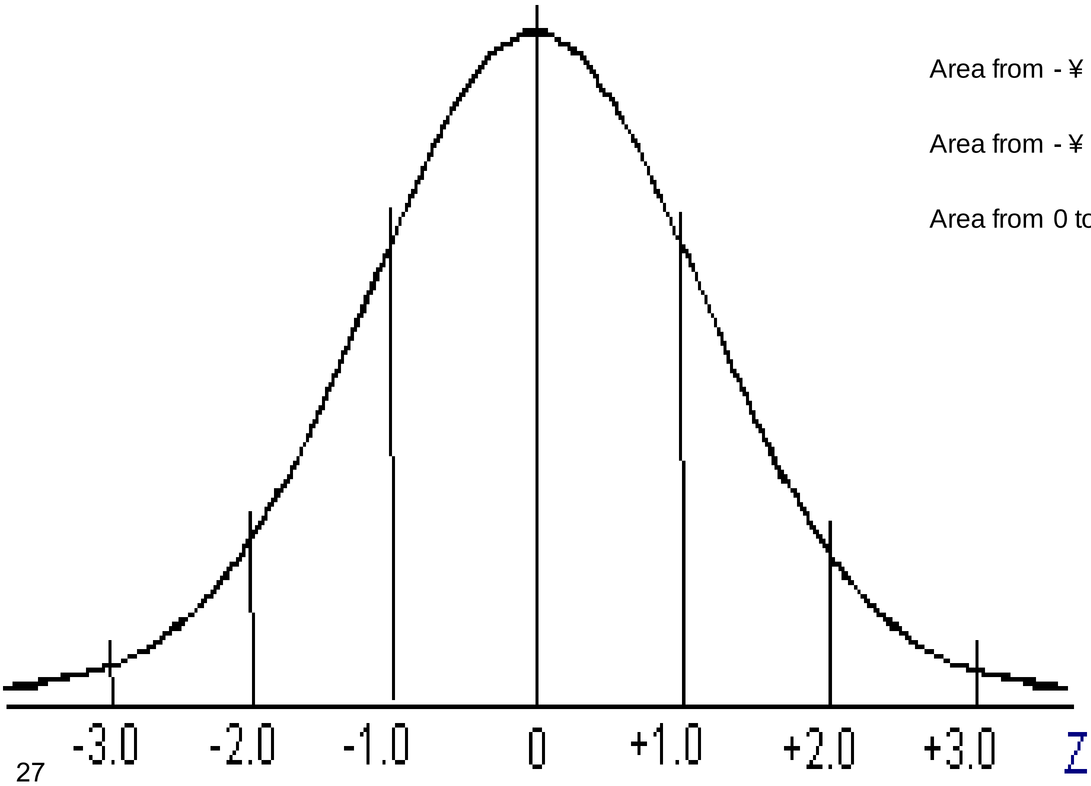
the x value 10.85 converts to Z = -1.0,

the x value 12 converts to Z = 0,

the x value 13.15 converts to Z = +1.0,

the x value 14.3 converts to the Z = +2.0, etc.

Normal curve



Area from - ¥ to + ¥ = 1.0

Area from - ¥ to 0 = .5

Area from 0 to + ¥ = .5

- **Example:**
- **Use the standard normal curve table to find the area between Z = +1.0 and Z = +2.0.**

- **Area from 0 to +2.0= .4772**

- **Area from 0 to +1.0= .3413**

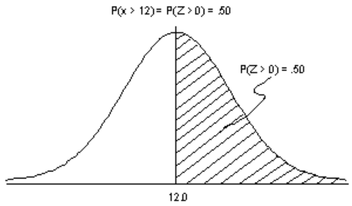
- **Area between +1.0 and +2.0= .4772 - .3413**

  = .1359

Example:

- For $\bar{x}$ = 12.0 and s = 1.15, find the probability that a measurement will be greater than 12.0.

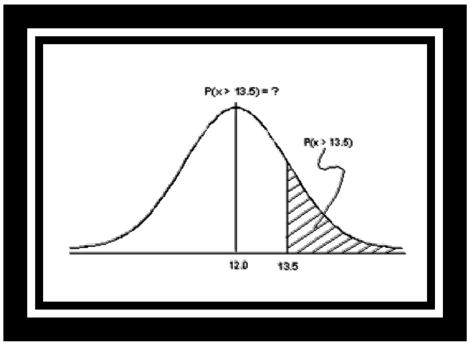
- This is written as P(x > 12).



- P(x > 12) = 0.50 which is the same as the probability that Z > 0 since the mean value on the x scale corresponds to 0 on the Z scale.

Example:

What is the probability that a part will have a measurement greater than 13.5?The first step is to draw a diagram indicating the area that represents the probability of a measurement greater than 13.5. This is a very important step because the areas under the normal curve are difficult to visualize and a diagram makes it easy.



The next step is to convert the x value into a Z value.

$$Z = \frac{(x_i - \bar{x})}{s} = \frac{(13.5 - 12.0)}{1.15} = +1.30$$

This is the area from Z = 0 to Z = +1.30, therefore P(x > 13.5) = P(Z > + 1.30) = (.5000 - .4032) = .0968.

Example:

What percentage of the population will have measurements between 9.0 and 10.0?

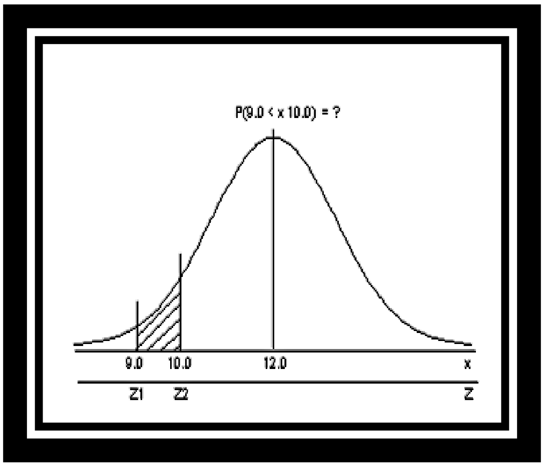
Area from Z_1 to 0	**= area from 9.0 to 12.0**	**= .4955**
Area from Z_2 to 0	**= area from 10.0 to 12.0**	**= .4591**
Area from Z_1 to Z_2	**= area from 9.0 to 10.0**	**= .4955 - .4591 = .0364**

$Z_1 = (9.0 - 12.0)/1.15 = -3.0/1.15 = -2.61$

$Z_2 = (10.0 - 12.0)/1.15 = -2.0/1.15 = -1.74$

The standard normal curve table gives the following results:

Therefore, 3.64% of the population will have measurements between 9.0 and 10.0.



Problems

Questions from 1 through 11 depend on the following situation

"820 ----8th grade students in a school district took a standardized social studies test that is normally distributed and has a mean of 340 and a variance of 256. Here are the scores for four of the students: **Salam** scored 364, **Magdy** scored 356, **Safaa** scored 344, and **Hanan** scored 332. "

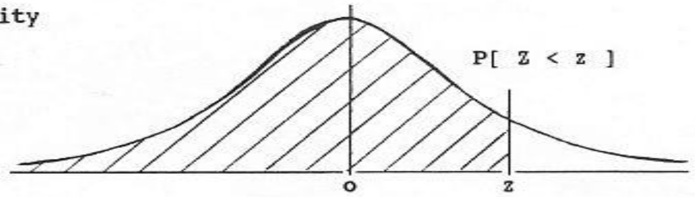
1. In this sample, how many students would be expected to score above Magdy?
2. What proportion of the students would be expected to score above Salam?
3. What percent of students would be expected to score above Hanan?
4. How many students would be expected to score below Hanan?
5. What percentage of students would be expected to score below Safaa?
6. What proportion of students would be expected to score below Salam?
7. How many students would be expected to score between Safaa and Hanan?
8. What proportion of students would be expected to score between Salam and Magdy?
9. What percentage of students would be expected to score between Safaa and Magdy?
10. 225 students score above Laith, what is his test score?
11. What are the approximate test scores for Q1 and Q3?
12. A middle-school student took two standardized tests. On the Language Proficiency test she scored 114.4 (mean = 100, sd = 16) and on Mathematical Reasoning she scored 61.7 ------- (mean = 50, sd = 9). On which test did she do better?

STANDARD STATISTICAL TABLES

1. Areas under the Normal Distribution

The table gives the cumulative probability up to the standardised normal value z i.e.

$$P[\, Z < z \,] = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}Z^2)\, dZ$$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5159	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7854
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8804	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9865	0.9868	0.9871	0.9874	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9980	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

z	3.00	3.10	3.20	3.30	3.40	3.50	3.60	3.70	3.80	3.90
P	0.9986	0.9990	0.9993	0.9995	0.9997	0.9998	0.9998	0.9999	0.9999	1.0000

Areas between 0 & Z of the Standard Normal Distribution

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.00	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.10	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.20	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.30	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.40	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.50	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.60	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.70	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.80	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.90	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.00	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621