

## 4.2 The Normal Distribution

Many physiological and psychological measurements are normally distributed; that is, a graph of the measurement looks like the familiar symmetrical, bell-shaped distribution (See Figure 4.2.1). If we know or can assume that our data is similar to that of a normal distribution, then there are many statistical conclusions or predictions we can make about our data. Many statistical calculations assume that the data is normally distributed. If the data is not normally distributed we can use non-parametric statistical tests to draw conclusions or make predictions about our data.

There is a whole family of normal curve, depending on the sample or population mean,  $M$  and the sample or population standard deviation,  $SD$ . The normal curve can also be drawn with the  $z$ -score instead of the actual data values. Figure 4.2.2 shows a graph of the **standard normal distribution** (a normal distribution with mean of zero and standard deviation of one).

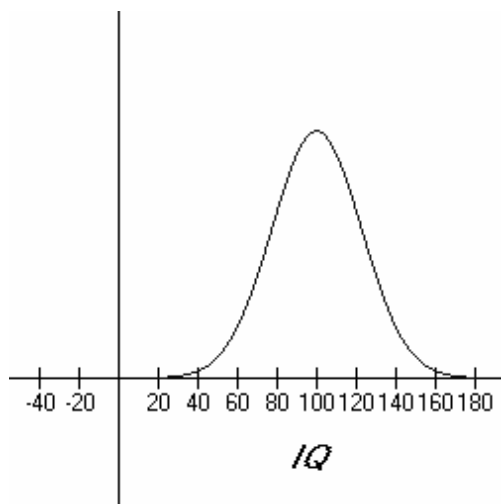


Figure 4.2. 1 IQ Normal Curve.

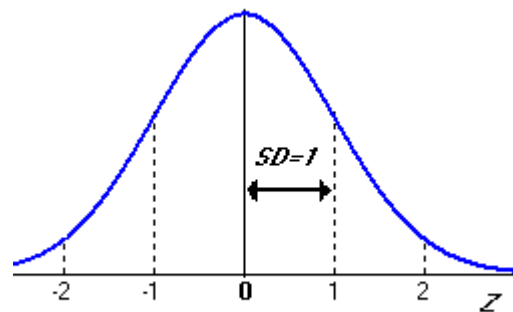


Figure 4.2.2 Standard Normal Distribution ( $M = 0$  and  $SD = 1$ ).

## Probability and the Normal Distribution

One important characteristic of the normal curve is the information that the area under the curve tells us about the probability of the set of data we are studying. We define the probability of a certain score,  $X$ , as the proportion of that score (frequency) relative to or divided by the proportion (or frequency) of all the other scores.

The **probability** of a score  $X$  is the proportion of the frequency of that score divided by the total frequency of all the scores.

The area under the normal curve when the  $z$ -score is 0 is 50%. This is indicated by the standard normal table in Appendix A has 0.50 (the decimal equivalence of the percent, i.e. the percent divided by 100). The standard normal curve of Appendix A is the cumulative proportion or probability of the distribution of scores that are less than a particular  $z$ -score value. The area under the normal curve when the  $z$ -score is +1.0 is 84.13% or 0.8413 from the  $z$ -score cumulative probability distribution table in the appendix. Figures 4.2.3 and 4.3.4 show the areas under the curve (cumulative probabilities or percentage frequencies) for  $z = 0$  and  $z = 1$  respectively.

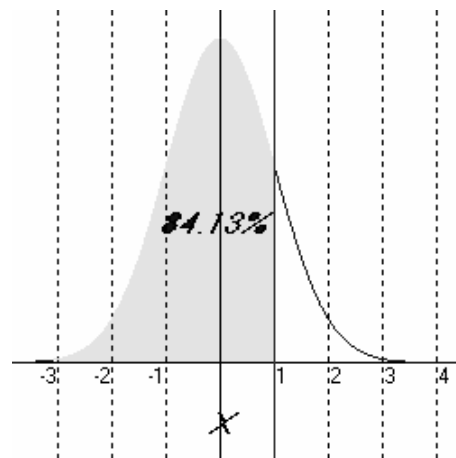
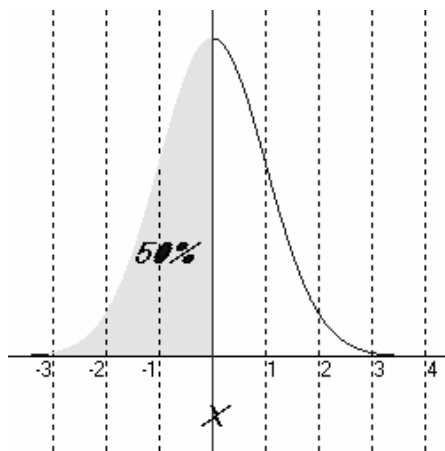


Figure 4.2.3. Area under curve when  $z = 0.0$ .      Figure 4.2.4. Area under curve when  $z = 1.0$ .

Using the standard normal ( $z$ -score) table in Appendix A, the percent of the sample or population between  $z$ -score of 0 and 1.0 is  $84.13\% - 50\% = 34.13\%$ . The percent of the sample distribution below a  $z$ -score of -1.0 is 15.87% (Figure 4.3.5). Therefore, the percent of the sample between a  $z$ -score of -1.0 and +1.0 is  $84.13\% - 15.87\% = 68.26\%$  or approximately **68%** (Figure 4.2.6). Note, that the percent of the population or sample between  $z = -2$  and  $z = +2$  is approximately **95%** and the percent between  $z = -3$  and  $z = +3$  is about **99%**.

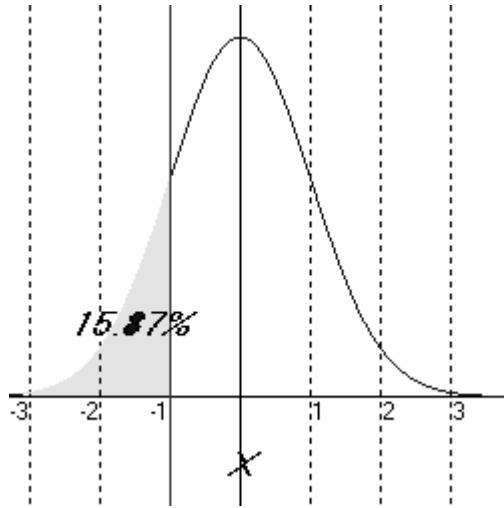


Figure 4.2. 5. Area under curve when  $z = -1$ .

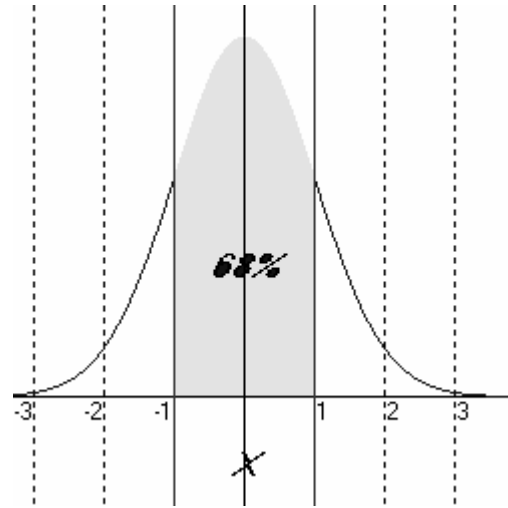
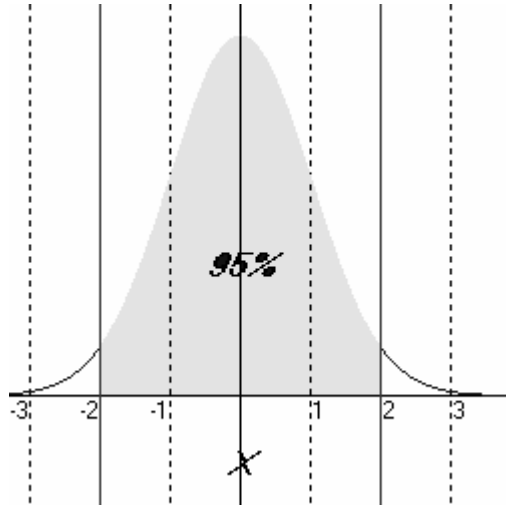
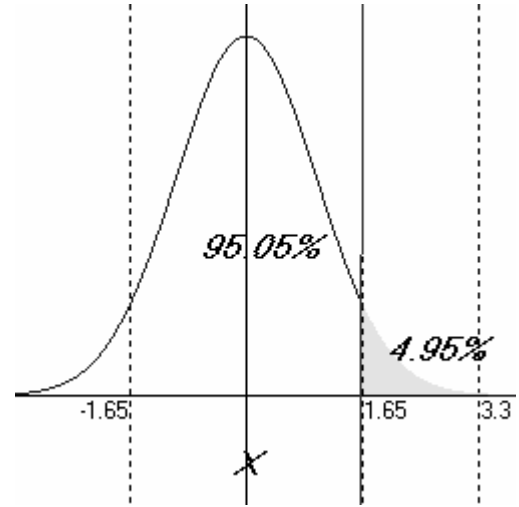


Figure 4.2. 6. Area between  $z = -1$  and  $z = +1$ .

If we wanted to know what percent of data is above a certain  $z$ -score value of 1.65, we would observed from the  $z$ -score table in the Appendix that the probability less than  $z = 1.65$  is 0.9505 or **95.05%** (Figure 4.2.7). Therefore, the percentage above  $z = 1.65$  is  $100 - 95.05 = 4.95\%$  (Figure 4.2.8).

Figure 4.2.7. Area between  $z = -2$  and  $z = +2$ .Figure 4.2.8. Area above  $z = 1.65$ .

In Example 4.1.1, using the ODE data table, we computed for the *pass9th* variable the following statistics:  $M=65.86$  and  $S=13.61$ . If we would like to know how the Arlington school did in proportion or percentage to the rest of the schools for the *pass9th* variable in ODE table we would first compute Arlington's  $z$ -score or standard score.

Arlington's observed score,  $X$  was 84, so its  $z$ -score is:

$$z = \frac{X - M}{S} = \frac{84 - 65.86}{13.61} = 1.33$$

The positive 1.33 tells us that Arlington school scored above the mean of all the schools or its mean score is **1.33** standard deviations above the mean. The percentile for this  $z$ -score (see the  $z$ -score cumulative table in Appendix A) is 0.9082 or **90.82%** ( $0.9082 \times 100$ ). This means that 90.82% of the rest of the schools had their *pass9th* scores below that of Arlington or that Arlington's *pass9th* score was 90.8% above that for the rest of the schools.

Similarly, if we would like to know how the Lima school did in proportion to the rest of the schools for the *pass9th* variable in the ODE table we would first compute Lima's  $z$ -score or standard score. Lima's observed score,  $X$  was 40, so its  $z$ -score is:

$$z = \frac{X - M}{S} = \frac{40 - 65.86}{13.61} = -1.90$$

The negative 1.90 tells us that Lima school scored below the mean of all the schools or its mean score is **-1.90** standard deviations below the mean. The percentile for this  $z$ -score (see the  $z$ -score cumulative table in Appendix) is 0.028716493 or **2.87%** (0.028716493 x 100). This means that 2.87% of the rest of the schools had their *pass9th* scores below that of Lima or that Lima's *pass9th* score was higher than only 2.87% of the rest of the schools. One may also state that 97.13% (100 - 2.87) of the schools scored higher than Lima for the *pass9th* variable.

SPSS can generate the normal curve (normal fit) for a set of data by selecting "Normal Curve" when drawing a frequency polygon (see Figure 4.2.9). The SPSS procedure for superimposing a normal fit upon a sample data histogram is shown in Figure 4.2.10.

The reasons why the normal distribution is one of the most important and useful distributions used in statistics are:

1. Many of the dependent variables examined in statistics are assumed to be normally distributed in the population. We would expect that a representative sample of the population would also be normally distributed, especially when the sample size is large or approaches the size of the population.

2. If we assume that a variable is at least approximately normally distributed, then we can make a number of inferences about scores or values of that variable.
3. The theoretical distribution of a set of sample means (repeatedly large or infinite samples) is approximately normal under a wide variety of conditions. Such a distribution is called the *sampling distribution* with a mean that approaches the population mean and standard deviation we called the standard error.
4. Most of the parametric statistical procedures assume that a variable is normally distributed.

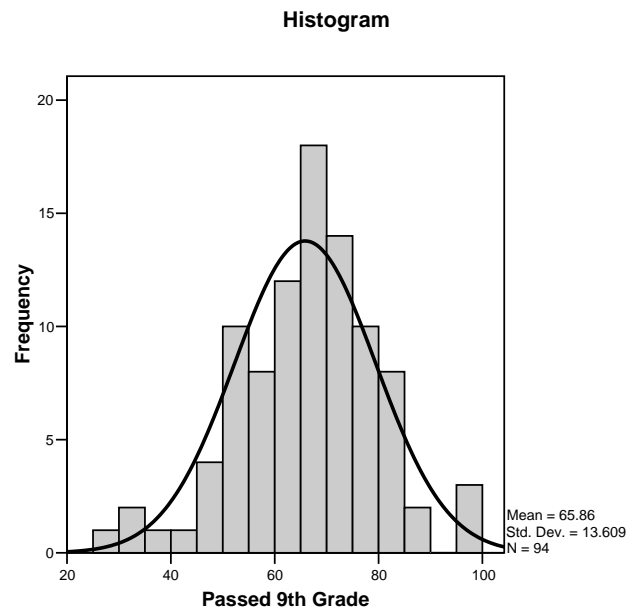


Figure 4.2.9 Normal curve fit for the pass9th variable from ODE dataset

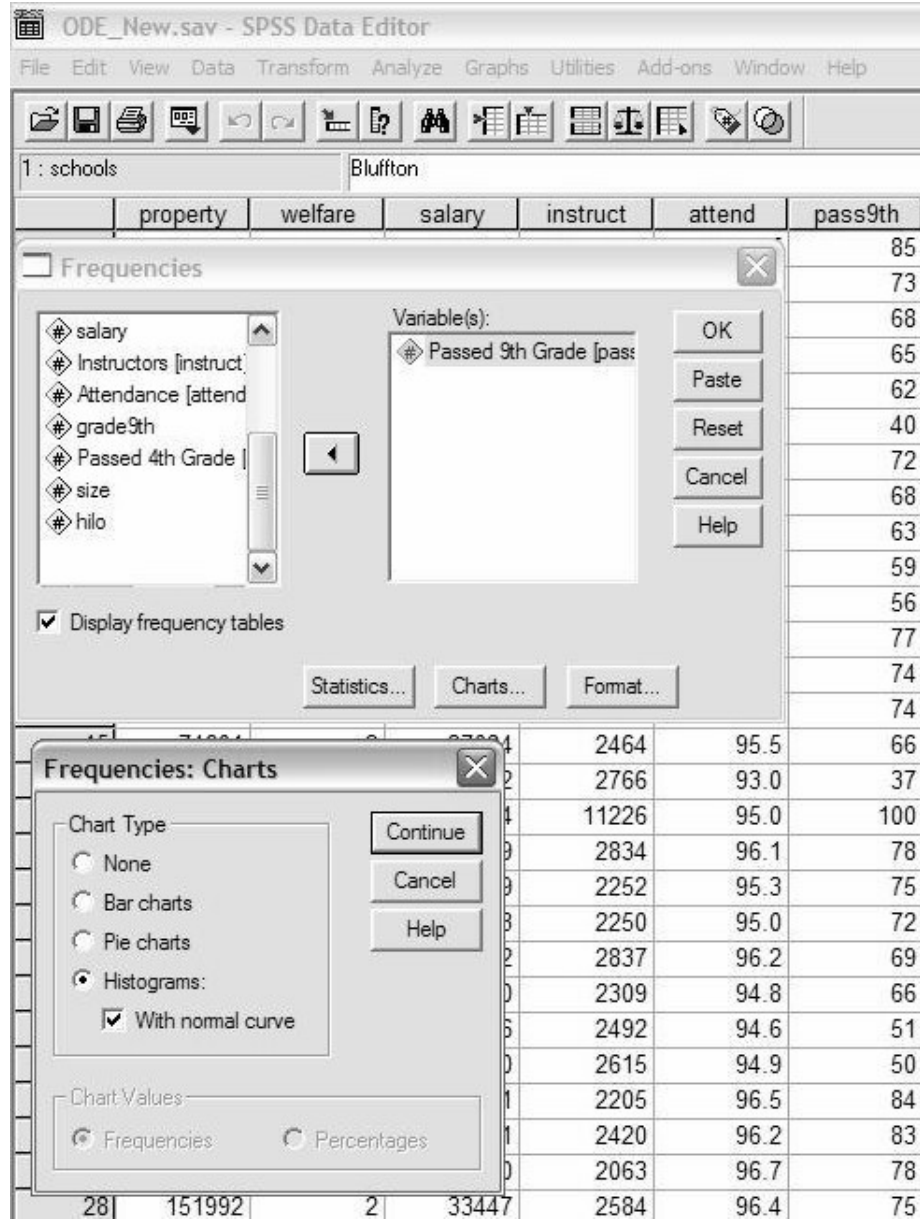


Figure 4.2.10 SPSS procedure for creating a normal curve fit.

### The Standard Normal Distribution

From any individual score,  $X$ , the  $z$ -score can be computed given the mean,  $\mu$  or  $M$  and standard deviation,  $s$  or  $SD$  by the formula  $z = (X - M)/SD$  or  $z = (X - \mu)/s$ .

The percentile for any  $z$ -score can be readily obtained from a cumulative standard normal probability table, since such table gives the probability or proportion of scores that are below any particular  $z$ -score. The percentile is the individual score,  $X$  of a distribution that gives a certain cumulative percentage. The percent rank is the percentile (cumulative percentage) of a given individual score,  $X$ .

The **percentile ( $P\%$ )** is the score,  $X$ , that a given percentage (%) of the distribution of scores are less than or equal to.

The **percent rank ( $Pr_x$ , Probability of  $X$ )** is the percentile or percentage or cumulative percentage of a given raw score,  $X$ .

Tables A1 through A4 (see Appendix A) show the percentiles of the  $z$ -scores of a symmetrical distribution (normal distribution). To find the percent rank of a given score,  $X$  we first use the mean and standard deviation of  $X$  to transform it into its corresponding  $z$ -score and then look up its cumulative probability from Tables A1 through A4. For example, the percent rank for an  $IQ$  score of 117 is 87.09% (see illustration below).

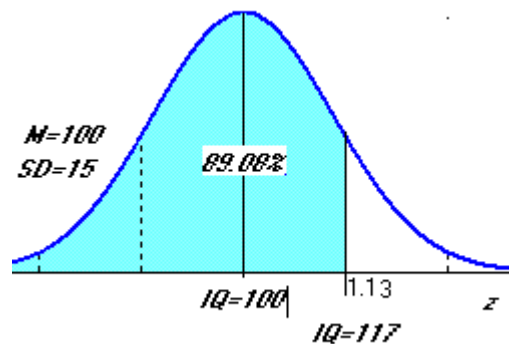
---

**Percent Rank,  $Pr_{IQ}$**

$$z = \frac{117 - 100}{15} = \frac{17}{15} = +1.1333$$

$Pr(z < 1.1333) = 0.8708$  or **87.08%**

**The percent rank of  $IQ = 117$  is 87.08%**



Portion of Table A1 thru A4 ( $z = 1.13$ )

$z$	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
<b>1.1</b>	0.8643	0.8665	0.8686	<b>0.8708</b>	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
<b>1.2</b>	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015



To find the 75 percentile for the  $IQ$  variable, we need to find the  $IQ$  score or value that 75% of the distribution is less than or equal to. First, we find the corresponding  $z$ -score for which the cumulative probability or percentage is equal to 0.75 (the decimal equivalence of 75%). When we search Tables A1 through A2 of Appendix A, we found the  $z$ -score for the 75 percentile (the value of  $z$  that gives a cumulative probability of 0.75) is 0.675 (we extrapolate between  $z = 0.67$  and 0.68). Now, to get the 75 percentile for  $IQ$  scores we manipulate the  $z$ -score formula to solve for  $X$  (see illustration below).

---

#### Percentile, $IQ_{\%}$

From Appendix A (Tables A1 to A4):

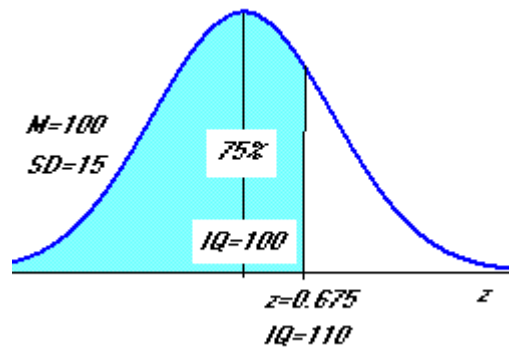
$$z_{75\%} = 0.675$$

$$z = \frac{IQ - M}{SD} = \frac{(IQ - 100)}{15} = 0.675$$

$$IQ - 100 = 15(z)$$

$$IQ = 15(0.675) + 100 = \mathbf{110.125}$$

$$\text{So } IQ_{75\%} = \mathbf{110}$$



Portion of Table A1 to A4:  $\Pr(z \leq 0.675) = 0.75$

$z$	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
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	<b>0.7454</b>	<b>0.7486</b>	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852

### Tony and Standard Scores

**Problem 4.2.1:** The mean score for an assessment exam of 500 subjects was calculated as 70 and the standard deviation was 10. Determine the following: (a) what percent of the subjects had a score of 50 or below? (b) what score did 80% of the subjects

scored less than? (c) what percent of the subjects had a score of 90 and above? (d) what percentage of subjects scored between 50 and 80 on the assessment exam?

**Tony:** “I think I know what is being asked, but how do these questions relate to percentile and percent rank concepts?”

**Rose:** “Remember that finding **percentile** is finding **the score** that a certain percentile (percentage of scores) is less than or equal to and finding **percent rank** is finding **the percentile** of a given score.” “We will need to use the  $z$ -score statistics or formula [  $z = (X - M)/SD$  ] to convert back and forth between individual score and their corresponding  $z$ -score.” “In all of these cases we either will use a cumulative percentage table, computed from the frequency distribution of the data as we did in the chapter on frequency distribution or we will use the standard normal cumulative tables to find  $z$  or its corresponding probabilities (Tables A1 to A3 of Appendix A).” “Let us examine each question individually using both a graph and computational formulas to illustrate solutions to each question.”

Given:  $M = 70$ ,  $SD = 10$  and  $N = 500$

---

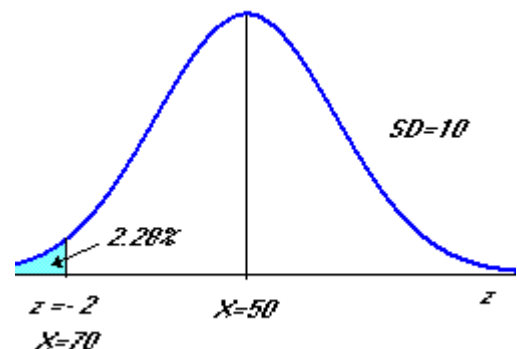
### Percent Rank

(a) What percent of the subjects had a score of 50 or below? (Probability of  $X \leq 50$ )

$$z = \frac{50 - 70}{10} = \frac{-20}{10} = -2$$

$\Pr(z \leq -2) = 0.0228$  or **2.28%**

**The percent rank of  $X = 50$  is 2.28%**



Portion of Table A1 thru A4 ( $z = -2$ )

$z$	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	<b>0.0228</b>	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233

**Percentile**

(b) What score did 80% of the subjects scored less than?

From Appendix A (Tables A1 to A4):

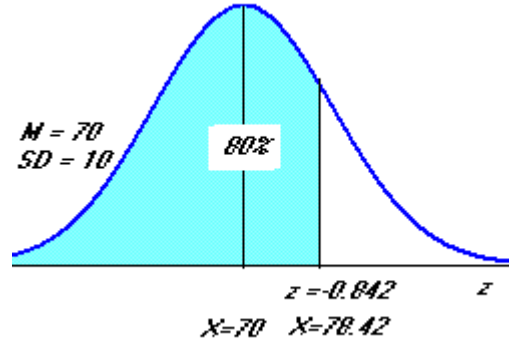
$Z_{80\%} = 0.842$

$$z = \frac{X - M}{SD} = \frac{(X - 70)}{10} = 0.842$$

$X - 70 = 10(z)$

$X = 10(0.842) + 70 = 78.42$

So  $X_{80\%} = 78.42$



Portion of Table A1 to A4:  $\Pr(z \leq 0.842) = 0.80$

<b>z</b>	<b>0</b>	<b>0.01</b>	<b>0.02</b>	<b>0.03</b>	<b>0.04</b>	<b>0.05</b>	<b>0.06</b>	<b>0.07</b>	<b>0.08</b>	<b>0.09</b>
<b>0.7</b>	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
<b>0.8</b>	0.7881	0.7910	0.7939	0.7967	<b>0.7995</b>	<b>0.8023</b>	0.8051	0.8078	0.8106	0.8133
<b>0.9</b>	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389

**Percent Rank**

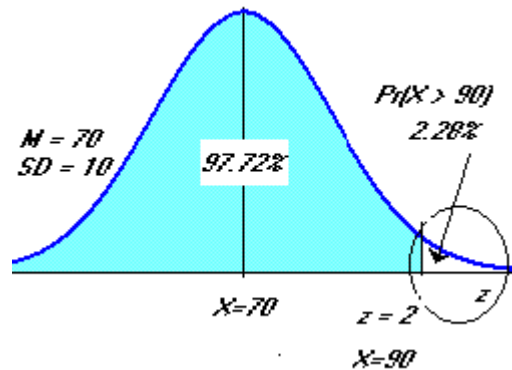
(c) What percent of the subjects had a score of 90 and above? (Probability of  $X \geq 90$ )

$$z = \frac{90 - 70}{10} = \frac{20}{10} = +2$$

$\Pr(z \leq +2) = 0.9772$  or **97.72%**

The percent rank of  $X = 90$  is 97.72%

So  $\Pr(z \geq 2) = 1 - 0.9772 = 0.0228$  or **2.28%**



Portion of Table A1 thru A4 ( $z = +2$ )

<b>z</b>	<b>0</b>	<b>0.01</b>	<b>0.02</b>	<b>0.03</b>	<b>0.04</b>	<b>0.05</b>	<b>0.06</b>	<b>0.07</b>	<b>0.08</b>	<b>0.09</b>
<b>1.9</b>	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
<b>2.0</b>	<b>0.9772</b>	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
<b>2.1</b>	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857

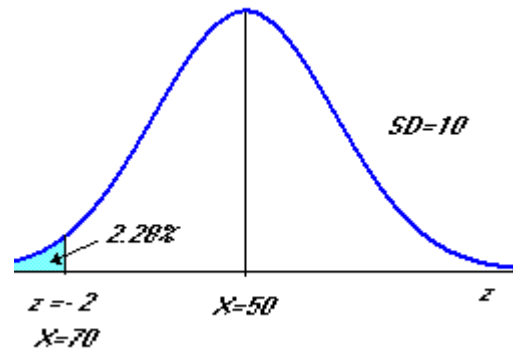
### Difference Between Percent Ranks

(d) What percentage of subjects scored between 50 and 80 on the assessment exam?

From (a) we found that Probability of  $X \leq 50 = 2.28\%$ , i.e.  $\Pr(z \leq -2)$  for calculation below

$$z = \frac{50 - 70}{10} = \frac{-20}{10} = -2$$

$\Pr(z \leq -2) = 0.0228$  or **2.28%**



$\Pr(X \leq 50) = 2.28\%$

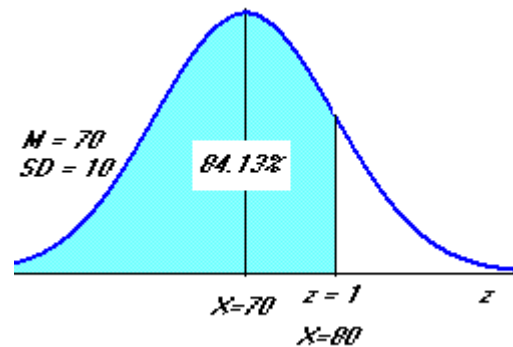
The  $\Pr(X \leq 80) = \Pr(z \leq +1)$

$$z = \frac{80 - 70}{10} = \frac{10}{10} = +1$$

$\Pr(z \leq 1) = 0.8413$  or **84.13%**

So Probability or percentage between  $X = 50$  and  $X = 80$  or  $z = -2$  and  $z = 1$  is

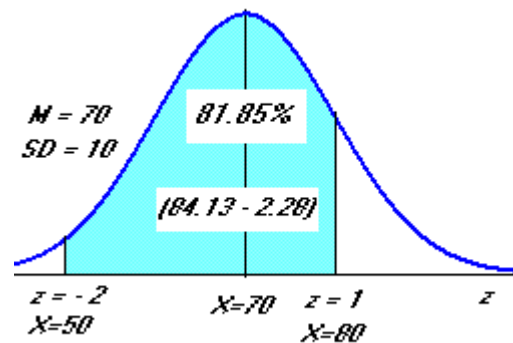
$$84.13\% - 2.28\% = \mathbf{81.85\%}$$



$\Pr(X \leq 80) = 84.13\%$

Portion of Table A1 thru A4 ( $z = 1$ )

$z$	0	0.01	0.02	0.03	0.04
0.9	0.8159	0.8186	0.8212	0.8238	0.8264
1.0	<b>0.8413</b>	0.8438	0.8461	0.8485	0.8508
1.1	0.8643	0.8665	0.8686	0.8708	0.8729



$\Pr(50 \leq X \leq 80) = 81.85\%$