Notes Chapter 6: The Normal distribution

A normal distribution is a continuous, symmetric, bell-shaped distribution of a variable.

Summary of the Properties of the Theoretical Normal Distribution:

1. A normal distribution curve is bell-shaped.

2. The mean, median, and mode are equal and are located at the center of the distribution.

3. A normal distribution curve is unimodal (i.e., it has only one mode).

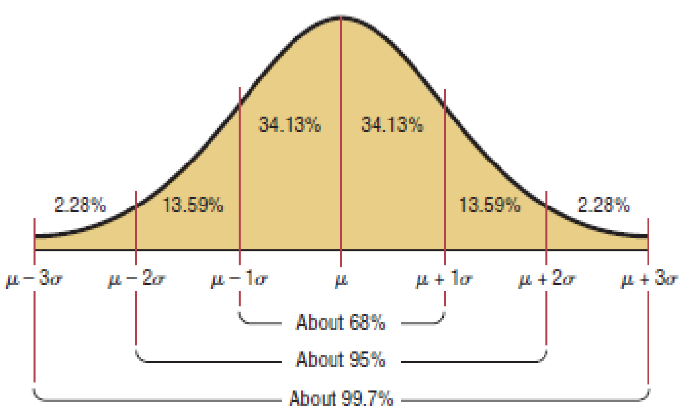
4. The curve is symmetric about the mean, which is equivalent to saying that its shape is the same on both sides of a vertical line passing through the center.

5. The curve is continuous; that is, there are no gaps or holes. For each value of X, there is a corresponding value of Y.

6. The curve never touches the x axis. Theoretically, no matter how far in either direction the curve extends, it never meets the x axis—but it gets increasingly closer.

7. The total area under a normal distribution curve is equal to 1.00, or 100%.

8. The area under the part of a normal curve that lies within 1 standard deviation of the mean is approximately 0.68, or 68%; within 2 standard deviations, about 0.95, or 95%; and within 3 standard deviations, about 0.997, or 99.7%. See the figure below, which also shows the area in each region.



The standard normal distribution is a normal distribution with a mean of 0 and a standard deviation of 1.

All normally distributed variables can be transformed into the standard normally distributed variable by using the formula for the standard score:

$$z = \frac{x - \mu}{\sigma}$$

Finding Areas Under the Standard Normal Distribution Curve:

Find the probability for each:

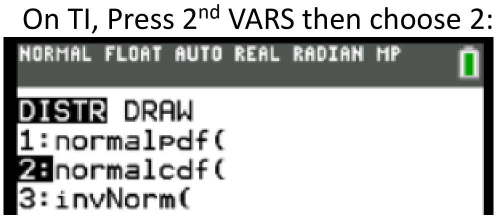
a. $P(0 < z < 2.32)$

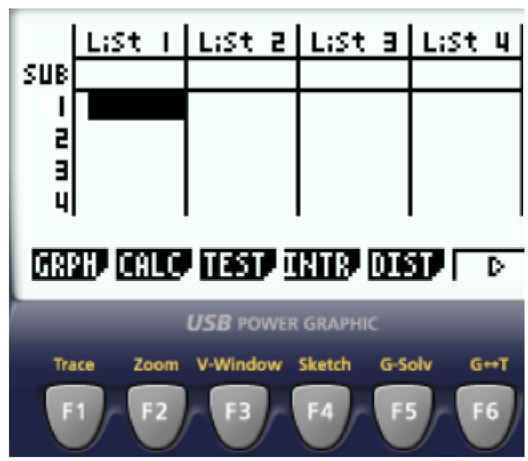
b. $P(z < 1.65)$

c. $P(z > 1.91)$

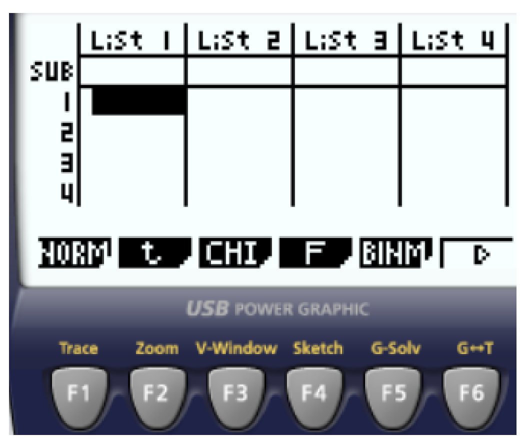
Using tables AND Using graphing Calculators

On TI, Press 2nd VARS then choose 2:





Casio, F5 for DISTR. Then F1 for NORM:



6.2 Applications of the Normal Distribution:

The standard normal distribution curve can be used to solve a wide variety of practical problems. The only requirement is that the variable be normally or approximately normally distributed. There are several mathematical tests to determine whether a variable is normally distributed.

Determining Normality:

A normally shaped or bell-shaped distribution is only one of many shapes that a distribution can assume; however, it is very important since many statistical methods require that the distribution of values (shown in subsequent chapters) be normally or approximately normally shaped.

There are several ways statisticians check for normality. The easiest way is to draw a histogram for the data and check its shape. If the histogram is not approximately bell-shaped, then the data are not normally distributed.

Skewness can be checked by using the Pearson coefficient of skewness (PC) also called Pearson's index of skewness. The formula is:

$$PC = \frac{3(\bar{x} - median)}{s}$$

If the index is greater than or equal to 1 or less than or equal to -1, it can be concluded that the data are significantly skewed.

For all the problems presented in this chapter, you can assume that the variable is normally or approximately normally distributed.

Examples:

1. Admission Charge for Movies The average early-bird special admission price for a movie is $5.81. If the distribution of movie admission charges is approximately normal with a standard deviation of $0.81, what is the probability that a randomly selected admission charge is less than $3.50? Ans: 0.0022

2. Population in U.S. Jails The average daily jail population in the United States is 706,242. If the distribution is normal and the standard deviation is 52,145, find the probability that on a randomly selected day, the jail population is
a. Greater than 750,000 Ans: 0.2007 b. Between 600,000 and 700,000 Ans: 0.4316

6.3 The Central Limit Theorem:

In addition to knowing how individual data values vary about the mean for a population, statisticians are interested in knowing how the means of samples of the same size taken from the same population vary about the population mean.

A sampling distribution of sample means is a distribution using the means computed from all possible random samples of a specific size taken from a population.

If the samples are randomly selected with replacement, the sample means, for the most part, will be somewhat different from the population mean m. These differences are caused by sampling error.

Sampling error is the difference between the sample measure and the corresponding population measure due to the fact that the sample is not a perfect representation of the population.

When all possible samples of a specific size are selected with replacement from a population, the distribution of the sample means for a variable has two important properties, which are explained next.

Properties of the Distribution of Sample Means:

1. The mean of the sample means will be the same as the population mean.

2. The standard deviation of the sample means will be smaller than the standard deviation of the population, and it will be equal to the population standard deviation divided by the square root of the sample size.

In summary, if all possible samples of size n are taken with replacement from the same population, the mean of the sample means, denoted by $\mu_{\bar{x}}$, equals the population mean μ; and the standard deviation of the sample means, denoted by $\sigma_{\bar{x}}$, equals σ.

The standard deviation of the sample means is called the standard error of the mean.

Hence,

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

The Central Limit Theorem

As the sample size n increases without limit, the shape of the distribution of the sample means taken with replacement from a population with mean μ and standard deviation σ will approach a normal distribution. As previously shown, this distribution will have a mean μ and a standard deviation $\sigma/\sqrt{n}$.

If the sample size is sufficiently large, the central limit theorem can be used to answer questions about sample means in the same manner that a normal distribution can be used to answer questions about individual values. The only difference is that a new formula must be used for the z values. It is

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

It's important to remember **two things** when you use the central limit theorem:

1. When the original variable is normally distributed, the distribution of the sample means will be normally distributed, for any sample size n.

2. When the distribution of the original variable might not be normal, a sample size of 30 or more is needed to use a normal distribution to approximate the distribution of the sample means. The larger the sample, the better the approximation will be.