Chapter 7—The Normal Probability Distribution

7.1 Properties of the Normal Distribution

***Definition*: Probability Density Function**

A probability density function is an equation used to compute probabilities of *continuous* random variables that must satisfy the following two properties.

1. Graph of the equation must be greater than or equal to zero for all possible values of the random variable.
2. Area under the curve equals 1.

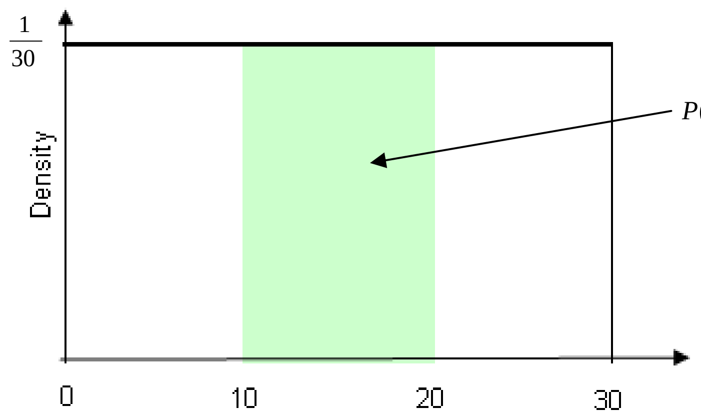
EXAMPLE: *Illustrating the Uniform Distribution*

Imagine that a friend of yours is always late. Let the random variable X represent the time from when you are supposed to meet your friend until he shows up. Further suppose that your friend could be on time (x=0) or up to 30 minutes late (x=30) with all 1-minute intervals of times between x=0 and x=30 equally likely. That is to say, your friend is just as likely to be from 3 to 4 minutes late as he is to be 25 to 26 minutes late. The random variable X can be any value in the interval from 0 to 30, that is, $0 \le X \le 30$. Because any two intervals of equal length between 0 and 30, inclusive, are equally likely, the random variable X is said to follow a **uniform probability distribution.**

Remember from Section 6.1 for a discrete probability distribution that probability was read from the vertical axis (similar to reading a relative frequency histogram).

By comparison, for a continuous random variable and a continuous probability density function (or probability distribution), the <u>area under the graph</u> of a density function over some interval represents the probability of observing a value of the random variable in that interval

Uniform Density Function



$$P(10 < X < 20) = \frac{1}{30} * (20 - 10)$$
$$= \frac{1}{30} * 10 = \frac{10}{30} = 0.33$$

X=time from when you are supposed to meet your friend until he shows up

Is this a probability density function?

What is the probability that your friend will show up from 0 to 30 minutes late?—i.e., P(0<X<30)?

What is the probability that your friend will show up between 10 and 20 minutes late?—i.e., P(10<X<20)?

What is the probability that your friend will show up 12.543 minutes late?—i.e., P(X=12.543)?

- Since there is no area under the curve at X=12.543, the P(X=12.543)=0.
- For the continuous variable X, no observation exists that is *exactly* 12.543.
- For a continuous variable X, with an infinite number of possible values, what is the probability of one particular value like X=12.543?—1/infinity = 0.

Relative frequency histograms that are approximately **symmetric and bell-shaped** are said to have the shape of a **normal probability density function (or normal probability curve)**. Many real-world variables are well-approximated by a normal probability curve (p. 277).

Figure 7 Heights of Three-Year-Old Females

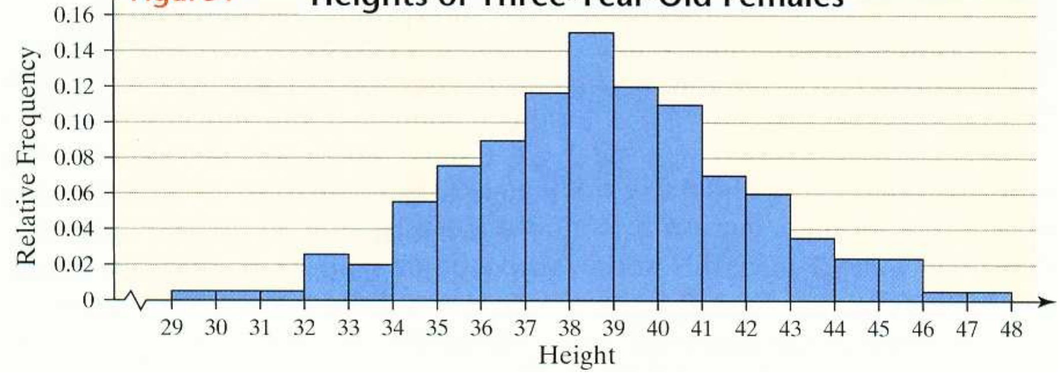
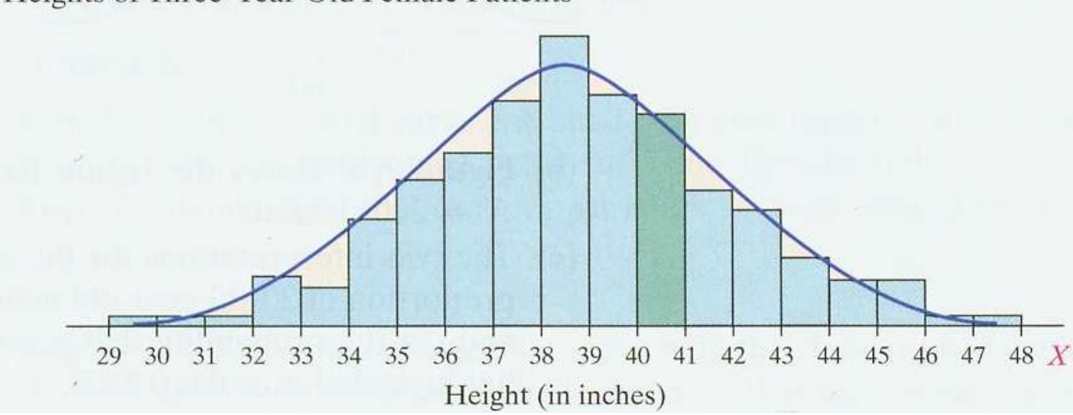


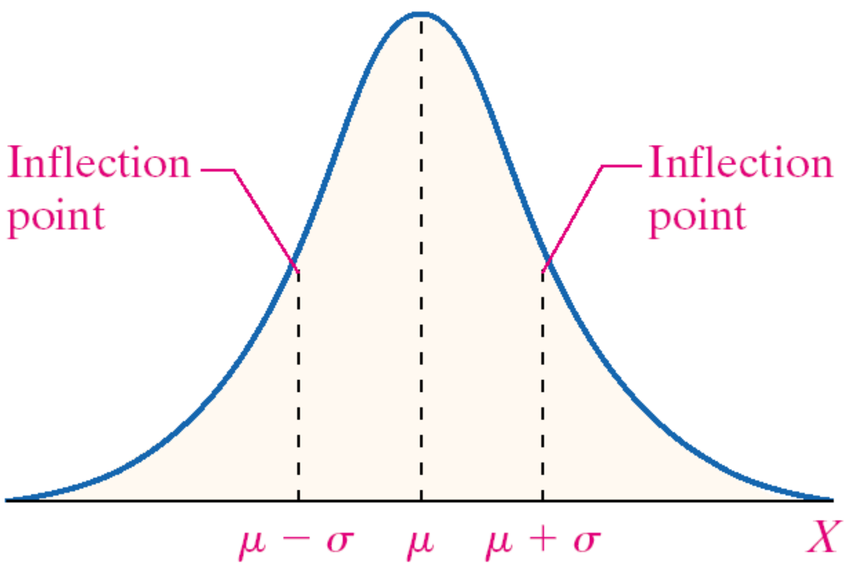
Figure 8
Heights of Three-Year-Old Female Patients



Properties of the Normal Probability Curve:

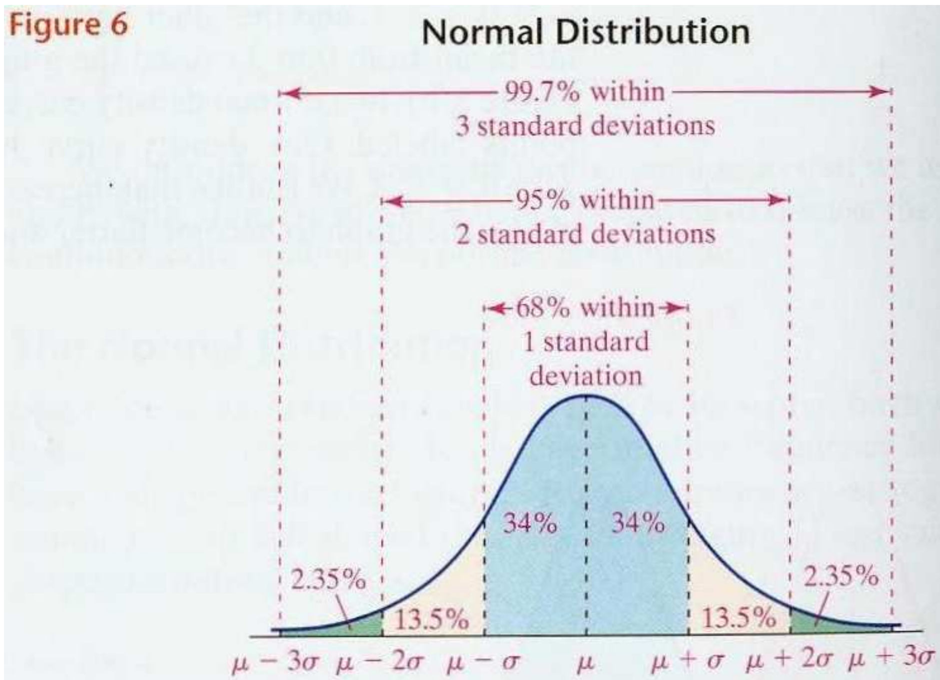
1. The highest point occurs at $x=\mu$.
2. It is symmetric about the mean, μ. One half of the curve is a mirror image of the other half, i.e., the area under the curve to the right of μ is equal to the area under the curve to the left of μ equals ½.
3. It has inflection points at $\mu-\sigma$ and $\mu+\sigma$.
4. The curve is asymptotic to the horizontal axis at the extremes.
5. The total area under the curve equals one.

Draw a normal (bell-shaped) curve to illustrate 1-5.



Properties of the Normal Probability Curve (continued):

6. Empirical Rule:
   - Approximately 68% of the area under the curve is between μ-σ and μ+σ.
   - Approximately 95% of the area under the curve is between μ-2σ and μ+2σ.
   - Approximately 99.7% of the area under the curve is between μ-3σ and μ+3σ.



A normal curve has two characteristics: mean (μ) and standard deviation (σ).

Example 1—normal curves for two populations with different means:

Population #1	Population #2
$\mu_1 = 50$	$\mu_2 = 70$
$\sigma = 4$	$\sigma = 4$

Draw the normal curves for both populations.

Summary: The two curves are exactly the same, except one curve is to the right of the other curve.

Example 2—normal curves for two populations with different standard deviations.

Population #1	Population #2
$\mu_1 = 50$	$\mu_2 = 50$
$\sigma_1 = 4$	$\sigma_2 = 7$

Draw the normal curves for both populations.

Summary: Increasing the standard deviation causes the curve for Population #2 to become flatter and more spread out. Comparing the two normal curves:

- For Population #1, there is more area under the curve within a given distance of the mean;
- For Population #2, there is more area under the curve away from the mean.

Standardized Variable—A variable is said to be standardized if it has been adjusted (or transformed) such that its mean equals 0 and its standard deviation equals 1.

Standardization can be accomplished using the formula for a z-score: $Z = \frac{X - \mu}{\sigma}$

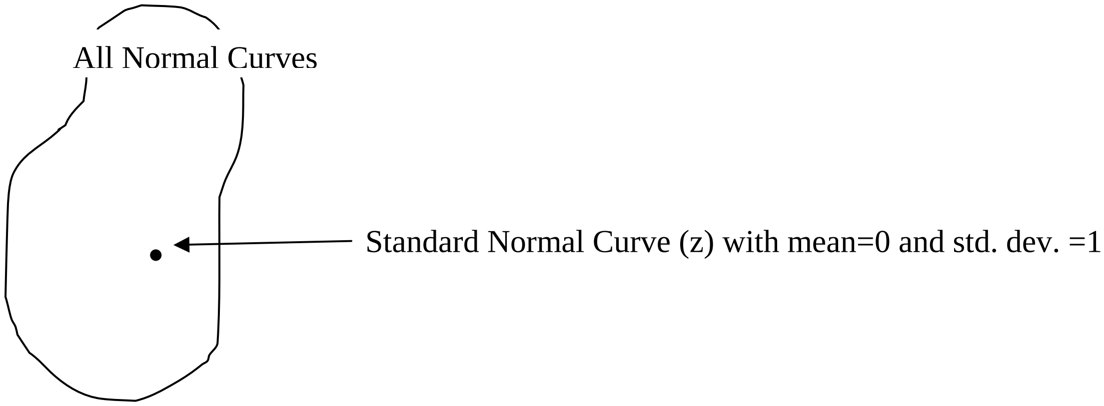
The z-score represents the number of standard deviations that a data value is away from the mean.

	X	$Z = \frac{(X - \mu)}{\sigma}$
	19.5	$= \frac{(19.5 - 19.5)}{6.21} = \frac{0}{6.21} = 0.00$
	22.0	$= \frac{(22.0 - 19.5)}{6.21} = 0.40$
	29.0	$= \frac{(29.0 - 19.5)}{6.21} = \frac{9.5}{6.21} = 1.53$
	10.0	$= \frac{(10.0 - 19.5)}{6.21} = \frac{-9.5}{6.21} = -1.53$
	17.0	$= \frac{(17 - 19.5)}{6.21} = -0.40$
N	5	
ΣX	97.5	
μ	19.5	
σ	6.21	

Note the following about the Z values:

- When the value of X is equal to the mean, then Z is equal to zero (yellow highlight).
- When the value of X deviates above the mean by the same amount that another X-value deviates below the mean, the two Z's are equal except for the difference in signs (+ or -)—see blue highlight.

Normal Probability Distributions (or Curves).



- A normal curve is characterized by its mean, μ, and standard deviation, σ.
- Since there are an infinite number of combinations of μ's and σ's, there are likewise an infinite number of normal curves.
- One particular type of normal curve is the **standard normal curve**…a normal curve with μ=0 and σ=1.

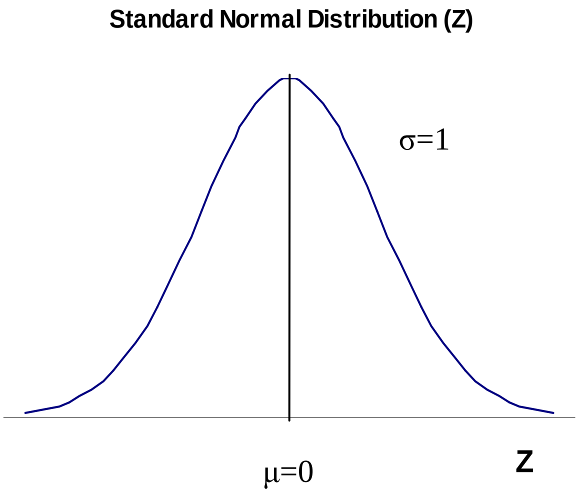
7.2 The Standard Normal Distribution

Standardizing a Normal Random Variable

Suppose the random variable X is normally distributed with mean μ and standard deviation σ. Then the random variable

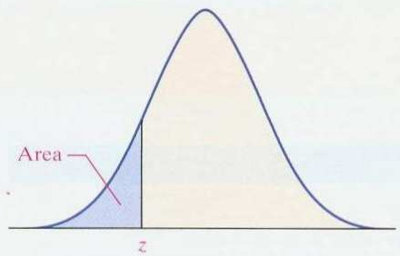
$$Z = \frac{X - \mu}{\sigma}$$

is normally distributed with mean μ=0 and standard deviation σ=1. The random variable Z is said to have the **standard normal distribution**.



Properties of the Standard Normal Curve (Z):

1. The highest point occurs at μ=0.
2. It is a bell-shaped curve that is symmetric about the mean, μ=0. One half of the curve is a mirror image of the other half, i.e., the area under the curve to the right of μ=0 is equal to the area under the curve to the left of μ=0 equals ½.
3. It has inflection points at μ-σ = 0-1 = -1 and μ+σ = 0+1 = +1.
4. The curve is asymptotic to the horizontal axis at the extremes.
5. The total area under the curve equals one.
6. Empirical Rule:
   - Approximately 68% of the area under the curve is between -1 and +1.
   - Approximately 95% of the area under the curve is between -2 and +2.
   - Approximately 99.7% of the area under the curve is between -3 and +3.



The table gives the area under the standard normal curve to the left of a specified Z-score as shown in the figure.

Z-values.

Numbers in the body of the table represent area under the standard normal curve.

TABLE II

Standard Normal Distribution

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.000[illegible]
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.000[illegible]
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.000[illegible]
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.000[illegible]
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.001[illegible]
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.001[illegible]
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.001[illegible]
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.002[illegible]
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.003[illegible]
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.004[illegible]
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	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
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4027	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

-0.2 0.4207 corrected area

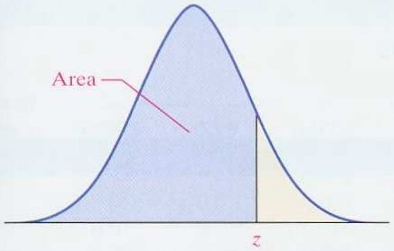


TABLE II *(continued)*

Standard Normal Distribution

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	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	0.7852
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	0.8810	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	0.9772	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	0.9864	0.9868	0.9871	0.9875	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	0.9925	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	0.9979	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
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Finding Area Under the Standard Normal Curve:

1. Find the area under the standard normal curve to the left of Z=-1.40.
2. Find the area under the standard normal curve to the right of Z=1.85.
3. Find the area under the standard normal curve between Z=0.50 and Z=2.25.

Notation for the Probability of a Standard Normal Random Variable	
$P(Z < a)$	represents the probability a standard normal random variable is less than *a*.
$P(Z > a)$	represents the probability a standard normal random variable is greater than a.
$P(a < Z < b)$	represents the probability a standard normal random variable is between *a* and *b*

Examples written using probability notation:

1. Find $P(Z < -1.40)$.
2. Find $P(Z > 1.85)$
3. Find $P(0.50 < Z < 2.25)$

+++

Finding Z-Scores for Given Areas:

1. Find the 85th percentile for the Z distribution—i.e., find z_0 such that $P(Z<z_0)=0.85$.
2. Find z_0 such that $P(Z>z_0)=0.25$.
3. Find the two values of Z (z_1 and z_2) that include the middle 95% of Z values.

+++

For any continuous random variable, the probability of observing a specific value of the random variable is 0. For example, for a standard normal random variable, $P(Z=a)=0$ for any value of a. Because of this, the following probabilities are equivalent:

$$P(a < Z < b) = P(a \le Z < b) = P(a < Z \le b) = P(a \le Z \le b)$$

7.3 Application of the Normal Distribution

Finding the Area under any Normal Curve:

Step 1: Draw the normal curve with the desired area shaded.

Step 2: Convert the values of X to Z-values using $Z = \frac{X - \mu}{\sigma}$.

Step 3: Draw a Z-axis under the X-axis on the normal curve in Step 1, and place the Z-values under the corresponding X-values.

Step 4: Find the area under the normal curve using the Z-values and the standard normal curve in Appendix Table II. *****Note that Z indicates the number of standard deviations a given data value is away from the mean.

The Area under a Normal Curve—Interpreting Your Answer

Suppose a random variable X is normally distributed with mean μ and standard deviation σ. The area under the normal curve for any range of values of the random variable X represents either:

- the proportion of the population with the characteristics described by the range, or
- the probability that a randomly selected individual from the population will have the characteristics described by the range.

Example 1: The random variable X is normally distributed with $\mu = 500$ and $\sigma = 100$. Find $P(X<400)$. Use a graph with labels to illustrate your answer. Interpret your answer (as explained above) in two ways.

Example 2: The random variable X is normally distributed with $\mu = 500$ and $\sigma = 100$. Find $P(X>620)$. Use a graph with labels to illustrate your answer.

Example 3: Scores on the SAT test are normally distributed with $\mu = 500$ and $\sigma = 100$. What score must a student make on the test to be at the 90^{th} percentile? Use a graph with labels to illustrate your answer.

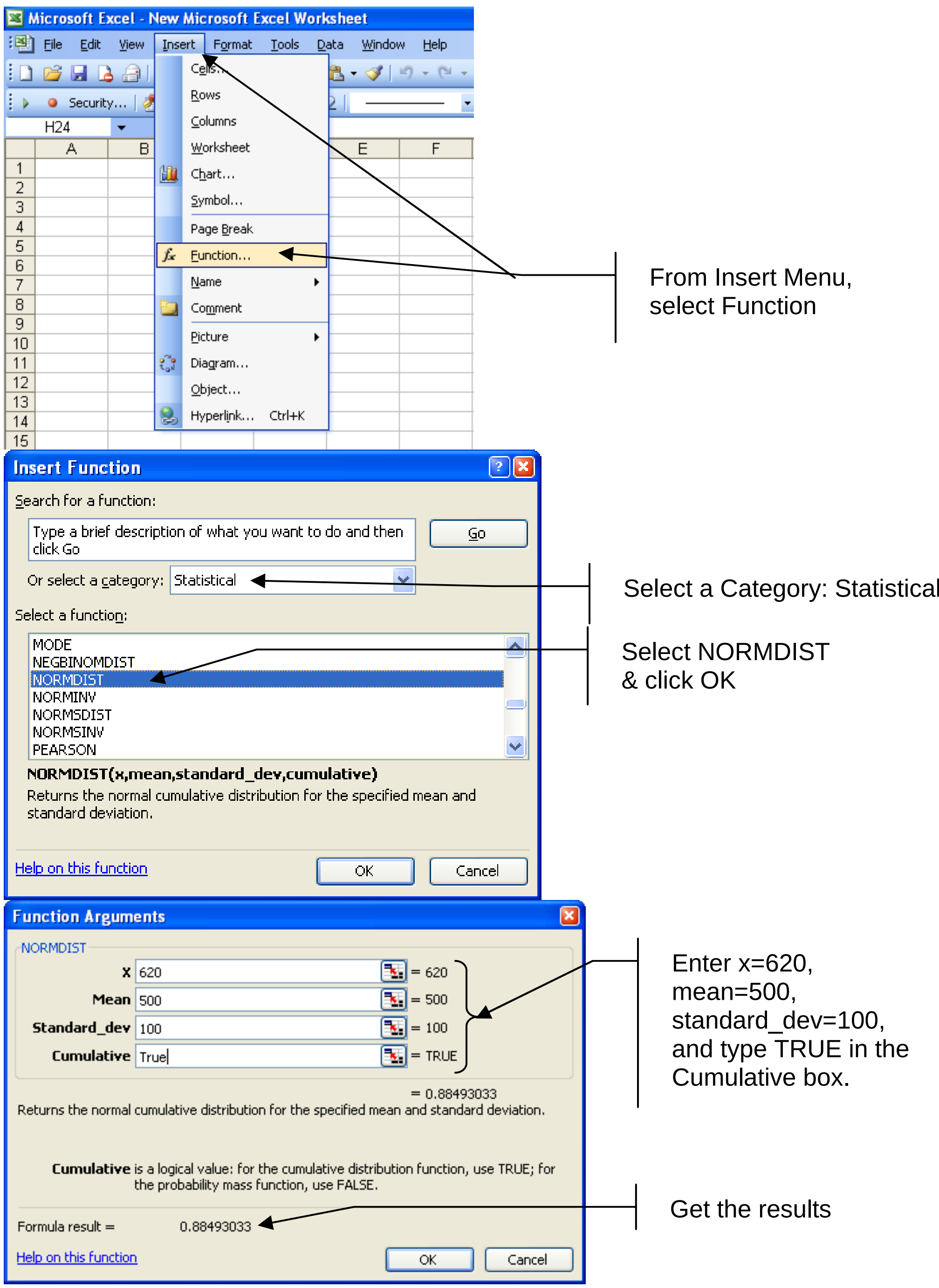
Example 4: Scores on the SAT test are normally distributed with $\mu = 500$ and $\sigma = 100$. What range in SAT scores (x_1 and x_2) includes the middle 50% of scores? Use a graph with labels to illustrate your answer.

Excel and the Normal Distribution: **Finding Areas under the Normal Curve:**

Step 1: Excel will find an area to the left of a specified value. Select **Insert/Function** (***f*$_x$**) from the Windows menubar. In the **Function Category**, select "Statistical." In the **Function Name**, select "NormDist." Click **OK**.

Step 2: Enter the specified value, μ and σ, and set cumulative to TRUE. Click **OK**.

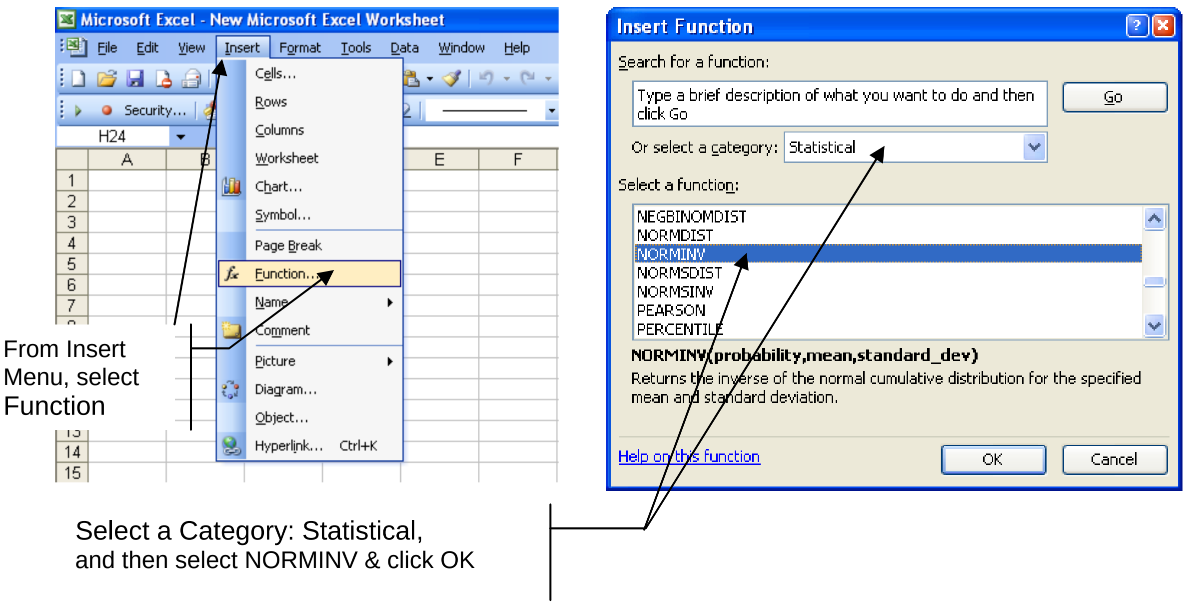
Example: For a normal distribution with μ = 500 and σ = 100, find P(X<620).

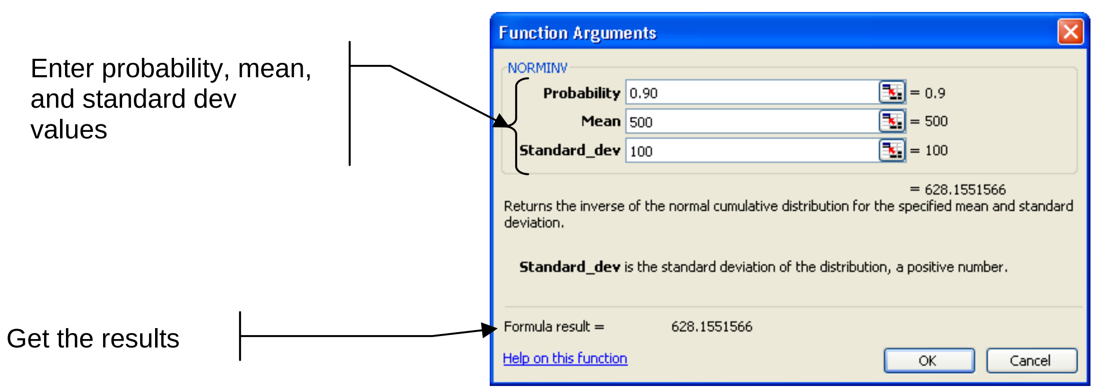


Excel and the Normal Distribution (continued): Finding Scores Corresponding to an Area:

Step 1: Select **Insert/Function** (f_x) from the Windows menubar. In **Function Category**, select "Statistical." In **Function Name**, select "NormInv." Click **OK**.
Step 2: Enter area to the left of the unknown score (i.e., probability), μ and σ**.** Click **OK**.

Example: For a normal distribution with μ = 500 and σ = 100, find the 90th percentile.





Excel and the Standard Normal Distribution (Z):
Finding Areas under the Standard Normal Curve (Z):

Use "Normsdist" and follow the example for "Normdist" above.

Finding Z-Scores Corresponding to an Area:

Use "Normsinv" and follow the example for "Norminv" above.

7.4 Assessing Normality

Suppose that we obtain a simple random sample from a population whose distribution is unknown. Many of the statistical tests that we perform on small data sets (sample size less than 30) require that the population from which the sample is drawn be normally distributed. Up to this point, we have said that a random variable *X* is normally distributed, or at least approximately normal, provided the histogram of the data is symmetric and bell-shaped. This method works well for large data sets, but the shape of a histogram drawn from a small sample of observations does not always accurately represent the shape of the population. For this reason, we need additional methods for assessing the normality of a random variable *X* when we are looking at sample data.

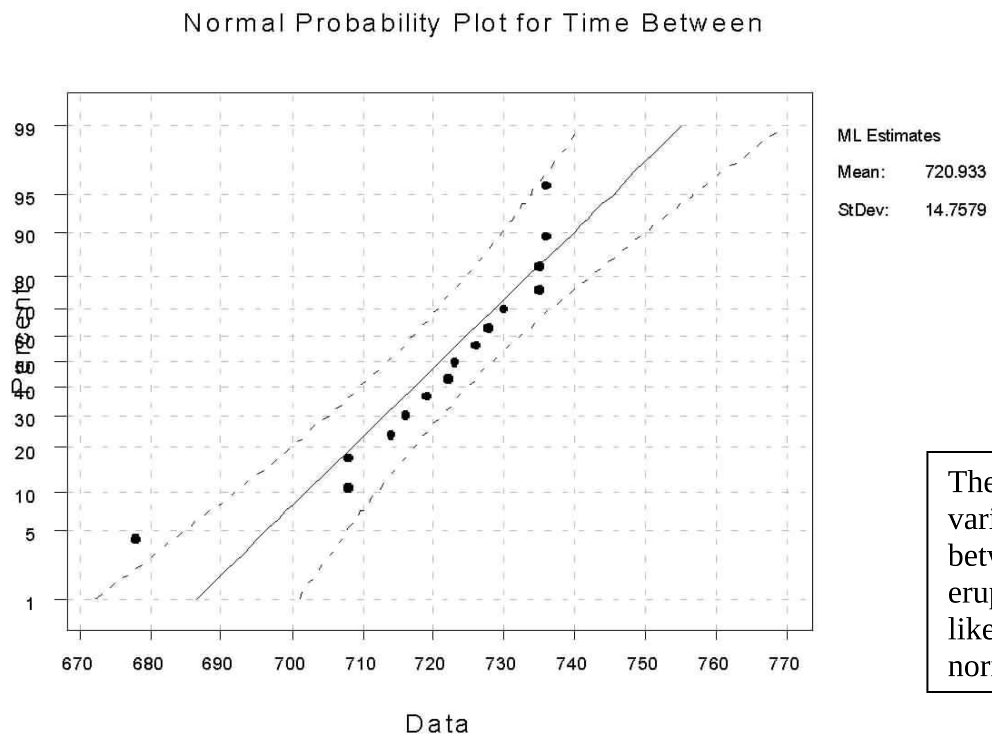
A **normal probability plot** plots observed data versus *normal scores*. A **normal score** is the expected *Z*-score of the data value if the distribution of the random variable is normal.

During this semester, we will be content in reading normal probability plots constructed using the statistical software package, Minitab. In Minitab, if the points plotted lie within the bounds provided in the graph, then we have reason to believe that the sample data come from a population that is normally distributed.

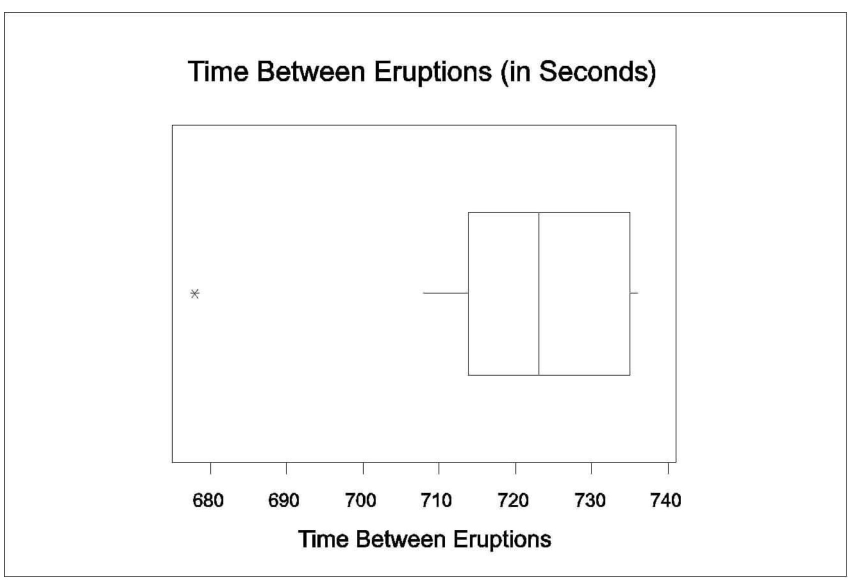
EXAMPLE: *Interpreting a Normal Probability Plot*

The following data represent the time between eruptions (in seconds) for a random sample of 15 eruptions at the Old Faithful Geyser in California. Is there reason to believe the time between eruptions is normally distributed?

728	678	723	735	735
730	722	708	708	714
726	716	736	736	719



The random variable "time between eruptions" is likely not normal.

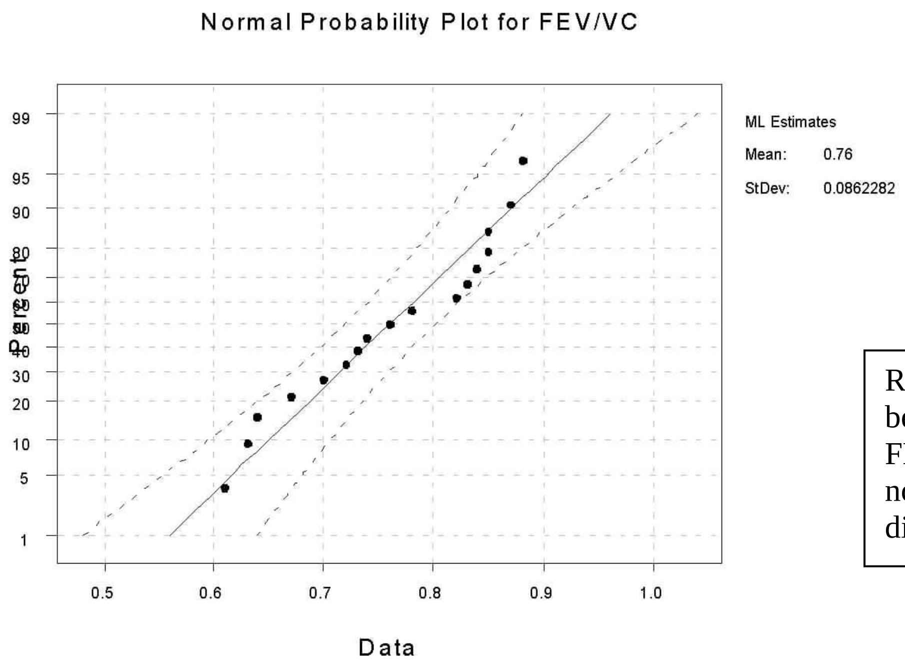


EXAMPLE: ***Assessing Normality***

Suppose that seventeen randomly selected workers at a detergent factory were tested for exposure to a Bacillus subtillis enzyme by measuring the ratio of forced expiratory volume (FEV) to vital capacity (VC). NOTE: FEV is the maximum volume of air a person can exhale in one second; VC is the maximum volume of air that a person can exhale after taking a deep breath. Is it reasonable to conclude that the FEV to VC (FEV/VC) ratio is normally distributed?

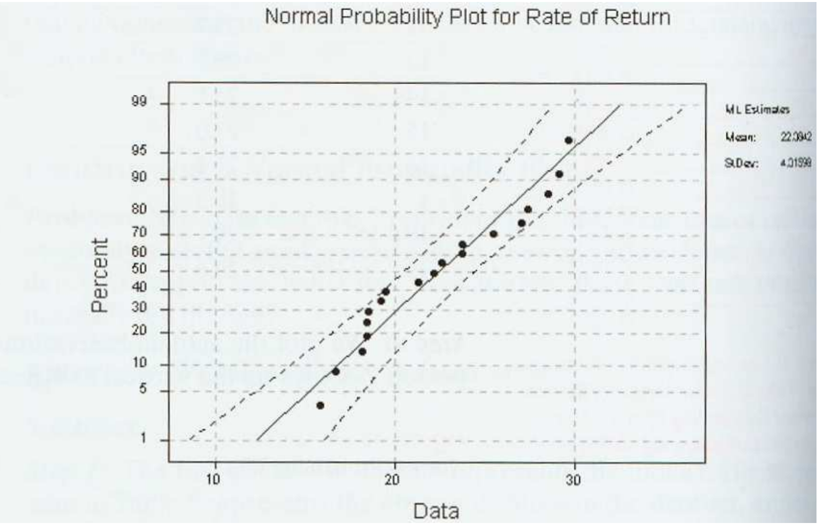
Shore, N.S.; Greene R.; and Kazemi, H. "Lung Dysfunction in Workers Exposed to *Bacillus subtillis* Enzyme," *Environmental Research*, 4 (1971), pp. 512 - 519.

0.61	0.70	0.76	0.84
0.63	0.72	0.78	0.85
0.64	0.73	0.82	0.85
0.67	0.74	0.83	0.87
0.88			



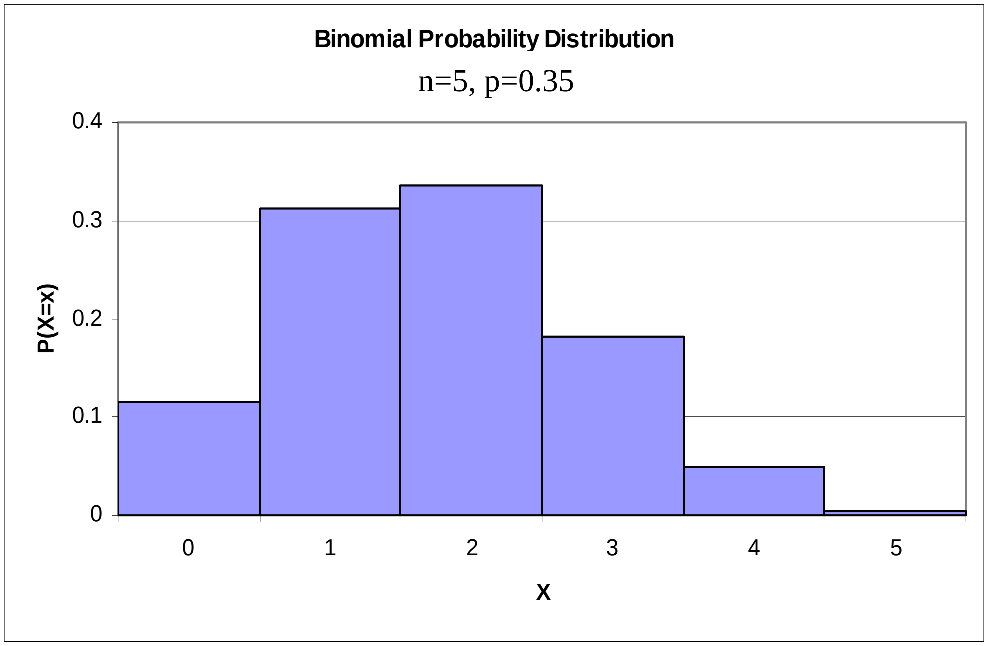
Reasonable to believe that FEV/VC is normally distributed.

Figure 44


Normal Probability Plot for Rate of Return
Percent
Data
99
95
90
80
70
60
50
40
30
20
10
5
1
10
20
30
ML Estimates
Mean: 22.0842
StDev: 4.01599

Putting It All Together (p. 271)

Probability for a discrete random variable is read from the vertical (y) axis, e.g., P(X=1) = 0.31



Notice that area is equal to probability in the binomial probability distribution:

- Area of a rectangle = height x width
- Width of each rectangle is 1
- P(X=1) is equal to the area 0.31 x 1