1. Random Variables and Distributions.

1. Random variables.

1. Random variable is a numerical event whose value is determined by a chance process. Notation: $X$.

2. Two types of random variables.

1. Discrete random variable: a variable that takes on a finite (countable) number of values over a given interval; expressed in whole numbers.

Examples:

* Number of persons in a household:
* Number of customers served in an hour:
* Number of RVs rented by a firm in a given period of time:

In each of these cases: $X \rightarrow 1, 2, 3 \ldots n$

2. Continuous random variable: a variable that takes on an infinite (uncountable) number of values over a given interval; expressed in fractional numbers.

Examples:

* Net weight of corn flakes in a box of Corn Flakes:
* Time required to prepare a given pizza at location A and to deliver it to location B:
* Length of time to failure of a piece of machinery:
* Dosage of medication to treat a given illness:

In each of these cases: $X \rightarrow$ Infinite number of possible values

In sum:

Discrete variables are counted quantities expressed in whole numbers.
Continuous variables are measured quantities having to do with length, weight, volume, velocity, etc.

5.1
2. Probability Distributions.

Corresponding to the two types of random variables there are two types of probability distributions:

(1) Discrete probability distribution; and
(2) Continuous probability distribution.
1. **Discrete probability distribution:**

A distribution which gives:

1. the possible values of the discrete random variable, and
2. the probabilities associated with each value.

That is, $X$ and $P(X)$.

**Properties of the discrete probability distribution:**

1. $P(X) \geq 0$ for all values of $X$.
2. $\Sigma P(X) = 1$.

**Examples of discrete probability distributions:**

1. Dice example: Table and graph.
2. Examples given in Topic 5 Homework Set.

**Note:** A graph of a discrete probability distribution is usually rendered as a bar graph of some form suggesting a variable that can take on only whole number values and not fractional values.

**A similarity:** A discrete probability distribution is much the same as a relative frequency distribution (Topic 2) for a discrete random variable.
2. **Continuous probability distribution:**

Note: Recall that a continuous random variable is one that can take on any value (including fractional values) over a given interval. Since there are an infinite number of such values, one cannot list every possible value in this case.

The continuous probability distribution takes the form of a probability density function (pdf), or its companion probability density curve, the representation which we will use in our subsequent analyses.

The probability density curve for a continuous random variable, \( X \), is a smooth curve that provides the probability for any interval of values of the random variable, \( X \).

The famous normal curve is such a probability density curve.

The probability density curve illustrated.

Properties of a probability density curve (with reference to the normal curve illustration).

1. \( P(0 \leq X \leq \infty) = 1 \)  The total area under the curve is equal to 1 representing 100%.

2. \( P(a \leq X \leq b) = \) the shaded area under the curve over the interval from \( a \) to \( b \).

3. \( P(X = a) = 0 \)  Since a line, such as that from \( a \) to \( c \), has zero width it will also have an area of zero. Thus, the probability of any given single value of \( X \) will be zero as well.

Note: In a probability density curve, probabilities are rendered by areas under the curve associated with some interval of values of \( X \).
5.5

Some technical aspects of the normal probability density curve for the student familiar with the calculus:

(1) The height of the curve for a given value of $X$ is given by:

$$f(X) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{(X-\mu)}{\sigma} \right)^2}$$

(2) The probability that $X$ lies between the values, $a$ and $b$, is given by the integral:

$$P(a \leq X \leq b) = \int_{a}^{b} f(X) \, dX$$
3. Expected value, variance, and standard deviation of a discrete random variable.

1. Expected value or mean:

\[ E(X) = \sum X \cdot P(X) \]

where: 
- \( E(X) \) = expected value (mean) of the discrete variable;
- \( X \) = a possible value of the variable;
- \( P(X) \) = probability of the possible value of the variable.

Note: This formula is a variation of the formula for the weighted mean:

\[ \frac{\sum Xw}{\sum w} \]

Equivalent to the term, \( \sum w \), in the weighted mean expression would be a denominator, \( \sum P(X) \), in the expected value expression. In the case of a probability distribution this latter term is dropped because it will always be equal to 1.

2. Variance:

\[ \sigma^2(X) = \sum \{X - E(X)\}^2 \cdot P(X) \]

where: 
- \( \sigma^2(X) \) = variance of the discrete variable;
- \( X \) = a possible value of the variable;
- \( E(X) \) = expected value (mean) of the variable;
- \( P(X) \) = probability of the possible value of the variable.

3. Standard deviation:

\[ \sigma(X) = \sqrt{\sigma^2(X)} \]

where: 
- \( \sigma(X) \) = standard deviation of the discrete variable;
- \( \sigma^2(X) \) = variance of the variable.
Example problem: Calculate the expected value, variance, and standard deviation of the discrete variable given in the probability distribution in Topic 5 Homework Set, problem 1.

<table>
<thead>
<tr>
<th>X</th>
<th>P(X)</th>
<th>X*P(X)</th>
<th>X-E(X)</th>
<th>(X-E(X))^2</th>
<th>[●]P(X)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>.2</td>
<td>1.0</td>
<td>-1.3</td>
<td>1.69</td>
<td>.338</td>
</tr>
<tr>
<td>6</td>
<td>.4</td>
<td>2.4</td>
<td>-0.3</td>
<td>0.09</td>
<td>.036</td>
</tr>
<tr>
<td>7</td>
<td>.3</td>
<td>2.1</td>
<td>0.7</td>
<td>0.49</td>
<td>.147</td>
</tr>
<tr>
<td>8</td>
<td>.1</td>
<td>0.8</td>
<td>1.7</td>
<td>2.89</td>
<td>.289</td>
</tr>
</tbody>
</table>

E(X) = 6.3  \quad \sigma^2(X) = .810
\quad \sigma(X) = \sqrt{.81} = .9

1. E(X) = \Sigma X*P(X) = 6.3
2. \sigma^2(X) = \Sigma (X-E(X))^2 * P(X) = .81
3. \sigma(X) = \sqrt{\sigma^2(X)} = .9

Histogram of this discrete distribution.
2. The Binomial Distribution.

1. The binomial distribution and the Bernoulli process.

The binomial distribution is used in the analysis of problems conforming to the Bernoulli process.

The Bernoulli process: A series of trials having the following three properties:

1. Each trial has two possible outcomes, "success" or "failure."
2. The probability of "success" is constant over trials; the same from one trial to the next.
3. The outcomes of the trials are independent of one another.

2. Determining binomial probabilities.

Binomial probabilities can be determined using:

1. The binomial formula;
2. A table of binomial probabilities;

Consider each of these tools.
5.9

1. Using the binomial formula.

1. The formula:

(1) \[ P(X|n, \pi) = \binom{n}{X} (\pi)^X (1-\pi)^{n-X} \]

where: \( X \) = number of "successes;"
\( n \) = number of trials (sample size);
\( \pi \) = probability of a "success;"

\( \binom{n}{X} \) = combination term, \( X \) successes in \( n \) trials.

Since \( \binom{n}{X} = \frac{n!}{X!(n-X)!} \) then:

(2) \[ P(X|n, \pi) = \frac{n!}{X!(n-X)!} (\pi)^X (1-\pi)^{n-X} \]

Note: In the use of this formula:

(1) ! is the factorial notation;

And by definition:

(2) 0! = 1; and

(3) Any \( n^0 = 1 \)
5.10

2. A problem:

Matilda Olson, an IBM representative selling the Abacus360 computer, calls on three clients in a given week. Based on past experience, she has a 20 percent chance of obtaining an order from a given client. What is the probability distribution for the number of orders she will receive?

Given: \( n = 3 \)
\( \pi = .2 \)
\( X \rightarrow 0,1,2,3 \)

Formula: \( P(X|n,\pi) = \frac{n!}{X!(n-X)!} \pi^X (1-\pi)^{n-X} \)
5.11

Solution:

Let $X = 0$:  $P(X=0|n=3, \pi=.2) = \frac{3!}{0!(3-0)!} \cdot (0.2)^0 \cdot (1-0.2)^{3-0}$

\[= (1)(1)(.512) = .512\]

Let $X = 1$:  $P(X=1|n=3, \pi=.2) = \frac{3!}{1!(3-1)!} \cdot (0.2)^1 \cdot (1-0.2)^{3-1}$

\[= (3)(.2)(.64) = .384\]

Let $X = 2$:  $P(X=2|n=3, \pi=.2) = \frac{3!}{2!(3-2)!} \cdot (0.2)^2 \cdot (1-0.2)^{3-2}$

\[= (3)(.04)(.8) = .096\]

Let $X = 3$:  $P(X=3|n=3, \pi=.2) = \frac{3!}{3!(3-3)!} \cdot (0.2)^3 \cdot (1-0.2)^{3-3}$

\[= (1)(.008)(1) = .008\]

Probability distribution for Olson's received orders:

<table>
<thead>
<tr>
<th>$X$</th>
<th>$P(X)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.512</td>
</tr>
<tr>
<td>1</td>
<td>.384</td>
</tr>
<tr>
<td>2</td>
<td>.096</td>
</tr>
<tr>
<td>3</td>
<td>.008</td>
</tr>
</tbody>
</table>

$\sum P(X) = 1.000$
2. Using the binomial table.

1. The table.

2. The Matilda Olson problem again using the binomial table.

Matilda Olson, an IBM representative selling the Abacus360 computer, calls on three clients in a given week. Based on past experience, she has a 20 percent chance of obtaining an order from a given client. What is the probability distribution for the number of orders she will receive?

Given: n = 3
\[ \pi = .2 \]
\[ X \rightarrow 0,1,2,3 \]

Solution:

Probability distribution for Olson's received orders:

<table>
<thead>
<tr>
<th>X</th>
<th>P(X)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.5120</td>
</tr>
<tr>
<td>1</td>
<td>.3840</td>
</tr>
<tr>
<td>2</td>
<td>.0960</td>
</tr>
<tr>
<td>3</td>
<td>.0080</td>
</tr>
</tbody>
</table>

\[ \sum P(X) = 1.0000 \]
3. Using Excel to determine binomial probabilities.

1. Accessing the binomial function in Excel.

   1. Click on paste function or function wizard ($f_x$) on the standard toolbar.
   2. Select the function category, Statistical.
   3. Select the function name, Binomdist.
   4. Input the requested information in the Binomdist menu.
   5. Excel syntax:

   \[
   \text{BINOMDIST}(\text{number}_s, \text{trials}, \text{probability}_s, \text{cumulative})
   \]

   \[
   \begin{align*}
   \text{number}_s &= X \\
   \text{trials} &= n \\
   \text{probability}_s &= \pi \\
   \text{cumulative} &= 1; \text{ noncumulative} = 0
   \end{align*}
   \]
3. Expected value, variance, and standard deviation of a binomial distribution.

1. Expected value or mean:

\[ E(X) = n\pi \]

where: \( E(X) \) = expected value (mean) of the binomial distribution;
\( n \) = number of trials (sample size);
\( \pi \) = probability of a "success."

2. Variance:

\[ \sigma^2(X) = n\pi(1-\pi) \]

where: \( \sigma^2(X) \) = variance of the binomial distribution;
\( n \) = number of trials (sample size);
\( \pi \) = probability of a "success."

3. Standard deviation:

\[ \sigma(X) = \sqrt{\sigma^2(X)} \]

where: \( \sigma(X) \) = standard deviation of the binomial distribution;
\( \sigma^2(X) \) = variance of the binomial distribution.
Example problem: Calculate the expected value, variance, and standard deviation of the Matilda Olson orders received distribution.

Given: n = 3  
\( \pi = .2 \)  
\( X \rightarrow 0,1,2,3 \)

Probability distribution for Matilda Olson's received orders:

<table>
<thead>
<tr>
<th>X</th>
<th>P(X)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.5120</td>
</tr>
<tr>
<td>1</td>
<td>.3840</td>
</tr>
<tr>
<td>2</td>
<td>.0960</td>
</tr>
<tr>
<td>3</td>
<td>.0080</td>
</tr>
</tbody>
</table>

\[ \Sigma P(X) = 1.0000 \]

(1) \( E(X) = n\pi = (3)(.2) = .6 \)
(2) \( \sigma^2(X) = n\pi(1-\pi) = (3)(.2)(.8) = .48 \)
(3) \( \sigma(X) = \sqrt{\sigma^2(X)} = \sqrt{.48} = .6928 \)

Histogram of this binomial distribution:
3. The Normal Distribution.

1. The importance of the normal distribution.

   1. The measurements obtained in many random processes are known to follow this distribution.

   - Distributions of the size and/or weight of virtually all plants and animals in the natural world.
   - Distributions around the starting times of the arrival of students in classes and of people at concerts or athletic events.
   - Distributions around peak load times of usage of public utilities like electricity and public facilities like commuter highways.

2. Normal probabilities can be used to approximate other probability distributions such as the binomial distribution.

3. Distributions of statistics like the sample mean and sample proportion tend to be normally distributed, when the sample size (n) is sufficiently large, regardless of the distribution of the parent population.

In sum, in the words of W. J. Youden (1900-1971):

THE NORMAL LAW OF ERROR STANDS OUT IN THE EXPERIENCE OF MANKIND AS ONE OF THE BROAREST GENERALIZATIONS OF NATURAL PHILOSOPHY ♦ IT SERVES AS THE GUIDING INSTRUMENT IN RESEARCHES IN THE PHYSICAL AND SOCIAL SCIENCES AND IN MEDICINE AGRICULTURE AND ENGINEERING ♦ IT IS AN INDISPENSABLE TOOL FOR THE ANALYSIS AND THE INTERPRETATION OF THE BASIC DATA OBTAINED BY OBSERVATION AND EXPERIMENT
2. The normal probability distribution.

1. The normal probability distribution can be rendered as a probability density curve for a normally distributed variable, $X$.

Probability density curve illustrated.

$X = \text{normally distributed continuous random variable};$

$f(X) = \text{height of the density curve for a given value of } X.$
3. Some observations about the normal distribution.

1. The normal curve is a bell-shaped symmetric curve with a mean of \( \mu \) and a standard deviation of \( \sigma \). The symmetry property means that the right side of the distribution has the same shape as the left side, and the three measures of central tendency are equal, that is, \( \mu = \text{Mdn} = \text{Mode} \).

2. The location of the curve (with reference to the horizontal axis) is determined by \( \mu \); the shape of the curve (its variation or dispersion) is determined by \( \sigma \).

3. The normal curve is asymptotic with respect to the horizontal axis. This means that either tail of the distribution continuously approaches the horizontal axis, but never quite touches it.

4. Since it is a probability density curve, probabilities or proportions, appear as areas under the curve where (with reference to the illustration):

1. \( P(-\infty \leq X \leq +\infty) = 1 \). The total area under the curve is equal to 1 or 100%.

2. \( P(a \leq X \leq b) \) = the shaded area under the curve between points a and b.

3. \( P(X = a) = 0 \). The line from point a to point c has zero width and, therefore, zero area. Thus, the probability that \( X \) will be equal to a single value, such as that represented by point a, will always be zero.
5.19

4. The standard normal probability distribution.

1. The standard normal probability distribution is the distribution of the standard normal variable:

\[ Z = \frac{X - \mu}{\sigma} \]

where: 
- \( Z \) = standard normal variable;
- \( X \) = value of the continuous random variable;
- \( \mu \) = population mean;
- \( \sigma \) = population standard deviation.

2. Some additional observations about the standard normal probability distribution.

Note: The observations about the normal distribution appearing on page 5.18 above apply as well to the standard normal probability distribution.

1. The standard normal variable, \( Z \), expresses a given value of \( X \) in standard deviation units.

- \( Z = +1.50 \): \( X \) is 1.50 standard deviations above the mean.
- \( Z = -2.33 \): \( X \) is 2.33 standard deviations below the mean.

2. The standard normal distribution, the distribution of \( Z \), has a mean of zero (\( \mu = 0 \)) and a standard deviation of one (\( \sigma = 1 \)).
3. The standard normal probability distribution illustrated.

Consider the birth weight of a population of newborn infants having a mean \((\mu)\) birth weight of 8 pounds with a standard deviation \((\sigma)\) of 1 pound.

Illustration.
5. Determining normal probabilities.

Normal probabilities can be determined using:

1. Integral calculus;
2. Standard normal distribution table;

Consider each of these tools.

1. Integral evaluation is used to produce the normal table.
2. Using the normal table to determine normal probabilities.

1. The table.
   Notations: \( Z \) is the standard normal variable defined above; top row provides the second digit to the right of the decimal for a given \( Z \).

2. Example problem: A large population of infants have a normally distributed birth weight with a mean (\( \mu \)) of 8 pounds with a standard deviation (\( \sigma \)) of 1 pound.

   1. What percentage of infants have a birth weight between 7 and 9 pounds?

      Solution:
      \[
      Z(7) = \frac{7-8}{1} = -1.00 \quad Z(9) = \frac{9-8}{1} = +1.00
      \]
      \[
      P(7 \leq X \leq 9) = (.3413)(2) = .6826 \approx 68.3 \text{ percent}
      \]
      Answer: 68.3 percent.

   2. What percentage of infants have a birth weight between 6 and 10 pounds?

      Solution:
      \[
      Z(6) = \frac{6-8}{1} = -2.00 \quad Z(10) = \frac{10-8}{1} = +2.00
      \]
      \[
      P(6 \leq X \leq 10) = (.4772)(2) = .9544 \approx 95.4 \text{ percent}
      \]
      Answer: 95.4 percent.

   3. What percentage of infants have a birth weight between 5 and 11 pounds?

      Solution:
      \[
      Z(5) = \frac{5-8}{1} = -3.00 \quad Z(11) = \frac{11-8}{1} = +3.00
      \]
      \[
      P(5 \leq X \leq 11) = (.4987)(2) = .9974 \approx 99.7 \text{ percent}
      \]
      Answer: 99.7 percent.

Note: These answers are consistent with what we found using the Empirical Rule back in Topic 3:

1. \( \pm 1\sigma \rightarrow 68\% \)
2. \( \pm 2\sigma \rightarrow 95\% \)
3. \( \pm 3\sigma \rightarrow 99.7\% \)
5.23

3. More practice in using the normal table.

Suppose the average high temperature in June in Las Vegas is normally distributed with a mean ($\mu$) of 100 degrees Fahrenheit and a standard deviation ($\sigma$) of 10 degrees. What is the probability that the temperature on a given June day will be:

(1) between 110 and 120 degrees?
(2) between 75 and 85 degrees?
(3) less than 115 degrees?
(4) more than 105 degrees?
(5) less than 113.5 degrees?

Solutions:

(1) $Z(110) = \frac{110-100}{10} = +1.00$  
$P(110 \leq X \leq 120) = .4772 - .3413 = .1359 = 13.59$ percent

(2) $Z(75) = \frac{75-100}{10} = -2.50$  
$P(75 \leq X \leq 85) = .4938 - .4332 = .0606 = 6.06$ percent

(3) $Z(115) = \frac{115-100}{10} = +1.50$  
$P(X < 115) = .4332 + .5000 = .9332 = 93.32$ percent

(4) $Z(105) = \frac{105-100}{10} = +0.50$  
$P(X > 105) = .5000 - .1915 = .3085 = 30.85$ percent

(5) $Z(113.5) = \frac{113.5-100}{10} = +1.35$  
$P(X < 113.5) = .5000 + .4115 = .9115 = 91.15$ percent

1. Accessing the normal function in Excel.

   1. Click on paste function or function wizard ($f_x$) on the standard toolbar.
   2. Select the function category, Statistical.
   3. Select function name from the following:
      - NORMDIST
        Returns cdf: $P(X \leq X_i | \mu_i, \sigma_i, \text{cumulative}=1)$
        Returns pdf: $f_x | X_i, \mu_i, \sigma_i, \text{cumulative}=0)$
      - NORMINV
        Returns invcdf: $X | P(\leq \mu_i), \mu_i, \sigma_i$
      - NORMSDIST
        Returns $Z(\text{standard normal})$cdf: $P(Z \leq Z_i | Z_i)$
      - NORMSINV
        Returns invZcdf: $Z | P(\leq \mu_i)$
6. The normal approximation of the binomial distribution.

The normal approximation of the binomial is useful in analyses of discrete data where \( n \) is large, beyond available binomial tables.

1. Condition for its use:

\[
\begin{align*}
n\pi & \geq 5 \\
n(1-\pi) & \geq 5
\end{align*}
\]

2. Correction for continuity:

\[
Z = \frac{X \pm 0.5 - \mu}{\sigma}
\]

Adding or subtracting .5:

<table>
<thead>
<tr>
<th>If you want this probability:</th>
<th>Then the appropriate correction is to:</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(X \geq X_l) )</td>
<td>Subtract .5 from X</td>
</tr>
<tr>
<td>( P(X &lt; X_l) )</td>
<td>Subtract .5 from X</td>
</tr>
<tr>
<td>( P(X \leq X_l) )</td>
<td>Add .5 to X</td>
</tr>
<tr>
<td>( P(X &gt; X_l) )</td>
<td>Add .5 to X</td>
</tr>
</tbody>
</table>

3. Example problem:

Suppose the Terrible Herbst Hotel on Paradise Road in Las Vegas has 200 rooms with an average occupancy rate of 80 percent.

1. What is the probability that at least 170 rooms will be occupied on a given day?

2. What is the probability that more than 155 rooms will be occupied on a given day?