**Statistics 10   Summer 2009**

**LAB 3**
**Probability Models: Binomial and Normal Distributions**

During the Lab 3 students should explore Binomial and Normal Distributions and understand when and how we can approximate Binomial Distribution using Normal Distribution.

**BINOMIAL  DISTRIBUTION: Simulation using the applet.**

Open the applet ***Binomial Distribution*.*** Go down to Applets and click on the red link "Binomial Coin Experiment". You will see the window with the Java applet.

**Simulation 1. Three-Child Family   (Do as a class. Do not include in the lab report)**

Suppose we want to describe the probability distribution for the number of girls in a three-child family.

1. Check that all Binomial conditions are satisfied.

2. Use three coins and default probability of success p = 0.5. Run the experiment 10 times.  Record the relative frequencies.

n = 3   p = 0.5            RUN = 10

| Number of girls | Distribution (Theoretical Probability) | Data (Relative Frequencies) |
|---|---|---|
| 0 | 0.125 | |
| 1 | 0.375 | |
| 2 | 0.375 | |
| 3 | 0.125 | |
| Mean | 1.5 | |
| St Dev | 0.86603 | |

3. Reset the applet. Use three coins and run the experiment 100 times.  Record the relative frequencies.

n = 3   p = 0.5            RUN = 100

| Number of girls | Distribution (Theoretical Probability) | Data (Relative Frequencies) |
|---|---|---|
| 0 | 0.125 | |
| 1 | 0.375 | |
| 2 | 0.375 | |
| 3 | 0.125 | |
| Mean | 1.5 | |
| St Dev | 0.86603 | |

4. Compare the results. Make your conclusion.

5. Run the experiment using the option Continuous and stop it by yourself when the empirical distribution will have exactly the same shape as theoretical distribution. How many trials were made?

**Simulation 2. . Multiple Choice Test. (On Your Own)**

Suppose a student is totally unprepared for a 5 question multiple-choice test, with four options for each question, and must guess the answer to each question.
We want to describe the probability distribution for the number of correct answers in a 5-questions test. Assume that coin we use is unbalanced and the probability of heads may be any, between 0 and 1.

1. Check that all Binomial conditions are satisfied. Determine the probability of success ***p***.

2. Use five coins and select the new value of the probability ***p***. Run the experiment 20 times. Record the relative frequencies.

n = 5 RUN = 20

| Number of correct answers | Distribution (Theoretical Probability) | Data (Relative Frequencies) |
|---|---|---|
| 0 | 0.2373 | |
| 1 | 0.39551 | |
| 2 | 0.26367 | |
| 3 | 0.08789 | |
| 4 | 0.01465 | |
| 5 | 0.00098 | |
| Mean | 1.25 | |
| St Dev | 0.96825 | |

3. Reset and run the experiment 200 times. Record the relative frequencies.

n =5 RUN = 200

| Number of correct answers | Distribution (Theoretical Probability) | Data (Relative Frequencies) |
|---|---|---|
| 0 | 0.2373 | |
| 1 | 0.39551 | |
| 2 | 0.26367 | |
| 3 | 0.08789 | |
| 4 | 0.01465 | |
| 5 | 0.00098 | |
| Mean | 1.25 | |
| St Dev | 0.96825 | |

4. Compare the results. Make your conclusion.

5. Run the experiment using the option Continuous and stop it by yourself when the empirical distribution will have exactly the same shape as theoretical distribution. How many trials were made?

**Properties of the Binomial Distribution.**

1. Begin with the default probability of heads (success) ***p = 0.50*** and number of coins (trials) ***n = 10***. Look at the shape of the distribution ***Binom (0.5, 10)***.

    a. Assume that coin we use is unbalanced and the probability of heads may be any, between 0 and 1. Change the probability from ***p = 0.04*** to ***p = 0.96.*** For each find and record the expected value (center) and the standard deviation (spread) of the binomial random variable on the table provided below. They are shown on the bottom of the column "Distribution" (scroll it down).

***Binom(p, 10)***

| ***n = 10*** | ***p = 0.04*** | ***p = 0.16*** | ***p = 0.24*** | ***p = 0.40*** | ***p = 0.50*** | ***p = 0.60*** | ***p = 0.76*** | ***p = 0.84*** | ***p = 0.96*** |
|---|---|---|---|---|---|---|---|---|---|
| ***E(X)*** | | | | | | | | | |
| ***σ(X)*** | | | | | | | | | |

    b. Describe the behavior of the shape of the distribution.
    c. What do you notice about the distribution for the pairs of the probabilities ***p = 0.04*** and ***p = 0.96***; ***p = 0.16*** and ***p = 0.84***; ***p = 0.24*** and ***p = 0.76***; ***p = 0.40*** and ***p = 0.60***?

2. Select ***p = 0.50*** and change the number of trials n ***= 10*** to ***n = 20***, ***n = 30, n = 40, n = 50***
    a. Change the sample size ***n = 10*** to ***n = 20*** and find and record the expected value, variance and standard deviation on the table.

***Binom(0.50, n)***

| ***p = 0.50*** | ***n = 10*** | ***n = 20*** | ***n = 30*** | ***n = 40*** | ***n = 50*** |
|---|---|---|---|---|---|
| ***E(X)*** | | | | | |
| ***σ(X)*** | | | | | |

    b. Describe the behavior of the shape of the distribution when the number of trial increases.

3. Select a small value of the probability of success, such as ***p = 0.20***.
    a. Increase the sample size to ***n = 20*** and find and record the expected value, variance and standard deviation.
    b. Notice the shape of distribution.

***Binom(0.20, n)***

| ***p = 0.20*** | ***n = 10*** | ***n = 20*** | ***n = 30*** | ***n = 40*** | ***n = 50*** |
|---|---|---|---|---|---|
| ***E(X)*** | | | | | |
| ***σ(X)*** | | | | | |

4. Make your own brief conclusion (3-4 sentences), comparing the shape of the Binomial Distribution for different values of parameters ***p*** and ***n***.

5. Answer two important questions (after finishing the Normal Distribution part)
    a. When does the Binomial Distribution look "nearly" Normal?
    b. When can we use a Normal Model to approximate a Binomial Probability model?

**NORMAL DISTRIBUTION: Stat 10 Student's Survey Data**

The first good way to see if a Normal model might work is drawing a histogram of the data. And the second way is to use more specialized graphical display – the Normal probability plot.

1. Open a collection: From **Moodle** open the **Survey Data** file that contains data from a sample of 1325 UCLA students who completed the Stat 10 survey.

2. On the **FATHOM** window double click on the **Collection** to open **Inspector**. Select **Cases** and check the **Attributes.**

3. The variable **Sleep** records the number of hours of sleep reported last night. Open a new **Graph** and make a histogram. You should try to vary the bin-widths to see how this affects the shape. Double click on graph and change the value of **binWidth**. Describe the shape and the characteristic features of the distribution. Do you think the Normal Model will fit these data?

4. Open the **Summary Table** for **Sleep** and add **Basic Statistics**. Delete unused stdError and count(missing). Record the values of the mean $\mu$ and standard deviation $\sigma$.

5. Change the scale for the histogram by selecting **Graph _ Scale _ Density.** The graph now shows the histogram using relative frequencies.

6. Add the graph of Normal density curve: **Graph _ Plot Function**.
   In a new window select
   **Functions** → **Distributions** → **Normal** → **normalDensity(Sleep, $\mu$, $\sigma$) –** use the values of the mean and standard deviation.
   Vary the width of the bins. Notice that sometimes the model seems to fit well, other times not.
   *Make your conclusion* how the Normal curve fits our data.

7. A more reliable way of checking to see whether the distribution is "nearly normal" is with a Normal Quantile Plot. If the distribution is normal, this plot is a straight line. If it is nearly normal, the plot will be nearly straight.
   Open a new **Graph** for **Sleep**. From the Graph pull-down menu, select **Normal Quantile Plot.**
   *Analyze the graph and make your conclusion*: is the distribution of the data approximately Normal or the Normal distribution is not reasonable.

8. Repeat **steps 3 - 7** for the variable **TravelTime** and make your conclusion.

9. **On Your Own**: select another numerical attribute. How well the Normal Distribution fits the Data?
   Is the Normal model appropriate? If not, explain, why.
   You have to find two variables: one of them shows the Normal distribution, and another doesn't.
   Repeat **steps 3 – 7** for both variables and give all explanations.

10. Prepare and print your lab report. Manage Fathom's screen to make your tables and graphs clear and understandable. Paste all graphs into your Word document. To paste a Fathom object to a World document, make an object active by clicking on it, select Edit – Copy As Picture. Write brief comments for all action taken, describe all graphs and make all conclusions.

10 extra points will be given for the report's excellence.

Use the Textbook: Chapter 6, Chapter 17.

* This applet is provided by the Virtual Laboratory in Probability and Statistics, University of Alabama, Huntsville: *http://www.math.uah.edu/stat/objects/distributions/BinomialDistribution.xhtml*