Illustrating Normal Distributions to Students in an Introductory Business Statistics Course via Visual Basic Macros in a Spreadsheet Application

Robert O. Nedigh, Shippensburg University

ABSTRACT: *This teaching application is used in an introductory Business Statistics class. Normal distributions are illustrated to the class using an Excel spreadsheet application.*

OVERVIEW

An understanding of sampling distributions is essential to understanding of hypothesis tests of a mean. Also, an understanding of normal distributions is needed to understand sampling distributions. Unfortunately, many students never grasp normal distributions and as a result, do not understand sampling distributions. Thus, many undergraduate students complete their introductory business statistics course frustrated and confused because they cannot fully understand the logic to hypothesis testing. In this paper, we will introduce an excel spreadsheet model that has been found to be successful in numerically and graphically illustrating the concepts of normal distributions.

BACKGROUND

In a hypothesis test of a mean, the hypothesis is either accepted or rejected by determining whether the sample mean is in the region of non-rejection or rejection (critical value method) or by comparing the p-value to the level of significance (p-value method). In both cases, the student needs to be able to use knowledge of sampling distributions to either determine the critical point(s) or calculate the p-value. Software is available to determine the p-value that allows our students to make the statistical decisions by comparing the p-value to the level of significance. Forcing students to calculate critical point(s) and/or p-values allows for enhanced learning and a better understanding of the theoretical underpinnings. When students grasp the logic of sampling distributions this enables them to better perform the calculations of critical point(s) and/or p-values and then perhaps more importantly understand their interpretation.

Probably the biggest impediment to students' comprehension of the critical values and p-values is the lack of knowledge of the normal distribution. Most students learn to use the formula to calculate a Z-value and then use the Standard Normal Table to determine a probability, but the majority of the students do not grasp the meaning of the Z-value and probability. Students struggle with equating an area underneath a curve to a probability but even more disturbing do not understand where the normal shape comes from.

In order to facilitate the comprehension of these concepts, I use an Excel spreadsheet model to select random numbers from a normal distribution. The model has proven effective in illustrating the properties of normal distributions in my business classes. In the next section, we will describe the workings of the spreadsheet model.

DESCRIPTION OF THE SPREADSHEET MODEL

The spreadsheet model contains two worksheets that perform different functions. Each worksheet has command buttons that are 'clicked' to perform actions. The command buttons run macros written in Visual Basic for Applications (VBA). The two worksheets are named GenerateNormalData and Histogram.

GenerateNormalData Worksheet

This worksheet is used to generate a sample of random numbers from a normal distribution with the mean, standard deviation, and sample size selected by the user. The worksheet is shown in Figure I.

The desired mean is entered in cell "B2" and the desired standard deviation is entered in cell "C2". Random numbers are then generated using Excel's "*RAND*" function in Cell "A1" and "*NORMINV*" function in Cell "A2". The formula in Cell "A1" is '=RAND()' and in Cell "A2" is '=NORMINV(A1,B2,C2)'. The *RAND* function generates random numbers uniformly distributed over the interval [0,1). The *NORMINV* function returns a z-value associated with the cumulative probability generated by the *RAND* function. The result is a value that is randomly generated from a normal distribution with the mean and standard deviation in Cells "B2" and "C2" that is z standard deviations away from the mean.

Using the above procedure, the user inputs into Cell "E2" how many numbers are to be generated from the normal distribution and then clicks the "Run" command button. The random numbers generated are placed in Column D starting in Cell "D1" of the GenerateNormalData worksheet and in Column A starting in Cell "A1" of the Histogram worksheet. This enables the students to view the distribution of numbers generated from the normal distribution.

Histogram Worksheet

This worksheet is used to display a histogram of the random numbers in Column A of this worksheet. The worksheet is shown in Figure II. The numbers in Column A were generated from the normal distribution in the GenerateNormalData worksheet. When the user clicks the "Generate Histogram" command button, bin ranges are automatically determined and a histogram chart is displayed in the worksheet. This histogram enables the students to see how closely the data in Column A follow a normal shape.

A user can also calculate what proportion of the values in Column A are between two values input by the user. The lower value is input into Cell "N2" and the upper value is input into Cell "N3". When the user clicks the "Probabilities" command button, the proportion is calculated and entered in Cell "N5". This enables students to compare the observed proportion to the theoretical proportion of the numbers being between the two values.

Figure I: GenerateNormalData Worksheet

	A	B	C	D	E	F	G
1	0.89021983	Mean	Standard Deviation	995.3716491	Trials		Average
2	1006.13849	1000	5	1006.60106	30		998.3386
3				993.7616788			
4				1008.914467	Samples		
5		Run		1002.588302	100		
6				989.5654232			

Figure II: Histogram Worksheet

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	995.3716	978.75		Mean	Standard Deviation									
2	1006.601	981.25		1000	5				Probabilities				Min	0
3	993.7617	983.75											Max	100
4	1008.914	986.25		Trials										
5	1002.588	988.75		30									Probability =	0.000%
6	989.5654	991.25						When the 'Probabilities' button is clicked, the percent of numbers in Column A that						
7	1000.579	993.75						are between the two values in Cells N2 and N3 is put in Cell N5.						
8	998.0917	996.25												
9	1011.903	998.75												
10	1001.619	1001.25							Generate Histogram					
11	997.8014	1003.75												
12	994.9016	1006.25												

APPLICATION OF THE SPREADSHEET MODEL

In teaching, one can use the GenerateNormalData worksheet to show how numbers chosen from a specific normal distribution are grouped above and below the specified mean. The instructor can change quickly from one normal distribution to another. Many students are unable to visualize how the magnitude of the standard deviation affects the grouping of the distribution. In my class presentation, one can choose a mean and then vary the magnitude of the standard deviation to demonstrate how the numbers range farther from the mean as the standard deviation increases. This in-class demonstration is a valuable tool to enhance the students' understanding of standard deviation.

After using the GenerateNormalData worksheet to generate numbers I then switch to the Histogram worksheet to demonstrate the normal shape. The "Generate Histogram" button displays a histogram of the numbers in Column A which were generated in the GenerateNormalData worksheet. The histogram enables the students to see how closely the numbers follow a normal shape. The students see the continuous normal curves in the text without truly understanding how they relate to the makeup of a group of numbers. The histograms with defined groupings better demonstrate the distribution groupings of the numbers. The instructor might start this demonstration with a fairly small sample size because the histogram usually varies from the normal shape. Then increase the sample size to demonstrate how much more closely the samples follow the normal shape with larger sample sizes. One can also use this exercise to show how larger samples tend to much more closely mimic the population which starts to set up one of the properties of sampling.

The instructor could next use the "Probabilities" button to calculate the proportion of the numbers in the sample that lie between two specified points. The students use the z-value calculations and standard normal table to determine the theoretical proportion. They can then compare the two proportions. Once again, one might start out with a small sample and progress to much larger samples to show how closely the proportions in large samples compare to the theoretical proportion. This in-class exercise shows the students how the calculated proportions are demonstrated in actual samples, which allows the students to see how they are realistic.

CONCLUSION

In this paper, we have described a spreadsheet application that enables an instructor to illustrate to students numerically and graphically the properties of a normal data set. The application has demonstrated in practice an enhanced understanding by students of the properties of normal distributions.

APPENDIX – VISUAL BASIC CODE

Visual Basic Code for "Run" Button in GenerateNormalData Worksheet

```
Private Sub CommandButton1_Click()
Dim Ronny As Variant
Dim Alyssa As Integer
Range("Runs").ClearContents
Range("Samples").ClearContents
Worksheets("Histogram").Range("Data").ClearContents
Cells(2, "G").Clear
Alyssa = 1
Do While Alyssa <= Cells(2, "E").Value
   Calculate
   Ronny = Cells(2, "A").Value
   Cells(Alyssa, "D").Value = Ronny
   Worksheets("Histogram").Cells(Alyssa, "A").Value = Ronny
   Alyssa = Alyssa + 1
Loop
Cells(2, "G").Value = "=Average(Runs)"
Worksheets("Histogram").Cells(2, "D").Value = Cells(2, "B").Value
Worksheets("Histogram").Cells(2, "E").Value = Cells(2, "C").Value
Worksheets("Histogram").Cells(5, "D").Value = Cells(2, "E").Value
End Sub
```

Visual Basic Code for "Generate Histogram" Button in Histogram Worksheet

```
Private Sub CommandButton2_Click()
Dim Counter As Integer
Counter = Cells(5, "D").Value
Range("HistogramData").ClearContents
ActiveSheet.ChartObjects("MyChart").Delete
Cells(1, "B").Value = Cells(2, "D").Value - (Cells(2, "E").Value / 4) - 8 * (Cells(2, "E").Value / 2)
Cells(2, "B").Value = Cells(2, "D").Value - (Cells(2, "E").Value / 4) - 7 * (Cells(2, "E").Value / 2)
Cells(3, "B").Value = Cells(2, "D").Value - (Cells(2, "E").Value / 4) - 6 * (Cells(2, "E").Value / 2)
Cells(4, "B").Value = Cells(2, "D").Value - (Cells(2, "E").Value / 4) - 5 * (Cells(2, "E").Value / 2)
Cells(5, "B").Value = Cells(2, "D").Value - (Cells(2, "E").Value / 4) - 4 * (Cells(2, "E").Value / 2)
Cells(6, "B").Value = Cells(2, "D").Value - (Cells(2, "E").Value / 4) - 3 * (Cells(2, "E").Value / 2)
Cells(7, "B").Value = Cells(2, "D").Value - (Cells(2, "E").Value / 4) - 2 * (Cells(2, "E").Value / 2)
Cells(8, "B").Value = Cells(2, "D").Value - (Cells(2, "E").Value / 4) - 1 * (Cells(2, "E").Value / 2)
Cells(9, "B").Value = Cells(2, "D").Value - (Cells(2, "E").Value / 4)
Cells(10, "B").Value = Cells(2, "D").Value + (Cells(2, "E").Value / 4)
Cells(11, "B").Value = Cells(2, "D").Value + (Cells(2, "E").Value / 4) + 1 * (Cells(2, "E").Value / 2)
Cells(12, "B").Value = Cells(2, "D").Value + (Cells(2, "E").Value / 4) + 2 * (Cells(2, "E").Value / 2)
Cells(13, "B").Value = Cells(2, "D").Value + (Cells(2, "E").Value / 4) + 3 * (Cells(2, "E").Value / 2)
Cells(14, "B").Value = Cells(2, "D").Value + (Cells(2, "E").Value / 4) + 4 * (Cells(2, "E").Value / 2)
Cells(15, "B").Value = Cells(2, "D").Value + (Cells(2, "E").Value / 4) + 5 * (Cells(2, "E").Value / 2)
Cells(16, "B").Value = Cells(2, "D").Value + (Cells(2, "E").Value / 4) + 6 * (Cells(2, "E").Value / 2)
```

```
Cells(17, "B").Value = Cells(2, "D").Value + (Cells(2, "E").Value / 4) + 7 * (Cells(2, "E").Value / 2)
Cells(18, "B").Value = Cells(2, "D").Value + (Cells(2, "E").Value / 4) + 8 * (Cells(2, "E").Value / 2)
Application.Run "ATPVBAEN.XLA!Histogram", ActiveSheet.Range("Data") _
    , ActiveSheet.Range("$D$18"), ActiveSheet.Range("$B$1:$B$18"), False, False _
    , True, False
  ActiveChart.Parent.Name = "MyChart"
  ActiveSheet.ChartObjects("MyChart").Activate
  ActiveChart.ChartArea.Select
  ActiveSheet.Shapes("MyChart").ScaleHeight 1.38, msoFalse, msoScaleFromTopLeft
  Windows("Normal.xls").SmallScroll ToRight:=6
  Windows("Normal.xls").SmallScroll Down:=16
  ActiveSheet.Shapes("MyChart").ScaleWidth 2.72, msoFalse, msoScaleFromTopLeft
  ActiveSheet.Shapes("MyChart").ScaleHeight 2.11, msoFalse, msoScaleFromTopLeft
  ActiveChart.Legend.Select
  Selection.Delete
  ActiveChart.SeriesCollection(1).Select
  With ActiveChart.ChartGroups(1)
    .Overlap = 0
    .GapWidth = 0
    .HasSeriesLines = False
    .VaryByCategories = False
  End With
  ActiveChart.Axes(xlCategory).Select
  Selection.TickLabels.AutoScaleFont = True
  With Selection.TickLabels.Font
    .Name = "Arial"
    .FontStyle = "Regular"
    .Size = 14
    .Strikethrough = False
    .Superscript = False
    .Subscript = False
    .OutlineFont = False
    .Shadow = False
    .Underline = xlUnderlineStyleNone
    .ColorIndex = xlAutomatic
    .Background = xlAutomatic
  End With
  ActiveChart.Axes(xlValue).Select
  Selection.TickLabels.AutoScaleFont = True
  With Selection.TickLabels.Font
    .Name = "Arial"
    .FontStyle = "Regular"
    .Size = 14
    .Strikethrough = False
    .Superscript = False
    .Subscript = False
    .OutlineFont = False
    .Shadow = False
    .Underline = xlUnderlineStyleNone
```

```
   .ColorIndex = xlAutomatic
   .Background = xlAutomatic
  End With
  ActiveWindow.Visible = False
  Windows("Normal.xls").Activate
  Range("A1").Select
  ActiveWindow.SmallScroll ToRight:=6
  ActiveWindow.SmallScroll Down:=17
End Sub
```

Visual Basic Code for "Probabilities" Button in Histogram Worksheet

```
Private Sub CommandButton1_Click()
Dim Min As Double
Dim Max As Double
Dim Counter As Integer
Dim Ronny As Double
Dim Probability As Double
Counter = 1
Ronny = 0
Min = Cells(2, "N").Value
Max = Cells(3, "N").Value
Do Until Cells(Counter, "A").Value = ""
  If Cells(Counter, "A").Value > Min And Cells(Counter, "A").Value < Max Then
    Ronny = Ronny + 1
  Else
  End If
  Cells(5, "N").Value = Ronny / Counter
  Counter = Counter + 1
Loop
End Sub
```

TEACHING NOTE

I use this application after I have already taught the normal distribution. The students have already learned the properties of the normal distribution, how to calculate Z-values, and how to use the standard normal table to determine probabilities of being within ranges. At this point, before progressing to sampling distributions, I use this spreadsheet model to reinforce the properties of the normal distribution to the students.

The application enables the teacher to:
- Generate and display numbers from a specified normal distribution
- Display a histogram graph of the generated numbers
- Determine the proportion of numbers in the set between two values

After demonstrating the application to the students, I usually generate samples for some text problems that we have already calculated and compare the theoretical probabilities to the observed proportions.