CONTINOUS RANDOM VARIABLES
EXCEL LAB #4

BUSN/ECON/FIN 130: Applied Statistics
Department of Economics and Business
Lake Forest College
Lake Forest, IL 60045

Overview

This lab is written for Excel 2010, which is available to students in the library. The notation => can be read as "go to" or "click on." This notation will most often be used when navigating the menu or toolbars in Excel. To indicate a command or icon that you might click on or search for in Excel, **bold** will be used. Likewise, anything that you are to type into Excel will be **bolded** in the instructions. Please do not enter such text as bolded text unless the instructions ask you to do so.

Tutorial

1. Open Excel. Rename worksheet **Sheet 1** to **Uniform Distribution**. Rename worksheet **Sheet 2** to **Normal Distribution**. Rename worksheet **Sheet 3** to **T Distribution**.

2. For all three worksheets, for center justify all cells in columns A through T and set the column width equal to 10, the font to Arial, and the font size to 9 point, and center-justify columns A through T.

3. Go to the **Uniform Distribution** worksheet. In cell A1 enter **Unif(0,1)**. In cell B1 enter **Unif(2,5)**. In cell C1 enter **Unif(8,14)**.

4. Set the formatting of cells A2:C47 to report 4 decimal places.

5. In cell A2 enter **=RAND()**. For Excel, the formula **RAND ()** is a random number generator from zero to one. That is, it returns a random draw on the Unif(0,1) distribution. One curious feature of the **RAND ()** function is that whenever anything in the worksheet changes, a new random draw will be made. For example, click on cell C5 and then press the **Delete** key on your keyboard several times while watching cell A2. When you pressed **Delete** you should notice that the value in cell A2 changed.

6. Copy and paste cell A2 to cells A3 through A41.

7. In cell A43 enter **=AVERAGE(A2:A41)**. In cell A44 enter **=VAR(A2:A41)**.

8. In cell A46 enter **0.5**. In cell A47 enter **0.0833**.

Notice, cells A43 and A44 calculate the average and variance of the 40 random draws on the Unif(0,1) listed above it. Cells A46 and A47 report the population mean and the population variance – the population mean of a Unif[0,1] is clearly ½, and recall that the population variance of a Unif(0,1) is 1/12 (or 0.0833).

Click on cell E43 again, and press the **Delete** key on the keyboard repeatedly, this time watching cells A43 and A44, comparing them to cells A46 and A47. You should notice that the sample average and sample variance jump around quite a bit, but they are in a vicinity around the population mean and population variance. This demonstrates the idea of random sampling.

In column B we want to list 40 draws from a Unif(2,5) random variable. Unfortunately Excel does not have a command for this directly. Rather, we will use the random probabilities in column A to generate random draws in column B. As column A is a Unif(0,1), to translate any random draw from a Unif(0,1)into a Unif(2,5), we must add 2 to all of the values in column A (to go from a minimum of 0 to a new minimum of 2). Moreover, as the range of draws in column A is 1 (from 0 to 1) and we want the range of draws in column B to be 3 (from 2 to 5), we need to "stretch" the probabilities in column A from 1 to 3 by multiplying the probability (in column A) by 3. Thus:

9. In cell B2 enter **=2+3*A2**. Copy cell B2, and paste it to cells B3 through B41.

10. In cell B43 enter **=AVERAGE(B2:B41)**, and in cell B44 enter **=VAR(B2:B41)**.

11. For a Unif(2,5) random variable, the population mean is the midpoint, or (2+5)/2 = 3.5, and the population variance is the length of the interval squared divided by 12, or $(5 - 2)^2 \div 12 = 9/12 = 0.75$. In cell B46 enter **3.5**, and in cell B47 enter **0.75**.

Once again cells B43 and B44 calculate the average and variance of the 40 random draws on the Unif(2,5) listed above it. Cells B46 and B47 report the population mean and population variance of the random variable. Once again click on cell E43, and press the **Delete** key on the keyboard repeatedly, this time watching cells B43 and B44, comparing them to cells B46 and B47. You should notice again that the sample average and sample variance jump around quite a bit, but they are in a vicinity around the population mean and population variance.

In column C we want to list 40 draws from a Unif(8,14) random variable. Again we will use the probabilities in column A to do this. As column A is a Unif(0,1), to translate any random draw into a Unif(8,14), we must add 8 to all of the values in column A (to go from a minimum of 0 to a new minimum of 8). Moreover, as the range of draws in column A is 1 (from 0 to 1) and we want the range of draws in column C to be 6 (from 8 to 14), we need to "stretch" the probabilities in column A from 1 to 6 by multiplying the probability (in column A) by 6. Thus:

12. In cell C2 enter **=8+6*A2**. Copy cell C2, and paste it into cells C3 through C41.

13. In cell C43 enter **=AVERAGE(C2:C41)**, and in cell C44 enter **=VAR(C2:C41)**.

14. For a Unif(8,14) random variable, the population mean is the midpoint, or (8+14)/2 = 11, and the population variance is the length of the interval squared divided by 12, or $(14 - 8)^2 \div 12 = 36/12 = 3$. In cell C46 enter **11**, and in cell C47 enter **3**.

Once again cells C43 and C44 calculate the average and variance of the 40 random draws on the Unif(8,14) listed above it. Cells C46 and C47 report the population mean and population variance of the random variable. Click on cell E43, and press the **Delete** key on the keyboard repeatedly while watching cells C43 and C44, comparing them to cells C46 and C47. You should notice again that the sample average and sample variance jump around quite a bit, but they are in a vicinity around the population mean and population variance.

15. Go to the **Normal Distribution** worksheet. In cell A1 enter **N(0,1)** in bold.

16. Set the width of columns C, F, and G to 3.

17. In cells A2, B2, D2, and E2 respectively, enter **z-values**, **F(z)**, **probabilities**, **z-values** (enter all underlined, but not in bold).

18. In cell A3 enter **-2.25**. In cell A4 enter **=A3+0.25**. Copy and paste cell A4 to cells A5 through A21. Cell A21 should have a value of **2.25**.

19. Make columns A and D report values to 2 decimal places; column B report probabilities to 4 decimal places; and column E report to 3 decimal places.

20. In column B we want Excel to report the probability a single draw on a standard normal distribution will return a value of the *z*-value or less. That is, we want Excel to report the CDF of a N(0,1) random variable. To do this, in cell B3 enter **=NORMSDIST(A3)**.

    Notice the **S** in the formula. It is this **S** that tells us (and Excel) that we are dealing with a N(0,1), and the **S** stands for **Standard**. The **NORMSDIST()** function in Excel returns the CDF for the N(0,1) for whatever value is placed in parentheses. For example, the probability that a N(0,1) returns a value less than -2.25 is only 1.22%, which is why cell B3 reports 0.0122.

    Copy and paste cell B3 into cells B4 through B21. Many of the numbers in column B should look familiar. We know that the N(0,1) is symmetrically distributed about 0, so half of the probability must fall below 0, which is why cell B12 reports 0.5000. We also know that there is about 2.5% probability in each tail outside of +/-2, which is why cell B4 reports 0.0228 (and cell B20 reports 0.9772).

21. In columns D and E we are going to essentially reverse the previous process. In cell D3 enter **0.05**. In cell D4 enter **=D3+0.05**. Copy and paste cell D4 into cells D5 through D21. Cell D21 should have a value of **0.95**.

22. In column E we want Excel to report the z-value that is associated with each of these probabilities. That is, in cell E3 we want Excel to report the number (z-value) for which 5% of all draws from a N(0,1) are beneath. To do this, in cell E3 enter **=NORMSINV(D3)**.

    Once again, the **S** tells us and Excel that we are dealing with a **standard** normal distribution. Notice too that instead of **NORMSDIST** the function is **NORMSINV**, the **INV** meaning to report a z-value rather than a probability. In particular, cell E3 reports -1.645 as 5% of all draws from a N(0,1) are below -1.645.

    Copy and paste cell E3 into cells E4 through E21. Many of the numbers in column E should look familiar. The N(0,1) is symmetrically distributed about 0, so half of the probability must fall below 0, which is why cell E12 reports 0.0000 as the z-value for which half of the probability falls beneath it is zero. In particular, the critical values of +/- 1.645 associated with 5% and 95% of the probability and the critical values of +/- 1.282 associated with 10% and 90% of the probability have been discussed in class.

23. In addition to using the standard normal distribution, Excel allows us to determine probabilities and z-values for any normal distribution. In cell A24 enter **N(5,100)**, which is a normal distribution with mean 5 and variance 100 (standard deviation 10).

24. In cells A25, B25, D25, and E25 respectively enter <u>**z-values**</u>, <u>**F(z)**</u>, <u>**probabilities**</u>, <u>**z-values**</u> (enter all underlined, but not in bold).

25. In cell A26 enter **-12.00**. In cell A27 enter **=A26+1.5**. Copy and paste cell A27 to cells A28 through A44. Cell A44 should have a value of **15.00**.

26. In column B we want Excel to report the probability a single draw on a N(5,100) will return a value of the *z*-value or less. That is, we want Excel to report the CDF of a N(5,100) random variable. To do this, in cell B26 enter **=NORMDIST(A26,5,10,TRUE)**.

    1. Notice that there is no **S** in the formula. Whenever, you are not dealing with a standard normal, you don't include the **S**, but you now need to tell Excel the mean and standard deviation of the distribution, which is the role of the **,5,10** in the formula. Notice that the N(5, 100) notation indicates that the variance of the random variable is 100. When using the **NORMDIST** function in Excel, however, you need to enter the standard deviation, which is the square root of the variance. This is why we enter **10** into the function rather than **100**.

    2. Excel also needs to know if you want the PDF or the CDF. (There will never be a need for us to know the PDF.) As we want the CDF, the last parameter in the formula is set to **TRUE**.

    3. For a N(5,100), we see in cell B26 that the probability of returning a value less than -12 is 4.46%.

    4. Copy and paste cell B26 into cells B27 through B44.

    The numbers in cells B26 through B44 should not look familiar as we are not familiar with a N(5,100) distribution. However, they should make sense. For one, the CDF probabilities increase as the z-values increase. Moreover, as the standard deviation is 10, and we know that roughly 16% of the distribution is below one standard below the mean and 16% of the distribution is above one standard deviation above the mean, the probabilities of 0.1711 in cell B31 and 0.8413 in cell B44 are believable.

27. In columns D and E we are going to essentially reverse the previous process. In cell D26 enter **0.05**. In cell D27 enter **=D26+0.05**. Copy and paste cell D27 into cells D28 through D44.

28. In column E we want Excel to report the z-value that is associated with each of these probabilities. That is, in cell E26 we want Excel to report the number (z-value) for which 5% of all draws from a N(5,100) are beneath. To do this, in cell E26 enter **=NORMINV(D26,5,10)**.

    1. Notice that there is no **S** in the formula. Whenever, you are not dealing with a standard normal, you don't include the **S**, but you now need to tell Excel the mean and standard deviation of the distribution, which is the role of the **,5,10** in the formula.

2. Once again, the **INV** means to report a z value rather than a probability. In particular, cell E26 reports -11.449 as 5% of all draws from a N(5,100) are below -11.449.

3. Copy and paste cell E26 into cells E27 through E44.

Lastly we are going to create a part of the N(0,1) table.

29. Set the number format to report just 1 decimal place for column G.

30. In cell G2 enter **0.0**. In cell G3 enter **=G2+0.1**. Copy and paste cell G3 into cells G4 through G32. Cells G2 through G32 should count from 0.0 to 3.0 by 0.1's. Bold cells G2 through G32.

31. Set the number format of row 1 to report to 2 decimal places: **right click on row 1** => **Format Cells** => **Number** (tab) => **Number** (category) => **2** => **OK**.

32. In cell H1 enter **0.00**. In cell I1 enter **=H1+0.01**. Copy and paste cell I1 into cells J1 through Q1. Cells H1 through Q1 should count from 0.00 to 0.09 by 0.01's. Bold cells H1 through Q1.

33. In cell H2 we want to enter **=NORMSDIST(G2+H1)** as we want Excel to return z-values (which it does with the **NORMSDIST** function, whereas it returns probabilities with the **NORMSINV** function). However, the problem with the above formula is that it cannot be copied and pasted correctly because of Excel's rolling the cells and rows. So, instead, in cell H2 enter **=NORMSDIST($G2+H$1)**. The **$G2** fixes the G column, but allows the row to scroll as we paste down. At the same time, the **H$1** fixes the row at 1 by allows the column to scroll as we paste over. So now copy and paste cell H2 into cells H2 through Q32.

34. Make cells H2 through Q32 report out to 4 decimals.

35. You should now have part of the N(0,1) table. For your own benefit, highlight the probabilities that correspond to 10%, 5%, 2.5%, 1%, and 0.5% of the probability being above a particular z-value (the corresponding z-values are 1.28, 1.645, 1.96, 2.33, and 2.575). Thus, you are bolding cells P14 (10%), L18 and M18 (5%), N21 (2.5%), K25 (1%), and O27 and P27 (0.5%).

Go to the **T Distribution** worksheet. The *t*-distribution (or student *t*-distribution) is a continuous distribution that involves degrees of freedom (a parameter) that approximates the standard normal distribution as the degrees of freedom approach infinity.

36. Set the width of column C to 3.

37. Merge cells A1 through E1, and enter **Degrees of Freedom = 10** (hint: to merge the cells, **left click on cell A1 and drag through cell E1** => **right click on the shaded area** => **Format Cells** => **Alignment** (tab) => click on **Merge Cells** => **OK**).

38. In cells A2, B2, D2, and E2 enter respectively **<u>t-values</u>**, **<u>alpha</u>**, **<u>probability</u>**, **<u>crit value</u>** (enter each with underlines, but not in bold).

39. Make the number format of all cells in columns A and D report to 2 decimal places; column B to 4 decimal places; and column E to 3 decimal places.

40. In cell A3 enter **0.00**. In cell A4 enter **=A3+0.15**. Copy and paste cell A4 into cells A5 through A27. Cells A3 through A27 should contain the numbers 0 to 3.6, counting by 0.15's.

41. Excel has a particular function that reports probabilities off of t-values (or critical values) called **TDIST()**. In particular, **TDIST()** reports the probability **above** a t-value, either as a two-sided probability or as a one-sided probability. As our class focuses on two-sided probabilities, we will elect to have Excel report two-sided probabilities. Therefore, in cell B3 enter **=TDIST(A3,10,2)**. In this formula, the **10** is the degrees of freedom, and the **2** tells Excel to report a two-sided probability.

42. Copy and paste cell B3 into cells B4 through B27. To check your work, the probability that a *t*-distribution with 10 degrees of freedom will return a number greater than 2.25 *in absolute value* is 4.82%.

43. In cell D3 enter **0.01**. In cell D4 enter **=D3+0.01**. Copy and paste cell D4 into cells D5 through D27. Cells D3 through D27 should contain the numbers 0.01 to 0.25, counting by 0.01's.

44. Excel has a particular function that reports critical values from probabilities called **TINV()**. In particular, **TINV()** reports the **two-sided** probability **above** a t-value. In cell E3 enter **=TINV(D3,10)**. In this formula, the **10** is the degrees of freedom.

45. Copy and paste cell E3 into cells E4 through E27. Therefore, the critical value to have 12% above the critical value in absolute value is 1.700 (cell E14).

At this point, part of the *t*-distribution results should look like the following:

Degrees of Freedom = 10

t-values	alpha	probability	crit value
0.00	1.0000	0.01	3.169
0.15	0.8837	0.02	2.764
0.30	0.7703	0.03	2.527
0.45	0.6623	0.04	2.359
0.60	0.5619	0.05	2.228
0.75	0.4705	0.06	2.120
0.90	0.3893	0.07	2.028
1.05	0.3184	0.08	1.948

The last thing we are going to use Excel's **TINV** function for is to demonstrate that *t*-distributions converge to the standard normal as the degrees of freedom approach infinity.

46. In cells H1, I1, and J1 respectively, enter **prob=1%**, **prob=5%**, and **prob=10%**. In cells G2, H2, I2, and J2 respectively, enter **dof**, **crit value**, **crit value**, **crit value** (enter each of these with the underline, but not the bold).

47. Right-justify columns G through J.

48. In cells G3 through G13 respectively, enter **10**, **20**, **50**, **100**, **200**, **500**, **1000**, **2500**, **5000**, **10000**, and **infinity**.

49. In cell H3 enter **=TINV(0.01,G3)**. The reference to cell G3 tells Excel the degrees of freedom. In cell I3 enter **=TINV(0.05,G3)**. In cell J3 enter **=TINV(0.1,G3)**. Copy and paste cells H3 through J3 to cells H4 through J12.

50. In cell H13 enter **=NORMSINV(0.995)**. Although **TINV** returns a two-sided probability, **NORMSINV** does not, and in fact it returns the CDF. So instead of giving it a probability in the upper tail, we must give it the entire probability below the critical value. As we want 1% in both tails combined, we want 0.5% above the critical value, which means we want 99.5% of the probability below the critical value. This accounts for the **0.995** in the function. We know that the two-sided critical value at the 1% significance level is 2.576 (we actually use 2.575, but it is indeed 2.576), and this is what appears in cell H13. In cell I13 enter **=NORMSINV(0.975)** and in cell J13 enter **=NORMSINV(0.95)**.

51. Lastly, make cells H3 through I13 report numbers out to 3 decimal places.

Notice how the critical values implied by the *t*-distribution approach the critical values of the standard normal distribution as the degrees of freedom get large.

Exercises

1. Insert a new worksheet in your Excel file called **T-Distribution Exercise**. Place it to the right of **T Distribution**.

2. In this worksheet, create a table that reports the two-sided critical values for the standard probabilities (10%, 5%, 2, and 1%) for degrees of freedom 1 through 30 and then 50, 100, 200, 500, 1000, and infinity.
    1. Put a title to this table in cell A1 that is centered, bolded, and merged with cells B1 through E1.
    2. Merge together cells A2 through E2, and put your name in the merged cells.
    3. Write **<u>DOF</u>** in cell A4, and then list the degrees of freedom (1 through 30 plus 50, 100, 200, 500, 1000, and infinity) in cells A5 through A40.
    4. Write **<u>10%</u>**, **<u>5%</u>**, **<u>2%</u>**, and **<u>1%</u>** respectively in cells B4 through E4.
    5. Enter the critical values in cells B4 through E41 to 3 decimal places using the **TINV** command.

3. Save your work (all four worksheets) as YourName_Lab4.xlsx.

Turning in your work

Email YourName_Lab4.xlsx to your professor as a file attachment to an email with the subject heading Excel Lab 4: Your Name. Also print and turn in the *t*-distribution table from the Exercise portion of the lab.