Question # 1. The Normal Distribution

Suppose $R \sim N(.1, .0016)$. (HINT: Use functions in Excel. Which one did you use?)

(a) What is $P(R < 0)$?

(b) What is $P(0 < R < 0.1)$?

(c) What is $P(R > 0.15)$?
Question # 2. The Normal Distribution

The problem is intended to learn more about the normal distribution.

(a) In Excel, evaluate the following:

1. =NORM.INV(.5,0,1)
2. =NORM.INV(.025,0,1)
3. =NORM.INV(.05,0,1)
4. =NORM.INV(0.16,0,1)
5. =NORM.INV(.84,0,1)
6. =NORM.INV(0.95,0,1)

(b) What does the function NORM.INV do? How does it relate to the standard normal c.d.f.?

Note: Excel also has a function NORM.S.INV which is only the standard normal.

(c) Without using Excel, what would NORM.INV(0.975,0,1) be?
Question # 3. The Law of Large Numbers

The goal of this question is to learn what the law of large numbers is. For this question, we will use the random number generators in Excel. There are two simple methods for doing this and you may use either option.

Option #1: An easy way to generate random numbers (from uniform, normal, binomial, Poisson distributions) is download the “Data Analysis” tool for Excel. There is a tutorial on the course webpage for PC users.

Option #2: Random numbers can also be generated from a distribution \( F(x) \) as long as we can draw uniform random numbers and invert the cumulative distribution function (CDF) of the distribution. (In fact, this is actually how Excel is doing it.) To draw from a distribution with CDF given by \( y = F(x) \) and inverse given by \( x = F^{-1}(y) \), all you do is:

1. Draw a uniform random number on the interval \([0, 1]\) given by \( y \sim U(0,1) \). To generate a uniform random number in Excel, you would type “=RAND()”.

2. Then, given the value of \( y \) from step 1, invert the CDF \( x = F^{-1}(y) \). The variable \( x \) is a random draw from the distribution \( F(x) \). For example, to draw from a normal distribution with mean \( \mu \) and standard deviation \( \sigma \) in Excel, you would type “=NORM.INV(y,\( \mu \),\( \sigma \))” where \( y \) is the uniform r.v. that was returned by “=RAND()”.

Questions:

(a) Draw \( n \) random numbers from the normal distribution \( x_i \sim N(\mu,\sigma^2) \). You may choose the values of \( \mu \) and \( \sigma \). Calculate the sample mean \( \bar{x} \) and sample variance \( s_x^2 \) for \( n = 100, 200, 500 \), and 1000 draws. Are the sample mean and variance getting closer to the true values of \( \mu \) and \( \sigma^2 \) as \( n \) gets larger?

(Do not hand in all the random numbers. Just hand in the sample means and variances with a description.)

---

1 Excel has functions to compute the inverse CDF for other distributions including: binomial (BINOM.INV), Chi-square (CHISQ.INV), and Student’s \( t \) (T.INV).
(b) Draw $n$ random numbers from the binomial distribution $x_i \sim \text{binomial}(20, 0.75)$. Calculate the sample mean $\bar{x}$ and sample variance $s^2_x$ for $n = 100, 200, 500, \text{ and } 1000$ draws. What values are $\bar{x}$ and $s^2_x$ converging to as $n$ becomes larger?

(Do not hand in all the random numbers. Just hand in the sample means and variances with a description.)
Question # 4. The Central Limit Theorem

The goal of this question is to learn what a central limit theorem is. For each part, you may hand in a histogram and a brief description of what you learned.

(a) The solution to this question proceeds in 3 steps:

1. In column A of Excel, draw \( n = 200 \) random numbers \( x_i \sim \text{binomial}(40, 0.35) \). In another excel spreadsheet, calculate the sample mean \( \bar{x} \) of these values. For example, my random numbers are in column A of Sheet1. In cell A1 on sheet 2, I typed: “=AVERAGE(Sheet1!A1:Sheet1!A200)” to calculate the sample mean of the 200 values.

2. Repeat the first step in columns B, C, D, E, etc. a large number of times, e.g. 500 columns. In other words, you are generating 500 data sets (one data set per column) each of size \( n = 200 \). And, for each of these 500 data sets, you are calculating the sample mean \( \bar{x} \).

3. Draw a histogram of the 500 values of \( \bar{x} \). What does the histogram look like? What is the variance of this distribution? What happens if \( n \) becomes large?

(b) Repeat question (a) above but draw the random variables from a uniform distribution \( x_i \sim \text{U}(0, 1) \).

(c) Repeat question (a) above but draw the random variables from another distribution of your choice other than the uniform or binomial distribution.

\(^2\text{HINT 1: If you are using the data analysis package to generate random numbers, you can draw all 500 datasets at once by selecting: “Number of Variables = 500”}

\(^2\text{HINT 2: If you are not using the data analysis package to generate random numbers, you can copy and drag the “=RAND()” function across all 500 columns and 200 rows. This will generate all the uniform variables that you need. Then, in the next excel spreadsheet (sheet2), you use the “=BINOM.INV” function and the uniform variables from the first spreadsheet (sheet1) as inputs. This will generate the binomial r.v.’s quickly.}
Question # 5. Confidence Intervals for voting

It is election night and you are working for a TV news service in Florida. You are covering a race between two candidates, George and Al. Assume that all voters in the state of Florida vote for one of these two candidates and let $p$ be equal to the proportion of people who vote for Al. Votes are now being counted, and the network executives want you to make a projection for which candidate will win the state. Though of course they want this projection as soon as possible, they are concerned with the network’s credibility and are willing to accept at most a 5% chance that the projection is wrong.

(a) Suppose that 500 votes from the state have been counted. The results are 263 votes for Al (the other 237 voted for George). Construct a 95% confidence interval for $p$. Can you project a winner in the election?

(b) Now suppose that 10,000 votes have been counted and 5,150 of them were votes for Al. Can you project a winner in the election?

(c) Now suppose you learn that the 10,000 voters in part (b) came from parts of the state known to support Al’s party, and that some parts of the state known to support George have not yet reported their results. Are you still willing to make a projection based on your confidence interval from (b)? Explain.
Question # 6. The Normal Distribution

Suppose you are in charge of the “machine that fills cereal boxes” from the notes. You have lots of data on your process and are sure that the amount (in grams) of cereal put in each box can be described as i.i.d. draws from the normal distribution with mean \( \mu = 345 \) and standard deviation \( \sigma = 25 \). You are about to be audited by an inspector who will take a sample of 25 boxes and compute the average amount put in the 25 boxes. If that average is greater than 350 or less than 340, you will fail the audit and be reprimanded by your superiors.

(a) The inspector is about to look at the average amount in 25 boxes. Before the inspector begins looking at boxes, what is the probability distribution of this average? (This is the “sampling distribution” we talked about in class.)

(b) What is the probability that you will pass the audit?

(c) How would your answer to part (b) change if the inspector took a sample of 100 boxes?

(d) If you are right about \( \mu \) and \( \sigma \), do you want the inspector to look at more boxes or fewer? Explain intuitively.
Question # 7. Confidence Intervals for the Normal Distribution

In this question, we compute confidence intervals using the Australian returns data.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>107.00</td>
</tr>
<tr>
<td>Mean</td>
<td>0.012</td>
</tr>
<tr>
<td>Stand. Dev.</td>
<td>0.054</td>
</tr>
</tbody>
</table>

(a) Above are summary statistics for the Australian returns (see conret.xls). Construct the 95% confidence interval for the mean of the Australian returns.

(b) Now open the country returns data in Excel. Select Data Analysis -> Descriptive Statistics... and from the menu, select “Confidence Level for the mean” and enter “95” as the confidence level. Verify that Excel’s answer matches yours from part (a).

(c) Using the numbers from part (a), construct a 90% confidence interval and a 99% confidence interval for the mean of the Australian returns (use Excel’s NORMSINV function to get the critical values as we discussed in class). Check your answers using Excel as in part (b).

Which interval is widest (90, 95, or 99%)? Give a brief intuitive explanation. In particular, does a wider interval mean we know less about the actual mean?