BIO 1101 LABORATORY: PROBABILITIES AND GRAPHING

Objectives:

To gain an appreciation for the concepts of variance and stochasticity, the basics of probability theory, the difference between a prediction and an observation, and to provide practice in the drawing of graphs and the generation of tables. This exercise will be handed in, fully completed, as this week's laboratory quiz.

Plate your name here: _____.

Important terms and concepts:

Dice (singular, die or dice): A cube upon which dots representing different numbers are present. A fair, that is, not-loaded die will typically display, facing upward, any one of six numbers with equal probability (and that probability is, in fact, equal to 1/6).

Probability when sequential order matters (Multiplication Rule): If one is to roll one die in two consecutive rolls then the likelihood of any ordered combination of numbers (e.g., a 3 and then a 2) is equal to (1/6) × (1/6). This occurs because each throw of the die is an independent action and, for example, a 3 and then a 2 (❸→❷) is a different outcome from a 2 and then a 3 (❷→❸). Conceptually, this is the easier situation to wrap one's mind around…

Probability when order does not matter (Addition Rule): If order does not matter then, for example, a 3 followed by a 2 *is* equivalent to a 2 that is followed by a 3 (❸,❷ = ❷, ❸). This situation complicates the calculation of probabilities. Here the likelihood is no longer 1/36 = (1/6) × (1/6) but instead is 1/18 (which happens to be twice as large as 1/36). Why is this so? The odds of rolling a 3 and then a 2 (❸→❷) is 1/36 while the odds of rolling a 2 and then a 3 (❷→❸) is also 1/36, while (1/36) + (1/36) = 2/36 = 1/18. The probability of having two independent events occurring is the product of the odds of each happening (e.g., 1/6 × 1/6), but when there exist different *paths* towards the same result (e.g., ❸→❷ or ❷→❸), then you have to add together the probabilities associated with each path. In the above example we had two paths, each of which consisted of two independent events, [(1/6) × (1/6)] + [(1/6) × (1/6)] = 1/18. Note the 36 possible throws:

❶❶ ❶❷ ❶❸ ❶❹ ❶❺ ❶❻ = 0
❷❶ ❷❷ ❷❸ ❷❹ ❷❺ ❷❻ = 1
❸❶ ❸❷ ❸❸ ❸❹ ❸❺ ❸❻ = 1
❹❶ ❹❷ ❹❸ ❹❹ ❹❺ ❹❻ = 0
❺❶ ❺❷ ❺❸ ❺❹ ❺❺ ❺❻ = 0
❻❶ ❻❷ ❻❸ ❻❹ ❻❺ ❻❻ = 0

Materials needed:

Calculators, smart phones, or laptops running a spreadsheet program. One die per student.

Prelab: See biologyaspoetry.com/textbooks/bio1101_prelabs/probabilities.html.

Exercise 1: Distinguishing a Predication from an Observation

Exercise 2: Practicing the Multiplication and Addition Rules

Perform the following calculations or experiments. *Record and circle your answers.*

1. What are the odds of throwing one 5 in one roll of a fair, six-sided die? [an expectation]
2. Given one roll of a fair, six-sided die, what would you *predict* would be the likelihood of throwing a 5? Throwing a one? Throwing a six? [these are expectations]
3. What are the odds of throwing one 2 in one roll of a fair, six-sided die? [expectation]
4. What are the odds of throwing one 2 and *then* one 5 in two rolls of a fair, six-sided die? [expectation]
5. Given two sequential rolls of a fair, six-sided die, what would you predict would be the expected number of throws of a 2 and then a 5? (hint: Multiplication Rule)
6. Given two sequential rolls of a fair, six-sided die, what would you *predict* would be the number of times you throw a 5 and then a 2? [expectation]
7. Given two sequential rolls of a fair, six-sided die, what would you predict would be the number of time you throw a 5 and then a 2 *or* a 2 and then a 5? (hint: Addition Rule)[expectation]
8. What are the odds of throwing one 2 and one 5 in one roll of *two* fair, six-sided dice? [expectation]
9. What would you *predict* would be the odds of throwing one 2 and one 5 in one roll of *two* fair, six-sided dice? [expectation]
10. If you were to roll *two* fair, six-sided dice a total of ten times, what would you *predict* would be the total number of 2 and 5 *or* 5 and 2 combinations? Answer this as a decimal. [expectation]
11. With your partner, roll two dice a total of ten times. How many 2 and 5 *or* 5 and 2 combinations did you *observe*? [observation]
12. In one roll of *two* fair, six-sided dice, what are the odds (*prediction*) of throwing a sum of 2?
13. In one roll of *two* fair, six-sided dice, what are the odds (*prediction*) of throwing a sum of 3?
14. In one roll of *two* fair, six-sided dice, what are the odds (*prediction*) of throwing a sum of 5?
15. In one roll of *two* fair, six-sided dice, what are the odds (*prediction*) of throwing a sum of 8?
16. Roll two dice once. How many rolls adding up to an 8 did you *observe*? [observation]
17. If you were to roll *two* fair, six-sided dice 20 times, what would you *predict* would be the total number of rolls that would sum up to an 8? [expectation]

Exercise 3: Probability with order mattering, example with smaller n

Find a partner to work with. Odd students out can observe and otherwise help pairs. Each student should roll their die 20 times, keeping track, in order, of the result of each roll. Fill in Table 1 in order with (1) the result of each roll by student "A", (2) the result of each roll by student "B", and (3) the sum of each pair of rolls. Thus, for example:

Roll	Student A	Student B	Sum
1	5	6	11
2	2	2	4

Use the following to record your actual rolls.

Table 1: Recording of Rolls of Pairs of Dice and Their Sums.

Roll	Student A	Student B	Sum
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			

Now calculate ***expected*** (i.e., ***predicted***) versus ***actual*** (i.e., ***observed***) sums by filling out Table 2 as based on 20 pairs of rolls as recorded in Table 1. "***Observed***" is found by counting the sums in Table 1 as found in the last column (e.g., if there are three 8s in the third column, then enter a "3"). "***Expected***" is equal to the probability of a given sum multiplied by the total number of rolls (20). Thus, a sum of 2 is expected to occur $1/6 \times 1/6 \times 20$ times = 0.56 out of 20 rolls; a sum of 8 is expected instead to be $\{(1/6 \times 1/6) + (1/6 \times 1/6) + (1/6 \times 1/6) + (1/6 \times 1/6) + (1/6 \times 1/6)\} \times 20$ times, as corresponding to the ***expected*** number of throws of the following rolls: (2,6), (6,2), (3,5), (5,3), or (4,4).

Table 2: Observed vs. Expected Number of Sums Found Rolling Two Dice 20 times.

	2	3	4	5	6	7	8	9	10	11	12
Observed											
Expected											

All of the "***Expected***" values (as equivalent to "***Predicted***" values) should together add up to a sum of 20, as so too should "***Observed***" (as equivalent to "***Actual***"). Do they? *Where does the 20 in that statement come from?*

Now regenerate Table 2 ☞☞☞☞☞ *but do so ordering the Expected/Predicted values from predicted smallest likelihood to predicted largest likelihood* ☜☜☜☜☜

Technically this is a sort on *Expected* values, increasing from left to right. **Try it on scrap paper or in the margins first.** The *Expected* row should start with smallest *Expected* number (or numbers) and end with the largest *Expected* number (or numbers). Don't be thrown off by the fact that different sums (e.g., as adding up to 2 and as adding up to 12) will have the same *Expected* values.

Shown is the beginning of this sort, e.g., the top row could/should end up reading 2, 12, 3, 11… After you are done, *Observed* and *Expected* rows each should still add to 20. Making sure the latter is still the case constitutes a check on your work. In science as in school as in life, you should always look out for and then perform easy checks on your work before proceeding to subsequent steps. That is, "Measure twice, cut once!"

Table 3: *Observed* versus *Expected* Number of Sums Found Rolling Two Dice 20 Times, Sorted so that the *Expected* Likelihoods are Increasing from Left to Right.

	2	12							6	8	7
Observed											
Expected											

Exercise 4: Drawing our first graph

On the following page you will be asked to draw a graph. On a graph typically the *independent variable* is drawn on the *x* axis (the horizontal direction) while the *dependent variable* is drawn on the *y* axis (the vertical direction). It is good practice to give variables meaningful names. For example, if the *x* axis on a graph were to represent time, then a good name for the *x* axis would be "Time" (without the quotes). One would then place "Time" (again without the quotes) below the lowest horizontal line on the graph, thereby labeling the *x* axis as Time. Similarly, the *y* axis should be given a name and that name should be placed to the left of the left-most vertical line found on the graph.

Here we will be treating the top row in Table 3 as the independent variable (i.e., 2, 12, 3, 11…).

What is the name of the independent variable and where should that name be placed?

Hint: What do the numbers 2, 12, 3, 11, etc. represent?

If these numbers were the words, Apple, Pear, Cherry, Lemon, Grape…, what word instead might you use to name the x axis?

Generally when graphing by hand it is best to make the graph as wide as can be conveniently done. Thus, how many different numbers are found in the top row of the Table 3? How many boxes instead are found going from left to right in the graph paper provided on the following page? What then would be the most convenient way to present the independent variable, i.e., 2, 12, 3, 11…, assuming that we want to fill up the graph as much as possible going from left to right?

Starting with "2" (without the quotes) under the second vertical line making up the graph paper, enter this string of numbers going from left to right.

Is this the correct order to enter these numbers 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12?

This exercise might be a little harder for the dependent variable. First you need to find, among the Observed and Expected values, what is the largest number. (See Table 2 or Table 3 to answer this question.) Second, how many boxes are present in the graph paper, as found on the following page, going from top to bottom? What would be the most convenient way, then, to graph the Observed and Expected values that would result in the most graphing "real estate" being taken up overall by the graph? Number the *y* axis, making sure that you stretch the graph vertically as much as is convenient. Now give the *y* axis a name.

What should you label the y axis as? That is, what name should it have?

Below we're going to graph the *Expected* and *Observed* values as found in Table 3 using open circles (= Expected values = ○) and closed circles (= Observed values = ●), respectively. The top row in Table 3, *in the order presented there*, is as noted the *independent variable* (*x* axis). The number of throws of each type (*Observed* or *Expected*) is the *dependent variable* (*y* axis). Remember to properly label axes in terms of both numbers and names (e.g., "number of throws" or "sum of the roll of two dice"). For values that are not whole numbers, approximate their position, and feel free to ask your professor about the "correctness" of your graph.



Figure 1: Expected and Observed Number of Times Specific Sums of Two Dice were Thrown in Twenty Rolls. (ABOVE, YOU DRAW THIS)

Exercise 5: Probability without order mattering, starting with larger n

We will now combine results for the entire class. This will be accomplished by filling out Table 4. All of the boxes below except for the rows marked "*Expected*" and "*Observed*" will be copied into Table 4 from the board. The numbers listed on the left in the table are numbers corresponding to groups, i.e., if we have a class of 24 students then where will 12 groups, with each group arbitrarily given a number of between, e.g., 1 and 12.

Each group will go up to the board and fill in a row with their *Observed* values, ordering the sum of throws from 2 to 12 (Table 2) rather than as reordered in Table 3.

The row marked "*Observed*" will be the sum of the columns (not including the immediately top row). *Expected* should be calculated as above but based on 20 throws multiplied by the total number of groups that presented their results on the board, and then multiplied again by the probability of receiving a given sum given a single throw. Thus a sum of 2 is expected to occur $1/6 \times 1/6 \times 20 \times 15$ times = 8.33 if there are 15 groups and 20 throws. Note that the actual order that groups enter their data into Table 4 or onto the board is not important, just so that you list everybody's results.

Table 4: *Observed* versus *Expected* Sums Found by Various Groups.

	2	3	4	5	6	7	8	9	10	11	12
1											
2											
3											
4											
5											
6											
7											
8											
9											
10											
11											
12											
13											
14											
Observed											
Expected											

Restated: The top row comes from Table 3. Both the "*Expected*" row and the "*Observed*" row are based upon the values calculated in Table 4. The next two rows are then calculated from these two other rows, in each case by dividing by the number of groups.

Table 5: *Observed* versus *Expected* Number of Sums Sorted so that the Expected Likelihoods are Increasing Going from Left to Right.

	2	12							6	8	7
Observed											
Expected											
Obs/#grps											
Exp/#grps											

Exercise 6: Drawing our second graph

On the following page, graph the information found in Table 5 as was done to generate Figure 1 as based on the data found in Table 3. To do this use the two last rows as presented in Table 5. Note that except for the actual data graphed, Figures 1 and 2 should be presented essentially identically, with mostly equivalent *x* and *y* axes. Once completed you thus should be able to flip back and forth between Figure 1 and Figure 2 in order to compare the two figures.

In what way does Figure 2, as generated on the following page, differ from Figure 1, as generated above? Hint: the Expected points should look the same while the Observed should differ. Second Hint: everything else about the two graphs should be more or less identical including the labeling of the x *and* y *axes.*

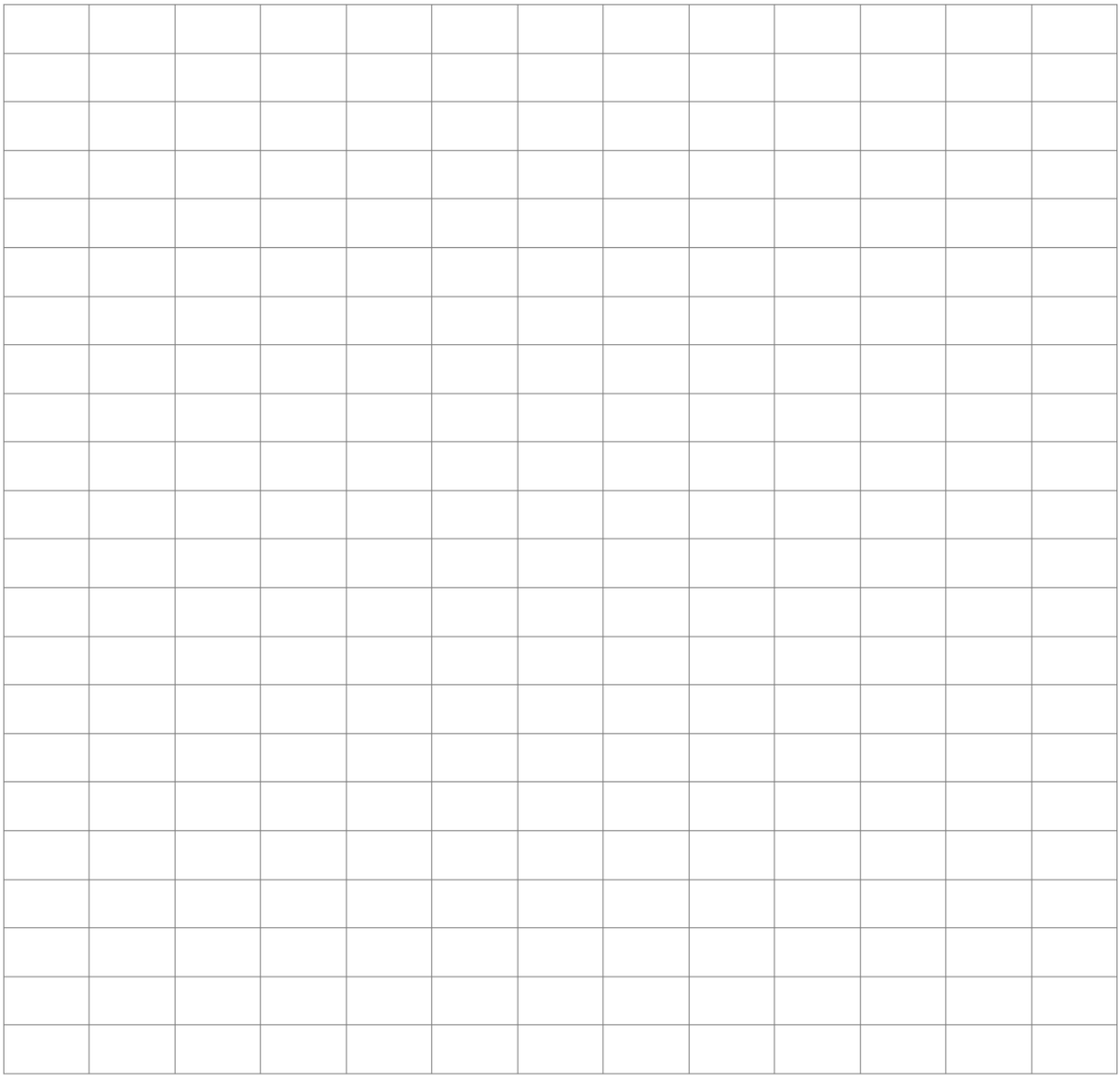


Figure 2: *Expected* and *Observed* Number of Times Specific Sums of Two Dice were Thrown in many more than Twenty Rolls. (ABOVE, YOU DRAW THIS)

Exercise 7: Probability and variance with roll order mattering, larger vs. smaller n

Starting with the data found Table 1, fill out Table 6 as found on the following page.

The Actual number of throws of each type as *Observed* in Table 1 should be placed to the left of the parentheses. Thus, if you saw five throws of a combination of a 4 by person A and a 2 by person B, then place "5" in the box corresponding to "A.4" and "B.2", but not

in the parentheses! Note that these numbers are not identical to the *Observed* values placed in previous tables, though all should still add up to 20.

In the parentheses, place the *Expected* values. Remember here that "order" is important since we are distinguishing throws by student A from those by student B rather than using the sum of those throws. Therefore the *Expected* values for each throw will *not* be the same as you calculated earlier. Instead, for example, row A.3 and column B.2 should be the number of times that person A threw a 3 in the same turn (row) as person B threw a 2 (column), with $1/6 \times 1/6 \times 20$ times = 0.56 therefore placed in the corresponding parenthesis.

Table 6: Observed versus Expected Combinations of Throws Where Order Matters.

	B.1	B.2	B.3	B.4	B.5	B.6
A.1	(___)	(___)	(___)	(___)	(___)	(___)
A.2	(___)	(___)	(___)	(___)	(___)	(___)
A.3	(___)	(___)	(___)	(___)	(___)	(___)
A.4	(___)	(___)	(___)	(___)	(___)	(___)
A.5	(___)	(___)	(___)	(___)	(___)	(___)
A.6	(___)	(___)	(___)	(___)	(___)	(___)

What value(s) should be placed in the rest of the parentheses?

Add up all of the values found in the parenthesis? What number do they add up to?

Add up all of the values found outside of the parenthesis? What number do they add up to?

Now roll another 20 pairs of dice, and with each member of the group rolling one of the die, filling out in the table found on the following page (Table 7).

Table 7: Recording of Rolls of Pairs of Dice and Their Sums (Second Round).

Roll	Student A	Student B	Sum
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			

Now, using both data sets as found in Tables 1 and 7 – that is from 40 pairs of rolls – again generate a table of the Table 6 type. This will *not* be identical to Table 6, neither *Observed* nor *Expected*, since Table 8 will be based on 40 pairs of throws rather than the 20 pairs of throws used to generate Table 6.

Table 8: *Observed* versus *Expected* Combinations of Throws Where Order Matters (Using Twice as many Throws).

	B.1	B.2	B.3	B.4	B.5	B.6
A.1	(___)	(___)	(___)	(___)	(___)	(___)
A.2	(___)	(___)	(___)	(___)	(___)	(___)
A.3	(___)	(___)	(___)	(___)	(___)	(___)
A.4	(___)	(___)	(___)	(___)	(___)	(___)
A.5	(___)	(___)	(___)	(___)	(___)	(___)
A.6	(___)	(___)	(___)	(___)	(___)	(___)

What is the number of times that you and your partner would expect to together throw a pair of sixes over forty pairs of rolls? Over twenty pairs of rolls instead?

The variance between *Expected* values and *Observed* values is some function of the difference between these numbers, and this variance should be greater overall in Table 6 than Table 8. To calculate the overall variance, minimally you need to determine the *absolute value* of the difference between *Expected* and *Observed* values. In the following two tables, fill in the blanks with these differences, first for the first 20 throws and then for all 40 throws. When you are done, add up all of these differences for each table, dividing the second number (based on 40 throws), and only the second number, by two when you are done to obtain comparable values.

Table 9: Absolute Value of Difference Between *Expected* and *Observed* in Table 6.

	B.1	B.2	B.3	B.4	B.5	B.6
A.1						
A.2						
A.3						
A.4						
A.5						
A.6						

Total variance for $n = 20$ is equal to _____ = $\{\sum|\text{Obs} - \text{Exp}|\}$ for 20 throws.

Table 10: Absolute Value of Difference Between *Expected* and *Observed* in Table 8.

	B.1	B.2	B.3	B.4	B.5	B.6
A.1						
A.2						
A.3						
A.4						
A.5						
A.6						

Total variance for $n = 40$ is equal to _____/2 = _____= $\{\sum|\text{Obs} - \text{Exp}|\}/2$.

Which data set has the greatest variance (that is, is the final result largest)?

Would you expect this general result every time you ran the experiment? Why or why not?