**MATH 2310**

**Discrete Distributions**

This lab will focus on the discrete probability distributions which we have been learning about in class.

In addition to a final answer, you are required to submit the code that you used to obtain the answer, with comments/print-outs as appropriate.

R has excellent built-in functionality for calculating probabilities. In this assignment, you will be using the functions for the Binomial and Poisson distributions (documentation [here](https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/Binomial), and [here](https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/Poisson)). Each distribution has four associated functions:

* Probability density, e.g., dbinom
* Cumulative distribution, e.g., pbinom
* Quantile function, e.g., qbinom
* Random sampling function, e.g., rbinom

To check your answers, feel free to use an online probability calculator, which has a more intuitive user interface than R. Here is one for [binomial distributions](https://stattrek.com/online-calculator/binomial), and here is one for [Poisson distributions](https://stattrek.com/online-calculator/poisson).

In most of the following problems, you are only trying to calculate a probability. However, in a few questions, rather than finding a probability, you are using a calculated probability to decide whether some assumption is true. This is an important principle of statistics that we will be examining in greater depth later in the quarter. This process is called hypothesis testing. In many situations, we may be interested in testing whether some assumption is in fact correct for our data. We look at whether our observed data seem reasonable, given some assumption. We go about this by seeing what the probability would be, if our assumption were correct, of seeing a result at least as far away from what we would expect as what was observed in our data. If this probability is low (a common rule of thumb is if it is below 5%), that tells us that we would not have expected our data to have occurred if our assumption were correct, and therefore the most likely explanation is that our assumption is incorrect. If the probability is not low (above 5%), then our data is not inconsistent with our assumption, and, while we cannot say for certain that our assumption is correct, we can at least say that there is no evidence it is wrong. Keep in mind, however, that **5% is only a guideline, not a hard rule**. If you find that the probability of a certain outcome is 7%, for instance, that should draw nearly as much scrutiny as 5%, especially if we are dealing with a small sample size.

For example, suppose you are flipping a coin 10 times. Our starting assumption is that the coin is fair, that is, that the probability of heads is ½. Now suppose we only get heads one time out of the ten flips. The chance of seeing one or fewer heads is about .011. That means that if we repeatedly flipped sets of ten coins, we would only expect to see one or fewer coins per ten about one time out of every hundred sets of ten coins - it would be very unusual to see so few heads if the coin were fair, and so we conclude that our assumption of a fair coin appears to be incorrect, and the probability of heads is likely actually something less than ½. If, on the other hand, we got heads three times, the chance of seeing three or fewer heads is about .172. This means it would not be that unusual, given our assumption, and so we would not see any reason to question our assumption – there is no evidence the coin is unfair. For these types of questions, our answer will always be one of these two types – either stating that our assumed value seems reasonable given the data, or stating that we would question our assumption, and conclude the value must actually be different in some way.

Goals for this assignment:

* Learn to use R to calculate binomial and Poisson probabilities
* Identify which distribution is appropriate for a given problem
* Understand how to use probabilities to answer questions about the reasonableness of assumptions

Grading: there are two possible points for each problem.

**Activity 1**

In the general population, about 10% of people are left-handed.

a) Suppose I randomly pick 200 people. What is the chance I will see fewer than 15 people who are left-handed?

b) Suppose I randomly pick 300 people. What is the chance I will see at least 40 people who are left-handed?

c) The Seattle Mariners have 21 pitchers. Of those 21 pitchers, 4 are left-handed. If we assume the probability of a pitcher being left-handed is the same as the probability of any randomly selected person from the general population being left-handed, what would be the probability of seeing at least 4 left-handed pitchers out of 21?

d) Based on your answer from part c), does our assumption about the probability of a pitcher being left-handed seem reasonable? If not, how does the probability differ from the probability for the general population?

e) The Seattle Mariners have 3 catchers. All 3 of them are right-handed. If we assume the probability of a catcher being left-handed is the same as the probability of any randomly selected person from the general population being left-handed, what would be the probability of seeing no left-handed catchers out of 3?

f) Based on your answer from part e), does our assumption about the probability of a catcher being left-handed seem reasonable? If not, how does the probability differ from the probability for the general population?

g) Suppose we were to look at a larger sample of catchers. If every catcher we sampled were right-handed, how many catchers would we need to sample before you would conclude that the true probability of a catcher being left-handed is less than 10%?

**Activity 2**

Suppose you are designing a computer server for students to log into to work remotely. You know that on average you will see ten students logging into the server per hour.

h) What is the chance that more than 15 students will log into the server in a particular hour?

i) What is the chance of seeing exactly 10 students log into the server in a particular hour?

j) What is the chance of fewer than 15 students logging into the server in a two-hour period?

k) In designing the server, you must decide the maximum number of students that it can accommodate at one time. The more students you allow it to accommodate, the more expensive it will be. But if more students attempt to log in during a single hour than it can accommodate, it will crash. How many students should you design it to accommodate if you want there to be at most a 1% chance that it will crash during any particular hour?