Problem 1
Each day a hospital records the number of people who come to the emergency room for treatment.

(a) In a particular week, the arrivals to the ER are:

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
10	8	14	7	21	44	60

Do you think that the Poisson distribution might describe the distribution of the number of daily arrivals (i.e., using an interval of one day)? Why or why not? Use graphical and/or analytical probability results to justify your answer.

No, I do not think the Poisson distribution would describe the distribution of the number of daily arrivals. The reason is because certain days of the week (Friday and Saturday) have significantly more arrivals than the other days. It is likely that those two days would continually be higher than the rest. While the majority of days have less than 15 arrivals the mean number of arrivals for the week is 23.4. Figure 1 shows what the PMF for a Poisson distribution with the given mean would look like. As mentioned the majority of days have less than 15 arrivals, but with the given distribution the probability of having less than 15 visits is 0.044. There are also two days with more than 40 arrivals, but the probability of having over 40 would be 6.3×10^{-4}. So this distribution doesn't really describe the data. We could used the median (14 arrivals) of the weekly numbers to build our distribution, which would make it fit the majority of the days better. However, the high numbers on Friday and Saturday would then be even less likely.

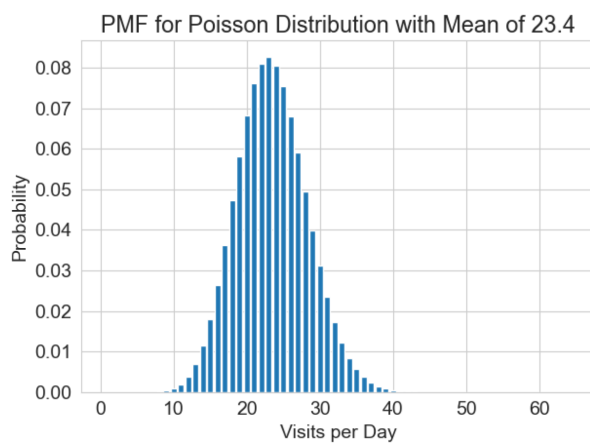

Figure 1: The probability of the number of ER visits per day based on a Poisson distribution with a mean of 23.4

(b) Building upon your answer the previous question, would you expect the Poisson distribution to better describe, or more poorly describe, the total number of weekly admissions to the ER (i.e., using an interval of one week)? Why?

I think the Poisson distribution would better describe the number of weekly admission rather than daily. The consistently high days of the week would no longer pull up the mean. The number of visits week to week would likely have less variance. I would really expect the number of visits per week to be a normal distribution. For a high mean number of events per interval, the Poisson distribution becomes similar to the normal distribution (as you can see from Figure 1).

Problem 2
Lake Wobegon Junior College admits students only if they score above 400 on a standardized achievement

test. Applicants from Group A have a mean of 500 and a standard deviation of 100 on this test, and applicants from Group B have a mean of 450 and a standard deviation of 100. Both distributions are approximately normal, and both groups have the same size.

Note: This problem builds on examples from the Progress Checks. If you haven't done those, this will be far more difficult.

(a) About 45% of accepted students come from Group B with the remainder coming from Group A (See Progress Check 14).

Assume that students with a score above 600 will decline their admission to Lake Wobegon Junior College in favor of another school. What is the expected proportion of those students who both (1) are accepted and (2) choose to attend who come from Group B?

Let A be the student being from Group A, B be the student being from Group B, C be the student being accepted and D be the student deciding to attend. We are determining the probability that a student is from Group B, given that they were both accepted and decided to attend. In mathematical terms we need to find $P(B|C \cap D)$. Because D is a subset of C, $C \cap D = D$ (all students that decided to attend were selected). We can apply Bayes's Rule to determine the probability:

$$P(B|C \cap D) = P(B|D) = \frac{P(B)P(D|B)}{P(B)P(D|B) + P(A)P(D|A)}. \tag{1}$$

Because the groups are the same size $P(B) = P(A) = 0.5$. To find $P(D|B)$ and $P(D|A)$ we use the normal distribution get the probability that the score was greater than 400 and less than 600, for each group respectively. Those probabilities are $P(D|B) = 0.625$ and $P(D|A) = 0.683$. Now we can solve equation 2.

$$P(B|C \cap D) = P(B|D) = \frac{(0.5)(0.625)}{(0.5)(0.625) + (0.5)(0.683)} = 0.478 \tag{2}$$

(b) Expanding on the last part, create a plot of the proportion of students from Group B as a function of the upper score cutoff (i.e., the score above which students will choose another school). At what upper cutoff score do you expect the proportion of students from each group to be equal?

Figure 2 shows the probability of students being from Group B given that they determined to attend based on the upper cutoff score. Because Group B has a lower mean score than group A as the upper cutoff score decreases, more students from Group A choose to attend another school, which increase the proportion of students from Group B. The point at which the proportion of students is equal is at an upper cutoff score of 550.

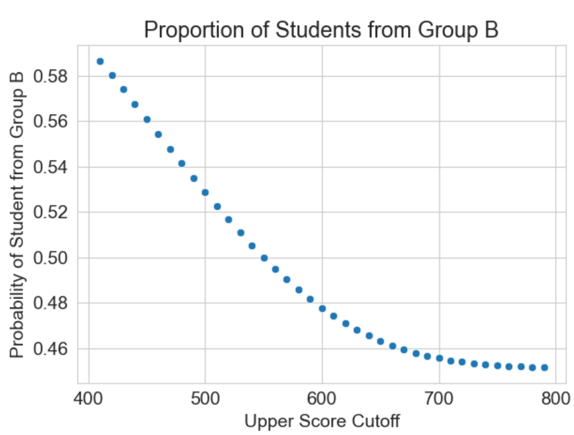

Figure 2

Problem 3

Demonstrate that the Central Limit Theorem works on a Poisson distributed variable with $\lambda = 2$. Use this as the population distribution for all sub-parts.

(a) Simulate taking the mean of 2 observations of the variable 1000 times. Make a histogram and boxplot of the means. Comment on the approximate 'normality' of the sampling distribution.

Figure 3 contains the histogram and boxplot of the means of 2 observations of the variable 1000 times. The black line is the normal distribution with the calculated mean and standared deviation from the sample means. From the boxplot you can see that the median of the means is almost exactly 2. The distribution of the means slightly resembles a normal distribution but has obvious right skewness.

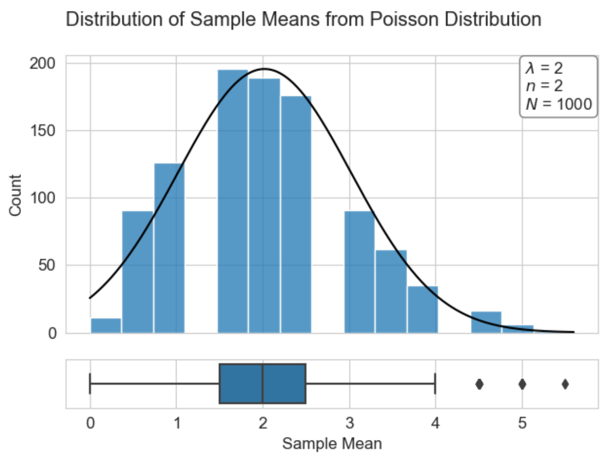

Figure 3

(b) Simulate taking the mean of 10 observations of the variable 1000 times. Make a histogram and boxplot of the means. Comment on the approximate 'normality' of the sampling distribution.

Figure 4 contains the histogram and boxplot of the means of 10 observations of the variable 1000 times. With the larger number of observations the distribution of the means does not have as much visible skewness. There are a few more outliers to the right, but it does resemble the normal distribution. There are still large dips or lack of smoothness to the distribution though.

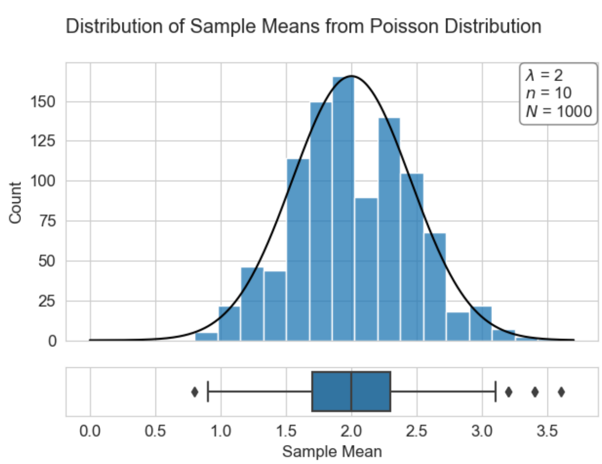

Figure 4

(c) How large does n (the number of observations used to calculate the mean) need to be for the sampling distribution to approximate normality? Justify your answer. Use logic to argue your point augmented by math as needed.

To determine this I continued to increase the number of observations and at $n = 30$ I would say the distribution was approximately normal. Figure 5 shows this distribution with a normal distribution as well. It looks symmetrical with a mean of 2. The histogram is not completely smooth, but that could be a result of the bin sizes (were values close to the bin edges get grouped).

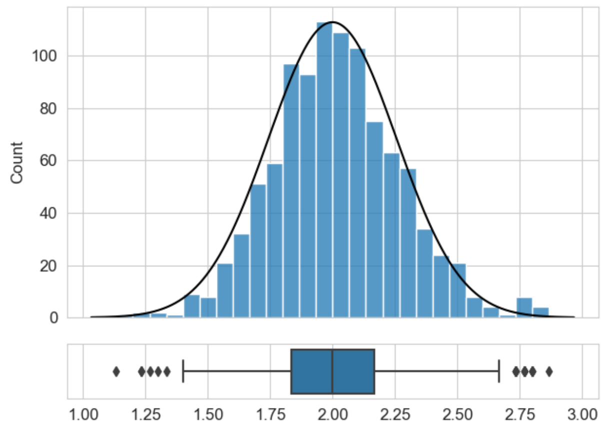

Figure 5