# UNL Statistics PhD Qualifying Exam - May 2024

**Print Your Qualifying Exam ID: _____**

**8:00 am – 1:00 pm**

**Open book, open note. Bring your resources with you. You will not be allowed to go back up to your office to get additional notes. You may not use your phones for any reason.**
**Bring a calculator.**
**You may bring a flash drive with any SAS and R programs that you may possibly want to use.**
**You will have access to software available on the computers in the lab. You may NOT bring your own computer. You will NOT have access to the internet. If there is an R module that you need, our IT person will download it for you.**

- **Answers to each problem should be saved as a single document in a directory on the USB drive you were provided.**
- **Any programs and output for a question should be included in the directory for that problem.**

# Day 2

1. (100 points) **[Estimation of $\pi$.]** There are many mathematical and statistical methods to find $\pi$. Here we consider two Monte Carlo methods.

   The average value method draws $n$ **uniform** random points from $[a, b]$ to estimate the mean value of a continuous function on $[a, b]$.

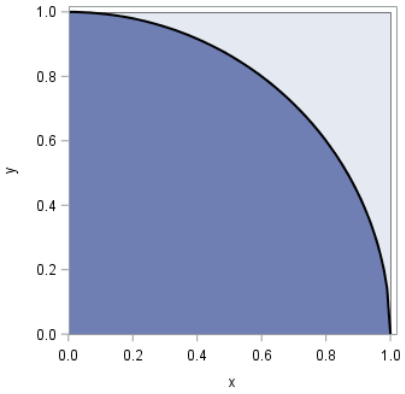


Figure 1: Average Value Method

   In calculus the mean value of a continuous function $f$ on $[a, b]$ is defined as

$$f_{\text{mean}} = \frac{1}{b-a} \int_a^b f(x)dx$$

   For example, $f(x) = \sqrt{1-x^2}$ and $a = 0$, $b = 1$, the true mean is $\frac{\pi}{4}$, because the integral is equal to the area under the curve (Figure 1), i.e.,

$$f_{\text{mean}} = \int_0^1 \sqrt{1-x^2}\, dx = \frac{\pi}{4}.$$

   Then we multiply $f_{\text{mean}}$ by 4 to estimate $\pi$.

   The area method estimates the area by generating a uniform sample of points and count how many fall into a planar region. We **uniformly** draw $n$ random points $(x, y)$ in a **2-D plane** with the domain as a square of side 2 centered at (0,0). We draw a circle inside the same domain with the radius of 1 and inscribed into the square. We use the ratio of number points that fell inside the circle to the total number of generated points to estimate $\pi$. See Figure 2.

Please read the questions carefully and complete the required tasks.

1) (40 points) Write two `R` functions, `avg.value(n)`$\{\cdots\}$ and `area(n)`$\{\cdots\}$ to implement two methods. Test your functions for $n = 5000$ and report the results. The methods must be coded as `R` functions.

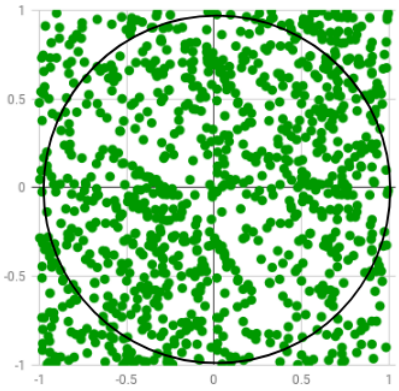


Figure 2: Area Method

2) (30 points) Simulate for $n = c(10^1, 10^2, 10^3, 10^4, 10^5, 10^6, 10^7, 10^8)$ draws. Plot the estimates of $\pi$ on $y$-axis versus $\log_{10} n$ on $x$-axis. `For` loop must be used in simulating for different $n$, and the simulation results of two methods must be plotted on the same figure.

3) (15 points) Comparing the estimates of two methods, what observations emerge? Why?

4) (15 points) Here's a mathematical method for computing $\pi$ (we skip the complex theory behind the math approach):

`Initialize` $M_0 = 3$

`for` $k$ from 0 to $K$:

$$M_{k+1} = 2 + \sqrt{M_k}, \qquad \pi_k = 3 \times 2^k \times \sqrt{4 - M_k}$$

`end`

Compute $\pi_1$ through $\pi_9$ using this method and compare the results to those obtained through two Monte Carlo methods. Then, discuss the strengths and weaknesses of statistical approaches for estimating $\pi$ to the mathematical approach.

2. Wheat is one of the most crucial crops worldwide, and the wheat yield performance must be improved to ensure food security for the growing human population. There are several biotic and abiotic stresses that are hindering the development of improved wheat cultivars. One of the major biotic stresses for wheat in the Great Plains is wheat curl mite, which is also able to transmit damaging plant viruses in wheat. Even without the presence of any virus, the wheat curl mite contributes to a significant amount of yield reduction in wheat. There are varieties that are more susceptible to the virus than others are. A research lab developed a chemical that have the potential to reduce the presence of the mites. They conducted a study where they used 10 different winter wheat varieties, and they applied their chemical treatment in two different doses. After

two weeks of the application, they counted the number of mites on the leaves and the stems of the plants within a plot. They also rated how yellow the leaves were.
For each variety x dose combination, the researchers replicated the experiment 6 times, and the experiment was replicated twice in 2021, and once in 2022.
The data file (it is on the pen drive) contains information about the replication, variety, dose level, and the two responses (mite count and yellowing rating) that the researchers collected. Please note that the researchers and the farmers prefer lower (zero) number of the mites, and lower number of the yellowing (zero means that the leaves are green and healthy looking).
The researchers contacts you because you are a statistical consultant, and they want you to analyze their data for them. In the process, there are several aspects you need to consider.
The client is interested in how the variety and the dose influence the mite count and the yellowing rate. Looking at the data, you will notice that there are missing values, which you need to address somehow, and you need to explain and justify your decision. You need to make recommendations to the client regarding their question (how do the different varieties and doses affect the two responses, and which variety or varieties and dose should they consider). Part of your job is to provide interpretation to the client who has minimal statistical background.
To present your work to the client, you need to write a report. In your report, you need to address the following:

(a) Provide exploratory analysis of the data, and present it with plots/tables that are easy to understand. Interpret your output.

(b) Suggest some statistical model(s) that you consider looking at the data. Explain the model in details, and justify your model choice.

(c) Implement your model(s) and conduct a data analysis. Do not forget to check model assumptions, and interpret your results.

(d) Discuss another way to analyze your data. You only need to implement one way (which you completed in part (c), but describe your "other" way in details without implementation).

Your report needs to contain a brief ***introduction***, a ***methods*** section where you discuss the data (with exploratory data analysis) and your model, a ***results*** section where you provide the model assumption checking, implementation of your model in terms of results and explanation of results. Your ***discussion*** section should provide information about the "other" way to analyze the data.
You need to have an ***appendix*** section with your code.