Morgan Watson, Sophie Salem, Swaycha Goli, Maya Yin, Maya Krishnamoorthy

**BIOS 664 Project Plan**

Topic: Proportion of available study rooms in UNC-CH campus libraries during different times of the day

## Overview

In this project, we aim to estimate the study room availability in libraries across the UNC-Chapel Hill campus at various times of the day. Through our investigation, we hope to find the trends and factors influencing the availability of study spaces on campus throughout the day. We seek to provide insight into when the university and its students can expect to access study spaces, and how to handle room allocation in the future.

We will estimate the proportion of available reservable study rooms across campus at various times of day. The population we will be sampling from will be limited to the three libraries with the greatest number of reservable study rooms at UNC on weekdays. Since all of these libraries on campus are open on weekdays, we believe that it will be more beneficial to estimate their availability than compare the availability of the select libraries that are open on weekends. We will examine their availability through a 12-hour period per day, between 8:00 AM and 8:00 PM.

Since we are estimating the proportion of available study rooms on UNC's campus, the members of our population are reservable study rooms during a designated hour block in the time periods specified above. The observation unit for our project is one study room within the hour block. We will specifically be observing whether the one room, or unit, is occupied or not. We will focus on specific times to determine changes in availability based on the time of day.

## Sample Design

We will be stratifying by library and time of day. The library strata will be Davis Library, the House Undergraduate Library, and the Health Sciences Library. The time of day strata will be 8 AM-12 PM, 12 PM-4 PM, and 4 PM-8 PM. The reason we are stratifying by library is because certain libraries may elicit different demands for study rooms by students, therefore we expect homogeneity within strata and heterogeneity between strata. Also, reservations may have great variance between times of day, as classes and work schedules probably fall during the morning and early afternoon and cause heterogeneity in reservation proportions between times of day.

To derive the sampling frame for each stage of selection, we will utilize the UNC Library website, specifically the reservation pages for each library. Since we are doing a stratified sample, we will sample from Davis Library, the House Undergraduate Library (UL), and the Health Sciences Library (HSL) within the different time strata by navigating their reservation websites. Cluster sampling will not be used. Because we have the ability to check all of the rooms in our population using the online reservation service, the sample will be representative of the population.

The sampling units for this study will be one room at a designated library within an hour time period. Sampling units will be selected through an SRS of the rooms available on the sampling frame within a specified time block. In this case, samples will be drawn from each of the library reservation websites.

The selection probabilities would be the same for each room in a particular library since we would be using stratified-proportional allocation to achieve an EPSEM design. The three libraries in the sampling frame have different numbers of reservable study rooms. Thus, each stratum would have a different size. We are selecting a sample size ($n_h$) proportional to the stratum size ($N_h$). The selection probability for each unit, a study room during a one-hour interval, would be $n_h/N_h$.

Design Summary Table

Observation Unit: a study room during a one-hour interval on a weekday

| Stage | Description of Sampling Unit (What is being sampled?) | Stratification (By what? How many strata? Is sample allocation balanced, proportionate and/or disproportionate?) | Type of Selection Method Used to Choose Sampling Units (e.g. SRS, PPSWOR, etc.) | Sample Size (How many sampling units were chosen?) |
|---|---|---|---|---|
| 1 | A study room during a one-hour interval on a weekday | 9 strata formed by crossing 3 libraries by 3 time periods. The sample is allocated proportionally across strata.<br><br>Strata: HSL in the morning, mid-day, and evening; the same applies for Davis and the UL | SRS | 357 |

Sample Size

In total, we have 77 reservable study rooms among all three libraries that we are interested in. Each room has 12 one–hour intervals for observation. In total, we have 4620 units in the population throughout 5 weekdays.

To calculate the sample size, we are controlling the margin of error at 0.05. The following values are specified in the sample size calculation: e=0.05 and $z_{1-\frac{\alpha}{2}}$=1.96. Because we are interested in the proportion of available rooms, it is reasonable to estimate the outcome to be

dichotomous. Thus, the conservative variance estimation would be
$p(1-p) = (0.5)(0.5) = 0.25$.

Assuming SRS with replacement of n' sampling units out of the total number of available rooms*hour*day population, the sample size needed is

$$n' = \frac{z_{1-\frac{\alpha}{2}}^2 s^2}{e^2} = \frac{1.96^2(0.5)(0.5)}{0.05^2} = 384.16$$

To adjust for finite population correction, we determine the sample size required to be

$$n = \frac{n'}{1+\frac{n'}{N}} = 354.67$$

In the first stage, we would be selecting units proportional to the strata size. The number of samples allocated to each stratum with sample calculations is shown in the table below:

| Library | Davis Library | | | Undergraduate Library | | | Health Science Library | | | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| Time interval | Morning | Midday | Evening | Morning | Midday | Evening | Morning | Midday | Evening | |
| Strata Size ($N_h$) | 42 rooms * 4 hours * 5 days = 840 | 840 | 840 | 240 | 240 | 240 | 440 | 440 | 440 | 4620 |
| Sample Size ($n_h$) | 355 * (840/4620) = 65 | 65 | 65 | 20* | 20* | 20* | 34 | 34 | 34 | 357 |

*To ensure ample sampling, this number is rounded up

The cost of collecting samples will only be checking the reservation website 24 hours before the selected one-hour period for a study room. We expect it to be manageable to collect 357 samples in 5 days. For this reason, we are not considering clustering in our sampling design. Since the project utilizes a stratification design with proportional allocation, the design effect is estimated to be less than or equal to 1.

Data Collection Plan

We first classify a study room to be "available" if there is no reservation on the room during the entire one-hour interval in each of our three blocks in the day (8-12, 12-4, 4-8). We will classify all reservable study spaces in the population to be available or unavailable depending on their reservation status online.

We will check the availability of each selected study room exactly 24 hours in advance to standardize our data collection. Another established criterion will be only marking a room as

available if it is available for the whole 1-hour time block. For example, if 10 AM- 11 AM is part of our secondary sampling unit, a team member is clear to only mark the room available if it is free from 10:00-11:00, not 10:30-11:00 or 10:30-11:30. To standardize data collection, we will only rely on the data provided to us on the UNC Libraries website. Checking each room in person is unfeasible given the time constraints. Moreover, each member of our group is not able to check the availability of the rooms between 10:30 AM and 12:30 PM on any weekdays because of class, which would limit our sample since no room can be checked in-person between 10 AM and 1 PM.

Our primary outcome, room availability, will be measured using a binary variable, where available is 1 and unavailable is 0. We plan to collect the data starting March 20th. Each team member will be assigned a weekday to do data collection for the next 5 days, and they will check the website every hour for the study space availability in HSL, Davis, and the UL for the following day.

The data collection team will use Microsoft Excel to enter data. In each column, members will enter the date, time of day, room, and availability. For quality purposes, we will do a pilot training session to ensure consistency across each member's data collection. By checking 24 hours before instead of measuring one day altogether, we are increasing the quality of our data collection because we are checking at a consistent rate for each sampled hour interval to maintain fairness. To estimate the range of available rooms 24 hours prior, we did a pilot study to check the number of all available rooms from 11 am to 12 pm on Thursday, which is exactly 24 hours from the Wednesday class time. We observed 7 available reservable rooms for 1 hour among all 3 major libraries. Since we are three days before spring break and students have already started to leave the campus, we assumed the number of available rooms would be on the higher end.

There are potential sources of measurement error in our study. There may be discrepancies between library reservations and whether or not students show up for their reservations. Thus, rooms may seem full when they are really empty. Alternatively, empty rooms may be taken by students who did not make a formal reservation. To minimize this error, we are utilizing the fact that many study rooms require one card access of the room reserver in order to unlock it. This will decrease the ability for people to enter rooms that are reserved by others. Additionally, we are counting on the reliability of the UNC website updating reservations on time. To minimize this, we practiced accessing the UNC library reservation website to make sure none of us had prior technical difficulties with it.

## Appendix

Data Collection Form
Data Collection on Google Sheets