

# STAT446 Project Proposal

Mark Braun, Harley Clifton

**Collecting the Data** In our sampling plan, individual units will be the 1600 5m by 5m quadrats enclosed in the university's agricultural research station. Since we have information about the severity types of each quadrat, we will stratify the region into four strata which correspond to the different categorical levels of severity: low (1), moderate (2), high (3), and very high (4). Each stratum has a different number of quadrats. In order to sample from each stratum, the units (quadrats) in each stratum will be numbered from one to that stratum's size. The first unit in each stratum will be the most northwest quadrat, which should be in the leftmost position in the first row in which that severity type appears. The numbering will continue by moving to the right across the rows, marking each subsequent unit in that stratum (severity type) with succeeding integers. Once no quadrats in the stratum remain in that row, the numbering will continue with the row directly below it, starting again at the leftmost unit in that stratum. Units not in that severity type will be skipped. This will continue until all quadrats of that severity type are numbered, which should match the respective stratum total.

A simple random sample will be taken from each stratum with the use of a random number generator. The units will be sampled without replacement to get unique pairs of data with nitrogen levels and number of Gudgrass plants. As discussed below, the number of units sampled from each stratum was determined based on cost and prior information about variability. This allows every sample within a stratum to have an equal chance of being selected, which results in a simple random sample.

**Sample Size Determination** The sample size from each stratum was calculated based on the sampling budget of \$7200 (after accounting for fixed costs included in the overall budget), the variable costs of sampling from each stratum, and prior information about variability within strata. If strata are expected to have larger variability, collecting a larger sample would allow for a more precise estimate of the mean number of Gudgrass plants per quadrat. However, cost also had to be taken into consideration, as collecting a sample from certain strata was more expensive than collecting a sample from another. The calculations used to arrive at these stratum sample sizes are attached in a separate file. The sample sizes we propose for the 4 strata are:

Stratum	$h_1$	$h_2$	$h_3$	$h_4$
Sample Size	$n = 62$	$n = 37$	$n = 28$	$n = 7$

The overall cost of this sampling plan is under our budget by \$10 with a total cost of \$7190.

**Plans for Analysis** Once we have the data on the number of Gudgrass plants and nitrogen levels for each sampling unit, we plan to use regression analysis to find the overall mean number of Gudgrass plants per quadrant at the mean nitrogen level. If we don't discover a significant relationship between nitrogen and Gudgrass plants with regression analysis, we plan to revise our analysis strategy. In this case, we will use a basic stratified SRS analysis strategy - find mean estimates for each of the strata and weight them based on stratum population size while summing them, resulting in an overall estimated mean count of Gudgrass plants per quadrat.