Introduction

The goal of this project is to compare the performances of three fertilizers in terms of their impact on wheat yield, with the aim of recommending the optimal fertilizer for future use on the farm. We will try to answer following three questions using Bayesian Statistics:

Q1: How do fertilisers fare against each other on average?

Q2: What is the probability for each fertiliser that it will on average give the largest yield (larger than the other two)?

Q3: What is the probability that a particular fertiliser will give the largest yield next year?

Methods

Data

In this homework, crop.csv dataset was used. The dataset contains information on used fertilizer (1, 2 or 3) and the corresponding yield measured in tons. Notably, there are 32 data points for each of the first two fertilizers and 7 data points for the third fertilizer.

Model

We assumed that the yields from each of the three fertilizers follow a normal distribution with different means and standard deviations. To represent this, three separate normal distribution models have been developed, one for each fertilizer, utilizing the Stan programming language.

$$y_i|\mu_i,\sigma_i \sim \mathcal{N}(\mu_i,\sigma_i) \tag{1}$$

where y corresponds to random variable (yield), while μ and σ are mean and standard deviation of the normal distribution.

Computation

We fitted normal model on the data of each fertilizer with the default set-up of 1000 warm-up and 1000 sampling iterations and performed convergence diagnostics. To answer Q1 and Q2 from the introduction part, we needed only means of the posterior distributions. Probability that one fertilizer outperforms the other (Q1) was computed for each two pairs of fertilizers, by comparing means of their posterior distributions. Similarly, the probability for each fertilizer that it will on average five the largest yield (Q2) was computed by comparing means of its posterior distribution with the means of other two fertilizers distributions. The probability that a particular fertilizer will give the largest yield next year (Q3) is influenced not only by the means of the posterior distributions but by the entire distribution. In this case we extracted samples from the posterior distributions with different parameters of each fertilizer with and subsequently compared those samples. Monte Carlo Standard Error (MCSE) was used for quantifying the uncertainty of our estimates.

Results

Diagnostics, including Gelman-Rubin (R-hat), trace plots, and effective sample size, all indicated satisfactory results, affirming that MCMC chains have effectively explored parameter space and converged to the posterior distribution.

Table 1 contains the probabilities, along with their associated MCSE, for pairwise comparisons of fertilizers in terms of average yield (Q1), and for the comparison of each fertilizer average yield with the other two (Q2). Fertilizer 3 has the highest probability (99.07%) to on average give the largest yield, with the MCSE of 0.31%. We can come to the same conclusion by looking at the Figure 1(a), that shows distribution of mean yields for each fertilizer.

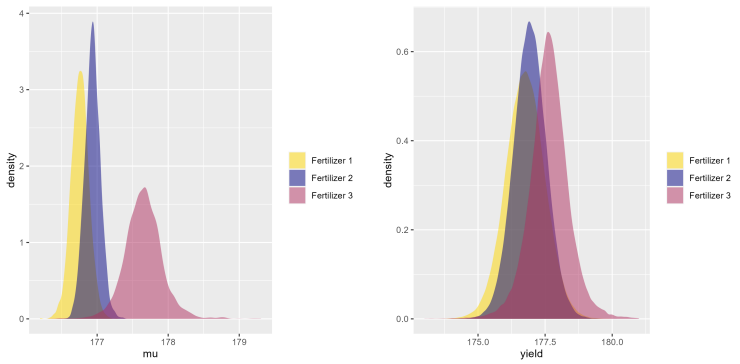
Table 1. Comparative probabilities of fertilizer average yields.

Q1: Pairwise comparison		**Q2: Largest yield probability**	
	P ± MCSE		P ± MCSE
$\mu_2 > \mu_1$	$(85.53 \pm 0.66)\%$	fertilizer 1	$(0.10 \pm 0.05)\%$
$\mu_3 > \mu_1$	$(99.68 \pm 0.17)\%$	fertilizer 2	$(0.83 \pm 0.22)\%$
$\mu_3 > \mu_2$	$(99.10 \pm 0.31)\%$	fertilizer 3	$(99.07 \pm 0.31)\%$

In Table 2, the response to the third question (Q3) is presented, i.e., the probability for each fertilizer to have the highest yield next year along with the corresponding MCSE. Fertilizer 3 has the highest probability (68.99%) to give the largest yield next year, with MCSE of 0.27%. The Figure 1(b) confirms our findings.

Table 2. Probability that a fertilizer will give the largest yield next year.

Q3	P ± MCSE
fertilizer 1	$(14.23 \pm 0.18)\%$
fertilizer 2	$(16.78 \pm 0.20)\%$
fertilizer 3	$(68.99 \pm 0.27)\%$



(a) Distributions of mean yields for each fertilizer

(b) Distributions of draws from posteriors

Figure 1

Discussion

All obtained results indicate that fertilizer 3 is by far the most favorable option for future wheat cultivation in terms of yield, while fertilizer 1 has shown to be the least favorable option. Additionally, we noticed that the results for fertilizer 3 are accompanied by higher uncertainty. This fertilizer was only recently introduced, resulting in a limited dataset available for analysis.