**STA 141A Fundamentals of Statistical Data Science - Homework 2**

**Introduction:**

The report presents the results of a Phase 1/2/3, placebo-controlled, randomized, observer-blind, dose-finding study that aims to evaluate the safety, tolerability, immunogenicity, and efficacy of SARS-CoV-2 RNA vaccine candidates against COVID-19 in healthy individuals. The study seeks to estimate the probability (θ) that a subject who fell ill with COVID-19 is from the treatment group and 1-θ is from the control group. The report adopts a Bayesian approach and assumes a Beta prior distribution for θ with a = 0.5 and b = 0.5. The report provides a detailed description of the methods used to calculate the posterior probability, credible and confidence intervals, and the posterior empirical predictive density. The results of the analysis are then discussed in the conclusion.

**Methods:**

The methods used in the analysis involve defining the prior distribution, likelihood function, and posterior distribution. The prior distribution is a Beta distribution with a shape parameter of a=0.5 and b=0.5. The likelihood function is a binomial distribution with the probability of success, θ, to be estimated from the data. Specifically, the likelihood function is defined as theta^k \* (1-theta)^(n-k), where k is the number of vaccinated subjects who fell ill with COVID-19, and n is the total number of subjects who fell ill with COVID-19. The posterior distribution is obtained by multiplying the prior distribution and the likelihood function, which is then normalized to obtain the posterior distribution. The posterior distribution is defined as dbeta(theta, a+k, b+n-k).

**Results:**

The plot of the prior, likelihood, and posterior distributions shows that the posterior distribution has shifted towards higher values of θ compared to the prior distribution, indicating that the data provides some evidence in favor of the vaccine being effective. The posterior probability of having a value of θ > 0.4118 is very low (5.826008e-11%), indicating strong evidence against this value of θ. The 95% credible interval for θ is approximately [0.0410614, 0.1541568], which means that there is a 95% probability that the true value of θ falls within this interval. Similarly, the 95% confidence interval for θ is approximately [0.02869607, 0.14151670], indicating that there is a 95% probability that the true value of θ falls within this interval. The posterior empirical predictive density plot shows that it is unlikely that a large number of subjects in a new sample of 94 subjects with COVID-19 would be vaccinated. The expected number of vaccinated subjects in a new sample of 94 subjects with COVID-19 is approximately 8, based on the posterior predictive distribution.

**Conclusion:**

In conclusion, the analysis suggests that the vaccine is effective. The posterior probability of having a value of θ > 0.4118 is very low, providing strong evidence against this value of θ. The credible and confidence intervals for θ suggest that the vaccine's efficacy is likely to be between 0.0410614 and 0.1541568 and 0.02869607 and 0.14151670, respectively. The posterior empirical predictive density plot indicates that it is unlikely that a large number of subjects in a new sample of 94 subjects with COVID-19 would be vaccinated. The expected number of vaccinated subjects in a new sample of 94 subjects with COVID-19 is approximately 8, based on the posterior predictive distribution. This Bayesian analysis provides a framework for estimating the probability of vaccine efficacy and can be used in future studies to evaluate the effectiveness of vaccines against COVID-19.

**Appendix:**

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated