Stat251 Final Project

Emily Liu

3/25/2022

Introduction

Alzheimer's is a brain disease where cells degenerate and cause memory loss. 40 million people worldwide suffer from this disease and a cure does not exist. Although there is no definitive cause of Alzheimer's, scientists speculate that genetics, aging, and environmental influences may affect the probability of developing the disease. Some specialists have found that the larger the brain, the more the brain may combat against the effects of cognitive atrophy. The total intracranial volume (TIV) is a way to quantify the size of the brain. TIV includes the volume of the cranium, brain, and spinal fluid. To discover how brain size relates to Alzheimer's, we will investigate the average TIV for patients 60 years and older of those with and without the disease. The parameter of interest is the average TIV for demented and non-demented patients.

Methods

Figure 1 is a plot of the TIV values for demented and nondemented groups. We see that these values have a uniform and bell-shaped distribution therefore, we use a normal distribution with parameters μ and σ^2 as the likelihood to model our data. The likelihood of our model is also listed below.

```r
# histogram for demented patients
ggplot(demented, aes(x = eTIV)) +
  geom_histogram() +
  theme_bw() +
  labs(title = 'Histogram of TIV for Demented Patients')

# histogram for nondemented patients
ggplot(nondemented, aes(x = eTIV)) +
  geom_histogram() +
  theme_bw() +
  labs(title = 'Histogram of TIV for Nondemented Patients')
```

$x_i|\mu, \sigma^2 \sim N(\mu, \sigma^2),\ i = 1, ...n$

We assume that the prior parameter μ which represents that average TIV for both populations, is normally distributed. We also assume that the variance denoted by σ^2 follows an inverse gamma distribution because the variance of TIV for both populations is positive and right-skewed. The distributions of the prior distributions for our parameters are listed below. We assume the prior distributions for both the demented (D) and nondemented (N) populations to be the same.

Prior distribution for demented popluation:

$\mu_D \sim N(\lambda, tau^2)$

$\sigma_D \sim IG(\gamma, \phi)$

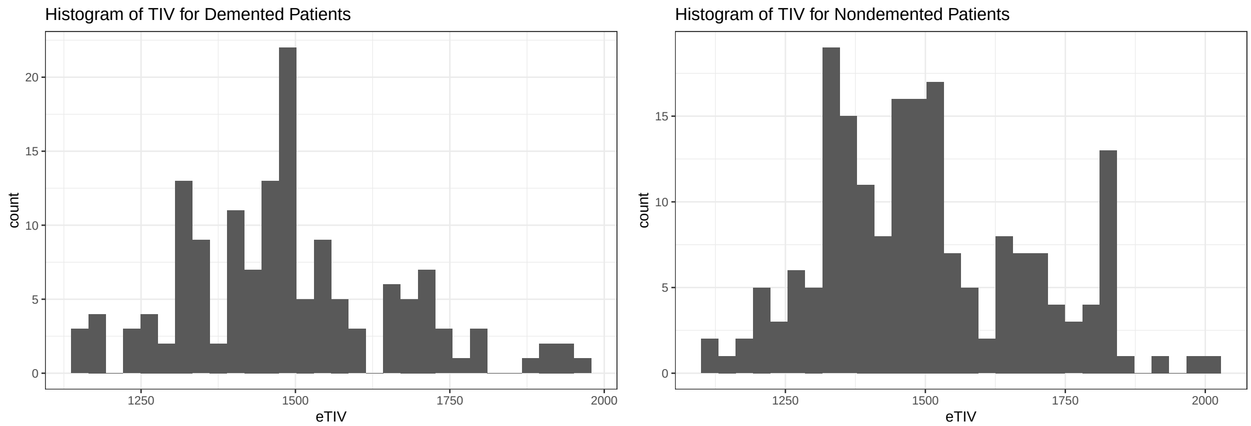


Figure 1: Histograms of TIV for Alzheimers patients.

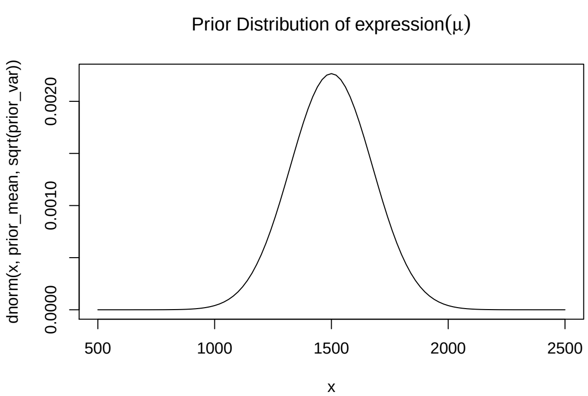
Prior distribution for non-demented population:

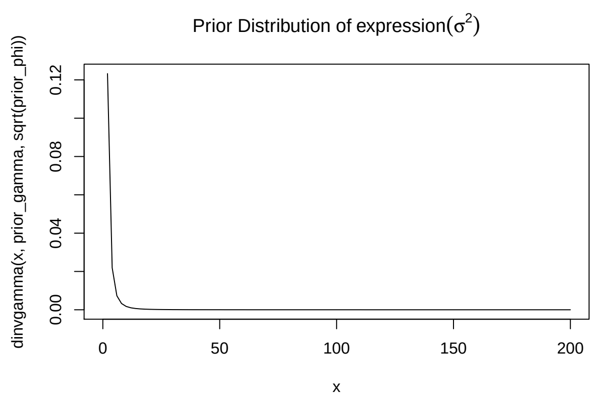
$\mu_N \sim N(\lambda, tau^2)$

$\sigma_N \sim IG(\gamma, \phi)$

```
# plot prior distribution of mu
curve(dnorm(x, prior_mean, sqrt(prior_var)), xlim = c(500, 2500),
      main = 'Prior Distribution of' ~ expression(mu))

# plot prior distribution of sigma2
curve(dinvgamma(x, prior_gamma, sqrt(prior_phi)), xlim = c(0, 200),
      main = 'Prior Distribution of' ~ expression(sigma^2))
```





In order to investigate this question, we analyzed a dataset taken from Kaggle that includes data from a sampled set of patients 60 years and older. Metrics collected on these individuals include whether they have Alzheimer's (our response variable), gender, age, education level, TIV measurements (our explanatory variable), and other physiological characteristics. The outcome variable of interest is whether the patient is demented or non-demented.

??put in kable

Results

Because of our likelihood and prior distributions, we can approximate a posterior distribution for both populations μ and σ^2 with Gibbs sampling.

The mean and variance of our posterior distribution for demented patients are 2885.4092 and 1.0102993×10^9 respectively.

The mean and variance of our posterior distribution for demented patients are 2834.1066 and 1.1783687×10^9 respectively.

We want to discover the difference between means

??Given our data, there is a 95% probability that the true difference between the average TIV for demented and non-demented patients is between -6683.0685 and 7304.5412.

We want to discover the difference between standard deviations