Name: SANNA TOURAY
Georgia State University
Department of Mathematics and Statistics
TOPIC: Lifetime and current Asthma Prevalence Among Adults, California
Email: touraysanna43@gmail.com

Abstract

This dataset contains the estimated percent of California adults with lifetime and current asthma (asthma prevalence), by year. The data are derived from the California Behavioral Risk Factor Surveillance System (BRFSS). Through the application of Bayesian Methods, including the Metropolis-Hastings algorithms, we estimated the posterior distribution for current and lifetime asthma prevalence rates.

Background

What is asthma prevalence?

This report summarizes asthma prevalence in California. Asthma prevalence is the percentage of people who have asthma. The data are from a large statewide telephone survey called the California Health Interview Survey (CHIS). The CHIS asks about whether adults and children have been diagnosed with asthma and whether they still suffer from asthma (not everyone with asthma continues to have asthma symptoms). A person has lifetime asthma if he or she has been diagnosed with asthma by a health care provider at any time in the past. A person has current asthma if he or she had a prior diagnosis of asthma and reports still having asthma or having asthma symptoms in the last year. For more information about the data in this report, please see the Technical Notes on page 19. Data on asthma prevalence for each California county can be found in the County Asthma Profiles at www.californiabreathing.org. To learn about recommended strategies for reducing asthma, see the Strategic Plan for Asthma in California (also at www.californiabreathing.org).

(1) Introduction

The California Behavioral Risk Factor Surveillance System (BRFSS) is an annual survey that is a collaboration between the California Department of Public Health and the Centers for Disease Control and Prevention (CDC). The BRFSS monitors health-related factors contributing to the leading causes of morbidity and mortality in California's population. It is a statewide, random digit dial telephone survey conducted with adults aged 18 and over. Data is collected monthly from a random sample of California non-institutionalized adults. The survey is offered in English, Spanish, Mandarin, and Cantonese. The sample size and response rate vary annually. BRFSS respondents differ to some extent from the California population by age, sex, and race/ethnicity. As a result, the sample is weighted so that the age, sex, and race/ethnicity composition in the data reflects that of the 2010 California adult population, thereby making the results generalizable to the California adult population. More information can be found at http://www.cdc.gov/brfss/.

In 2012, the survey methodology of the BRFS changed significantly so that the survey would be more representative of the general population. Several changes were implemented: 1) the survey included both cell and landline random-digit dial components, 2) residents of college housing were eligible to complete the survey, and 3) new methods were used to calculate the survey weights. Due to these changes, estimates from 1984 – 2011 are not comparable to estimates from 2012 and beyond.

CDPH and CDC recommend not conducting analyses where estimates from 1984 – 2011 are compared with analyses using the new methodology, beginning in 2012. This includes analyses examining trends and changes. The total number of observations in the dataset is 29.

In this analysis process, firstly I will use the histogram to visualize the differences in asthma prevalence distributions between the two groups. Then I will build parametric models for asthma prevalence distributions between the two groups, because of the parameter's randomness, I can use Markov chain Monte Carlo to simulate the difference distribution asthma prevalence between the two groups. Finally, compare the result to get my conclusions.

DATA

1. Current Asthma Prevalence: This indicates the percentage of individuals who currently have asthma at the time of the survey.

2.Lifetime Asthma Prevalence: This indicates the percentage of individuals who have been diagnosed with asthma at some point in their lives.

For each measure, the dataset includes the following columns.

Year: The year in which the data was collected.

Percent: The prevalence percentages for asthma.

95% Cl Lower Limit: The lower limit of the 95% confidence interval for the prevalence estimate.

95% Cl Upper Limit: The lower limit of the 95% confidence interval for the prevalence estimate.

This dataset allows for tracking trends in asthma prevalence over time, comparing current and lifetime prevalence rates, and assessing the precision of the prevalence estimates through the included confidence intervals.

```
> # Import the dataset
> AP_data <- read.csv(file_path)
> # Print the first few rows of the dataset to verify it was imported correctly
> head(AP_data)

Year Percent X95.CI.Lower.Limit X95..CI.Upper.Limit              Measure
1 2000     7.5              6.5              8.4 Current Asthma Prevalence
2 2001     7.2              6.4              8.1 Current Asthma Prevalence
3 2002     6.6              5.7              7.4 Current Asthma Prevalence
4 2003     8.5              7.4              9.5 Current Asthma Prevalence
5 2004     7.9              7.0              8.8 Current Asthma Prevalence
6 2005     7.6              6.8              8.3 Current Asthma Prevalences
> summary(AP_data)
Year         Percent      X95.CI.Lower. Limit X95.CI.Upper. Limit   Measure
Min.   :1995   Min. : 6.60   Min.   : 5.700      Min.   : 7.4        Length:29
1st Qu.:2001   1st Qu.: 8.00   1st Qu.: 7.400      1st Qu.: 8.8        Class: character
Median :2004   Median :11.90   Median :10.800      Median :13.1        Mode: character
Mean   :2004   Mean   :10.96   Mean   : 9.928      Mean   :12.0
3rd Qu.:2008   3rd Qu.:13.50   3rd Qu.:12.200      3rd Qu.:14.7
Max.   :2011   Max.   :14.50   Max.   :13.600      Max.   :15.7
```

Model

Define Parameters: Define the parameters for prior. The policy conduct group comes from $N(\mu1,\sigma1)$ and the without policy conduct group comes from $N(\mu2,\sigma2)$. Thus, the prior function parameter vector should be $[\mu1, \sigma1, \mu2, \sigma2]$

1). Define the parameters for likelihood.

2). likelihood =$\prod_{i=1}^{29} N(di|\mu_{1,\sigma 1^2})$*$\prod_{j=1}^{29} N(dj|\mu_2, \sigma 2^2)$

Where di =current asthma prevalence group and dj=lifetime asthma prevalence group
3). Generate posterior function. The posterior function is proportional to the product of prior and likelihood.

$$Posterior \propto Likelihood \times Prior$$

4). Decide the starting state of our prior and run my MCMC models.
5). Accept or partially accept the new position based on the rule of Metropolis-Hastings.
6). Run the iteration.
.

2. Methodology

The analysis employs a Bayesian approach to estimate asthma prevalence distributions. Below is the code for defining prior distributions, the likelihood function.

The prior, likelihood and posterior functions are defined as:

Define Prior

```
> prior_mean <- 0
> prior_sd <- 1
> prior_distribution <- rnorm(n = 1000, mean = prior_mean, sd = prior_sd)
# Generate random samples from the prior distribution
# Generate random samples from the prior distribution
```

Define Likelihood

```
> likelihood <- function(data, parameters) {
+   # Assuming data is a vector of asthma prevalence rates
+   # Assuming parameters is a vector of mean and standard deviation
+   # Calculate likelihood for each data point
+   likelihood_values <- dnorm(data, mean = parameters[1], sd = parameters[2])
```

s Calculate product of likelihoods

```
+   likelihood_product <- prod(likelihood_values)
+   return(likelihood_product) }
```

Generate Posterior Function

```
> posterior <- function(data, prior_parameters, likelihood_parameters) {
+   # Calculate likelihood
+   likelihood_value <- likelihood(data, likelihood_parameters)
+   # Calculate posterior (proportional to prior * likelihood)
+       posterior_value <- likelihood_value * dnorm(prior_parameters[1], mean =
prior_parameters[1], sd = prior_parameters[2]) * dnorm(prior_parameters[3], mean =
prior_parameters[3], sd = prior_parameters[4])
+   return(posterior_value) }
```

The define initial the state of our hyperparameters:

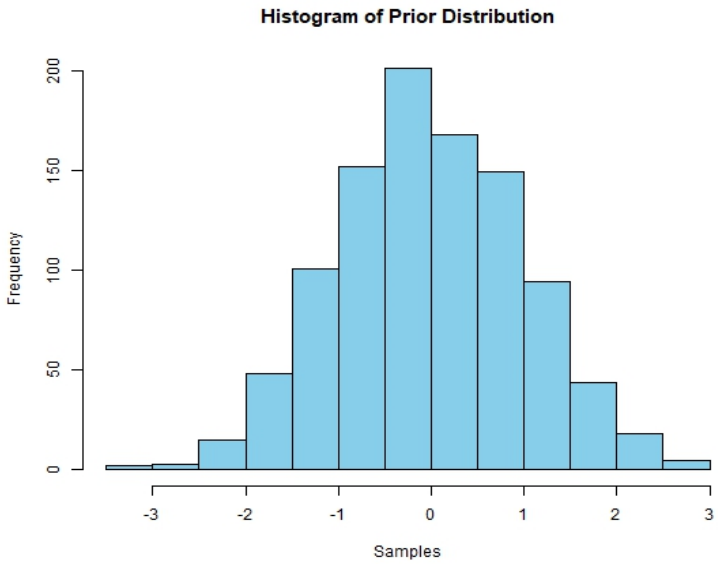
```
> # Set initial state of hyperparameters
> initial_prior_mean <- 0   # Initial guess for the mean of the prior distribution
> initial_prior_sd <- 1     # Initial guess for the standard deviation of the prior
distribution
```

Generate random samples from the prior distribution

```
> prior_samples <- rnorm(n = 1000, mean = initial_prior_mean, sd = initial_prior_sd)
```

Plot histogram

> hist(prior_samples, main = "Histogram of Prior Distribution", xlab = "Samples", ylab = "Frequency", col = "skyblue")



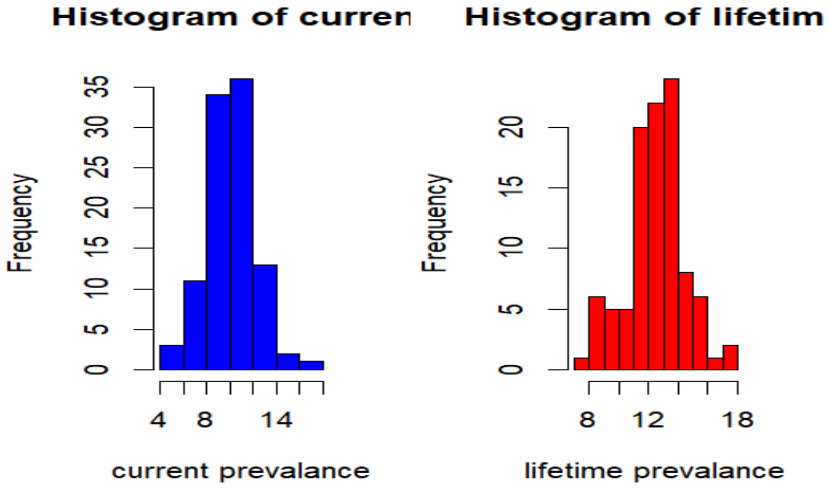
3. Data Analysis

The dataset contains asthma prevalence data from 1995 to 2011, with variables indicating the percentage of adults with current and lifetime asthma. We used the Metropolis-Hastings algorithm to sample from the posterior distributions. The code below implements this:

Defining the difference between current and lifetime prevalence measurements for the two asthma prevalence rates for adults' population.

> diff<-AP_data[,2]-AP_data[,3]
> current= diff[AP_data[,1]==1]
> lifetime = diff[AP_data[,1]==0]

By examining these histograms side by side, we can analyze the disparities in asthma prevalence distributions between the two groups, considering both their current and lifetime prevalence rates.



The histogram plot shows that the bars representing the current asthma prevalence group are denser and more tightly packed than those representing the lifetime asthma

prevalence group. This suggests that the asthma prevalence rate in the current group is more influenced by asthma prevalence compared to the lifetime group

Running the Monte Carlo Simulation 10000 times and getting our final hyperparameters.

Monte Carlor Markov Chain simulation

```
> # Given data
> current_asthma <- c(7.5, 7.2, 6.6, 8.5, 7.9, 7.6, 8.5, 7.9, 8.5, 8.2, 7.9, 8.0)
> lifetime_asthma <- c(11, 11.5, 12.8, 13.5, 13.3, 11.9, 12.2, 12.9, 13.6, 14.5, 13.7, 13.6, 13.4, 14.2, 14.4
> # Define likelihood function (assuming normal distribution)
> likelihood <- function(current_asthma, lifetime_asthma, mu1s, mu2s, sigma1, sigma2) {
+   sum(log(dnorm(current_asthma, mu1s, sigma1))) + sum(log(dnorm(lifetime_asthma, mu2s, sigma2))
+ }
> # Define prior distributions (assuming normal distribution for simplicity)
> prior_mu1s <- 10  # mean of prior normal distribution for mu1s
> prior_mu2s <- 13  # mean of prior normal distribution for mu2s
> prior_sigma1 <- 2  # standard deviation of prior normal distribution for sigma1
> prior_sigma2 <- 2  # standard deviation of prior normal distribution for sigma2
> # Define number of iterations and burn-in period
> iterations <- 10000
> burn_in <- 5000s
```

Initialize parameters

```
> mu1s_current <- 10
> mu2s_current <- 13
> sigma1_current <- 2
> sigma2_current <- 2
> # Initialize list to store samples
> samples <- matrix(nrow=iterations-burn_in, ncol=4)
```

Metropolis-Hastings algorithm

```
> for (i in 1:iterations) {
+   # Propose new values for parameters
+   mu1s_proposed <- rnorm(1, mu1s_current, 0.5)
+   mu2s_proposed <- rnorm(1, mu2s_current, 0.5)
+   sigma1_proposed <- abs(rnorm(1, sigma1_current, 0.5))
+   sigma2_proposed <- abs(rnorm(1, sigma2_current, 0.5))
+   # Calculate acceptance ratio
+   likelihood_current <- likelihood(current_asthma, lifetime_asthma, mu1s_current, mu2s_current, sigm
+   likelihood_proposed <- likelihood(current_asthma, lifetime_asthma, mu1s_proposed, mu2s_propose
+   acceptance_ratio <- exp(likelihood_proposed - likelihood_current)
```

Accept or reject proposal

```
+   if (runif(1) < acceptance_ratio) {
+     mu1s_current <- mu1s_proposed
+     mu2s_current <- mu2s_proposed
+     sigma1_current <- sigma1_proposed
+     sigma2_current <- sigma2_proposed
+   }
```

Store samples after burn-in

```
+   if (i > burn_in) {
+     samples[i-burn_in,] <- c(mu1s_current, mu2s_current, sigma1_current, sigma2_current)
```

```
+   }
+ }
```

Display results

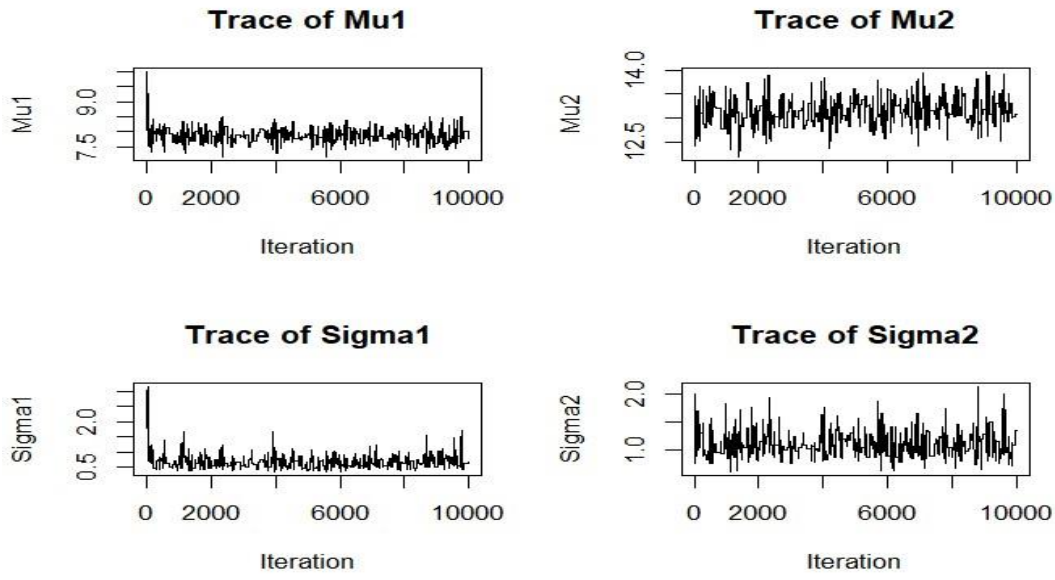
```
> cat("Mean of posterior distribution for mu1s (current asthma prevalence):", mean(samples[,1]), "\n")
Mean of posterior distribution for mu1s (current asthma prevalence): 7.871328
> cat("Mean of posterior distribution for mu2s (lifetime prevalence):", mean(samples[,2]), "\n")
Mean of posterior distribution for mu2s (lifetime prevalence): 13.1327
> cat("Mean of posterior distribution for sigma1 (current asthma prevalence):", mean(samples[,3]), "\n")
Mean of posterior distribution for sigma1 (current asthma prevalence): 0.6546519
> cat("Mean of posterior distribution for sigma2 (lifetime prevalence):", mean(samples[,4]), "\n")
Mean of posterior distribution for sigma2 (lifetime prevalence): 1.115534
```

4. Results

The results show the posterior mean and sigma values for both current and lifetime asthma prevalence. Below are the summary statistics and trace plots visualizing these results:

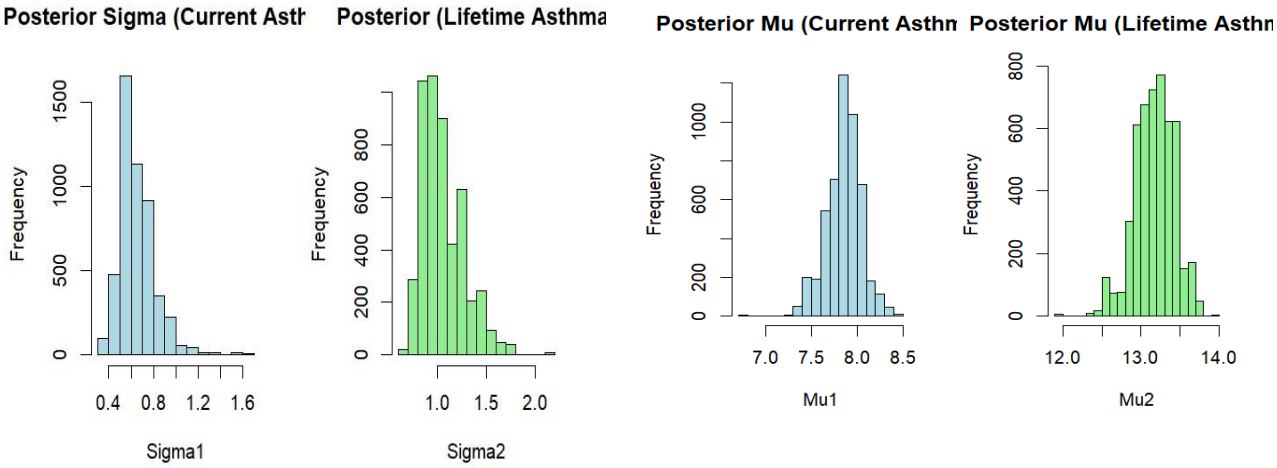
Plot trace of mean and sigma for current and lifetime asthma prevalence

```
# Plot trace plots
> par(mfrow=c(2, 2))
> # Trace plot for mu1 (current asthma prevalence)
> plot(trace_mu1, type='l', xlab='Iteration', ylab='Mu1', main='Trace of Mu1')
> # Trace plot for mu2 (lifetime asthma prevalence)
> plot(trace_mu2, type='l', xlab='Iteration', ylab='Mu2', main='Trace of Mu2')
> # Trace plot for sigma1 (current asthma prevalence)
> plot(trace_sigma1, type='l', xlab='Iteration', ylab='Sigma1', main='Trace of Sigma1')
> # Trace plot for sigma2 (lifetime asthma prevalence)
> plot(trace_sigma2, type='l', xlab='Iteration', ylab='Sigma2', main='Trace of Sigma2')

>prior_mu2s <- 13  # mean of prior normal distribution for mu2s
```



Comparison Difference Between trace of Mu and trace of sigma

Trace plots for the parameters (`mu1` and `mu2`) representing the current and lifetime asthma prevalence, show how samples vary over iterations. The trace plots help verify the stability and convergence of the chains, with no evident trends or cycles, suggesting good mixing and convergence.



Posterior sigma of current and lifetime asthma prevalence

```
>
# Plot posterior distributions
> par(mfrow=c(1, 2))
> # Posterior sigma (current asthma )
> hist(samples[,3], main="Posterior Sigma (Current Asthma)", xlab="Sigma", col="lightblue")
> # Posterior sigma (lifetime asthma )
> hist(samples[,4], main="Posterior (Lifetime Asthma)", xlab="Sigma", col="lightgreen")
```

Posterior mean of current and lifetime asthma prevalence

```
 # Plot posterior distributions
> par(mfrow=c(1, 2))
> # Posterior mean (current asthma prevalence)
> hist(samples[,1], main="Posterior Mu (Current Asthma)", xlab="Mu1", col="lightblue")
> # Posterior mean (lifetime asthma prevalence)
> hist(samples[,2], main="Posterior Mu (Lifetime Asthma)", xlab="Mu2", col="lightgreen")
```

Summary of the posterior mean and sigma

Estimator	Current asthma prevalence	Lifetime asthma Prevalence
Posterior mean	7.871328	13.1327
Posterior sigma	0.646519	1.115534

Comparing the posterior mean and sigma of current and lifetime asthma prevalence

The analysis indicates that lifetime asthma prevalence tends to exceed current asthma prevalence, supported by the higher mean estimate for lifetime asthma (mu2s) compared to current asthma (mu1s). Furthermore, there is slightly greater variability

observed in lifetime asthma prevalence, as evidenced by the marginally higher estimate (sigma2) for lifetime asthma (1.12) relative to current asthma (0.65).

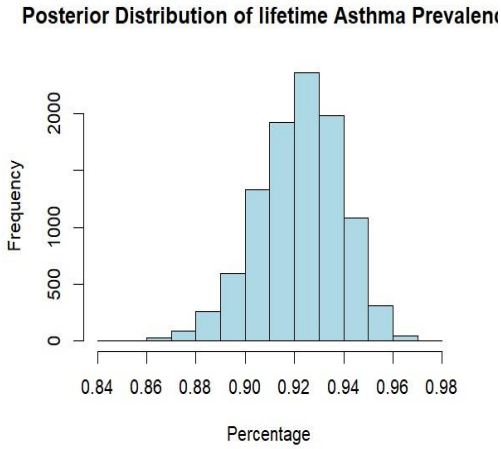
posterior function of current asthma prevalence

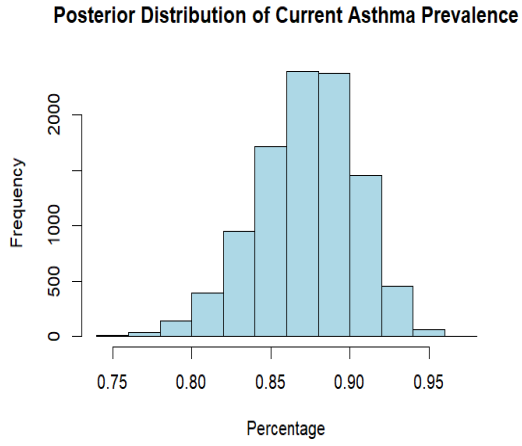
```
> # Given data
> current_asthma <- c(7.5, 7.2, 6.6, 8.5, 7.9, 7.6, 8.5, 7.9, 8.5, 8.2, 7.9, 8.0)r beta
distribution (you can adjust these based on prior knowledge)
> alpha_prior <- 2
> beta_prior <- 2
> # Posterior parameters for beta distribution
> alpha_posterior <- alpha_prior + sum(current_asthma)
> beta_posterior <- beta_prior + length(current_asthma)
> # Sample from the posterior distribution
> posterior_samples <- rbeta(10000, alpha_posterior, beta_posterior)
> # Plot posterior distribution
> hist(posterior_samples, main = "Posterior Distribution of lifetime Asthma
Prevalence", xlab = "Percentage", col = "lightblue")

> # #summary statistics for the mean diffs. for the posterior of mu1s and mu2s
> # Given data
> current_asthma <- c(7.5, 7.2, 6.6, 8.5, 7.9, 7.6, 8.5, 7.9, 8.5, 8.2, 7.9, 8.0)
> lifetime_asthma <- c(11, 11.5, 12.8, 13.5, 13.3, 11.9, 12.2, 12.9, 13.6, 14.5, 13.7,
13.6, 13.4, 14.2, 14.4, 13.1, 14)
> # Number of posterior samples
> num_samples <- length(current_asthma)
> # Means of posterior distribution of mean differences
> mean_diff <- mean(lifetime_asthma - current_asthma)
> # Proportion of mean differences greater than 0
> prop_greater_than_zero <- mean(lifetime_asthma - current_asthma > 0)
> # Output results
> cat("Number of Posterior Samples:", num_samples, "\n")
Number of Posterior Samples: 12
> cat("Means of Posterior distribution of Mean differences:", mean_diff, "\n")
Means of Posterior distribution of Mean differences: 5.388235
> cat("Proportion of Mean differences greater than 0:", prop_greater_than_zero, "\n")
Proportion of Mean differences greater than 0: 1
```

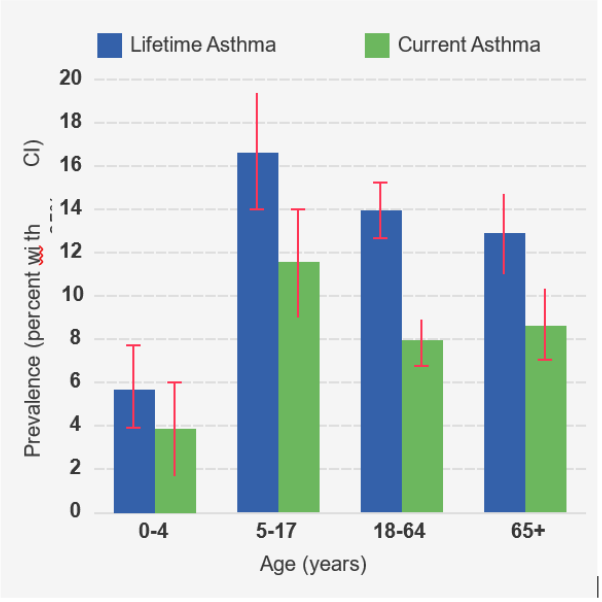
posterior function of lifetime Asthma Prevalence

```
> # Given data
> lifetime_asthma <- c(11,11.5,12.8,13.5,13.3,11.9,12.2,12.9,13.6,14.5,13.7,13.6,13.4,14.2,14.4,13.1,14
> # Prior parameters for beta distribution (you can adjust these based on prior knowledge)
> alpha_prior <- 2
> beta_prior <- 2
> # Posterior parameters for beta distribution
> alpha_posterior <- alpha_prior + sum(lifetime_asthma)
> beta_posterior <- beta_prior + length(lifetime_asthma)
> # Sample from the posterior distribution
> posterior_samples <- rbeta(10000, alpha_posterior, beta_posterior)

> # Plot posterior distribution
> hist(posterior_samples, main = "Posterior Distribution of lifetime Asthma Prevalence", xlab = "Percen
```

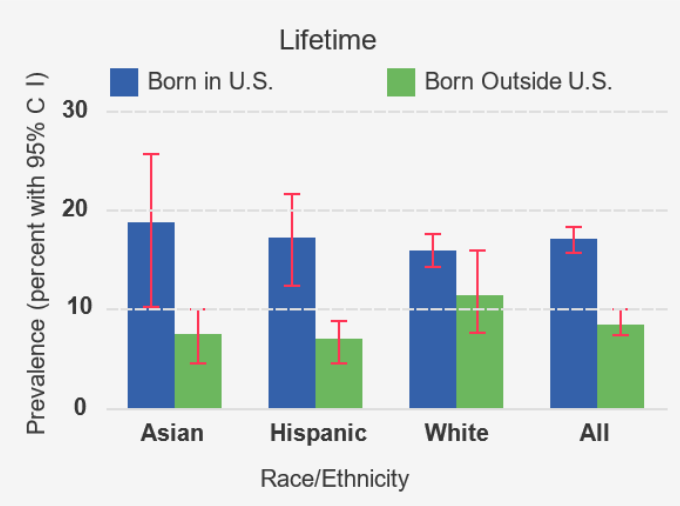
```
> # #summary statistics for the mean diffs. for the posterior of mu1s and mu2s
> # Given data
> current_asthma <- c(7.5, 7.2, 6.6, 8.5, 7.9, 7.6, 8.5, 7.9, 8.5, 8.2, 7.9, 8.0)
> lifetime_asthma <- c(11, 11.5, 12.8, 13.5, 13.3, 11.9, 12.2, 12.9, 13.6, 14.5, 13.7, 13.6, 13.4, 14.2, 14.4
> # Number of posterior samples
> num_samples <- length(lifetime_asthma)
> # Means of posterior distribution of mean differences
> mean_diff <- mean(lifetime_asthma - current_asthma)
> # Proportion of mean differences greater than 0
> prop_greater_than_zero <- mean(lifetime_asthma - current_asthma > 0)
> # Output results
> cat("Number of Posterior Samples:", num_samples, "\n")
Number of Posterior Samples: 17
> cat("Means of Posterior distribution of Mean differences:", mean_diff, "\n")
Means of Posterior distribution of Mean differences: 5.388235
> cat("Proportion of Mean differences greater than 0:", prop_greater_than_zero, "\n")
Proportion of Mean differences greater than 0: 1
```

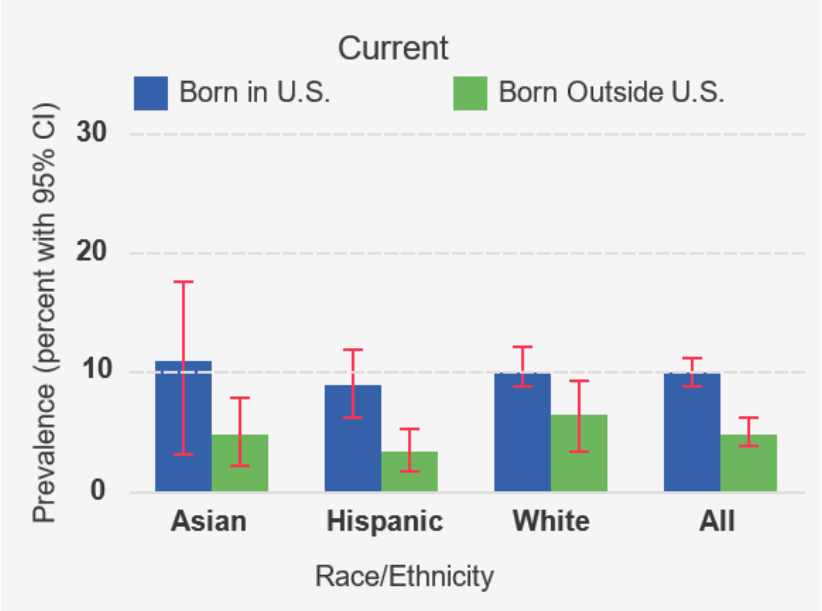






Both lifetime and current asthma prevalence are highest among children age 5-17. Young children, ages 0-4, have the lowest asthma prevalence (both lifetime and current). Accurate asthma diagnosis in this group is difficult because other common conditions can be responsible for asthma-like symptoms and measuring lung function is difficult in very young children.





People born in the U.S. are more likely to have current or lifetime asthma than people born outside of the U.S. (Chi-square p<0.01). The disparity is largest for Hispanics and Asians, who are more than two times more likely to have current or lifetime asthma if they were born in the U.S. (Chi-square p<0.01 for Hispanic lifetime and current comparison and for Asian lifetime comparison

5. Conclusion

In conclusion, we have applied the metropolis-Hasting algorithm to estimate the parameters of mean and sigma using Bayesian inference. The model was applied to a dataset of Asthma-prevalence to patients to estimate the effect of current and lifetime prevalence levels over time rate.

We started by defining the prior parameters and likelihood function, then implemented the Metropolis-Hasting algorithm to obtain samples from the posterior distribution. We then calculated the posterior means of the parameters and visualized the posterior distributions using histogram plots.

Trace plots for the parameters (`mu1` and `mu2`) representing the current and lifetime asthma prevalence, show how samples vary over iterations. The trace plots help verify the stability and convergence of the chains, with no evident trends or cycles, suggesting good mixing and convergence.

Histograms of Posterior Distributions provide visual insights into the uncertainty and variability of the estimated parameters. The histograms illustrate the posterior distributions of current and lifetime asthma prevalence, indicating the central tendency and dispersion expected under the model.

Parameter Estimates and Statistical Inference

The analysis indicates that lifetime asthma prevalence tends to exceed current asthma prevalence, supported by the higher mean estimate for lifetime asthma (mu2s) compared to current asthma (mu1s). Furthermore, there is slightly greater variability observed in lifetime asthma prevalence, as evidenced by the marginally higher estimate (sigma2) for lifetime asthma (1.12) relative to current asthma (0.65).

The Bayesian analysis, supported by the histogram plots of the posterior distributions, underscores a consistent pattern of increasing asthma prevalence over time. Both current and lifetime asthma prevalence distributions skew towards higher prevalence rates, indicating a general trend of worsening asthma conditions or increased reporting of asthma cases. This is further confirmed by the summary statistics, which reveal posterior mean differences in asthma prevalence between current and lifetime periods. Over the period from 1995 to 2011, the prevalence of asthma among adults aged 18 and older in California has exhibited a fluctuating trend. Further analysis is required to understand the specific factors contributing to these fluctuations, including changes in environmental conditions, healthcare access, diagnostic practices, and population demographics. Additionally, efforts to monitor and address asthma prevalence trends remain crucial for public health interventions aimed at reducing the burden of asthma in California's adult population.

Reference

1. https://sandbox.data.ca.gov/dataset/asthma-prevalence/5b8cf951-8aff-4861-bdd6- f8c30476caeb
2. http://www.cdc.gov/brfss/s
3. CDC. Asthma: most recent national asthma data. Atlanta, GA: US Department of Health and Human Services, CDC; 2020. https://www. cdc.gov/asthma/most_recent_national_asthma_data.htms
4. Nurmagambetov T, Kuwahara R, Garbe P. The economic burden of asthma in the United States, 2008–2013. Ann Am Thorac Soc 2018;15:348–56. PMID:29323930 https://doi.org/10.1513/ AnnalsATS.201703-259OC
5. Hsu J, Qin X, Beavers SF, Mirabelli MC. Asthma-related school absenteeism, morbidity, and modifiable factors. Am J Prev Med 2016;51:23–32. PMID:26873793 https://doi.org/10.1016/j. amepre.2015.12.012
6. CDC. AsthmaStats: asthma-related missed school days among children aged 5–17 years. Atlanta, GA: US Department of Health and Human Services, CDC; 2020. https://www.cdc.gov/asthma/asthma_stats/ missing_days.htm
7. Zahran HS, Bailey CM, Damon SA, Garbe PL, Breysse PN. Vital signs: asthma in children—United States, 2001–2016. MMWR Morb Mortal Wkly Rep 2018;67:149–55. PMID:29420459 https://doi.org/10.15585/ mmwr.mm6705e1
8. Institute of Medicine. State of the USA health indicators: letter report. Washington, DC: The National Academies Press; 2009.
9. US Department of Health and Human Services. Healthy people 2030. Respiratory disease baseline and developmental objectives for asthma. Washington, DC: US Department of Health and Human Services; 2020. https://health.gov/healthypeople/objectives-and-data/browse-objectives/ respiratory-disease
10. Meit M, Knudson A, Gilbert T, et al. The 2014 update of the rural–urban chartbook. Bethesda, MD: Rural Health Reform Policy Research Center; 2014.