

I. Introduction

Marijuana, also known as cannabis, is a psychoactive drug derived from the Cannabis plant. The plant contains delta-9-tetrahydrocannabinol(THC), a mind-altering chemical that is primarily responsible for the intoxicating effects sought by users.¹ It has been used for both medicinal and recreational purposes. People use marijuana by smoking, vaporizing, or mixing it with edibles. The abuse of Marijuana can lead to various effects such as alter perceptions, mood changes, impaired body movement, and impaired memory and learning.

According to 2021 NSDUH Annual National Report², marijuana is one of the most commonly used drugs in the United States, and its use is widespread among young people. In 2023, 29% of 12th-grade students reported using marijuana in the past year, and 6.5% reported using it daily, according to the Monitoring the Future survey.³ The survey results also demonstrate that drug use among secondary school students has remained steady since the late 1990s.

The use of marijuana during adolescence is a significant concern due to its potential effects on the developing brain. It is also linked to an increased risk of psychiatric disorders, including depression and anxiety, and can lead to substance dependency in adulthood.

II. Data

The data used in this report is from the file marijuanause.csv, which contains survey data on reported marijuana use during adolescence. The sample contains 236 rows, each of which represents a person in the sample. Each row contains 6 columns as follows:

female its value indicates the person is female(1) or male(0).
use1976 its value indicates if the person used marijuana in year 1976(1) or not(0).
use1977 its value indicates if the person used marijuana in year 1977(1) or not(0).
use1978 its value indicates if the person used marijuana in year 1978(1) or not(0).
use1979 its value indicates if the person used marijuana in year 1979(1) or not(0).
use1980 its value indicates if the person used marijuana in year 1980(1) or not(0).

Table 1 shows the total number of adolescents and female/male users that used marijuana in a particular year and their percentages of sample population they represent.

Table 1: Data Summary

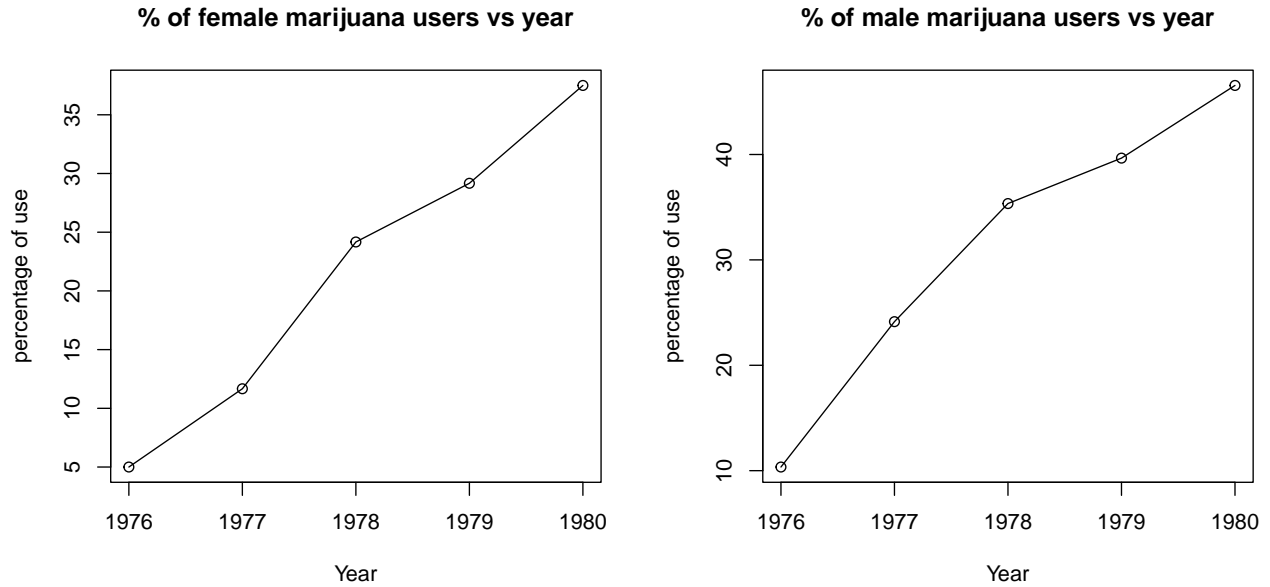
	Year	Total_users	Percentage	Female	Female_percentage	Male	Male_percentage
use1976	1976	18	7.627119	6	5.00000	12	10.34483
use1977	1977	42	17.796610	14	11.66667	28	24.13793
use1978	1978	70	29.661017	29	24.16667	41	35.34483

¹NIDA. "What is marijuana? ." National Institute on Drug Abuse, 17 Apr. 2023, <https://nida.nih.gov/publications/research-reports/marijuana/what-marijuana> Accessed 25 Apr. 2024.

²2021 NSDUH Annual National Report, <https://www.samhsa.gov/data/report/2021-nsduh-annual-national-report>.

³Miech, R. A., Johnston, L. D., Patrick, M. E., O'Malley, P. M., & Bachman, J. G. (2024). Monitoring the Future national survey results on drug use, 1975–2023: Overview and detailed results for secondary school students (PDF). Monitoring the Future Monograph Series. Ann Arbor, MI: Institute for Social Research, University of Michigan.

	Year	Total_users	Percentage	Female	Female_percentage	Male	Male_percentage
use1979	1979	81	34.322034	35	29.16667	46	39.65517
use1980	1980	99	41.949152	45	37.50000	54	46.55172



III. First Model

a.

```

model {
  for (i in 1: 236) {
    for (j in 1: 5) {
      use[i, j] ~ dbern(prob[i, j])
      logit(prob[i, j]) <- beta0 + betaYear * yearCentered[j] + betaFemale * femaleCentered[i]
      userep[i, j] ~ dbern(prob[i, j])
    }
  }

  beta0 ~ dt(0, 1/(10^2), 1)
  betaYear ~ dt(0, 1/(2.5^2), 1)
  betaFemale ~ dt(0, 1/(2.5^2), 1)
}

```

b.

The first model will run in 4 chains, 3,000 burn-in iterations, and 10,000 iterations used per chain. Gelman Rubin diagnostics and trace plots show the model reach to convergence. For each of the parameter, the effective number of parameters is sufficient.

Table 2: MCMC Summary for Model 1

Number.of.chains	Length.of.burn.in	Number.of.iterations.used.per.chain	Thinning
4	3000	10000	1

Table 3: Effective sample sizes of all parameters for Model 1

effective sample sizes	
beta0	20229.01
betaFemale	25451.98
betaYear	19735.11

c.

Table 4: Approximate posterior data for each parameter in Model 1

	Mean	SD	X2.5..quantile	X97.5..quantile
beta0	-1.1544632	0.0731993	-1.2990662	-1.0131191
betaFemale	-0.5448286	0.1397764	-0.8197678	-0.2714591
betaYear	1.5024296	0.1654687	1.1799749	1.8317042

d. The posterior probability that the coefficient for the (centered) indicator of being female exceeds zero ($Pr(\beta_{female} > 0)$) is equal to 2.5×10^{-5} . This indicates females are less likely to use marijuana than males.

e. The value of (Plummer's) DIC is approximated by using 100,000 iterations. The approximate associated effective number of parameters is 2.9956439, and the actual number of parameters is 1255.186291.

IV. Second Model

a.

```

model {
  for (i in 1:236) {
    for (j in 1:5) {
      use[i, j] ~ dbern(prob[i, j])
      logit(prob[i, j]) <- beta0 + betaYear*yearCentered[j] + betaFemale*femaleCentered[i] + epsilon[i]
      userep[i, j] ~ dbern(prob[i, j])
    }

    epsilon[i] ~ dnorm(0, 1/sigmaepsilon^2)
  }

  beta0 ~ dt(0, 1/(10^2), 1)
  betaYear ~ dt(0, 1/(2.5^2), 1)
  betaFemale ~ dt(0, 1/(2.5^2), 1)
  sigmaepsilon ~ dunif(0,10)
}

```

b. The second model runs in 4 chains, 5,000 burn-in iterations, and 30,000 iterations used per chain. Gelman Rubin diagnostics and trace plots show the model reach to convergence. For each of the parameter, the effective number of parameters is sufficient.

Table 5: MCMC Summary for Model 2

Number.of.chains	Length.of.burn.in	Number.of.iterations.used.per.chain	Thinning
4	5000	30000	1

Table 6: Effective sample sizes of all parameters for Model 2

	effective sample sizes
beta0	3250.085
betaFemale	5780.248
betaYear	8550.227
sigmaepsilon	3393.558

c.

Table 7: Approximate posterior data for each parameter in Model 2

	Mean	SD	X2.5..quantile	X97.5..quantile
beta0	-2.3981036	0.2823925	-2.981297	-1.8856917
betaFemale	-0.9995406	0.4591210	-1.923899	-0.1211583
betaYear	2.9628381	0.2872354	2.420861	3.5398470
sigmaepsilon	2.9434871	0.3083499	2.385341	3.5762956

d. The value of (Plummer's) DIC is approximated by using 100,000 iterations. The approximate associated effective number of parameters is 167.7788574, and the actual number of parameters is 674.4496793. The DIC value of model 2 is much smaller than model 1.

V. Conclusions

Model 2 is the preferred model as it has a significantly lower (Plummer's) DIC value compared to model 1. This suggests that including a random effect term that varies by individual in the linear component of the model improves the fit to the data. According to the output results, there is evidence that marijuana use increases over time. Furthermore, the results indicate that males have a higher likelihood of using marijuana compared to females.

VI. Appendix

This section provides the R code used to conduct the analysis.

```
# Part II: data summary
library(knitr)

# read the data
data<- read.csv("./marijuanause.csv", header=TRUE)

# get the statistics
total_use <- colSums(data[, -1])
female_use <- colSums(data[data$female == 1, -1])
male_use <- colSums(data[data$female == 0, -1])

sample_size <- nrow(data)
female_sample_size <- sum(data$female)
male_sample_size <- sum(data$female == 0)

percent_use <- (total_use / sample_size) * 100
female_percent_use <- (female_use / female_sample_size) * 100
male_percent_use <- (male_use / male_sample_size) * 100
```

```

# create the table
summary_table <- data.frame(
  Year = 1976:1980,
  Total_users = total_use,
  Percentage = percent_use,
  Female = female_use,
  Female_percentage = female_percent_use,
  Male = male_use,
  Male_percentage = male_percent_use
)

kable(summary_table, caption = "Data Summary")

# Part II: two plots of percentage of use versus year
par(mfrow=c(1, 2))

# % of female marijuana users vs year
plot(summary_table$Female_percentage, type = 'o', main="% of female marijuana users vs year",
      ylab="percentage of use", xlab="Year", xaxt = 'n')
axis(side=1, at=c(1:5), labels=c("1976", "1977", "1978", "1979", "1980"))

# % of male marijuana users vs year
plot(summary_table$Male_percentage, type = 'o', main = "% of male marijuana users vs year",
      ylab="percentage of use", xlab="Year", xaxt='n')
axis(side=1, at=c(1:5), labels=c("1976", "1977", "1978", "1979", "1980"))

# Part III(a): run jags
years = c(1976, 1977, 1978, 1979, 1980)
centered_years = years - mean(years)
d1 <- list(use = as.matrix(data[,2:6]),
           femaleCentered = as.vector(scale(data$female, scale=FALSE)),
           yearCentered = as.vector(centered_years / (2 * sd(centered_years))))

# 4 chains
inits1 <- list(list(beta0=10, betaYear=10, betaFemale=-10),
              list(beta0=10, betaYear=-10, betaFemale=10),
              list(beta0=-10, betaYear=10, betaFemale=-10),
              list(beta0=-10, betaYear=-10, betaFemale=10))

library(rjags)
m1 <- jags.model("model_1.bug", d1, inits1, n.chains=4, n.adapt=5000, quiet = TRUE)

# burn-in iterations
update(m1, 3000)
# get samples
x1 <- coda.samples(m1, c("beta0", "betaYear", "betaFemale"), n.iter=10000)

# Gelman Rubin diagnostics
gelman.diag(x1, autoburnin=FALSE)
# trace plot
plot(x1)

# Part III(b): summarize the details of MCMC computation
x1_summary = summary(x1)

```

```

chains = x1_summary$nchain
burn_in = 3000
iterations = x1_summary$end - x1_summary$start + 1

combined = data.frame('Number of chains' = chains,
                      'Length of burn-in' = burn_in,
                      'Number of iterations used per chain' = iterations,
                      'Thinning' = 1)

kable(combined, caption = "MCMC Summary for Model 1")

# Part III(b): effective sample sizes of all parameters
eff = data.frame(effectiveSize(x1[,c("beta0", "betaFemale", "betaYear")]))
colnames(eff) = "effective sample sizes"
kable(eff, caption = "Effective sample sizes of all parameters for Model 1")

# Part III(c): approximate posterior data for each parameter
posterior.mean = x1_summary$statistics[, "Mean"]
posterior.sd = x1_summary$statistics[, "SD"]
posterior.interval = x1_summary$quantiles[, c("2.5%", "97.5%")]

combined = data.frame(Mean = posterior.mean, SD = posterior.sd,
                      `2.5% quantile` = posterior.interval[,1],
                      `97.5% quantile` = posterior.interval[,2])
row.names(combined) = c("beta0", "betaFemale", "betaYear")
kable(combined, caption="Approximate posterior data for each parameter in Model 1")

# Part III(d): approximate the posterior probability
betaFemale = as.matrix(x1[, "betaFemale"])
posterior.prop = mean(betaFemale > 0)
densplot(x1[, "betaFemale"], main = "Density of the coefficient for the (centered) indicator of being fer

# Part III(e): approximate the value of (Plummer's) DIC
load.module("dic")
dic1 = dic.samples(m1, 100000)

# Part IV(a): run jags
d2 <- list(use = as.matrix(data[,2:6]),
          femaleCentered = as.vector(scale(data$female, scale=FALSE)),
          yearCentered = as.vector(centered_years / (2 * sd(centered_years))))

# 4 chains
inits2 <- list(list(beta0=10, betaYear=10, betaFemale=-10, sigmaepsilon = 0.01),
              list(beta0=10, betaYear=-10, betaFemale=10, sigmaepsilon = 9),
              list(beta0=-10, betaYear=10, betaFemale=-10, sigmaepsilon = 9),
              list(beta0=-10, betaYear=-10, betaFemale=10, sigmaepsilon = 0.01))

library(rjags)
m2 <- jags.model("model_2.bug", d2, inits2, n.chains=4, n.adapt=5000, quiet = TRUE)

# burn-in iterations
update(m2, 5000)
# get samples
x2 <- coda.samples(m2, c("beta0", "betaFemale", "betaYear", "sigmaepsilon"), n.iter=30000)

```

```

# Gelman Rubin diagnostics
gelman.diag(x2, autoburnin=FALSE)
# trace plot
plot(x2)

# Part IV(b): summarize the details of MCMC computation
x2_summary = summary(x2)
chains = x2_summary$nchain
burn_in = 5000
iterations = x2_summary$end - x2_summary$start + 1

combined_2 = data.frame('Number of chains' = chains,
                        'Length of burn-in' = burn_in,
                        'Number of iterations used per chain' = iterations,
                        'Thinning' = 1)

kable(combined_2, caption = "MCMC Summary for Model 2")

# Part IV(b): effective sample sizes of all parameters
eff_2 = data.frame(effectiveSize(x2[,c("beta0", "betaFemale", "betaYear", "sigmaepsilon")]))
colnames(eff_2) = "effective sample sizes"
kable(eff_2, caption = "Effective sample sizes of all parameters for Model 2")

# Part IV(c): approximate posterior data for each parameter
posterior.mean = x2_summary$statistics[, "Mean"]
posterior.sd = x2_summary$statistics[, "SD"]
posterior.interval = x2_summary$quantiles[, c("2.5%", "97.5%")]

combined_2 = data.frame(Mean = posterior.mean, SD = posterior.sd,
                        `2.5% quantile` = posterior.interval[,1],
                        `97.5% quantile` = posterior.interval[,2])
row.names(combined_2) = c("beta0", "betaFemale", "betaYear", "sigmaepsilon")
kable(combined_2, caption="Approximate posterior data for each parameter in Model 2")

# Part IV(d): approximate the value of (Plummer's) DIC
load.module("dic")
dic2 = dic.samples(m2, 100000)

```