Group 3:
Matthew Anthony Tjahjadi - 2702404576
James Andersen - 2702350402

1. Introduction

In this report, we analyze sales performance data from a retail giant, Walmart, to explore the factors influencing weekly sales across various stores. The dataset, titled Walmart Sales, contains detailed information on weekly sales metrics and other key predictors that may impact overall sales performance.
The dataset include:

- Store: Identifier for the store (1 to 45).
- Date: Week-ending date of the observation.
- Weekly_Sales: Total sales for the store in the specified week.
- Holiday_Flag: Binary indicator (0 or 1), where 1 indicates a special holiday week.
- Temperature: Average temperature for that week (in Fahrenheit).
- Fuel_Price: Price of fuel per gallon in the region.
- CPI: Consumer Price Index, indicating changes in prices.
- Unemployment: Unemployment rate in the region

Retail performance is a critical measure of success for businesses like Walmart, where understanding sales drivers is essential for strategic decision-making. By identifying how external factors (e.g., unemployment, fuel price, temperature) and internal business cycles (e.g., holiday weeks) influence weekly sales, we can provide actionable insights to improve operational efficiency, resource allocation, and sales forecasting.

2. Model

In this study, we will implement 2 different models.

Model 1: Linear Regression

```
model_string <- textConnection("model {
   # Likelihood
   for (i in 1:n) {
      y[i] ~ dnorm(mu[i], tau)
      mu[i] <- alpha + inprod(x[i,], beta[])
      like[i] <- logdensity.norm(y[i], mu[i], tau)
   }

   # Priors
   for (j in 1:p) {
      beta[j] ~ dnorm(0, 0.001)
   }
   alpha ~ dnorm(0, 0.001)
   tau ~ dgamma(0.1, 0.1)
}")
```

Likelihood: Normal distribution

Prior: Weakly informative Gaussian priors for alpha and beta, and a Gamma prior for tau

Model 2: Logistic Regression

```
model_string <- textConnection("model{
  # Likelihood
  for(i in 1:n){
    y[i] ~ dbern(pi[i])
    logit(pi[i]) <- alpha + inprod(x[i,], beta[])
    like[i] <- logdensity.bern(y[i], pi[i])
  }

  # Priors
  alpha ~ dnorm(0, 0.01)
  for(j in 1:p){
    beta[j] ~ dnorm(0, 0.01)
  }
}")
```

Likelihood: Bernoullii distribution

Prior: Weakly informative Gaussian priors for alpha and beta

Algorithm

We use Markov Chain Monte Carlo (MCMC) with the following settings for both models:

- Burn-in: 1,000 iterations.
- Total Iterations: 10,000.
- Number of Chains: 3.

Convergence of model 1

From the graphical diagnostics of model1, every covariate in the trace plot shows convergence, and every chain mixes well, overlapping over each other. The numerical diagnostics for this model, taken using autocorr, ESS, gelman and geweke, all collectively show that each model achieves convergence

Convergence of model 2

Similarly to model1, every covariate in the plots converges. However there is a slight difference in the autocorr plot, where some covariates of model2 take a slightly longer time to converge. Numerical diagnostics for this model also collectively show that the model achieves convergence for every statistical test

Results

In order to decide which model is best fit for our data, we compared both models using Deviance Information Criterion (DIC) and Watanabe Akaike Information Criterion (WAIC)

```
[1] "DIC Model1"
Mean deviance:  198300
penalty 1.009
Penalized deviance: 198301
```

```
[1] "DIC Model3"
Mean deviance:  8815
penalty 6.032
Penalized deviance: 8821
```

```
$waic
[1] 918497.4

$lppd
[1] -459248.2

$p_waic
[1] 0.5247965
```

```
$waic
[1] 40571.66

$lppd
[1] -20279.85

$p_waic
[1] 5.985729
```

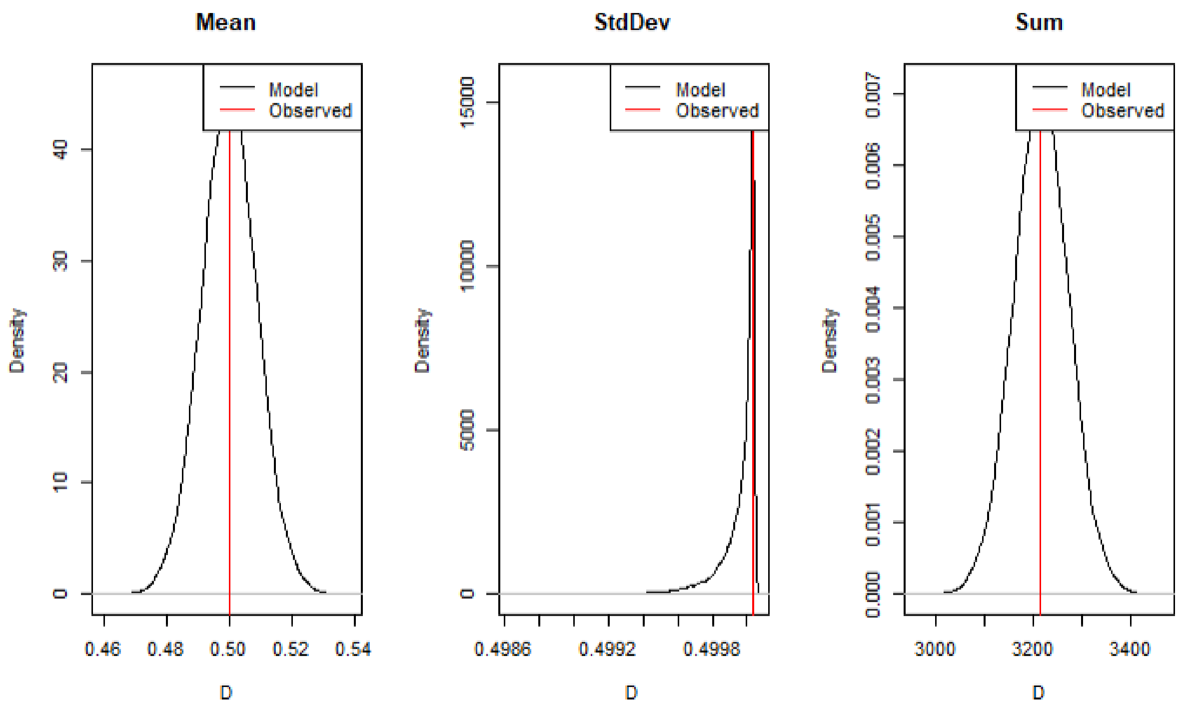
The results of both information criteria favors the logistic regression model, hence we choose this as the preferred model going forward.

A statistical summary is then made in order for us to interpret said model

```
Iterations = 2001:12000
Thinning interval = 1
Number of chains = 3
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

            Mean      SD  Naive SE Time-series SE
beta[1]  0.05316 0.02557 0.0001476      0.0001924
beta[2] -0.02462 0.02696 0.0001556      0.0002220
beta[3]  0.01077 0.02624 0.0001515      0.0002068
beta[4] -0.25309 0.02788 0.0001610      0.0002465
beta[5] -0.15010 0.02736 0.0001580      0.0002331

2. Quantiles for each variable:

            2.5%       25%      50%       75%    97.5%
beta[1]  0.00289  0.035987  0.05332  0.070589  0.10330
beta[2] -0.07692 -0.042860 -0.02467 -0.006107  0.02786
beta[3] -0.04062 -0.006951  0.01078  0.028399  0.06236
beta[4] -0.30752 -0.271953 -0.25317 -0.234190 -0.19845
beta[5] -0.20418 -0.168623 -0.14996 -0.131682 -0.09724
```

We can conclude from the means and the confidence interval of each covariate, that beta4 (Consumer Price Index) and beta5 (Unemployment) negatively affects our target variable While beta1 (presence of holiday flag) positively affects our target variable

Lastly, posterior predictive checks were done to evaluate how well the model fits the observed data, this was done by replicating data from the posterior distribution and comparing it with simulations from actual observed data. The statistics we used to compare are mean, standard deviation and sum



For each statistic, our model is able to replicate the values of observed data well. This can be seen by the red that crosses through the centre of every distribution

Conclusion

In this study, we analyzed Walmart's sales performance data using two models: a Linear Regression model and a Logistic Regression model. The goal was to understand the factors influencing weekly sales and identify the model that best fits the data.

Overall, the Logistic Regression model provides robust insights into the relationship between weekly sales and key predictors. These findings can be used to inform strategic decisions, such as optimizing operations during holiday weeks or mitigating the impact of external economic factors like unemployment and fuel prices.

This analysis highlights the importance of selecting an appropriate model for sales forecasting, which can improve business decision-making and operational efficiency for retail giants like Walmart.