---

title: "FinalProject"

author: "Oscar Sucre"

output: html_document

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
library(tidyverse)
library(rstanarm)
library(bayesplot)
library(bayestestR)
library(parameters)
library(knitr)
library(magrittr)
library(ISLR)
library(quantmod)
```

data comes from ISLR library package, the description is Wage and other data for a group of 3000 male workers in the Mid-Atlantic region. factors that impact it are year,age,maritl,race,education,region,jobclass,health,and health_ins.

```{r data}
fpdata <- Wage
```

predicting the wage, I used log wage because it made the model work better as the peak was much lower. also included all the factors that I thought would have the biggest impact. Multiplying them made the model worse so I sticked to adding. The prior Summary is to check what the scaling is and to see if adjustments could be useful.

```{r fit,echo=FALSE,include=FALSE}

wagefit <- stan_glm(logwage~age+education+race+year,data = fpdata, adapt_delta = .99)



```

```{r prior_summary}

prior_summary(wagefit)

```

posterior predictive check: A bit of a higher peak and a spike on the bottom right so it fits decently

```{r posterior predictive check plot}

pp_check(wagefit)

```

trace plots for very important factors to wage

```{r}

plot(wagefit, plotfun = "trace",pars = c("age","education5. Advanced Degree"))

```

posterior density of the same factors for the trace plots

```{r dens, echo=FALSE}

plot(wagefit, plotfun = "dens",pars = c("age","education5. Advanced Degree"))

```

```{r posterior}

describe_posterior(wagefit, ci = .95, centrality = "mean")

```

This Proves that Advanced Degrees are very helpful to earning more

```{r probability question,echo=FALSE}
wagefit %>%
  as.data.frame() %$%
  mean(`education5. Advanced Degree` > .5)
```

This fit is adding grouping to the first fit. Had to add more iterations and chains because it needed more ESS. Also it takes around 15 minutes to compile

```{r fit2, include=FALSE}
wagefit2 <- stan_glmer(logwage ~ age+education+(1|race)+(1|year),data = fpdata,adapt_delta = .9999,iter = 4000,chains = 8,cores = 8)


```

```{r prior_summary 2}
prior_summary(wagefit2)
```

Model Comparison

second fit is better but not by a huge margin, So I did the same analysis to see if there was a difference

```{r loo fit 2}
l1 <- loo(wagefit)
l2 <- loo(wagefit2)

loo_compare(l1, l2)

```

pp check is very simiplar

```{r pp check 2}
pp_check(wagefit2)
```

Much more changes in the posterior because of the grouping.

```{r mean 2}
describe_posterior(wagefit2, ci = .95, centrality = "mean")
```

Density Plot

```{r dens plot 2}
plot(wagefit2, plotfun = "dens",pars = c("age","education5. Advanced Degree"))

```

Trace Plot

```{r trace plot 2 }
plot(wagefit, plotfun = "trace",pars = c("age","education5. Advanced Degree"))
```

Advanced Degree proves to be more prevalent.

```{r probability question 2}
wagefit2 %>%
  as.data.frame() %$%
  mean(`education5. Advanced Degree` > .5)
```

On my honor, I have

neither received nor given any unauthorized assistance on this project

Oscar Sucre