

Baysian Statistics Project

By: Nick Luerkens and Josh Gunderson

Research Question: Which Statistic in major league baseball has more of an impact on a team's success, batting average or era?

Variables: We obtained data from all 30 major league baseball teams during the 2004 major league season.

- Predictor variables: batting average (x) and era (y)
- Response variable: season winning percentage (z)

Baysian Model: We specified our model by assuming normal distributions for all three of our variables. We will analyze our data using multiple linear regression: $\mu(z) = \alpha + \beta(x) + \gamma(y)$.

Winbugs Code:

```
model
{
for (i in 1:N) {
  wins[i] ~ dnorm( mu[i], tau ) ;
  mu[i] <- alpha + beta * batting[i] + gamma * era[i] ;
}

tau ~ dgamma(.01, .01) ;
alpha ~ dnorm (0 , .01) ;
beta ~ dnorm ( .954, .01) ;
gamma ~ dnorm (-.057, .01) ;
}
```

list(N = 30)

Prior parameters:

* We must guess what E (x) and E (y) will be non-informatively. E (x) = 0.262 and E (y) = 4.39.

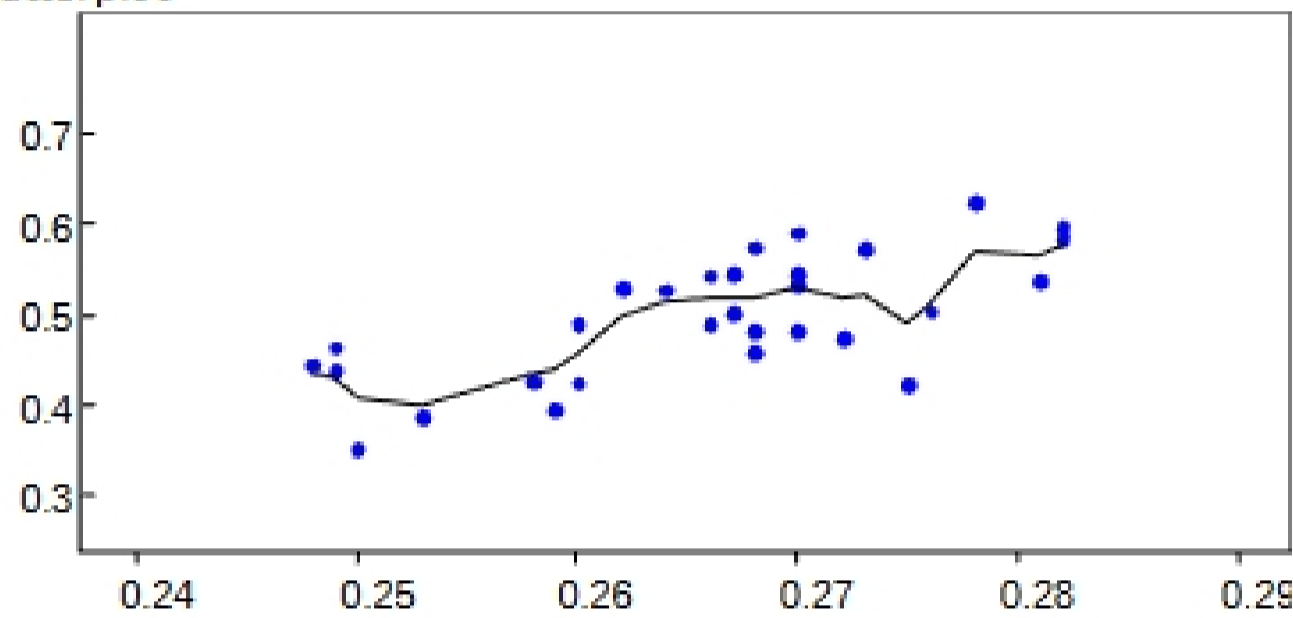
- win%: mu [i] for win% will be 0.5. The reason is because we are assuming the average team to win half of their games.
- alpha: alpha indicates the intercept for our multiple regression equation. We will set it at zero.

- Beta and Gamma: $0.5 = \alpha + \text{Beta} (.262) + \text{-Gamma} (4.39)$. Assume that average and era have the same impact on wins. $.25 = \text{Beta} (.262)$; Beta = .954
 $.25 = \text{-Gamma} (4.39)$; Gamma = -.057 (This needs to be negative because the slope of era vs. wins is negative in simple regression).

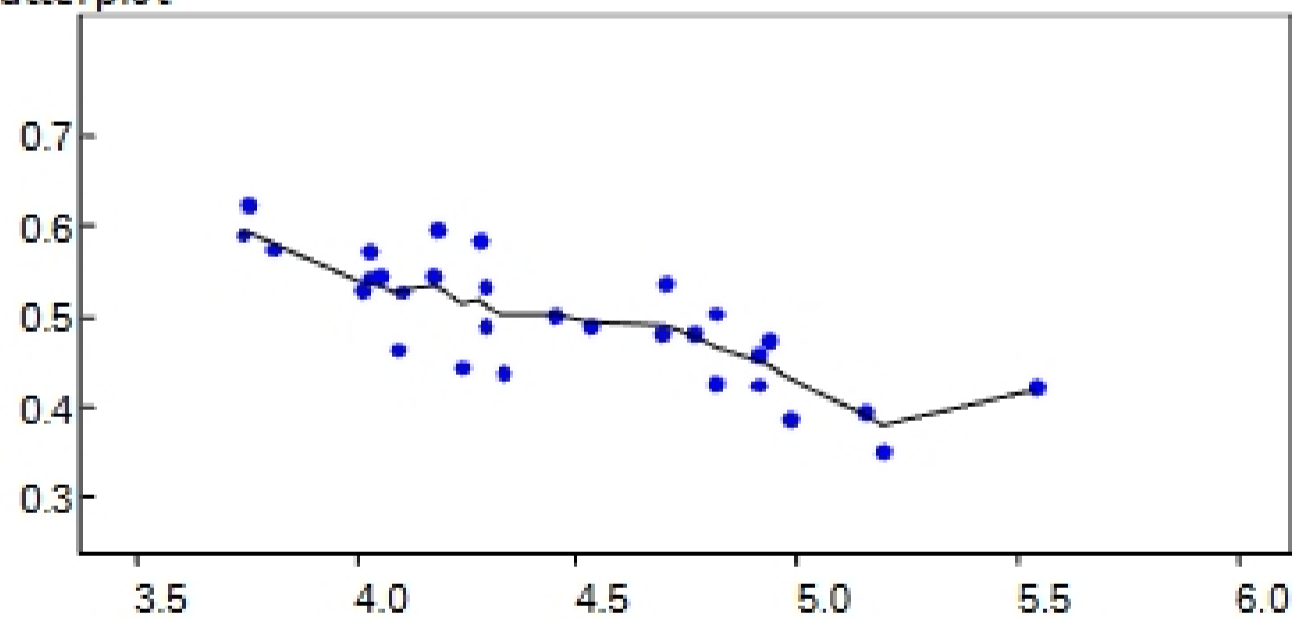
Winbugs Output and Interpretation:

- The 1st is a scatter plot of era (x – axis) vs. winning % (y – axis).
- The 2nd is a plot of batting average (x – axis) vs. winning % (y – axis).

scatterplot



scatterplot



Bivariate posterior catte plots

