Savannah Curtis (Bates)
ST 540 - Applied Bayesian Analysis
April 15, 2020

Midterm Exam: Vaporfly shoes analysis

1 Introduction

Nike recently released a new running shoe called the Vaporfly that has made a huge splash in the marathon running community. Several studies have reported dramatic speed improvements for runners wearing these shoes. The objective of this study is to determine the magnitude of the improvement and whether the improvement varies across (1) gender, (2) runner and/or (3) course.

2 Methods

The response variable, Y_i, represents the scaled time in minutes of performance i by a certain runner at a specific marathon. The response was scaled to have mean zero and variance one and then modeled using a Gaussian likelihood. Checks for normality (eg. comparing density plots) were performed to make sure this was a reasonable assumption. We concluded it is reasonable, especially given the robustness of the normality assumption (though men seemed to contradict this assumption more). Having a Gaussian likelihood was preferred (over a gamma for example) because of the ease of interpreting the parameters in a normal distribution. Recall that data was separated by gender. We chose to keep this separation and model the data for each gender separately because the combined data was clearly bimodal. Therefore, each model below was fit to two data sets: one for men and one for women. Model 1 is given by

$$Y_i \sim \text{Normal}(\beta_0 + \beta_1 x_i, \tau) \text{ for } i = 1, \ldots, N$$

$$\beta_j \sim \text{Normal}(0, 0.01) \text{ for } j = 0, 1$$

$$\tau \sim \text{Gamma}(0.1, 0.1)$$

where x_i is the indicator covariate of vaporfly use for performance i. Model 2 is given by

$$Y_i \sim \text{Normal}\left(\beta_0 + \beta_{0j}^{(r)} + \beta_{0k}^{(m)} + \left(\beta_1 + \beta_{1j}^{(r)} + \beta_{1k}^{(m)}\right) x_i, \tau\right) \text{ for } i = 1, \ldots, N$$

$$\beta_i \sim \text{Normal}(0, 0.01) \text{ for } i = 0, 1$$

$$\beta_{ij}^{(r)} \sim \text{Normal}\left(0, \tau_i^{(r)}\right) \text{ for } i = 0, 1 \text{ and } j = 1, \ldots, R$$
$$\beta_{ik}^{(m)} \sim \text{Normal}\left(0, \tau_i^{(m)}\right) \text{ for } i = 0, 1 \text{ and } k = 1, \ldots, M$$
$$\tau \sim \text{Gamma}(0.1, 0.1)$$
$$\tau_i^{(r)} \sim \text{Gamma}(0.1, 0.1) \text{ for } i = 0, 1$$
$$\tau_i^{(m)} \sim \text{Gamma}(0.1, 0.1) \text{ for } i = 0, 1$$

where $\beta_{ij}^{(r)}$ are random effects (slope and intercept) for each runner and $\beta_{ik}^{(m)}$ are random effects (slope and intercept) for each marathon. Model 3 is the same as model 2 except the random effects, $\beta_{ij}^{(r)}$ and $\beta_{ij}^{(r)}$, are modeled as a double-exponential random variables. For each model, the number of runners is $R = 68$ for men and $R = 56$ for women. Also, the number of marathons is $M = 15$ for men and women. You may recall that this is far fewer marathons and runners than were included in the full dataset. However, I believe in order to properly include "random slopes" for each runner and marathon, we must observe each runner/marathon both with and without the vaporfly shoes (models converged when I didn't do this but the values for random effects didn't make sense). Thus, after only including data that fit this criteria, we were left with $N = 287$ total observations for men and $N = 240$ for women. All three models were fit with the same cleaned data to keep the comparison equal.

3 Computation

Each model was fit for the male and female data separately using `JAGS` in R. The first 50,000 samples were discarded as a burn-in then 200,000 samples were drawn with a thinning factor of 10 to retain 20,000 samples in total for each of 2 chains. The Gelman-Rubin statistics were computed for each parameter in each model and none were above 1.1. Specifically, all parameters in all models had a Gelman-Rubin statistic below 1.005 for both men and women. Additionally, the effective sample sizes were all above 1000 (smallest effective sample sizes were for model 3 and were 5,703 for men and 10,669 for women). Originally, models 2 and 3 didn't contain β_j's and instead had $\beta_{ij}^{(r)} \sim \text{Normal}\left(\mu_i^{(r)}, \tau_i^{(r)}\right)$ for $i = 0, 1$ and $j = 1, \ldots, R$ (or double-exponential for model 3) and similarly for the $\beta_{ik}^{(m)}$'s. However, this model didn't coverge (effective sample sizes were between 4 and 12 for the random effects). I suspect that this is due to identifiability issues because the random effects are additive. However, no formal evaluation of identifiability was performed.

4 Model Comparisons

Models were compared using the deviance information criterion (DIC) and the Watanabe-Akaike information criterion (WAIC). We visually checked to ensure the posterior densities were approximately normal and ensured all models had the same likelihood for valid comparison. The results for men can be seen in Table 1 and the results for women are shown in Table 2. Interestingly, DIC and WAIC don't agree on the best model for men. However, we note that the DIC values for models 2 and 3 for both genders are fairly close. We conclude that the best model for both genders is model 2 (Gaussian random effects).

Model	Mean Deviance	DIC Penalty	Penalized Deviance (DIC)	WAIC Penalty	WAIC
1	803.3	3.05	806.4	4.16	807.61
2	576.7	80.87	657.5	69.22	**659.97**
3	575.2	77.05	**652.3**	78.05	672.94

Table 1: Model comparisons for models 1-3 fit to the data for men.

Model	Mean Deviance	DIC Penalty	Penalized Deviance (DIC)	WAIC Penalty	WAIC
1	673.4	3.02	676.4	3.87	677.33
2	492	67.37	**559.4**	55.24	**558.67**
3	500.4	60.34	560.8	52.52	562.75

Table 2: Model comparisons for models 1-3 fit to the data for women.

5 Results

The final estimate of the improvement given to men running a marathon in the vaporfly shoes is $\hat{\beta}_1 = -0.32$ (95% CI is $(-0.58, -0.10)$). For women, this estimate is $\hat{\beta}_1 = -0.28$ (95% CI is $(-0.59, 0.03)$). Note that the 95% CI for women contains zero. However, we can conclude for men that the improvement for wearing the vaporfly shoes is significant. This would lead me to conclude that the improvement for wearing vaporfly shoes varies by gender (men have a higher and more significant improvement than women). However, we did have more data for men and the credible intervals have a lot of overlap so perhaps we just didn't have enough power to find a significant result for women and the results are actually the same. Figure 1 shows the boxplot of the posterior means for each random slope variable.

From this figure, we can conclude that the average improvement due to wearing the shoes doesn't show much relative variation across runners or marathons. For the most part, the effect of the shoes on an individual runner or on a specific course are small, after averaging over the fixed effects. However, there are a couple of interesting outliers in this figure that are fun to explore further. We note that there aren't any specific marathons for men where the vaporfly shoe had an exceptional impact on performance. However, for the women, $\hat{\beta}^{(m)}_{1,9} = -0.29$ which represents the New York City Marathon stood out. The mean time of women running this race in the vaporfly shoes was over 2.5 minutes faster than women running the race in other shoes. For the male data, three men stood out with respect to the effect the vaporfly shoe had on their performance. Two of these men are Mason Frank ($\hat{\beta}^{(r)}_{1,44} = 0.53$) who ran the California International Marathon 23 minutes slower while wearing the vaporfly shoes and Teklu Deneke ($\hat{\beta}^{(r)}_{1,63} = -0.29$) who ran the Eugene Marathon 29 minutes faster when he wore the shoes. For the ladies, Serena Burla stood out ($\hat{\beta}^{(r)}_{1,47} = 0.32$). While she ran different races, the one race where she wore the vaporfly shoes was about 19 minutes slower than other races she ran. While the effects we are seeing here are almost certainly attributed to some other factor besides the vaporfly shoes that the data just didn't capture, it is still reaffirming to see that the random effects estimates do make sense given the performance data we do have. In conclusion, we believe the vaporfly shoe does improve a runners performance overall (but it's certainly not going to redeem my terrible running skills).

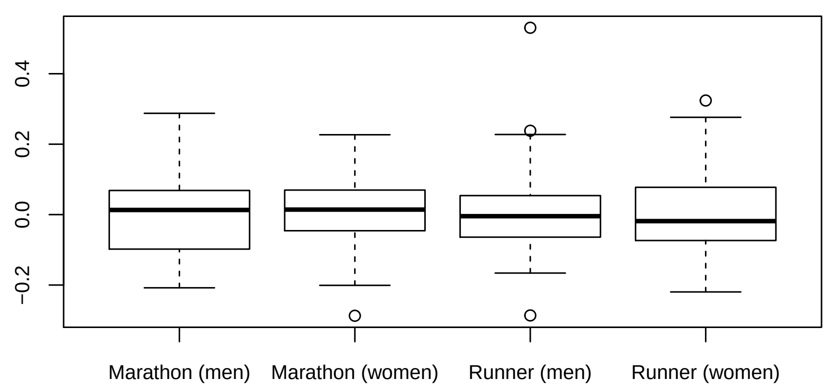

Figure 1: Boxplot of the posterior means for each random slope variable. Marathon (men) is the boxplot of posterior means of the $\beta^{(m)}_{1k}$'s for each $k = 1, \ldots, M$. Runner (men) is the boxplot of posterior means of the $\beta^{(r)}_{1j}$'s for each $j = 1, \ldots, R$. This is also true for the women.

Code

```
##### ----- Midterm.R (portion of code) ----- #####

##### ---- MODEL 2: Gaussian random effects for runner and marathon  ----- #####

# Define covariates
X_men <- cbind(men$match_name, men$marathon, men$vaporfly)
R_men <- max(X_men[,1])
M_men <- max(X_men[,2])
X_women <- cbind(women$match_name, women$marathon, women$vaporfly)
R_women <- max(X_women[,1])
M_women <- max(X_women[,2])

# Define model as string (same for men and women)
model2_string <- textConnection("model{
    # Likelihood
    for (i in 1:N)
    {
        mu[i] <- beta0 + beta0_r[X[i,1]] + beta0_m[X[i,2]]
                 + (beta1 + beta1_r[X[i,1]] + beta1_m[X[i,2]]) * X[i,3]
        Y[i] ~ dnorm(mu[i], tau)
        zlikelihood[i] <- dnorm(Y[i],mu[i],tau) # for WAIC computation
    }

    # Priors for fixed effects
    beta0 ~ dnorm(0, 0.01)
    beta1 ~ dnorm(0, 0.01)

    # Random effects
    for(i in 1:R)
    {
        beta0_r[i] ~ dnorm(0, tau0_r)
        beta1_r[i] ~ dnorm(0, tau1_r)
    }
    tau0_r ~ dgamma(0.1, 0.1)
    tau1_r ~ dgamma(0.1, 0.1)
    for(i in 1:M)
    {
        beta0_m[i] ~ dnorm(0, tau0_m)
        beta1_m[i] ~ dnorm(0, tau1_m)
    }
    tau0_m ~ dgamma(0.1, 0.1)
    tau1_m ~ dgamma(0.1, 0.1)

    # Prior for variance
    tau ~ dgamma(0.1, 0.1)
    sigma <- sqrt(1/tau) }")
```