ST540: Final Exam

Rachel Broughton

April 15th, 2020

Introduction

Nike released a new pair of shoes, the vaporfly, to improve distance running performance. Marathon times of professional runners with and without the shoes were compiled from 2016-2019 for several different races. This data will be used to evaluate the effect on marathon time of wearing the vaporfly shoes and whether that improvement varies with gender, runner, or course.

Models

Linear models were used with an intercept, one covariate being whether or not vaporfly shoes were worn, and a model variance. Each course is defined by an index that is unique for every race name (c_i) and each runner is also given a label (r_i). Y_i represents the marathon time in minutes for the i^{th} data entry. X_i is the indicator for whether the Vaporfly shoes were worn in the i^{th} data entry. Different years are not considered to be different courses. In the analysis, data for men and women are analyzed separately for all models. There are 270 women runners, 578 men runners, and 23 total courses. This is done to simplify the analysis and because men and women have clearly different intercepts and slopes as well as to avoid any compound effect of a participant as a woman/man and an individual.

1) Model 1

Both the slope and the intercept are treated as fixed effects that are the same for every runner and course, but separated as two different models by gender.

$$Y_i \sim \text{Normal}(b_0 + b_1 X_i, \sigma^2)$$

Uninformative priors are used, which are set as: $b_0 \sim \text{Normal}(0, 1000)$, $b_1 \sim \text{Normal}(0, 1000)$, $\sigma \sim$ Half-Cauchy.

2) Model 2

The slopes are fixed and equal for all participants and courses. The intercepts are treated as mixed effects with random effects for the participant number and course number.

$$Y_i \sim \text{Normal}(b_0 + b_1 X_i + A_{r_i} + Cc_i, \sigma^2)$$

The random effects are defined as:

A_{r_i} = random effect on intercept for runner of i^{th} entry, $A_r \sim \text{Normal}(\mu_A, \sigma_A^2)$

C_{c_i} = random effect on intercept for course of i^{th} entry, $C_c \sim \text{Normal}(\mu_C, \sigma_C^2)$.

The priors for b_0 and b_1 are defined the same as in Model 1 with the exception that information from the vaporfly paper was given to the b_0 prior by setting the prior mean to 160 for women and 140 for men. Priors for the additional parameter are $\mu_A, \mu_C \sim \text{Normal}(0, 1)$, $\sigma_A, \sigma_C \sim \text{Half-Cauchy}$.

3) Model 3

The third model allows the slope to vary based on the participant and the course run of the data entry. The intercepts remain the same as Model 2.

$$Y_i \sim \text{Normal}(b_0 + \text{slope}_i * X_i + A_{r_i} + Cc_i, \sigma^2)$$

$$\text{slope}_i = b_1 + b_{r_i} + b_{c_i}$$

The priors are the same as Model 2. The component b_{r_i} allows the value of the slope to change for each runner and b_{c_i} allows the slope to change with a given course with $b_{r_i}, b_{c_i} \sim \text{N}(0, 100)$.

4)Model 4

Independent random effects are incorporated into the slope and intercept based on course and runner groups.

$$Y_i \sim \text{Normal}(b_0 + \text{slope}_i * X_i + A_{r_i} + Cc_i, \sigma^2)$$

$$\text{slope}_i = b_1 + b_{r_i} + b_{c_i}$$

The random effects are defined as $b_r \sim \text{Normal}(\mu_1, \sigma_1^2)$, $b_r \sim \text{Normal}(\mu_2, \sigma_2^2)$. The priors for these new parameters are $\mu_1, \mu_1 \sim \text{Normal}()$ and $\sigma_1, \sigma_2 \sim \text{Half-Cauchy}$.

5) Model 5

The final model treats both the slopes and the intercepts as correlated random effects with separate parameters for the runner effect and the course effect. An interaction term for the course and runner indices is added.

$$Y_i \sim \text{Normal}(\alpha_{i,1} + \alpha_{i,2} * X_i, \sigma^2)$$

$$\alpha_{i,j} = \alpha 1_{r(i),j} + \alpha 2_{c(i),j} + \alpha 12_{r(i),c(i),j}$$

$$\alpha\mathbf{1_i} \sim \text{Normal}(\mathbf{B1}, \mathbf{\Omega 1}),\ \alpha\mathbf{2_i} \sim \text{Normal}(\mathbf{B2}, \mathbf{\Omega 2}),\ \alpha\mathbf{12_i} \sim \text{Normal}(\mathbf{B12}, \mathbf{\Omega 12})$$

The vectors $\alpha_{\mathbf{i}}, \mathbf{B}$ and the matrix $\mathbf{\Omega}$ all have entries for (1)intercept and (2)slope. The priors for the means and covariances of $\alpha 1, \alpha 2$, and $\alpha 12$ are set as: $B[1] \sim \text{Normal}(50,\ 100)$, $B[2] \sim$

Normal(-5, 100), $\mathbf{\Omega} \sim$ InvWishart(2.1, $\mathbf{I_2}$/2.1). Slightly more informative priors with mean values approximated from the paper results and tighter variances are used to help improve convergence. The Inverse Wishart prior parameters were selected to set the variances at 1 and the correlations at 0.

Computation

JAGS was used to perform the MCMC computation in R. 100,000 iterations were run for each model with 50,000 burnin. The convergence of the models was determined by examining some of the parameter traceplots and checking the effective sample size. The effective sample sizes for investigated parameters of all models were all above 1000, and so it can be said that convergence was achieved. However, the convergence of Model 5 was notably worse than that of models 1-4.

Model Comparison

The models for the women were compared using the Deviance Information Criteria command (DIC), which are reported in Table 1. Based on the DIC results, the selected model to use is Model 5. While this model was the most complicated with the highest penalty, its overall DIC was noticeably lower than the others. However, with the exception of the simplest Model 1, the DIC comparison values were all fairly close together. Nonetheless, Model 5 was still selected as the model to use for women and, consequently, also men as the comparison results are expected to be similar for both genders.

Model Number	Penalty	Deviance	Total BIC Value
1	3	5641	5644
2	215	4836	5051
3	265	4810	5075
4	219	4833	5052
5	275	4759	5033

Table 1: DIC results for models

Results

By applying Model 5, it was found that the vaporfly shoes improve marathon performance for both men and women. The posterior samples of the mean slopes of all the course and runner fits are centered around negative values with only the tails reaching positive values (Figures 1 and 2), suggesting that the vaporflies improve marathon time. The mean of the posterior distribution of the mean slope values were found to be different for both men and women (see Table 2). By examining the posterior mean of the runner contribution to the slope random effects for each fit (Fig.3), it was found that the effect varies greatly by runner: zero slopes for participants that never wore vaporfly shoes, positive values for runners with worse performances in the shoes, and negative results for those that improved. The posterior distributions for the course component of the random effect of slope do not exhibit much difference between courses (Fig. 4). It can be concluded that the vaporfly effect, though negative overall, varies greatly by individual runner and gender, but is not significantly impacted by the course.

	Mean Slope Posterior Mean	95% CI	Mean Intercept Posterior Mean	95% CI
Women	-1.81	[-3.93, 0.25]	159.71	[158.75, 160.71]
Men	-2.1	[-3.77, -0.49]	139.42	[138.74, 140.14]

Table 2: Means and credible intervals for mean slopes and intercepts from all fits.

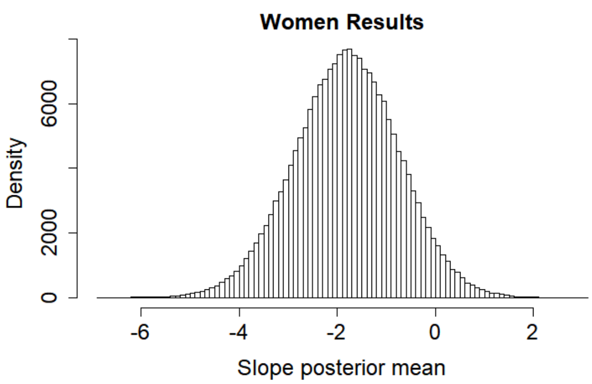

Figure 1: Women posterior mean slopes.

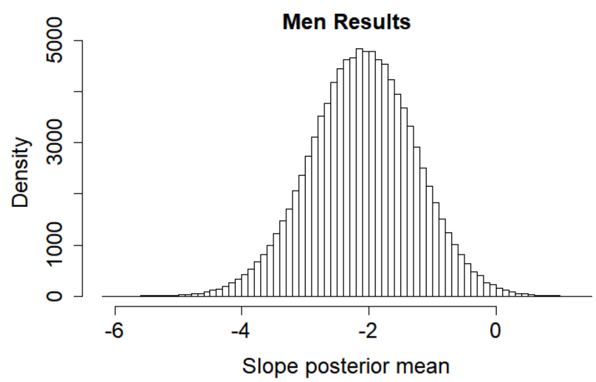

Figure 2: Men posterior mean slopes.

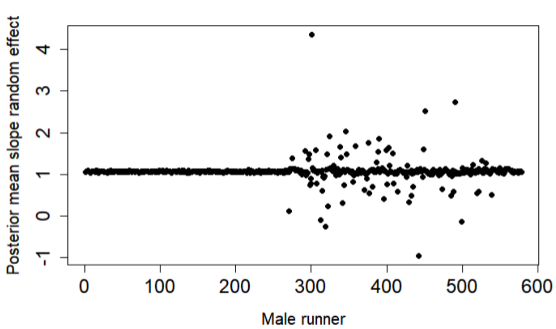

Figure 3: Male runner effect means.

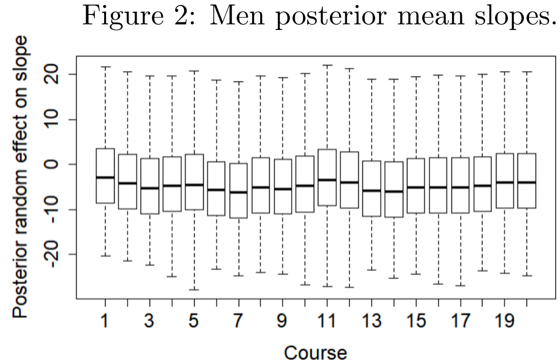

Figure 4: Course effect male posteriors.

```
# Model 5: Men
# Random effects: course + participant, with interaction

data <- list(Y=Y, X=X, n=n, c=c, r=r, course=course.men, run_num=number.men)
burn     <- 50000
n.iter   <- 100000
n.chains <- 2
params <- c("mu1", "mu2", "mu12", "Omega1", "Omega2", "Omega12", "mean_slope", "mean_intercept")

model_string <- textConnection("model{
  # Likelihood
  for(i in 1:n){
    Y[i] ~ dnorm(alpha[i,1] + alpha[i,2]*X[i], taue) }
  # Random Effects
  for (jj in 1:n){
    alpha[jj,1:2] <- alpha1[run_num[jj], 1:2] + alpha2[course[jj], 1:2] + alpha12[run_num[jj], course[jj], 1:2] }
  for(kk in 1:r){
    alpha1[kk, 1:2] ~ dmnorm(mu1[1:2], Omega1[1:2,1:2]) }
  for(ii in 1:c){
    alpha2[ii, 1:2] ~ dmnorm(mu2[1:2], Omega2[1:2,1:2])}
  for(qq in 1:r){
    for(pp in 1:c){
      alpha12[qq, pp, 1:2] ~ dmnorm(mu12[1:2], Omega12[1:2,1:2])}}
  # output mean slope of all entries for every iteration
  mean_slope <- mean(alpha1[,2]) + mean(alpha2[,2]) + mean(alpha12[,,2])
  mean_intercept <- mean(alpha1[,1]) + mean(alpha2[,1]) + mean(alpha12[,,1])
  # Priors
  #for(j in 1:2){mu1[j] ~ dnorm(0, 0.1)}
  #for(j in 1:2){mu2[j] ~ dnorm(0, 0.1)}
  #for(j in 1:2){mu12[j] ~ dnorm(0, 0.1)}
  mu1[1] ~ dnorm(50,0.01)
  mu2[1] ~ dnorm(50,0.01)
  mu12[1] ~ dnorm(50,0.01)
  mu1[2] ~ dnorm(-5,0.01)
  mu2[2] ~ dnorm(-5,0.01)
  mu12[2] ~ dnorm(-5,0.01)
  Omega1[1:2,1:2] ~ dwish(R[,], 2.1)
  Omega2[1:2,1:2] ~ dwish(R[,], 2.1)
  Omega12[1:2,1:2] ~ dwish(R[,], 2.1)
  taue <- pow(sigmae, -2) # Half cauchy priors on sd
  sigmae ~ dt(0,1,1)T(0,)
  R[1,1]<-1/2.1
  R[1,2]<-0
  R[2,1]<-0
  R[2,2]<-1/2.1 }")

model <- jags.model(model_string,data = data, n.chains=n.chains,quiet=TRUE)
update(model, burn, progress.bar="none")
samples5.men <- coda.samples(model, variable.names=params, n.iter=n.iter, progress.bar="none")
```