A Bayesian Hierarchical Model for Estimating the Cost of Postponing the Cyclocross National Championships

J.T. Fry*[1], Andy Hoegh[2], Scotland Leman[3] and Matthew Montesano[4]

[1]Department of Statistics, Virginia Tech

[2]Department of Mathematical Sciences, Montana State University

[3]Department of Statistics, Virginia Tech

December 20, 2016

Abstract

In early January 2015, the multi-event national cyclo-cross bicycle races were set to take place in Austin, Texas. Cyclo-cross has a rich history in this country, and throughout the world, attracting huge crowds and competitors. Being primarily a winter sport, these athletes often compete in harsh conditions, which include rain, snow, mud, and revel in the excitement that comes with such elements. Unfortunately, the competition was postponed mid-event when a local arborist group protested to the parks department. The issue: there was too much mud, in an event where many spectators and racers alike hope for such conditions. For many competitors, the postponement generated additional expenses, such as flights, hotels, and car rentals. Although people on opposite sides of the debate may greatly disagree, we instead focus on the competitors themselves. We analyze the financial

*Contact: fryjt1@vt.edu

Hutcheson Hall, RM 406-A,

250 Drillfield Drive Blacksburg, VA 24061

impact of the disagreement using a hierarchical Bayesian mixed model which accounts for heterogeneity within the costs endured by the event's participants.

1 Introduction

January 7, 2015: the U.S.A. Cycling Cyclo-cross National Championships returned to Zilker Park in Austin, Texas. It has been 37 years since the town of Austin hosted the prestigious five-day, multi-event bicycle race. Cyclo-cross is a curious form of bicycle racing, where racers must dismount and carry their bicycle over and across obstacles, utilizing both cycling and running abilities. Being a primarily winter sport, athletes often compete in harsh conditions such as rain, snow, and mud and revel in the excitement that accompanies such elements. Crowds gather to watch participants race down dirt paths, run through sandpits, slide through mud, hurdle barriers, and occasionally wipeout. Hundreds of competitors, young and old, amateur and professional, travel from all over the United States to race. Most of these athletes cover their own expenses, including airline tickets, hotel, car rental, and food. For such a long, expensive trip, most competitors budgeted and prepared well in advance.

Unfortunately, many of these racers' preparations were in vain. On Saturday, January 10 Zilker Park received 0.3 inches of rainfall (Accuweather 2015). What would have been seen as a blessing by many cyclo-cross enthusiasts, turned out to be a relative nightmare. Beginning as early as 7:15 AM on Sunday, January 11, police were turning away competitors from the park. After several hours of confusion, the Austin Parks and Recreation Department (APRD) announced that the eight races, or fields, scheduled for that day would be postponed until Monday. The issue: mud.

In a sport beloved by competitors and fans alike for mud, and all the excitement it brings, the National Championships were postponed for this very reason. Following the rain on Saturday night, the Austin Heritage Tree Foundation (AHTF) posted a "9-1-1 Tree Emergency" on its website (Austin Heritage Tree Foundation 2015). The group was concerned that racers could damage the tree roots by compacting the soil with their bicycles

and potentially kill the 300-year-old oak trees which adorn the park. Within a few hours, the Austin Parks and Recreation Department was bombarded with phone calls and emails from various fans, residents, and arborist groups, arguing *both* for and against the AHTF's concerns. The AHTF posted on their website, "The USA Cyclo-cross championship race is totally inappropriate for Zilker Park." After an investigation on March 18, consulting arborist Patrick Wentworth wrote, "No detectable or significant soil compaction was found to have been caused by the USA Cyclo-cross Race" (Wentworth 2015). Regardless, amidst the controversy the APRD sided with the AHTF and prohibited the competitors from entering the park. However, given the efforts and sacrifices made by the competitors, ultimately the final decision was to re-schedule the final eight races until Monday.

The 358 competitors whose events were rescheduled had to make an important decision: pack up and head home or make arrangements to spend an extra day in Austin. The vast majority of the competitors are not full-time professional athletes. They have nine-to-five jobs, families, and fixed budgets. To make the race, they would need to change airline tickets, extend hotel stays, and incur other miscellaneous expenses. Of the original 358, 276 decided to stay for Monday's race, while the remaining 82 individuals headed home. Following the race, the big question to be answered was: How much extra money did these competitors spend because of the postponement?

Our cost analysis utilizes a Bayesian Hierarchical model for modeling/estimating both fixed and random effects. Such frameworks have widespread use in sports analyses, where both predictive forecasting and policy making are concerned. Perhaps most common, Bayesian hierarchical modeling is now standard practice in sabermetric analyses for studying both offensive (Albert 2006, 1994) and defensive (Jensen et al. 2009) characteristics of players. Also, concerning basketball, Bayesian hierarchical modeling has been employed for studying player's shot accuracy at varying spatial locations (Carlin 1996). For forecasting game outcomes, Glickman & Hennessy (2015) uses a Bayesian state-space representation to model competitor's abilities. Also, Hoegh et al. (2015) uses a Bayesian hierarchical framework for understanding the 'matchup' effect between teams within the NCAA basketball tournament. Additionally, Leman et al. (2014) employs a Bayesian hierarchical model for

studying committee decision-making trends for selecting teams into the NCAA basketball tournament. Section 2.2 details our usage of Bayesian hierarchical modeling, and its relative advantages for understanding the costs incurred by the postponement of the 2015 Cyclo-cross National Championships.

2 Data Analysis and Modeling

Cyclo-cross enthusiast Matthew Montesano issued a survey via Twitter to estimate the cost of the postponement to the athletes. The survey included data on four categories: airline, car rental, hotel, and other miscellaneous expenses. Matthew's initial work utilized point estimates to extrapolate the total cost of the postponement. In a Velonews article discussing his findings, Matthew states,

> "I know how to clean up an Excel spreadsheet and generate some numbers, but I'm sure that there are people out there who could do a more sophisticated analysis and generate some numbers that could be more powerful. I'm happy to share anonymized data with anybody who's interested in doing that." (VeloNews 2015)

This began the collaborative effort between cyclist and statisticians to estimate the cost of the postponement to competitors and quantify the uncertainty of these estimates. The survey responses were cross-referenced with the list of competitors to ensure reliability and to determine each competitor's home state. The remaining data featured a 26 percent response rate that included the field, home state, and expenses for the competitors.

2.1 Distribution of Expenses

We group the four types of expenses into two categories: *travel* and *other*. Travel expenses are comprised of airline and car rental costs, whereas other expenses consist of hotel and miscellaneous expenses. Since travel expenses are dependent on the point of travel origin, our estimation of total travel expense groups states with similar travel expenditures using K-Means clustering, see Figure 1 and Table 1 for the clusters. Other expenses, however,

do not depend on the originating state; but, rather personal information such as income level and age. For example, competitors with steady careers may be more willing to spend on hotel rooms than college students. While this explicit data is unavailable, competitor field presents a reasonable stratification mechanism since field controls for the age and experience of the competitor. The fields included *Junior Females, Junior Males, Men's Elite 23 + Pro/1/2, Women's Elite/U23 19 + Pro/1/2/3*, and *Men's U23* (Table 2).

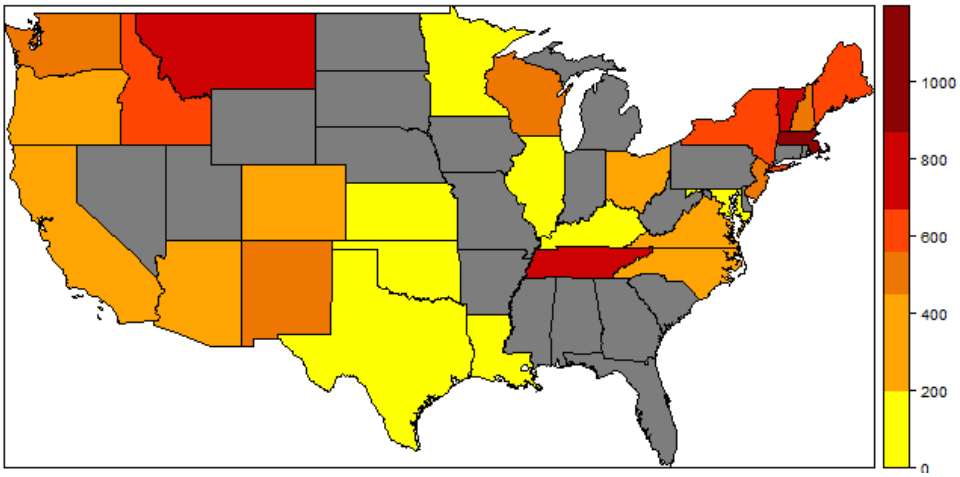

Figure 1: Map of state groups for travel expense. States in grey had no respondents in the survey.

Group	States	Mean Travel Expense	St. Dev Travel Expense
1	MA	1,195	488
2	MT, TN, VT	761	218
3	ID, ME, NY	626	187
4	NH, NJ, NM, WA, WI	538	435
5	AZ, CA, CO, NC, OH, OR, VA	351	258
6	IL, KS, KY, LA, MD, MN, OK, TX	67	113
7	AL, AR, CT, FL, GA, IA, IN, MI, MO, ND, PA, UT	Not Available	Not Available

Table 1: Mean Travel Expense for Clustered States

Group	Mean Other Expense	St. Dev Other Expense
Junior Females	547	461
Junior Males	468	369
Men's Elite 23+Pro/1/2	626	409
Women's Elite/U23 19+ Pro/1/2/3	335	202
Men's U23	164	97

Table 2: Mean Other Expense for Field Groups

Both travel and other expenses come from heavy-tailed distributions and are uncorrelated with each other (Correlation = .113). Figures 2 and 3 display normal quantile plots to demonstrate the dispersion.

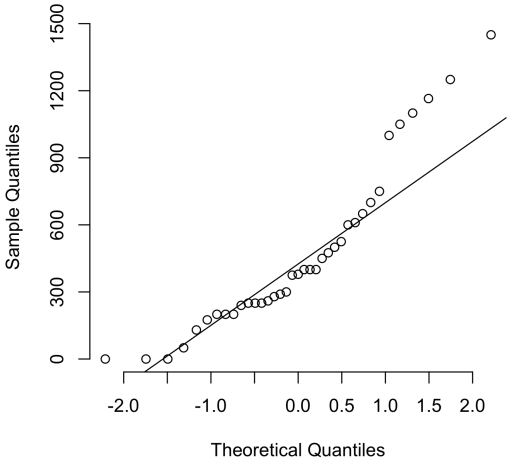

Figure 2: QQ Plot comparing State Group 5 travel expense quantiles with theoretical normal quantiles.

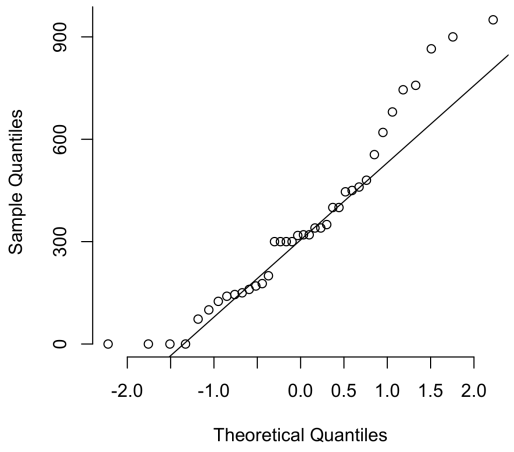

Figure 3: QQ Plot comparing Men Juniors other expense quantiles with theoretical normal quantiles.

2.2 Expense Modeling

We denote each expense type by $y^{(c)}$, where $c \in \{travel, other\}$. As previously mentioned, travel expenses ($y^{(travel)}$) are comprised of airline and car rental costs, whereas other ex-

penses ($y^{(other)}$) consist of lodging and all other miscellaneous costs. We utilize a random effect to control for variability in these costs across fields and spatial clusters. Using our desired random effects structure, our expense model is given through the hierarchical representation:

$$y_{kj_c}^{(c)} = \mu^{(c)} + \alpha_{j_c}^{(c)} + \epsilon_{kj_c}(\nu^{(c)}), \tag{1}$$

$$(\alpha_{j_c}^{(c)} | \mu^{(c)}) \sim Normal(0, 1/\psi^{(c)}), \qquad \text{where } \mu^{(c)} + \alpha_{j_c}^{(c)} \geq 0. \tag{2}$$

In Equations (1) and (2), $y_{kj_c}^{(c)}$ represents the expense for competitor k, of class $c \in \{travel, other\}$, with associated random effects j_c for $j_{travel} = 1, \ldots, 6$ and $j_{other} = 1, \ldots, 5$. $\mu^{(c)}$ represents the overall mean expense for category c, and the random effect $\alpha_{j_c}^{(c)}$ represents the deviation from $\mu^{(c)}$ for group j_c. Given the dispersion of expenses, model errors ($\epsilon_{kj_c}(\nu^{(c)})$) are specified using a T-distribution with $\nu^{(c)}$ degrees of freedom. The precision parameters for the random effects are denoted by $\psi^{(c)}$.

For this analysis, we employ a Bayesian framework (albeit, a classical analysis would provide quite similar estimates, yet model constraints would cause some minor issues), and require prior distributions for the parameters: $\{\mu^{(c)}, \phi^{(c)}, \psi^{(c)}\}$, for $c \in \{travel, other\}$. Prior distributions were specified by:

$$\begin{aligned} P(\mu^{(c)}) &\propto 1, \qquad \text{where } \mu^{(c)} \geq 0, \\ P(\phi_{j_c}^{(c)}) &\propto 1/\phi_{j_c}^{(c)}, \\ P(\psi^{(c)}) &\propto (1/\psi^{(c)})^{3/2}. \end{aligned} \tag{3}$$

The priors chosen for $\mu^{(c)}$, and $\phi_{j_c}^{(c)}$ represent standard reference priors; whereas, the prior for the random effects precision parameter has the form ${\psi^{(c)}}^{-(k+1)}$. While Hobert and Casella (Hobert & Casella 1996) show necessary and sufficient conditions for choosing k to ensure propriety of the model, Gelman (Gelman 2006) found $k = 1/2$ to perform well in general settings. Utilizing these hierarchical models, we are able to compute point estimates and obtain distributions for the cost of postponement for each distinct state group and field.

3 Results

For competitors who responded to the survey, we use the amounts reported as the estimated. For all other competitors who stayed an additional day, as determined by the race results, we use the model-predicted value. In total, the postponement of the event cost competitors an estimated $213,193 with a 95% credible interval of ($191,361 , $235,430). Figure 4 and Table 3 display the total additional expense breakdown by state group and Figure 5 and Table 4 show expenses by field.

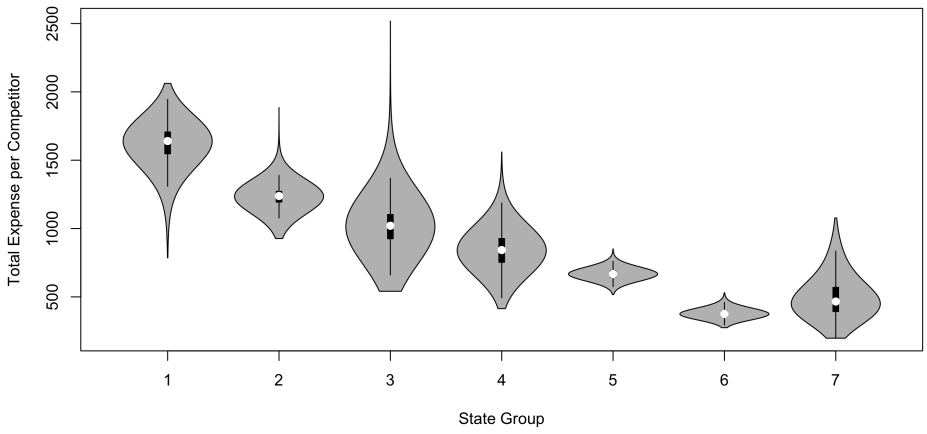

Figure 4: Posterior density of cost per competitor by state group. The white point represents the posterior mean.

Group	Mean	Median	95% Credible Interval
1	1,613	1,640	(1,270 , 1,789)
2	1,229	1,241	(1,051 , 1,356)
3	1,016	1,019	(692 , 1,354)
4	844	843	(586 , 1,116)
5	667	668	(599 , 737)
6	375	375	(318 , 443)
7	474	480	(302 , 760)

Table 3: Mean cost per competitor with 95% credible bounds by state group.

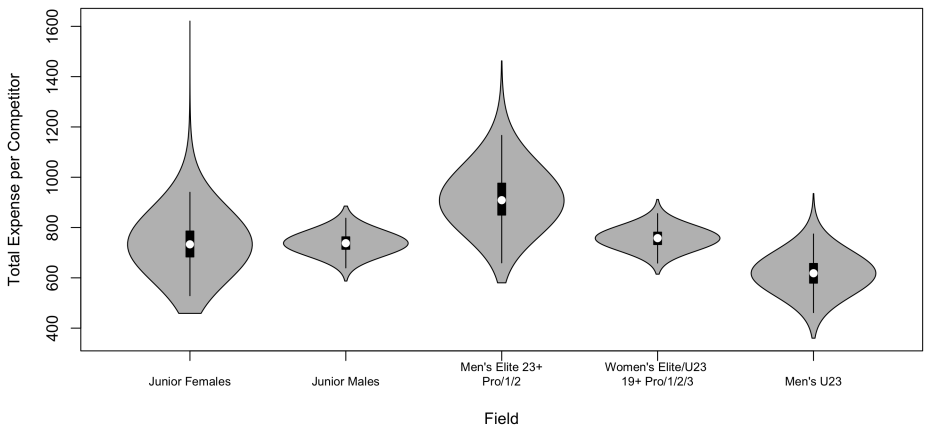

Figure 5: Posterior density of cost per competitor by field. The white point represents the posterior mean.

Group	Mean	Median	95% Credible Interval
Junior Females	737	735	(580 , 925)
Junior Males	736	738	(665 , 813)
Men's Elite 23+Pro/1/2	913	910	(748 , 1,122)
Women's Elite/U23 19+ Pro/1/2/3	757	758	(687 , 827)
Men's U23	620	619	(496 , 741)

Table 4: Mean cost per competitor with 95% credible bounds by field.

4 Conclusion

For many of the competitors, the 2015 Cyclo-cross National Championships was the biggest event of the year. Traveling from as far as Maine and Washington, many athletes made special arrangements to take time off from work, purchase expensive airline tickets, and book a hotel for an entire week. The excitement to compete in Austin, Texas, however, was matched only by the great disappointment of the postponement on Sunday, January 11. Collectively, 276 competitors incurred an estimated additional $213,193 in expenses. If only more careful steps were taken to see that both the APRD and the AHTF were on the

same page, each of the unlucky 276 would have returned home with an additional $753, on average. Hopefully, the shocking postponement in Zigler Park will lead the way for more communication in and planning of future cyclo-cross events.

References

Accuweather (2015), 'Austin,TX', www.accuweather.com/en/us/austin-tx/78701/january-weather/351193?monyr=1/1/2015&view=table. Accessed: 2015-09-15.

Albert, J. (1994), 'Exploring baseball hitting data: what about those breakdown statistics?', *Journal of the American Statistical Association* **89**(427), 1066–1074.

Albert, J. (2006), 'Pitching statistics, talent and luck, and the best strikeout seasons of all-time', *Journal of Quantitative Analysis in Sports* **2**(1).

Austin Heritage Tree Foundation (2015), 'Please help the 20 Zilker heritage trees impacted by the USA cyclo cross bike championship', www.austinheritagetreefoundation.com/Tree_Alerts.html. Accessed: 2015-09-15.

Carlin, B. P. (1996), 'Improved NCAA basketball tournament modeling via point spread and team strength information', *The American Statistician* **50**(1), 39–43.

Gelman, A. (2006), 'Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper)', *Bayesian Anal.* **1**(3), 515–534.

Glickman, M. E. & Hennessy, J. (2015), 'A stochastic rank ordered logit model for rating multi-competitor games and sports', *Journal of Quantitative Analysis in Sports* **11**(3), 131–144.

Hobert, J. P. & Casella, G. (1996), 'The effect of improper priors on Gibbs sampling in hierarchical linear mixed models', *Journal of the American Statistical Association* **91**(436), pp. 1461–1473.

Hoegh, A., Carzolio, M., Crandell, I., Hu, X., Roberts, L., Song, Y. & Leman, S. C. (2015), 'Nearest-neighbor matchup effects: accounting for team matchups for predicting March Madness', *Journal of Quantitative Analysis in Sports* **11**(1), 29–37.

Jensen, S. T., Shirley, K. E. & Wyner, A. J. (2009), 'Bayesball: A Bayesian hierarchical model for evaluating fielding in major league baseball', *Ann. Appl. Stat.* **3**(2), 491–520.

Leman, S. C., House, L., Szarka, J. & Nelson, H. (2014), 'Life on the bubble: Who's in and who's out of March Madness?', *Journal of Quantitative Analysis in Sports* **10**(3), 315–328.

VeloNews (2015), 'Cross nationals postponement cost over $250,000, according to survey', http://velonews.competitor.com/2015/01/news/cross-nationals-postponement-cost-250000-according-survey_359043. Accessed: 2015-09-15.

Wentworth, P. (2015), 'Tree Care Remediation Report USA Cyclocross Championship Zilker Park Austin, TX'.