ST440/540 – Mid-term exam 1 - Due February 18

THIS IS AN EXAM - DO NOT DISCUSS THE PROBLEM WITH ANYONE (INCLUDING OTHER STUDENTS OR THE TA)! If you have questions, please email me.

The Olympics are for mathletes too. In this exam you will explore the home-country advantage in the Summer Olympics by the rate of medals per participant in the host year compared to their previous Olympics. The data[1] are:

		Medals won		Participants	
Host country	Year	Previous	Host	Previous	Host
Finland	1952	24	22	129	258
Australia	1956	11	35	81	294
Italy	1960	25	36	135	280
Japan	1964	18	29	162	328
Mexico	1968	1	9	94	275
West Germany	1972	26	40	275	423
Canada	1976	5	11	208	385
Soviet Union	1980	125	195	410	489
United States	1984	94	174	396	522
South Korea	1988	19	33	175	401
Spain	1992	4	22	229	422
United States	1996	108	101	545	647
Australia	2000	41	58	417	617
Greece	2004	13	16	140	426
China	2008	63	100	384	599
Great Britain	2012	47	65	304	530
Brazil	2016	17	19	236	462
Japan	2021	41	51	395	621

Let Y_{i1} be the number of medals won by the host country during the Olympics i and Y_{i0} be the numer of medals won by the county in the previous Olympics. Similary, let N_{i0} and N_{i1} be the number of participates from the country in the corresponding Olympics. For example, Finland hosted the Olympics in 1952. They had $N_{11} = 258$ partipicants in 1952 and $N_{10} = 129$ participants in the 1948; they won $Y_{11} = 22$ medals in 1952 and $Y_{10} = 24$ medals in the 1948.

Analyze these data in the following sections:

1. **Aggregate analysis**: Combine data across all years for the host county in the host year, $Y_1 = \sum_{i=1}^{18} Y_{i1}$ and $N_1 = \sum_{i=1}^{18} N_{i1}$. Conduct a Bayesian analysis of λ_1, the expected number of medals per participant in their home country. State a reasonable likelihood, an uninformative conjugate prior distribution and give the posterior distribution. Repeat this analysis using the data from the previous year, $Y_0 = \sum_{i=1}^{18} Y_{i0}$ and

[1] Taken from NPR and fivethirtyeight.com

$N_0 = \sum_{i=1}^{18} N_{i0}$, to estimate λ_0. Compare these two posterior distributions in a figure. What are the main assumptions in your analysis and do you think they are valid?

2. **Hypothesis test**: Conduct a Bayesian test of the hypothesis that there is a home-county advantage, i.e., $\lambda_1 > \lambda_0$. Clearly state your hypotheses and describe methods you are using for the test. Are your results sensitive to the prior?

3. **Prediction**: The next Olympics will be held in France in 2024. In 2021, France had 398 partipicants and won 33 medals (we do not know the number of participants in 2024 so you will have to predict this value). Predict the number medals France will win in the 2024 Olympics and quantify your uncertainty about this prediction. Clearly describe the methods you are using to make this prediction.

4. **Country-specfic analysis**: Conduct an analysis separately by country (combine the data across the two Olympics for AU, Japan and USA, so there are a total of 15 counties). Describe the likelihood, prior and posterior. Compare the posterior distribution of the ratio $r = \lambda_1/\lambda_0$ for each county. Is there evidence that the home-country advantage differs by country?

5. **Conclusions**: Summarize your main findings and enumerate at least two limitations to the your analysis, and suggest how these could be addressed in future work.

Your paper should be written as a professional document with full sentences and paragraphs, clearly labeled and numbered figures and tables, and few spelling/grammar errors. Organize your report with five sections, labelled as above. You should include enough detail that another student in class could reproduce your results. Summarize your analysis in a PDF document that is **no more than four pages long** (excluding code). Append your code to the end of this document and submit a single document on moodle.

HAVE FUN!