IST772 Week 11: Final Examination

Instructions: You have received three data sets, which are described below. Each dataset is in RData format, which means that you can simply open it with the open dialog on the Environment tab and it will read in as an R object. This will save time and effort in preparing your data.

Your goal for this final exam is to conduct the necessary analyses and then write up a technical report for a scientifically knowledgeable staff member in a state legislator's office. Thus, you should provide sufficient numeric and graphical detail that the staff member can create a comprehensive briefing for a legislator. You can assume that the staff member understands the concept of statistical significance. Your report should include a few graphics created by R, keeping in mind that you must provide some accompanying text to explain each graphic that you include in your report.

This exam is open book and open notes, but you may not receive assistance, help, coaching, guidance, or support from any human except your section instructor. Your section instructor will be available by email throughout the exam period: If you are stuck on an R code problem, make sure to include your complete code in the email, preferably as a file attachment.

These three data sets all pertain to vaccinations. The first and second datasets are the same for everyone and are mainly included to provide context for interpretation of the results. Most of the substantive analyses occur in reference to the third dataset. This third dataset is different for every student and results will vary depending upon the sample the student received.

The datasets are:

- usVaccines.Rdata
- allSchoolsReportStatus.RData
- districtsX.RData

Here is a description of each dataset:

usVaccines.Rdata – Time series data from the World Health Organization reporting vaccination rates in the U.S. for five common vaccines

```
Time-Series [1:38, 1:5] from 1980 to 2017:
 - attr(*, "dimnames")=List of 2
  ..$ : NULL
  ..$ : chr [1:5] "DTP1" "HepB_BD" "Pol3" "Hib3" "MCV1"...
```

(Note: DTP1 = First dose of Diphtheria/Pertussis/Tetanus vaccine; HepB_BD = Hepatitis B, Birth Dose; Pol3 = Polio third dose; Hib3 – Influenza third dose; MCV1 = Measles first dose)

allSchoolsReportStatus.RData – A list of California kindergartens and whether they reported vaccination data to the state in 2013

'data.frame': 7381 obs. of 3 variables:
$ name : Name of the school
$ pubpriv : "PUBLIC" or "PRIVATE"
$ reported: "Y" or "N"

districtsX.RData – (Where X is the number of your particular dataset) A sample of California public school districts from the 2013 data collection, along with specific numbers and percentages for each district:

'data.frame': 700 obs. of 13 variables:
$ DistrictName : Name of the district
$ WithoutDTP : Percentage of students without the DTP vaccine
$ WithoutPolio : Percentage of students without the Polio vaccine
$ WithoutMMR : Percentage of students without the MMR vaccine
$ WithoutHepB : Percentage of students without the Hepatitis B vaccine
$ PctUpToDate : Percentage of all enrolled students with completely up-to-date vaccines
$ DistrictComplete: Boolean indicating whether or not the district's reporting was complete
$ PctBeliefExempt : Percentage of all enrolled students with belief exceptions
$ PctChildPoverty : Percentage of children in the district living below the poverty line
$ PctFreeMeal : Percentage of children in the district eligible for free student meals
$ PctFamilyPoverty: num Percentage of familes in the district living below the poverty line
$ Enrolled : Total number of enrolled students in the district
$ TotalSchools : Total number of different schools in the district

The research questions for you to explore with these three data sets are as follows:

Introductory/Descriptive Reports:

1. How have U.S. vaccination rates varied over time? Are vaccination rates increasing or decreasing? Which vaccination has the highest rate at the conclusion of the time series? Which vaccination has the lowest rate at the conclusion of the time series? Which vaccine has the greatest volatility?
2. What proportion of public schools reported vaccination data? What proportion of private schools reported vaccination data? Was there any credible difference in overall reporting proportions between public and private schools?
3. What are 2013 vaccination rates for individual vaccines (i.e., DOT, Polio, MMR, and HepB) in California public schools? How do these rates for individual vaccines in California districts compare with overall US vaccination rates (make an informal comparison to the final observations in the time series)?
4. Among districts, how are the vaccination rates for individual vaccines related? In other words, if students are missing one vaccine are they missing all of the others?

Predictive Analyses:
(For all of these analyses, use PctChildPoverty, PctFreeMeal, PctFamilyPoverty, Enrolled, and TotalSchools as predictors. Transform variables as necessary to improve prediction and/or interpretability. In general, if there is a Bayesian version of an analysis available, you are expected to run that analysis in addition to the frequentist version of the analysis.), response

5. What variables predict whether or not a district's reporting was complete?
6. What variables predict the percentage of all enrolled students with completely up-to-date vaccines?
7. What variables predict the percentage of all enrolled students with belief exceptions?
8. What's the big picture, based on all of the foregoing analyses? The staff member in the state legislator's office is **interested to know how to allocate financial assistance to school districts to improve both their vaccination rates and their reporting compliance**. What have you learned from the data and analyses that might inform this question?

Submit your exam as a PDF prior to the deadline established by your instructor. If you used Markdown, make sure to include your Markdown file. Otherwise, your submission should include, as an appendix, complete and commented R code that reproduces all of your analyses. Your exam will be graded on the basis of clarity; conciseness; inclusion and explanation of specific and appropriate statistical values; inclusion of both traditional and Bayesian inferential evidence; inclusion of suitable tabular and graphical displays.