

## 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 should provide some accompanying text to explain each graphic.

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. Most of the substantive analyses occur with the third dataset. This third dataset is different for every student and results will vary depending upon the sample the student received. Here is a description of each dataset:

**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"

**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)

**reportSampleX.RData** – (Where X is the number of your particular dataset) A sample of California kindergartens that reported vaccination data, along with specific numbers and percentages for each school in the sample:

'data.frame': 698 obs. of 13 variables:

\$ code : CA School Code Number (included for completeness, can be ignored)  
\$ name : Name of the school  
\$ pubpriv : Factor w/ 2 levels "PRIVATE","PUBLIC"  
\$ enrollment : An integer indicating the number of students enrolled at the school  
\$ allvaccs : The percent of enrolled students who had documented all required vaccinations  
\$ conditional: The percent of enrolled students needed follow-up on their records  
\$ medical : The percent of enrolled students with a medical exemption  
\$ religious : The percent of enrolled students with a religious/belief exemption  
\$ dptMiss : The percent of students missing the Diphtheria/Pertussis/Tetanus vaccine  
\$ polMiss : The percent of students missing the Polio vaccine  
\$ mmrMiss : The percent of students missing the Measles/Mumps/Rubella vaccine  
\$ hepMiss : The percent of students missing the Hepatitis vaccine  
\$ varMiss : The percent of students missing the Varicella (chickenpox) vaccine

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

*Introductory/Descriptive Reports:*

1. What proportion of public schools reported vaccination data?
2. What proportion of private schools reported vaccination data?
3. Have U.S. vaccinations rates been stable over time?
4. Are there any notable patterns in U.S. vaccinations rates over time?

*Public vs. Private School Comparisons:*

5. Was there any credible difference in overall reporting proportions between public and private schools?
6. Compare overall vaccination rates (allvaccs) between public and private schools. Are there any credible differences?
7. Compare medical exemptions between public and private schools. Are there any credible differences?
8. Compare religious/belief exemptions between public and private schools. Are there any credible differences?

*Predictive Analyses:*

9. Is it possible to predict whether a school is public or private based on conditional, medical, and religious percentages? If so, what are the specifics?
10. Is it possible to predict conditional percentage, based on the percentages of specific vaccines that are missing? If so, what are the specifics?

11. Is it possible to predict medical percentage, based on the percentages of specific vaccines that are missing? If so, what are the specifics?
12. Is it possible to predict religious percentage, based on the percentages of specific vaccines that are missing? If so, what are the specifics?
13. What's the big picture, based on all of the foregoing analyses?

Develop your responses in an RStudio notebook or word processor file and submit your exam as a PDF file (only) prior to the deadline established in the syllabus. Your submission should include complete R code to reproduce all of your analyses. Your exam will be graded on the basis of completeness; 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.