

Fall 2020 - Final Examination Specification

Instructions: You have an .RData available on Blackboard in the Groups area that contains two data sets, which are described below. Because your custom data is in the .RData format, you can simply open it with the open dialog on the Environment tab and it will read in as two R objects. The first block of code in the associated Markdown file does this for you.

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 California 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 and other basic concepts like mean, standard deviation, and correlation. Your report can include graphics created by R, keeping in mind that if you do include a graphic, you will have to provide some accompanying narrative text to explain what it is doing in your report.

Because the analysis portion of the exam (conducted in class during Week 13) begins with a Markdown file, you will need your laptop with a working copy of R and R-Studio installed. The Markdown file also depends upon the following packages (and their various dependencies), so install these packages now if you do not have them:

- readr
- changepoint
- BayesFactor
- MCMCpack

You may not receive assistance, help, coaching, guidance, or support from any human except your instructor at any point during this exam, including during the report writing period that follows the in-class exam session. You will (obviously) have your computer open during the exam, but there should be little need to look up anything in the book or elsewhere, as the Markdown file contains examples that you can adapt.

After the in-class portion of the exam is complete and you have submitted your analytic results to Blackboard, you will have one week to complete a written report using the instructions below. Your instructor will be available by email throughout the report writing period if you have questions, but don't wait until the last minute!

Your data sets pertain to vaccinations for the U.S. as a whole and for Californian school districts. The U.S. vaccine data is a time series and the California data is a sample of end of year vaccination reports from n=700 school districts. Here is a description of both datasets:

usVaccines – 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)

districts – A sample of California public school districts from the 2017 data collection, along with specific numbers and percentages for each district:

```
'data.frame': 700 obs. of 12 variables:
 $ DistrictName      : Name of the district
 $ WithoutDTP        : Percentage of students in the district without the DTP vaccine
 $ WithoutPolio      : Percentage of students in the district without the Polio vaccine
 $ WithoutMMR        : Percentage of students in the district without the MMR vaccine
 $ WithoutHepB       : Percentage of students in district without Hepatitis B vaccine
 $ PctUpToDate       : Percentage of students with completely up-to-date vaccines
 $ DistrictComplete  : Boolean showing whether or not district's reporting was complete
 $ PctBeliefExempt   : Percentage of all enrolled students with belief exceptions
 $ PctChildPoverty   : Percentage of children in district living below the poverty line
 $ PctFamilyPoverty  : Percentage of families in 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 Markdown file contains lots of instructions and suggestions for descriptive and inferential statistics that you can add or modify before knitting your file, transforming it to PDF, and submitting it to Blackboard. Take note of the following exam questions that will guide the development of your report. Keep in mind that your Blackboard submission of the Markdown file must contain everything that you need in order to write your report. The list below contains 14 elements. Each of these elements will count for two points in your grade for the final.

Descriptive Reporting:

1. Basic Introductory Paragraph: In your own words, write about three sentences of introduction addressing the staff member in the state legislator's office. Frame the problem/topic that your report addresses.
2. Descriptive Overview of U.S. Vaccinations: You have U.S. vaccination data going back 38 years, but the staff member is only interested in recent vaccination rates as a basis of comparison with California schools.
 - a. How have U.S. vaccination rates varied over time?
 - b. Are there notable trends or cyclical variation in U.S. vaccination rates?
 - c. What are the mean U.S. vaccination rates *when including only recent years* in the calculation of the mean?
3. Descriptive Overview of California Vaccinations: Your **districts** dataset contains four variables that capture the individual vaccination rates by district: WithoutDTP, WithoutPolio, WithoutMMR, and WithoutHepB.
 - a. What are the mean levels of these variables across districts?

- b. 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?
 - c. Descriptively speaking, how do these Californian vaccination levels compare to U.S. vaccination levels (recent years only). Note that no inferential tests are necessary to answer this question descriptively.
4. Conclusion Paragraph for Vaccination Rates: Provide one or two sentences of your professional judgment about where California school districts stand with respect to vaccination rates and in the larger context of the U.S.

Inferential Reporting:

For every item below except 7, use PctChildPoverty, PctFamilyPoverty, Enrolled, and TotalSchools as the four predictors. Transform variables as necessary to improve prediction and/or interpretability. In general, if there are Bayesian versions of an analysis available, you are expected to run those in addition to the frequentist version of the analysis. In your final report for each question, make sure you write a narrative with complete sentences. You can choose to put important statistical values into a table for readability, or you can include the statistics within your narrative. Make sure to include enough statistical information so that another analytics professional could review your work.

5. Which of the four predictor variables predicts *the percentage of all enrolled students with belief exceptions*?
6. Which of the four predictor variables predicts *the percentage of all enrolled students with completely up-to-date vaccines*?
7. Using any set of predictors that you want to use, what's the best R-squared you can achieve in predicting *the percentage of all enrolled students with completely up-to-date vaccines*?
8. In predicting *the percentage of all enrolled students with completely up-to-date vaccines*, is there an interaction between PctChildPoverty and Enrolled?
9. Which, if any, of the four predictor variables predict *whether or not a district's reporting was complete*?
10. Concluding Paragraph: Describe your conclusions, 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**. Make sure you have at least one sentence that makes a recommendation about improving vaccination rates. Make sure you have at least one sentence that makes a recommendation about improving reporting rates.

For this exam, the report writing is very important: Your responses will be graded on the basis of clarity; conciseness; inclusion and explanation of specific and appropriate statistical values; inclusion of both frequentist and Bayesian inferential evidence; explanation of any included tabular material and the appropriate use of graphical displays when/if necessary.