**AML 612 Homework #7**

**For all questions, please submit your R code, and a doc file with a copy of the screen output and plots (where applicable), and all other relevant files to** **smtowers@asu.edu****.**

**Due Wed April 11th at noon**

**The code in your R files should exhibit all of the good coding practices mentioned in** [**http://sherrytowers.com/2012/12/14/good-programming-practices-in-any-language/**](http://sherrytowers.com/2012/12/14/good-programming-practices-in-any-language/)

**Plots should exhibit all of the good plotting practices mentioned in** [**http://sherrytowers.com/2013/01/04/good-practices-in-producing-plots/**](http://sherrytowers.com/2013/01/04/good-practices-in-producing-plots/)

**Please submit your files in a format hwk7\_<last name>.xxx where xxx is the format of the document (R, doc, tex, pdf, bib, etc)**

**Question 1)**

The project groups:

Crime and student performance

   Rick

   Bechir

   Cesar

State alcohol laws and DUI fatalities

   Karen

   Juan

Bullying and academic achievement

   Mugdha

   Caleb

   Josean

**a)** From now until the end of the semester, each project group must continue to schedule a regular weekly meeting with me to consult on their project. Based on the consultations, project groups will write the R code to perform the statistical analyses associated with their project. The analysis will expand upon the Results section, and include relevant plots overlaying model fits on data, and relevant tables. All figures and tables have to be properly captioned, and all figures must be properly labelled and of publication quality.

b) Based on the results of the analysis, fill in the Abstract and Results sections, and begin filling in the Discussion section of paper, comparing and contrasting the results of the analysis to those of prior publications in the literature, and explaining how the results expand upon the prior literature.

**Question 2)**

For our class publication project, I have decided that we will be examining metrics for unsafe firearm storage and carry practices in the US. We will be aiming this at a public health journal, and I expect it will be a high impact publication.

Note: participating as a co-author in this publication is not required, but is an option for interested students once the course ends. It is, however, required of all students that they complete these homework exercises whether they intend to participate as a co-author on the future paper or not.

**TSA firearm detection data:**

Since October, 2011 the TSA has maintained a website providing weekly updates on the prohibited items discovered in carry-on baggage at TSA airport security passenger

checkpoints (available at [www.tsa.gov/blog](http://www.tsa.gov/blog)).

The data include the details of individual incidents of firearms detected, including the airport at which each firearm was detected, the date it was found, the caliber of the firearm, whether it was loaded, and whether it also had a round chambered.

It was only after partway through December, 2011, however, that the TSA listed the

details of unloaded firearms in addition to the loaded firearms.

In the file tsa\_data\_by\_state\_by\_month.csv I give the total number of firearms detected at airports in the US by state and month, the number that were loaded, and the number that were loaded with a round chambered. The data can be found at:

<https://www.dropbox.com/s/ydt30ytenh9oh1a/tsa_data_by_state_by_month.csv?dl=0>

a) Write the R code to read in the TSA data, create a new field in the TSA data set that is the date expressed in fractions of a year: date=year+(month-0.5)/12

Note that shifting the month by 0.5 ensures that all months during the year 2013 (for example) will have dates larger than 2013.0, and smaller than 2014.0. Thus, these dates represent the middle of each month over which the data are aggregated.

Using this date, write the R code to aggregate the data and produce the plots below showing the temporal change in the probability of finding a firearm to be loaded, and the probability of finding a loaded firearm to also have a round chambered. Note: you can introduce a line break to have multi-line axis labels using “\n”.

Using the R glm() method, regress the logarithm of the odds on the date, and overlay the fitted regression lines over the data, using different colours than I used.



b) based on the results of the regressions, insert numbers to complete the following statements, paying attention to significant figures. As a general rule of thumb, include enough numbers past the decimal point to have two significant figures in the one standard deviation uncertainty. These statements are how such results are quoted in a manuscript:

The log-odds of the TSA detecting a firearm to be loaded increased by XXX per year, 95% CI [XXX,XXX] (logistic linear regression p-value p=XXX).

The log-odds of the TSA detecting a loaded firearm to also have a round chambered increased by XXX per year, 95% CI[XXX,XXX] (logistic linear regression p-value p=XXX).

Note: if the p-value is less than 0.001, do not use the = sign in the above. Instead, state p<0.001.

c) Tests of TSA screening in the past have revealed a 95% failure rate in detecting prohibited items. TSA detection efficiency may have changed over time as their screening methods have changed in light of those tests.

<http://www.latimes.com/opinion/topoftheticket/la-na-tt-tsa-airport-security-charade-20150608-story.html>

Does that affect this analysis?

d) Comparison to past data assessing prevalence of firearm ownership and risky firearm storage practices in the US:

Every year the CDC publishes the results of their Behavioural Risk Factor Surveillance System (BRFSS). The survey consists of thousands of telephone calls to people in all 50 states, asking respondents about health risk factors like seatbelt use, smoking, obesity, etc:

<https://en.wikipedia.org/wiki/Behavioral_Risk_Factor_Surveillance_System>

<https://www.cdc.gov/brfss/index.html>

Up until 2004, the CDC also included questions about firearm ownership and storage practices in the core questions on the survey. Since the lapse of the inclusion of those questions, it has been difficult to determine whether or not risky storage practices (such as storing firearms loaded and/or unlocked) have changed.

The most recent publication involving BRFSS firearm storage data appears to be a 2005 study based on the 2002 BRFSS data:

[Okoro, Catherine A., et al. "Prevalence of household firearms and firearm-storage practices in the 50 states and the District of Columbia: findings from the Behavioral Risk Factor Surveillance System, 2002." *Pediatrics* 116.3 (2005): e370-e376.](http://pediatrics.aappublications.org/content/116/3/e370.full)

In Table 1 of the Okoro et al paper, they show the survey estimates of household firearm ownership by state, and the estimate of households that store firearms loaded, and the estimate of households that store firearms loaded and unlocked.

I have put this data in the file okoros\_et\_al.csv.

The data can be found in the file: <https://www.dropbox.com/s/sphkcync7g721ry/okoros_et_al.csv?dl=0>

The TSA has stated that most people who are caught with firearms at security checkpoints claim that they had stored the firearm in their luggage at home, and had forgotten about it. If this is true, then (naively at least) we would expect that the fraction of firearms by state that the TSA finds loaded would be correlated to the percentage of households that stored their firearms loaded in 2002 divided by the percentage of households that owned firearms. While the Orokos et al 2002 data are from 10 to 15 years prior to the TSA data, most of the same people from 2002 still live within the same state, and still own firearms…

Aggregate the TSA data set total firearms and loaded firearms by state. For states where at least 100 firearms were found between 2012 to 2017, and for which the Orokos et al paper had data for the fraction of households with loaded firearms, make the following plot, using different colours than I used.

We require at least 100 firearms to be found to ensure statistically reliable estimates of the fraction found loaded (ie; it helps reduce the stochasticity on the estimated fraction).



e) Because we require that there be at least 100 firearms detected by the TSA, and because thousands of households are surveyed in the BRFSS survey, the two sets of fractions in the data sets have approximately Normal stochasticity ([in the limit of large N\*p\*(1-p) the Binomial distribution approaches the Normal](https://en.wikipedia.org/wiki/Binomial_distribution#Normal_approximation)). We can thus examine the correlation of these two data sets.

Using the R cor.test() function, calculate the Spearman rho correlation between the two data sets, and complete the following statement (which is how it would be presented in a paper):

The correlation between the two data sets is rho=XXX (Spearman rho p-value p=XXX).

(again, if the p-value is less than 0.001, put p<0.001).

**Note, if your data set has something other than 32 rows in the above calculation, you’ve done something wrong.**

**Question 3)**

Find at least three papers in the literature on three different subjects selected from the following list (online reports from organisations like the Pew Research Center or Gallup are acceptable, if properly documented)

1. firearm ownership estimates in the US, by state (even better if also by year)
2. estimates of firearm storage practices by state: firearms stored loaded
3. estimates of firearm storage practices by state: firearms stored unlocked
4. estimates of the public health burden of firearm mortality and injury in the US (deaths, hospitalisations, cost to society, etc)
5. risk factors for firearm injury and mortality in the general population
6. risk factors for firearm injury and mortality in children
7. estimates of what fraction of the population carries or stores firearms with a round chambered

Create a latex file summarizing the documents according to the rubric of Lacum et al. Also submit an annotated bibtex file for the documents you find. These bibtex files will be combined to form the bibtex file for the paper.