STAT 740 – Spring 2019 – Exam 2

*Directions:*

* *This exam consists of four (4) questions that are not equally weighted; the number of points a problem is worth is indicated next to the item stem.*
* *Sub-parts may also be worth different point values, but these will be determined during grading.*
* *Each problem is on a different page for convenience.*
* *You should provide your answers in this file, in the space below each sub-part. It would be helpful if you would change the font color of the questions or answers to something else (e.g. red) to help with scoring, but this is optional.*
* *You may submit your answers in LaTeX form instead of using this file if you so choose.*
* *You should submit your answer file (e.g. this file) as well as a separate file containing your R code.*
* *If necessary, you may insert photographs of handwritten work in lieu of a typed answer. This should be rare.*
* *You have 72 hours from when you were emailed this exam to submit it.*
* *Submissions should be made to the Dropbox on D2L.*
* *You may reference our course materials (e.g. the book, slides, notes, etc.), but you may not use outside resources (e.g. Google).*
* *If you have any questions, email me (*[whitakerdo@uwstout.edu](mailto:whitakerdo@uwstout.edu)*).*
* *The total exam is graded out of* ***25 points****.*

1. [4 points] An economist compiled data on productivity improvements last year for a sample of firms producing electronic computing equipment. The firms were classified according to the level of their average expenditures for research and development in the past three years (low, moderate, high). The results of the study are in the file prodimp.csv (productivity improvement is measured on a scale from a to 100). You may assume the one-way ANOVA model is appropriate.
   1. Obtain the ANOVA table.
   2. Test whether or not the mean productivity improvement differs according to the level of research and development expenditures. Use .
   3. If you determine that productivity improvement differs, produce appropriate post-hoc confidence intervals for the differences.
   4. Interpret your findings in the context of the problem.
2. [6 points] As part of a study on its sheet metal assembly process, a major automobile manufacturer uses sensors that record the deviation from the nominal thickness (millimeters) at six locations on a car. The first four are measured when the car body is complete and the last two are measured on the underbody at an earlier stage of assembly. Data on 50 cars are given in the file T5-14.dat.
   1. The process seems stable for the first 30 cases. Use these cases to estimate and  . Then construct a chart using all of the variables. Include all 50 cases.
   2. Which individual locations seem to show a cause for concern?
   3. These are all measured deviations from target value so it is appropriate to test the null hypothesis that the mean vector is zero. Using the first 30 cases, test at .
3. [10 points] The dataset for this problem comes from the State of New York. Their description of the dataset is:

*The Behavioral Health Organization (BHO) initiative oversees the transition to managed care for Medicaid recipients who receive mental health (MH) and substance use disorder (SUD) services in New York State. The metrics emphasize improving rates of timely follow-up treatment post discharge, timely filling of appropriate medication prescriptions post discharge, and reducing rates of readmission. The BHO Medication Fill dataset is designed to assess the frequency with which individuals receiving Mental Health or SUD treatment fill an initial prescription for a mood stabilizer, anti-psychotic, psychotropic, or anti-addition medication within 30 days of discharge from a mental health inpatient hospitalization or SUD inpatient stay and refill those prescriptions within 100 days of discharge. The year 2015 saw the conclusion of the first phase of the Behavioral Health Organization initiative (BHO). A new Behavioral Health Managed Care Transition phase II is underway. The data contained in the BHO metrics span 2010 to 2014, using the 2010 calendar year for a baseline. Earlier in the program (2011‐2012) the metrics were calculated quarterly and on a year‐to‐date basis, later in (2013‐2014), New York State Office of Mental Health opted for semi‐annual and year‐to‐date aggregations.* (<https://data.ny.gov/Human-Services/BHO-Medication-Fill-Data-2010-2014/cc9j-4ujx>)

For this problem, determine if there is evidence that the true rate for each of the following metrics is different from 75%:

* Percentage of MH Discharges where a Prescription for a Mood Stabilizer/Antidepressant Medication was Filled within 30 Days
* Percentage of MH Discharges where a Prescription for a Psychotropic Medication was Filled within 30 Days
* Percentage of MH Discharges where a Prescription for an Antipsychotic Medication was Filled within 30 Days

Conduct a full analysis including checking assumptions. (Examining “first-fill” rate is important because not filling prescriptions is associated with worse patient outcomes, e.g. <http://www.healthaffairs.org/do/10.1377/hblog20150401.046033/full/> The choice to compare the above metrics to 75% is based on failure rates described at the beginning of this article.)

Note: The data file (BHO\_Medication\_Fill\_Data\_\_2010-2014.csv) is available from the link above but is also provided with the exam. Because of the format of the raw data, R code (BHO-loading.R) is provided for loading the dataset into an appropriate format for analysis.

1. [5 points] The dataset for this problem comes from the City of Bloomington, Indiana. Their description of the dataset is:

*Two disinfectant byproducts (DBPs), Total Trihalomethanes (TTHMs) and Haloacedic Acids (HAA5s) are regulated by the Safe Drinking Water Act through the Stage 2 Disinfectant Byproduct Rule. The City of Bloomington Utilities samples for DBPs at 8 testing locations throughout the distribution system on a quarterly compliance schedule. Compliance is calculated on a Locational Running Annual Average (LRAA). The Maximum Contaminant Level for TTHMs is 80 ug/l and 60 ug/l for HAA5s. Micrograms per liter (ug/l) is equal to parts per billion (ppb). One part per billion is roughly equal to a teaspoon of water in an Olympic-size swimming pool*. (<https://catalog.data.gov/dataset/disinfectant-byproduct-data>)

Note: The raw data is available online at the link above, but two data files with appropriate subsets of the data and associated R code for merging them are included with the exam. Only the “Hospital” location is used for this problem. (For example, other locations and sampling time periods that are not of interest for this problem are excluded.)

* Files: TTHM.csv, HAA5.csv, DBP-loading.R

Construct appropriate multivariate quality control charts for these data. Interpret them.