STAT 830 Due 16 September

This first problem set is partly review. I want to see how you answer relatively elementary problems. I don't plan to discuss these with anyone before they are handed in and I want complete clear explanations about what you are doing and assuming. Nothing I have said in class is particularly relevant to problems 1 or 2.

This assignment will be evaluated for writing and for quality of graphs. Throughout this course certain questions will have an extra part: why did Richard ask this question? For this assignment I don't need you to write answer this meta-question but I do ask you to think about it.

Problems: Assignment 1

1. The concentration of cadmium in a lake is measured 17 times. The measurements average 211 parts per billion with an SD of 15 parts per billion. Could the real concentration of cadmium be below the standard of 200 ppb? I want an answer in the form of a paragraph with NO formulas, no Greek letters. An answer is not 1 word long. I also want this turned in by email in the form of a pdf document produced in LaTeX.

2. Suppose X and Y are independent Geometric(p) random variables. In other words for non-negative integers j and k

$$\mathrm{P}(X = j \text{ and } Y = k) = \mathrm{P}(X = j)\mathrm{P}(Y = k) = p^2(1-p)^{j+k}\,.$$

   WARNING: there are two standard definitions of Geometric distributions. The formula above specifies which I am talking about.

   (a) Let $U = \min(X, Y)$, $V = \max(X, Y)$ and $W = V - U$. Express the event $U = j$ and $W = k$ in terms of X and Y.

   (b) Compute $\mathrm{P}(U = j)$ and $\mathrm{P}(W = k)$ and prove that the event $U = j$ and the event $W = k$ are independent.

3. Each month Statistics Canada publishes data on employment and unemployment in Canada arising from the Labour Force Survey (LFS). My home page has a link to the Daily. On September 11 there will be a release of August data. Navigate from the Daily page to the LFS

page and download the pdf – about 360 Kilobytes. If necessary try the August release by going back to the second Friday ini August. I want you each to get one estimate from the tables in that document and compare it to the standard error for that estimate as follows:

- Use the last digit of the day of the month which is your birthday to pick a province: 0 for Newfoundland, 1 for PEI, 2 for Nova Scotia, 3 for New Brunswick, 4 for Quebec, 5 for Ontario, 6 for Manitoba, 7 for Saskatchewan, 8 for Alberta, and 9 for BC. The data for that province is in one of tables 4 or 5. Go to the section for men, 25 years and over, or women, 25 years and over according to whether or not you are male or female OR toss a coin to pick between those two possibilities. Please tell me what table and row you end up at.
- Get the estimated change, August minus July, and the associated Standard Error, the column labelled 'S.E.'.
- Does a 1 standard error confidence interval include the value 'no change'?
- Suppose that X has a normal distribution with mean μ and standard deviation σ. (In this question I have in mind that σ is known and you are making a confidence interval for μ which is unknown.) Give a formula, as a function of μ and σ, for the probability that a one standard deviation confidence interval includes the value 0. Your answer should be expressed as an integral involving the standard normal density

$$\phi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}.$$

- Show that this probability is maximized when $\mu = 0$ and tell me what the maximum probability is.

4. The summary to lecture 1 contains a link to the `R` code I used for the Pearson-Lee height data. I used vertical strips centred at round numbers of inches for the fathers' heights in approximating the graph of $\mathrm{E}(S|F)$. I want you to try bars with widths 0.2, 0.4 ,0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8 and 2.0. Present one graph, properly labelled and captioned

for the width 0.2 case which is otherwise like the one I showed in class. Add to that graph the corresponding approximation to $\mathrm{E}(F|S)$. (This is a different line than the other one.) Present another graph which permits comparison of the effects of bin width on the plots; you might use colours or line types in `R`, for instance. You can use my R code to get the data I used.

5. Monte Hall was a television game show host from North Vancouver. He hosted, long ago, a popular show called Let's Make a Deal. In the show he would give contestants small prizes and offer to trade them the prize for some other thing that might or might not be more valuable. At the end of the show the two contestants who had one the most valuable prizes were allowed to trade in those prizes for a chance to play for the big prize of the day which was hidden behind one of three doors.

   This setting prompted someone, I don't know who, to dream up the following simplified version — a version which does not match the TV show as I remember it. In this version a single contestant picks a door, Monte then opens another door to show the contestant that this other door does not have a prize and then offers the contestant the opportunity to switch to the third door – the one which is not the one the contestant first picked and not the one Monte Hall opened.

   Please describe a sample space for this experiment and make a choice of probabilities for all the elementary outcomes. Explain clearly exactly why you choose these probabilities. Are they based on long run relative frequency ideas or are they subjectively motivated?

   Finally: should the contestant take the offer? Explain carefully why. I want to see the formal rules of probability used here — no hand waving.

Due date: 16 September 2020; please email me a scan or a photograph or a pdf or . . . by the end of the day.