WINTER 2019 TERM 2   EPSE 482: ASSIGNMENT #1
DUE DATE: FEB 8th (Section 075)

General Instructions: Please print your answers on separate pages. You don't need to include the question (saves paper) but do include the number and letter of the question you are answering following my system (i.e. 1A, 1B,..., 2A, 2B..., etc.). For answers where computations are asked for, I'd recommend you show work. It's O.K. if you don't (UNLESS I ASK YOU TO), but then you won't get partial marks if your answer is wrong. You don't need to type out the answers to the math questions. I'll take hand-written answers IF they are neatly organized and I can follow the work easily. You can work in groups but I need individual assignments handed-in.

Question 1: A researcher is interested in investigating the eating habits of women with children aged 12 months or less in the town where she lives. She is specifically interested in knowing whether or not women with infant children follow diets higher in iron compared to the national average. She begins by obtaining a list of all the women in the town from the census office. She then proceeds to roll a fair die and only selects the names of women if it the die falls on an even number.

Once she has obtained a list of these women she calls them on their phone and asks the following 4 questions:

  (a) *Do you have children aged 12 months or less? Yes/No*

  (b) *Health Canada recognizes 4 food groups. Order them in preference from your most favourite one using the number "1" and your least favourite one using "4":*
      _____*Vegetables & Fruits* _____*Grains & Cereals* _____*Dairy* _____*Meat*

  (c) *What percentage of an average working day for you is spent cooking?*

  (d) *How much money do you spend each week on your groceries?*

From those women who answered 'yes' to the first question, she asks for a follow-up interview. She proceeds to interview the women who were O.K. with her request.

1A. Does the group of women the researcher called on their telephones constitute a random sample? **WHY?**

1B. Each one of the 4 questions in our researcher's survey constitute an example from a different type of variable as defined by Steven's levels of measurement. Identify which level of measurement corresponds to which question (Hint: This time there is only ONE right answer for each option).

1C. Is the hypothesis related to iron consumption scientific or unscientific? **WHY?** How could she, potentially, test this hypothesis?

1D. What is the sample and what is the population of the women selected for the telephone interview?

1E. Can our researcher generalize her findings from the follow-up, in-person interview to all women with infant children in her town? **WHY?**

1F. Can our researcher generalize her findings from the telephone interview to all women with infant children in her town? **WHY?**

1G. Can our researcher generalize her findings from the follow-up, in-person interview to ALL women her town? **WHY?**

1H. Can our researcher generalize her findings from the telephone interview to ALL women her town? **WHY?**

Question 2: Assume you have **3 fair coins** that you are tossing. (Hint: It might be useful to have a table or list that presents all possible combinations of "heads" and "tails").

      2A. What is the sample space?

      2B. What is the probability of obtaining 3 consecutive heads?

      2C. What is the probability of NOT landing any heads?

      2D. What is the probability of landing the combination heads–tails–heads OR tails–heads–tails?

      2E. What is the probability of landing the combination heads–tails–heads AND tails–heads–tails?

      2F. Which combination (or combinations, if more than one) of outcomes are the most likely?

      2G. What is the probability of landing 3 heads GIVEN that 2 coins have landed on heads?

      2H. What is the probability of landing NO heads GIVEN that 2 coins have landed on tails?

Question 3: A study was conducted on 100 adults to document their smoking habits. Out of 65 surveyed men, 35 of them also smoked. Out of all interviewed women, only 25 did not smoke.

      3A. Draw a 2×2 table where you calculate both the joint and marginal probabilities of the people in the study. Trust me, getting this on a table is going to make the following questions *a lot* easier.

      3B. Pr(Selecting a woman $\cap$ she smokes)?

      3C. Pr(Selecting a man $\cap$ he does not smoke)?

      3D. Pr(Selecting a smoker)?

      3E. Pr(NOT selecting a smoker)?

      3F. Pr(Selecting a woman|she smokes)?

      3G. Are the events of being a man and smoking independent? Use the definition of independence we saw in class to justify your answer!

      3H. Are the events of being a woman and NOT smoking independent? Use the definition of independence we saw in class to justify your answer!

Question 4: A four-sided die (yes! they do exist!) is called a "D4" and they are used in some board games such as *Dungeons & Dragons* (brownie points if you've seen them in the popular hit Netflix series *Stranger Things*). Because of their shape, whichever corner is pointing upwards reflects the number the die landed on. In the following example, the die would have landed on 4:

Consider the experiment of rolling 2 fair D4 dice each numbered $\{1, 2, 3, 4\}$. Let the even of interest $X=$(sum of the 2 dice).

4A. Is X a continuous or a discrete random variable? **WHY?**

4B. Fill out the PMF of X with the proper probabilities.(Hint: We did something similar with two 6-sided dice in class. There's more than one way to get a 3, 4, 5,...7 and your answers need to reflect that. There's only one way to get a 2 and an 8 though, ;)

x	$Pr(X = x)$
2	
3	
4	
5	
6	
7	
8	

4C. Calculate the average of X like you would usually do (show work!)

4D. Calculate the expected value of X using the information from the PMF. (show work!)

4E. Calculate the median of X

4F. Calculate the variance of X

4G. Calculate the mean absolute deviation of X

4H. Obtaining the mode does not make much sense in this scenario because none of the values repeat more than once. However, using the definition of the mode and the information in the $Pr(X = x)$ column, which value of X would you expect to be the mode if you were rolling your 4D dice over and over again? **WHY?**

Question 5: Assume $Z \sim \mathcal{N}(0, 1)$. Using the 68-95-99.7 rule either give the correct probability or write "not possible" if it is, indeed, either impossible or if it does not make sense to calculate it.

5A. $Pr(Z \geqslant 0)$

5B. $Pr(Z \leqslant -1)$

5C. $Pr(-2 \leqslant Z \leqslant 1)$

5D. $Pr(Z \leqslant -2 \cup Z \geqslant 2)$

5E. $Pr(Z \leqslant -2 \cap Z \geqslant 2)$

5F. $Pr(Z \leqslant -5)$

5G. $Pr(-3 \leqslant Z \leqslant -2 \cup 0 \leqslant Z \leqslant 1 \cup Z \geqslant 3)$

5H. $Pr(Z = 0)$

Question 6: Immigration Canada accepts a few tests to demonstrate a new applicant's command of the English or French language. Let's focus on the *Canadian English Language Proficiency Index Program* for English, CELPIP (administered by UBC's company Paragon Testing) and Test d'Évaluation de Français, TEF.

I don't really know how these tests are scored, but assume $CELPIP \sim \mathcal{N}(85, 15)$ and $TEF \sim \mathcal{N}(9.7, 5.7)$. Say you score 112 on the CELPIP and 6 on the TEF.

6A. On which score did you get a higher mark?

6B. Let's pretend for a moment that Immigration Canada only keeps scores on the CELPIP metric. What would be your TEF score in CELPIP units?