STAT 614 - HW 2

1.

(a) Which of these studies, if any, is an example of a randomized experiment?

Study 2

(b) In which of these studies, if any, were subjects randomly chosen to participate?

Study 1

(c) Which study provides the strongest evidence that a vegetarian diet causes lower blood pressure? Explain your reasoning.

In Study 2 the researchers randomly assigned participants to the two diet groups, vegetarian and non-vegetarian. Thus, the difference in blood pressure between the two groups can be attributed to the different diets.

(d) For the two remaining studies (that you did not select for part c), pick one and briefly describe why you did not select it in part c.

Both studies 1 and 3 are observational studies. In Study 1, the vegetarians in the randomly selected sample had lower blood pressure but this could be due to confounding with lurking variables that were not measured as part of the study. For example, perhaps vegetarians are overall more health-conscious and thus smoke less and exercise more than non-vegetarians. These confounding variables might actually explain the difference in blood pressure, and not the diet. In Study 3, the researchers take advantage of two special groups that should be equivalent in some important confounding variables, such as smoking. But, we cannot say that there are no other systematic differences between Trappist and Benedictine lifestyles that might explain the difference in blood pressure. That limits the conclusions we can draw from this study. Note, only one study needed to be discussed to receive full credit.

2. From *The Statistical Sleuth, Third Edition*, Chapter 1, problems 19 & 20.

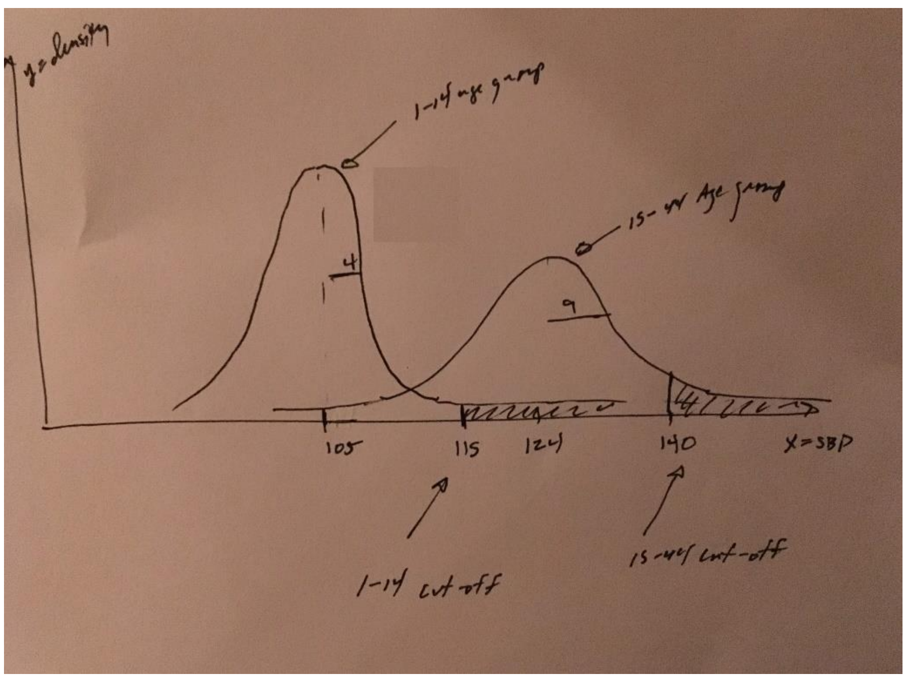
For #19: Each group should have a similar average distribution. Thus, one group should not tend to have most of the older observations than the other (compare the average ages of the two groups, the medians, the quartiles, etc.) There is no way to predict which group would have the higher average age in advance as the ages were assigned randomly.

For #20: I would put the names of the ten people in a hat and blindly select five. Like in 19, each group should have a similar average distribution. Thus, one group should not tend to have most of the older observations than the other (compare the average ages of the two groups, the medians, the quartiles, etc.) There is no way to predict which group would have the higher average age in advance as the ages were assigned randomly.

3. People are classified as hypertensive if their systolic blood pressure is higher than a specified level for their age group, according to the last column of the following table. Assume systolic blood pressure is *Normally distributed* with mean and standard deviation given in the table for the age groups 1-14 and 15-44, respectively.

Age group	Mean	St. Dev	Level
1-14	105.0	4.0	115.0
15-44	124.0	9.0	140.0

(a) Sketch a graph of the distribution of systolic blood pressure for both groups (on the same plot). Label the x- and y-axes, the mean, the standard deviation, and indicate the hypertensive cut-off for each group. You may take a picture of your graph and copy-and-paste it into your homework to include it with your solution. DO NOT submit a separate image file! Paste it into your homework document (and I recommend saving your homework document as a pdf to make sure it is readable in Blackboard).



(b) What proportion of 1- to 14-year-olds are hypertensive?

The proportion of 1-14-year-old hypertensive = $P(X \geq 115) = 1 - P(X < 115)$ $1 - 0.9938 = 0.0062$ = 1 – pnorm(115,105,4)

(c) What proportion of 15- to 44-year-olds are hypertensive?

The proportion of 15-44-year-old hypertensive = $P(Y \geq 140) = 1 - P(Y < 140) = 1 - 0.9623 = 0.0377$

(d) What is the probability that a randomly select 1-14 year old is not hypertensive?

The probability of NOT be hypertensive is 1 – the probability of being hypertensive. That is $P(Y < 140) = 0.9938$

(e) Suppose two people are selected at random from 1-14 year old age group. What is the probability that at least one of them is not hypertensive?

We need to translate “at least one of two randomly selection 1-14 year olds is hypertensive.” Let H = hypertensive and N = not hypertensive. Then, based on (c) and (d) we know for a single randomly selected 1-14 year old that $P(H) = 0.0062$ and $P(N) = 0.9938$. At least one of two hypertensive means we could have both hypertensive, HH, the first but not the second, HN, or the second but not the first, NH. This is similar in spirit to the coin toss example we did in class. We can find P(at least one out of two hypertensive) = P(HH) + P(HN) + P(NH) using our probability rules. But, notice that at least one hypertensive is the complement of NONE hypertensive (NN). So, equivalently,
P(at least one of two hypertensive) = 1 – P(NN) = 1 – P(N)P(N) as the individuals are independent due to random selection. So….

P(at least one of two hypertensive) = 1 – (0.9938)(0.9938) = 0.012

(f) Suppose ten people are selected at random from 1-14-year-old age group. What is the probability that all of them are not hypertensive?

We need to translate “all are not hypertensive” into a probability expression. Using the same notation as in (e), N = not hypertensive, we want P(NNNNNNNNNN) (all of these are “ands”, i.e first person is not hypertensive AND 2nd person is not hypertensive AND 3rd person is not hypertensive AND … AND 10th person is not hypertensive). Due to the random selection these are all independent events so P(NNNNNNNNNN) = P(N) x P(N) x P(N) x P(N) x P(N) x P(N) x P(N) x P(N) x P(N) x P(N) = $P(N)^{10} = (0.9938)^{10} = 0.939705$ or about 94% chance!

4. Find **one** scholarly article which uses **one** of the following probability distributions:

a. Normal
b. Poisson
c. Binomial
d. Any other distribution of interest to you.

Give the following information from your article:

(a) The full citation of the article (authors, title of article, journal name, year, volume, page #s).
(b) A brief summary of the purpose of the article based on the information in the abstract.
(c) A brief summary of how the distribution is used in the article. You do not need to go into a lot of detail here. Just try to get a sense of what they are using the particular distribution for.
(d) How you found the article (for example, through a PubMed or Google Scholar search).
(e) Why you picked that article.

Here is an example answer to this problem (but do not use this article!):

The Exponential distribution is used in:

a. Roccato A, Uyttendaele M, Membré JM. [Analysis of domestic refrigerator temperatures and home storage time distributions for shelf-life studies and food safety risk assessment.]() Food Res Int. 2017 Jun;96:171-181. doi: 10.1016/j.foodres.2017.02.017. Epub 2017 Mar 6. Review.

b. I'll quote the article: "The aim of this study was to analyse data on domestic refrigerator temperatures and storage times of chilled food in European countries in order to draw general rules which could be used either in shelf-life testing or risk assessment." So basically, they want to understand the distribution of refrigerator temperatures and storage time of chilled foods in European countries.

c. The authors state that "[d]ata fitting showed the exponential distribution was the most appropriate distribution to describe the time that food spent at consumer's place." They further conclude that time-to-consumption should be modeled (in European markets) using an Exponential law (while refrigeration temperature distributions are normal.)

a. A search in PubMed of "exponential distribution" and "health."

b. Honestly, it was because it was about refrigerators and I've never seen an article on refrigerators before! But food safety is important area of research so that totally makes sense.