Stat 509/Econ 580: Homework 1

This Homework is due in GradeScope by Friday, October 9 at 8AM.

1. A and B are two events with $P(A) = 0.6$ and $P(B) = 0.45$.

(a) Can A and B be mutually exclusive? Explain.

(b) Can A and B^C be mutually exclusive? Explain.

(c) Can A^C and B be mutually exclusive? Explain.

(d) Could A and B be independent?

(e) Can A and B be both mutually exclusive and independent?

2. A and B are events with $P(A) = 0.3$, $P(A \cup B) = 0.44$ and $P(A \cap B) = 0.06$.

(a) Are A and B mutually exclusive?

(b) What can be said about $P(B)$?

(c) Can A and B be *independent*, so that $P(A \cap B) = P(A)P(B)$?

3. If Z is a random variable taking integer values, if $P(Z < 1) = a$, $P(Z \geq 0) = b$, find $P(Z = 0)$ in terms of a and b.

4. Suppose that A, B and C are three events. $P(A \cup B) = r$, $P(A \cup C) = s$.

(a) If we have no additional information, what can be said about $P(A)$? *Hint: What is the largest or smallest $P(A)$ could be?*

(b) Suppose that we learn that $P(A \cap B \cap C) = t$. What further can be said about $P(A)$?

Hint: consider all of the different logically possible values for r, s and t in your answers.

5. Consider the following table relating to hypothetical data for four patients:

Patient	Y_C	Y_T
1	3	2
2	5	10
3	2	6
4	7	9

The variables Y_C and Y_T record the number of years that the patient *would* survive if the patient received control C, or treatment T, respectively.

(a) Researchers will randomly assign two patients to receive treatment and two to receive control. Concretely, the four numbers 1, 2, 3, 4 are put into a bag. Two numbers $\{t_1, t_2\}$ are pulled out and these patients receive treatment. The other two receive control.

Write down all of the different possible pairs $\{t_1, t_2\}$ of patients who could receive Treatment. *Hint: It might be useful to construct a spreadsheet with a column for t_1 & t_2 and a separate row for each pair. For the number of rows, consider the number of ways of choosing two people from a set of four.*

(b) The researchers get to see the Y_T entries in the table for those patients assigned to Treatment (for those in Control, the Y_T values will remain unknown). For each of the pairs listed in (a) find the average of the two Y_T values for these pairs. *Hint: Add columns for the Y_T values for t_1 and t_2 and for the average of these.*

(c) For each of the pairs $\{t_1, t_2\}$ found in (a), consider the corresponding pair $\{c_1, c_2\}$ corresponding to the remaining two individuals who receive Control. The researchers get to see the Y_C entries in the table for each of these pairs $\{c_1, c_2\}$ (whereas the Y_C values for those in Treatment are unknown). For each of the pairs $\{c_1, c_2\}$ find the average of the two corresponding Y_C values.

(d) To assess the effectiveness of treatment the researchers will look at the difference between the average of the two Y_T values and the average of the two Y_C values. Compute this for each of the assignments to T and C (in other words for each pair obtained in (a)). The researchers carrying out the experiment will get to see only the difference corresponding to the specific assignment of two patients to treatment and two to control.

(e) Since the $\{t_1, t_2\}$ pair was selected at random, they are all equally likely. Find the average of the differences obtained in (d). Here we are averaging over *hypothetical* replications of the experiment.

(f) Now consider the difference between the average of the *four* Y_T values and the *four* Y_C values in the original table.

(g) What do you notice when you compare your answers to (e) and (f) ? The difference found in (f) is something that the researchers cannot (in principle) observe because, for a given patient, they can only record *either* Y_T (if the patient receives Treatment) *or* Y_C if the patient receives control.

6. A population of n people contains k individuals with a disease. A sample of m are selected at random (without replacement) from the n in the population.

(a) For a given sample size m, find the probability that there is at least one person with the disease.

Hint: calculate the probability that the first person sampled does not have the disease; then that the second also does not, given that the first does not. . . .

(b) If $n = 1000$ and $k = 20$, find the smallest integer m such that the probability in (a) is ≥ 0.95.

(c) If $n = 10,000$ and $k = 200$, find the smallest integer m such that the probability in (a) is ≥ 0.95.

You may want to use `R` or a spreadsheet for parts (b) and (c).

In `R` you may find the functions `choose`, `factorial` and `lfactorial` useful.

(d) Use a Binomial probability to obtain an approximate answer to (c) by a simple calculation.

Hint: Intuitively, as $n, k \to \infty$, $k/n \to c$, *for some constant* $c \in (0, 1)$, *will it make a difference whether we are sampling 'with replacement' or 'without replacement'?*

7. Goldberger Qu. 2.8