Statistical Methods in Education Research
Vanderbilt University   Due September 27, 2022
Prof. Sean P. Corcoran   **61 total points**

Problem Set 4

Instructions: Answer the following questions in their entirety in a separate document. Submit your completed problem set as a PDF document via email to `sean.corcoran@vanderbilt.edu`. Use your last name and problem set number as the filename (e.g., *Ginsburg Problem Set 4.pdf*). Working together is encouraged, but it is expected that all submitted work be that of the individual student.

1. (**16 points—2 each**) In a population of students, the number of absences during the school year ranges from 3 to 7. The probabilities of a randomly drawn student from this population having 3, 4, 5, 6, or 7 absences are shown in the table below. Define the event A as the student being absent *more than* 4 days, and the event B as the student being absent *fewer than* 6 days.

# of Days	3	4	5	6	7
Probability	0.08	0.24	0.41	0.20	0.07

   (a) What is the probability of event A?

   (b) What is the probability of event B?

   (c) What is the probability of $\sim A$?

   (d) Are events A and B mutually exclusive? Explain why or why not.

   (e) What is the probability of $A \cap B$?

   (f) What is the probability of $A \cup B$?

   (g) Show that $P((A \cap B) \cup (\sim A \cap B)) = P(B)$.

   (h) Show that $P(A \cup (\sim A \cap B)) = P(A \cup B)$.

2. (**6 points—3 each**) Using the probability distribution in Question 1, find the following (and show your work):

   (a) $E(\#$ of absences$)$

   (b) $Var(\#$ of absences$)$

3. (**8 points—2 each**) Shown below is a 2 x 2 table that reports the fraction of the population in each cell:

		Education level		
		HS	<HS	Totals
Current smoker:	NO	0.614	0.130	0.744
	YES	0.194	0.062	0.256
	Totals	0.808	0.192	1.000

   (a) For a randomly drawn person, what is P(smoker)?

   (b) For a randomly drawn person, what is P(smoker | <HS diploma)?

   (c) For a randomly drawn person, what is P(smoker | HS diploma+)?

   (d) Are education and smoking status independent? Why or why not?

4. (**5 points**) Shown below is a 2 x 2 table. In Period 1, events A or B can happen. In Period 2, outcome C or D will result. If $P(C|B) = 0.150$ and $P(D|A) = 0.7$, then fill in the missing boxes below:

		Period 1	
		Event A	Event B
Period 2	Event C		0.030
	Event D		

5. (**4 points**) After the attacks of September 11, 2001, the TSA implemented a program called SPOT (Screening of Passengers by Observation Techniques) in which passengers were flagged for suspicious behavior and given additional searching or screening. Suppose that:

   • There are 2 billion plus 100 passenger trips per year (2,000,000,100).

   • 100 of these passengers are terrorists (i.e., less than 0.00000001%).

   • Nearly all (99%) terrorists exhibit the kinds of behaviors that were flagged.

   • Some non-terrorists exhibit these suspicious behaviors, but it is rare (1%).

   The SPOT test has low false negative and false positive rates, suggesting it is an effective way to catch would-be terrorists. Use Bayes' Theorem to calculate the probability that a flagged passenger is, in fact, a terrorist.

6. (**6 points—3 each**) Paul and Natasha live in Los Angeles. Paul hates cold weather but Natasha has been transferred to a cold Northeastern city. Paul notes that he cannot move go to a city where more than 30% of the days have an average daily high below freezing. Suppose the average daily high temperatures (X) in a city can be described by a uniform distribution where the minimum and maximum average daily highs are -2 and 105, respectively.

   (a) What is the PDF for X, and what is $P(x \leq 32)$? Should Natasha look for a one or a two bedroom apartment? (Hint: you do not need calculus to find the requested probability).

   (b) What are $E(X)$ and $Var(X)$?

7. (**4 points**) Assume the random variable z has a standard normal distribution. Use Stata, an online calculator, or a textbook table to answer the following:

   (a) The probability is 0.70 that z is less than what number?

   (b) The probability is 0.25 that z is less than what number?

   (c) The probability is 0.20 that z is greater than what number?

   (d) The probability is 0.60 that z is greater than what number?

8. (**6 points**) To graduate with honors, you must be in the top 2 percent (*summa cum laude*), 3 percent (*magna cum laude*) or 5 percent (*cum laude*) of your class. Suppose GPAs are distributed normally with a mean of 2.6 and a standard deviation of 0.65. What GPA will you need in order to graduate at each of these three levels?

9. (**6 points—3 each**) On the midterm exam in introductory statistics, an instructor always gives a grade of B to students who score between 80 and 90. One year, the scores have an approximately normal distribution with a mean $\mu = 83$ and a standard deviation $\sigma = 5$. About what fraction of the students get a B?

   (a) First, answer this question using what you know about the normal distribution.

   (b) Now use simulated data in Stata. Generate 1,000 student exam scores—this instructor has a big class!—from a normal distribution with the above parameters. Then answer the question based on the data you drew. Are there any differences between your two answers?