Problem Set 1

PUBHLTH 490Z

Due 20 September 2021

Problem set policies. *Please provide concise, clear answers for each question. Note that only writing the result of a calculation (e.g., "SD = 3.3") without explanation is not sufficient. For problems involving* R*, include the code in your solution, along with any plots.*

Each problem set is due by 11:00 am on the due date; please submit your problem set via Moodle as a PDF.

We encourage you to discuss problems with other students (and, of course, with the TA and instructor), but you must write your final answer in your own words. Solutions prepared "in committee" are not acceptable. If you do collaborate with classmates on a problem, please list your collaborators on your solution.

Problem 1.

(IPS 7^{th} edition, 4.60) The Census Bureau reports that 27% of California residents are foreign-born. Suppose that you choose three Californians at random, so that each has probability 0.27 of being foreign-born and the three are independent of each other. Let the random variable W be the number of foreign-born people you chose.

a) What are the possible values of W?
b) There are eight possible arrangements of foreign (F) and domestic (D) birth. For example, FFD means the first two are foreign-born and the third is not. What is the probability of each possible sequence? What is the value of W for each arrangement?
c) What is the probability of each possible value of W?

Problem 2.

(IPS 7^{th} ed. 4.74) The table below gives the distribution of the number of servings of fruits and vegetables per day in a population. Find the mean and the standard deviation for this random variable.

Value of X	0	1	2	3	4
Probability	0.05	0.04	0.20	0.40	0.31

Problem 3.

Suppose that you are interested in monitoring air pollution in Los Angeles, California, over a one-week period. Let X be a random variable that represents the number of days out of the seven on which the concentration of carbon monoxide surpasses a specified level. Do you believe that X has a binomial distribution? Explain.

Problem 4.

According to data from the CDC, about 43% of adults in the United States (individuals 18 years of age or older) received a flu vaccine during the 2014-2015 flu season.

a) Consider a random sample of 50 adults.

   i. Calculate the probability that exactly 20 adults received a flu vaccine.

   ii. Calculate the probability that exactly 30 adults did not receive a flu vaccine.

b) Consider a random sample of 20 adults.

   i. What is the probability that at most 10 adults received a flu vaccine?

   ii. What is the probability that at least 11 adults received a flu vaccine?

c) State any assumptions you needed to make in order to answer parts a) and b).

Problem 5.

This is a simple exercise in computing probabilities for a Poisson random variable. Suppose that X is a Poisson random variable with rate parameter $\lambda = 2$.

a) By hand (i.e., without using R), calculate $\Pr(X = 2)$, $\Pr(X \leq 2)$, and $\Pr(X \geq 3)$. Be sure to show your calculations in your solution.

b) Now use R to calculate each of the 3 probabilities in part (a). Be sure to show the R commands and output in your PDF file.

Problem 6.

Osteosarcoma is a relatively rare type of bone cancer. It occurs most often in young adults, age 10 - 19; it is diagnosed in approximately 8 per 1,000,000 individuals per year in that age group. In New York City (including all five boroughs), the number of young adults in this age range is approximately 1,400,000.

a) What is the expected number of cases of osteosarcoma in NYC in a given year?

b) What is the probability that 15 or more cases will be diagnosed in a given year?

c) The largest concentration of young adults in NYC is in the borough of Brooklyn, where the population in that age range is approximately 450,000. What is the probability of 10 or more cases in Brooklyn in a given year?

d) Suppose that in a given year, 10 cases of osteosarcoma were observed in NYC, with all 10 cases occurring among young adults living in Brooklyn. An official from the NYC Public Health Department claims that the probability of this event (that is, the probability of 10 or more cases being observed, and all of them occurring in Brooklyn) is what was calculated in part c). Is the official correct? Explain your answer. You may assume that your answer to part c) is correct. This question can be answered without doing any calculations.

Problem 7.

Consider the standard normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$.

a) What is the probability that an outcome z is greater than 2.60?

b) What is the probability that z is less than 1.35?

c) What is the probability that z is between -1.70 and 3.10?

d) What value of z cuts off the upper 15% of the standard normal distribution?

e) What value of z marks off the lower 20% of the distribution?

Problem 8.

The World Health Organization defines osteoporosis in young adults as a measured bone mineral density 2.5 or more standard deviations below the mean for young adults. Assume that bone mineral density follows a normal distribution in young adults. What percentage of young adults suffer from osteoporosis according to this criterion?

Problem 9.

Arsenic blood concentration is normally distributed with mean $\mu = 3.2$ μg/dl and standard deviation $\sigma = 1.5$ μg/dl. What range of arsenic blood concentration defines the middle 95% of this distribution?

Problem 10.

(Problem 1.136 in IPS, 6^{th} edition.) High blood cholesterol levels increase the risk of heart disease. Young women are generally less afflicted with high cholesterol than other groups. The cholesterol levels for women aged 20 to 34 follow an approximately normal distribution with mean 185 milligrams per deciliter (mg/dl) and standard deviation 39 mg/dl.

a) Cholesterol levels above 240 mg/dl demand medical attention. What percent of young women have levels above 240 mg/dl?

b) Levels above 200 mg/dl are considered borderline high. What percent of young women have blood cholesterol between 200 and 240 mg/dl?

Problem 11.

Hemophilia is a sex-linked bleeding disorder that slows the blood clotting process. In severe cases of hemophilia, continued bleeding occurs after minor trauma or even in the absence of injury. Hemophilia affects 1 in 5,000 male births. In the United States, there are approximately 4,000,000 births per year. Assume that there are equal numbers of males and females born each year.

a) What is the probability that at most 380 newborns in a year are born with hemophilia?

b) What is the probability that 450 or more newborns in a year are born with hemophilia?

c) Consider a hypothetical country in which there are approximately 1.5 million births per year. If the incidence rate of hemophilia is equal to that in the US, as well as the sex ratio at birth, how many newborns are expected to have hemophilia over five years, and with what standard deviation?