The Basics of Mathematical Statistics

Home assignment 1

1. A graduating engineer has signed up for three job interviews. She intends to categorise each one as being either a "success" or a "failure" depending on whether it leads to a plant trip. Write an appropriate sample space. What outcomes are in the event A: Second success occurs on third interview? In B: First success never occurs?

2. How many different letter arrangements can be obtained from the letters of the word *bioinformatics*, using all the letters?

3. Three green balls, two blue balls and seven yellow balls are put into a bag.

   (a) Three balls are taken out, one at a time, and each ball taken out is put back in the bag again before the next ball is chosen. What is the chance that one green, one blue and one yellow ball is drawn in that order? What is the probability that we draw one green, one blue and one yellow in any order?

   (b) What is the probability of choosing one green, followed by one blue, followed by one yellow, not replacing the balls?

4. A fire insurance company has high-risk, medium risk, and low-risk clients, who have, respectively, probabilities 0.02, 0.01, and 0.0025 of filing claims within a given year. The proportions of the numbers of clients in the three categories are 0.10, 0.20, and 0.70, respectively. What proportion of the claims filed each year come from high-risk clients?

Home assignment 2

1. A family has two children. What is the probability that both are boys given that at least one is a boy? What is the probability that both are boys given that the younger is a boy?

2. Spike's chances of passing chemistry are 0.35; mathematics, 0.40; both, 0.12. Are the events "Spike passes chemistry" and "Spike passes mathematics" independent? What is the probability that he fails both subjects?

3. Urn A has two white and three blue balls. Urn B has three white and four blue balls. A ball is drawn from urn A and put into urn B, and then a ball is drawn from urn B. What is the probability that it is blue?

4. The following table shows the cumulative distribution function of a discrete random variable. Find the frequency function.

x	$F(x)$
0	0
1	0.1
2	0.3
3	0.7
4	0.8
5	1.0

Home assignment 3

1. A multiple-choice test consists of 20 items, each with four choices. A student is able to eliminate one of the choices on each question and chooses randomly from the remaining three choices. A passing grade is 12 items or more correct. What is the probability that the student passes?

2. Phone calls are received at a certain residence as a Poisson process with parameter $\lambda = 2$ per hour. If Diane takes a 10-min shower, what is the probability that the phone rings during that time?

3. A fair die is tossed. What is the probability the first 5 occurs on the fourth roll?

4. Suppose that a rare disease has an incidence of 1 in 1000. Assuming that members of the population are affected independently, find the probability of k cases in a population of 100 000 for $k = 0, 1, 2$.

5. Five individuals from an animal population thought to be near extinction in a certain region have been caught, tagged, and released to mix into the population. After they have had an opportunity to mix, a random sample of 10 of these animals is selected. Let $X =$ the number of tagged animals in the second sample. If there are actually 25 animals of this type in the region, what is the probability that

   (a) $X = 2$?

   (b) $X \leq 2$?

Home assignment 4

1. Over a recent 90-year period, temperatures recorded in Bismarck, North Dakota, in December ranged from $-36°F$ to $+72°F$. Assume that the standard deviation of the distribution of daily extremes is approximately $18°F$. What is the corresponding standard deviation in degrees *Celsius*? [*Note*: If X is a temperature recorded in degrees of Fahrenheit and Y is the same temperature expressed in degrees Celsius, then $Y = \frac{5}{9}(X - 32)$.]

2. A new type of electronic flash for cameras will last an average of 5000 hours with a standard deviation of 500 hours. A company quality control engineer intends to select a random sample of 100 of these flashes and use them until they fail. What is the probability that the mean lifetime of the 100 flashes will be less than 4928 hours?

3. Let X denote the amount of radiation that can be absorbed by an individual before death ensues. Assume that X is normal with mean of 500 roentgens and a standard deviation of 150 roentgens. Above what dosage level will only 5% of those exposed survive?

4. In healthy individual age 20 to 29 years, the calcium level in the blood, X, is usually between 8.5 and 10.5 milligrams per deciliter (mg/dl) and the cholesterol level, Y, is usually between 120 and 240 mg/dl. Assume that for a healthy individual the joint density for (X, Y) is

   $f_{XY}(x, y) = \frac{1}{240}$, $8.5 \leq x \leq 10.5$, $120 \leq y \leq 240$

   ($f_Y(y) = \frac{1}{120}$, $f_X(x) = \frac{1}{2}$)

   (a) Find the probability that a healthy individual has a cholesterol level between 150 and 200.

   (b) Are the random variables X, an individual's blood calcium level, and Y, his or her blood cholesterol level, independent?

   (c) Find the covariance between X, an individual's blood calcium level, and Y, his or her blood cholesterol level.

Home assignment 5

1. Suppose that we wish to compare a new drug to that of a problem 2 in exercise session 4. Let X denote the number of heartbeats per minute obtained using the old drug and Y the number of heartbeats per minute obtained with the new drug. The hypothetical frequency of each of these variables is given in table below

x	40	60	68	70	72	80	100
$p(x)$	0.01	0.04	0.05	0.80	0.05	0.04	0.01

y	40	60	68	70	72	80	100
$p(y)$	0.40	0.05	0.04	0.02	0.04	0.05	0.40

Inspection shows that $E(X) = E(Y) = 70$. Each drug produces *on the average* the same number of heartbeats per minute. However, there is obviously a drastic difference between the two drugs that is not being detected by mean. The old drug produces fairly consistent reactions in patients, with 90% differing from the mean by at most 2; very few (2%) have an extreme reaction to the drug. However the new drug produces highly diverse responses. Only 10% of the patients have heart rates within 2 units of the mean, whereas 80% show an extreme reaction. If we examined only the mean, we would conclude that the two drugs had identical effects — but nothing could be further from the truth! It's obvious that something is not being measured by the mean. That something is variability. Find the variances of the variables X and Y.

2. A sociologist wishes to estimate the mean number of hours, μ, that children in a city watch TV during a particular week. A random sample of n children is selected and the number of hours spent watching TV recorded. Assume that the standard deviation of the time spent watching TV over all children in the city is $\sigma = 2.5$ hours. If the sample mean is $\bar{x} = 14.7$ hours, find a 90% confidence interval for μ if the sample size is

    (a) $n = 64$;

    (b) $n = 100$;

(c) $n = 400$.

3. If gene frequences are in equilibrium, the genotypes AA, Aa and aa occur with probabilities $(1-\theta)^2$, $2\theta(1-\theta)$, and θ^2, respectively. Plato et al. (1964) published the following data on haptoglobin type in sample of 190 people:

Haptoglobin Type		
Hp1-1	Hp1-2	Hp2-2
10	68	112

Find the mle of θ.

Home assignment 6

1. A coin is thrown independently 10 times to test the hypothesis that the probability of heads is 0.5 versus the alternative that the probability is not 0.5. The test rejects if either 0 or 10 heads are observed.

   (a) What is the significance level of the test?

   (b) If in fact the probability of heads is 0.1, what is the power of the test?

2. A particular IQ test was taken by sample of 22 students with the sample mean and sample variance computed to be 107 and 241. To see if the population mean is significantly different from 100

   (a) state a simple null hypothesis and a two sided alternative hypothesis;

   (b) apply the one-sample t-test at 5 % significance level;

   (c) compute a p-value of the test;

   (d) using an appropriate confidence interval find if the population IQ-mean is significantly different from 100 at 1 % significance level.

3. Criminologists have long debated whether there is a relationship between weather conditions and the incidence of violent crime. The author of the article "Is there a Season for Homicide?" (*Criminology*, 1988:287-296) classified 1361 homicides according to season, resulting in the accompanying data. Test the null hypothesis of equal proportions using $\alpha = 0.01$ by using the chi-squared table to say as much as possible about the p-value.

Winter	Spring	Summer	Fall
328	334	372	327

4. Suppose that a parameter, Θ, takes on values $\theta_1 = 1$, $\theta_2 = 10$, and $\theta_3 = 20$. The distribution of X is discrete and depends on Θ as shown in the following table:

	θ_1	θ_2	θ_3
x_1	0.1	0.2	0.4
x_2	0.1	0.2	0.2
x_3	0.2	0.2	0.2
x_4	0.6	0.4	0.2

Assume a prior distribution of Θ: $P(\theta_1) = 0.5$, $P(\theta_2) = 0.25$, and $P(\theta_3) = 0.25$.

(a) Suppose that x_2 is observed. What is the posterior distribution of Θ?

(b) What is the Bayes estimate under squared error loss in this case?