

## Data Analysis: Assignment 02

1. The below table gives the number of commercial airline accidents and the total number of resulting fatalities in the United States in the years from 1985 to 2006.

Year	Departures (millions)	Accidents	Fatalities	Year	Departures (millions)	Accidents	Fatalities
1985	6.1	4	197	1996	7.9	3	342
1986	6.4	2	5	1997	9.9	3	3
1987	6.6	4	231	1998	10.5	1	1
1988	6.7	3	285	1999	10.9	2	12
1989	6.6	11	278	2000	11.1	2	89
1990	7.8	6	39	2001	10.6	6	531
1991	7.5	4	62	2002	10.3	0	0
1992	7.5	4	33	2003	10.2	2	22
1993	7.7	1	1	2004	10.8	1	13
1994	7.8	4	239	2005	10.9	3	22
1995	8.1	2	166	2006	11.2	2	50

*Source: National Transportation Safety Board.*

- a) Find the sample mean of the number of yearly airline accidents.  
b) Find the sample median of the number of yearly airline accidents.  
c) Find the sample standard deviation of the number of yearly airline accidents.
2. The sample mean of the weights of the adult women of town A is larger than the sample mean of the weights of the adult women of town B. Moreover, the sample mean of the weights of the adult men of town A is larger than the sample mean of the weights of the adult men of town B. Can we conclude that the sample mean of the weights of the adults of town A is larger than the sample mean of the weights of the adults of town B? Explain your answer.

3. The following are the heights and starting salaries of 12 law school classmates whose law school examination scores were roughly the same. Find the sample correlation coefficient.

Height	Salary
64	91
65	94
66	88
67	103
69	77
70	96
72	105
72	88
74	122
74	102
75	90
76	114

4. There are two local factories that produce microwaves. Each microwave produced at factory A is defective with probability .05, whereas each one produced at factory B is defective with probability .01. Suppose you purchase two microwaves that were produced at the same factory, which is equally likely to have been either factory A or factory B. If the first microwave that you check is defective, what is the conditional probability that the other one is also defective?
5. Three prisoners are informed by their jailer that one of them has been chosen at random to be executed, and the other two are to be freed. Prisoner A asks the jailer to tell him privately which of his fellow prisoners will be set free, claiming that there would be no harm in divulging this information because he already knows that at least one of the two will go free. The jailer refuses to answer this question, pointing out that if A knew which of his fellow prisoners were to be set free, then his own probability of being executed would rise from  $\frac{1}{3}$  to  $\frac{1}{2}$  because he would then be one of two prisoners. What do you think of the jailer's reasoning?
6. Each night different meteorologists give us the "probability" that it will rain the next day. To judge how well these people predict, we will score each of them as follows: If a meteorologist says that it will rain with probability  $p$ , then he or she will receive a score of

$$\begin{array}{ll} 1 - (1 - p)^2 & \text{if it does rain} \\ 1 - p^2 & \text{if it does not rain} \end{array}$$

We will then keep track of scores over a certain time span and conclude that the meteorologist with the highest average score is the best predictor of weather. Suppose now that a given meteorologist is aware of this and so wants to maximize his or her expected score. If this individual truly believes that it will rain tomorrow with probability  $p^*$ , what value of  $p$  should he or she assert so as to maximize the expected score?

7. If  $E[X] = 2$  and  $E[X^2] = 8$ , calculate **(a)**  $E[(2 + 4X)^2]$  and **(b)**  $E[X^2 + (X + 1)^2]$ .

8. If  $X_1$  and  $X_2$  have the same probability distribution function, show that

$$\text{Cov}(X_1 - X_2, X_1 + X_2) = 0$$

Note that independence is not being assumed.

9. Let  $X$  and  $Y$  have respective distribution functions  $F_X$  and  $F_Y$ , and suppose that for some constants  $a$  and  $b > 0$ ,

$$F_X(x) = F_Y\left(\frac{x - a}{b}\right)$$

a) Determine  $E[X]$  in terms of  $E[Y]$ .

b) Determine  $\text{Var}(X)$  in terms of  $\text{Var}(Y)$ .

10. Let  $X_1$  and  $X_2$  be independent normal random variables, each having mean 10 and variance  $\sigma^2$ . Which probability is larger:

a)  $P(X_1 > 15)$  or  $P(X_1 + X_2 > 25)$ ?

b)  $P(X_1 > 15)$  or  $P(X_1 + X_2 > 30)$ ?

c) Find  $x$  such that  $P(X_1 + X_2 > x) = P(X_1 > 15)$ .

11. Let  $\Phi$  be the standard normal distribution function. If, for constants  $a$  and  $b > 0$ ,

$$P\{X \leq x\} = \Phi\left(\frac{x - a}{b}\right)$$

Characterize the distribution of  $X$  (i.e., describe what distribution random variable  $X$  follows, and what the mean and variance of  $X$  are).