Dealing with Data II Homework 5 Solutions

Joshua D. Ingram

2022-11-29

Reading

a) On Probability, Basic Rules of Probability, and Conditional Probability: Sections 5.1, 5.2 and 5.3.

b) On Random Variables and Probability Distribution, especially the normal (review) and binomal distribution: Section 6.1, 6.2 and 6.3.

Also, make sure you study the slides that I uploaded to Canvas which summarize these topics, and the very brief end-of-chapter summaries for chapter 5 and chapter 6 in the textbook.

Note: This HW has no R component, so you can either type it up and then turn in a pdf file, or you can write it on a piece of paper (legible, please!), and then take a picture and turn it in.

Textbook Exercises Chapter 5

5.22 - Cell phone and case

At the local cell phone store, the probability that a customer who walks in will purchase a new cell phone is 0.2. The probability that the customer will purchase a new cell phone protective case is 0.25. Is this information sufficient to determine the probability that a customer will purchase a new cell phone and a new cell phone protective case? If so, find the probability. If not, explain why not. (*Hint:* Considering the buying practices of consumers in this context, is it reasonable to assume these events are independent?)

Answer: This information is not sufficient to determine the probability that both A (customer purchases a new cell phone) and B (customer purchases a new cell phone protective case) occur. This is because it it is unlikely that these events are independent, as a customer coming to the store buying a new phone is probably going to purchase a phone case because of them purchasing a new phone.

5.29

Because of the increasing nuisance of spam e-mail messages, many start-up companies have emerged to develop e-mail filters. One such filter was recently advertised as being 95% accurate. The way the advertisement is worded, 95% accurate could mean that (a) 95% of spam is blocked, (b) 95% of valid e-mail is allowed through, (c) 95% of the e-mail allowed through is valid, or (d) 95% of the blocked e-mail is spam. Let S denote {message is spam}, and let B denote {filter blocks message}. Using these events and their complements, identify each of these four possibilities as a conditional probability.

Answer:

(a) $P(B|S) = 0.95$
(b) $P(B^c|S^c) = 0.95$
(c) $P(S^c|B^c) = 0.95$

(d) $P(S|B) = 0.95$

5.39 - Happiness in relationship

Are people happy in their romantic relationships? The table shows results from the 2012 General Social Survey for adults classified by gender and happiness. (Source: 2012 General Social Survey 2011-2019 NORC.)

a) Estimate the probability that an adult is very happy in their romantic relationship.

Answer:

$$P(\text{Very Happy}) = \frac{147}{317} = 0.4637224$$

b) Estimate the probability that an adult is very happy (i) given that the adult identifies as male and (ii) given that the adult identifies as female.

Answer:

(i)

$$P(\text{Very Happy}|\text{Male}) = \frac{69}{146} = 0.4726027$$

(ii)

$$P(\text{Very Happy}|\text{Female}) = \frac{78}{171} = 0.4561404$$

c) For these subjects, are the events being very happy and identifying as female independent?

Answer: We can determine if these events are independent by comparing the joint probability to the product of the probabilities of the individual events. Where $P(\text{Very Happy}) = \frac{147}{317} = 0.4637224$, $P(\text{Female}) = \frac{171}{317} = 0.5394322$, and $P(\text{Very Happy} \cap \text{Female}) = \frac{78}{317} = 0.2460568$. Thus,

$$P(\text{Very Happy}) \times P(\text{Female}) = 0.4637224 \times 0.5394322 = 0.2501468.$$

The difference in the joint probability and the product is less than 1 percentage point, so it is reasonable to state that being very happy in a relationship and identifying as female are independent.

5.91 - HIV blood test, parts a, c, d

For a combined ELISA-Western blood test for HIV-positive status, the sensitivity is about 0.999 and the specificity is about 0.9999.

a) Consider a high-risk group in which about 10% are truly HIV positive. Construct a tree diagram to summarize this diagnostic test.

Answer:

c) A person from the high-risk group has a positive test result. Using the tree diagram or the contingency table, find the probability that the person really is HIV positive.

Answer:

$$P(\text{True HIV Postive}|\text{Postive HIV Test}) = 0.1080.999 = 0.0999$$

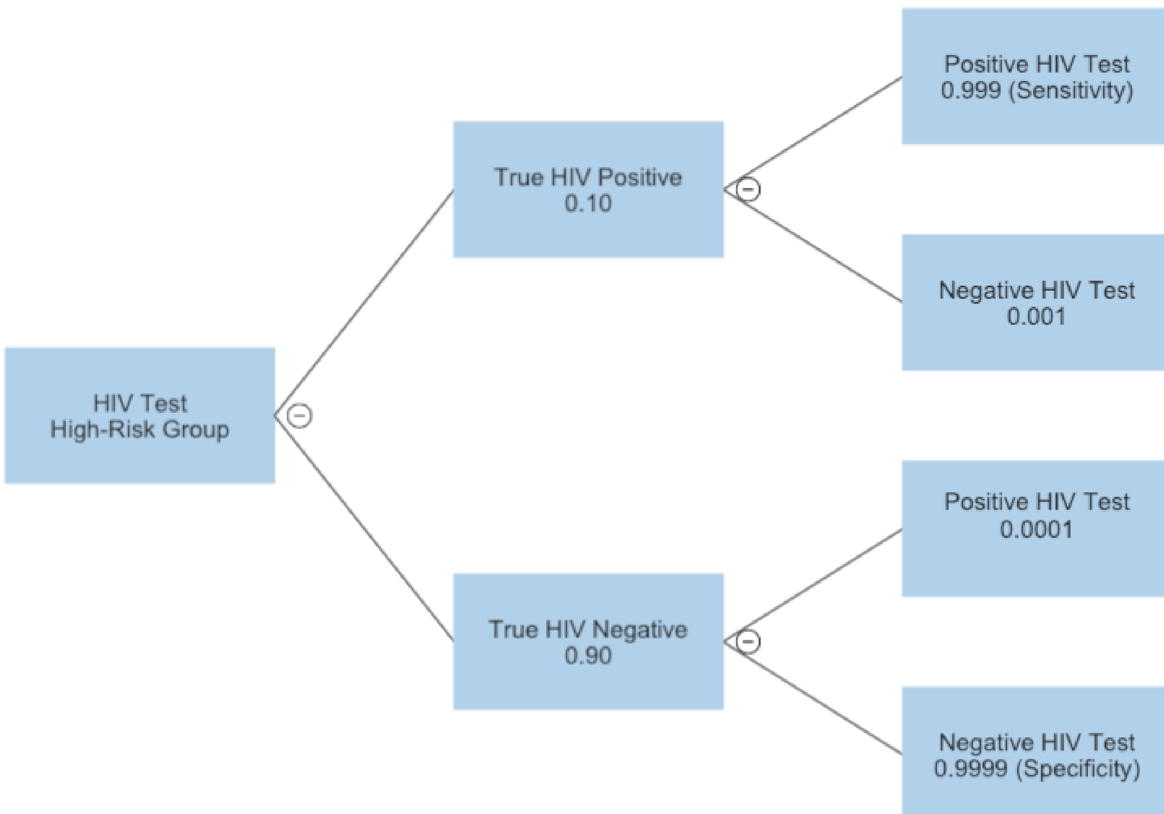

Figure 1: Tree diagram for HIV diagnostic test in high-risk group

d) Explain why a positive test result is more likely to be in error when the prevalence is lower. Use tree diagrams or contingency tables with frequencies for 10,000 people with 10% and 1% rates to illustrate your point.

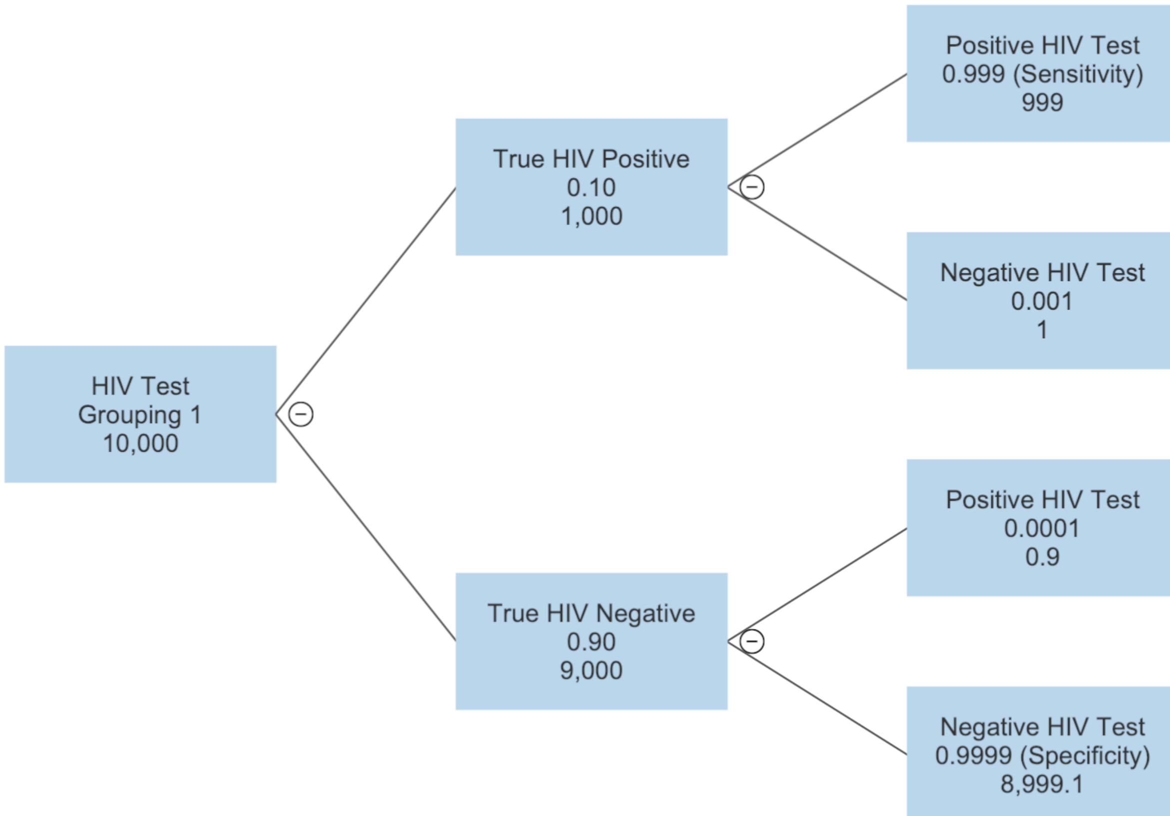

Figure 2: Tree diagram for HIV diagnostic test in high-risk group

Answer:

5.102 - Hillary Clinton's Tweet, parts b, c

During the 2016 presidential campaign, Hillary Clinton tweeted the following about background checks for gun ownership. It is an example of the wrong use of Venn diagrams and not using conditional probabilities. Let A = {gun owner} and B = {support background checks}

b) According to polls, 43% of Americans personally own guns. Of those personally owning guns, 87% favor universal background checks. Based on these numbers, state P(A) and P(B | A) and create a tweet that describes these findings.

Answer:

$$P(A) = 0.43, \; P(B|A) = 0.87$$

c) The polls show that overall, 93% of Americans favor background checks. Given this information, find P(A|B^c).

Answer:

$$P(A|B^c) = \frac{P(B^c|A)P(A)}{P(B^c)} = \frac{0.13 \times 0.43}{0.07} = 0.7985714$$

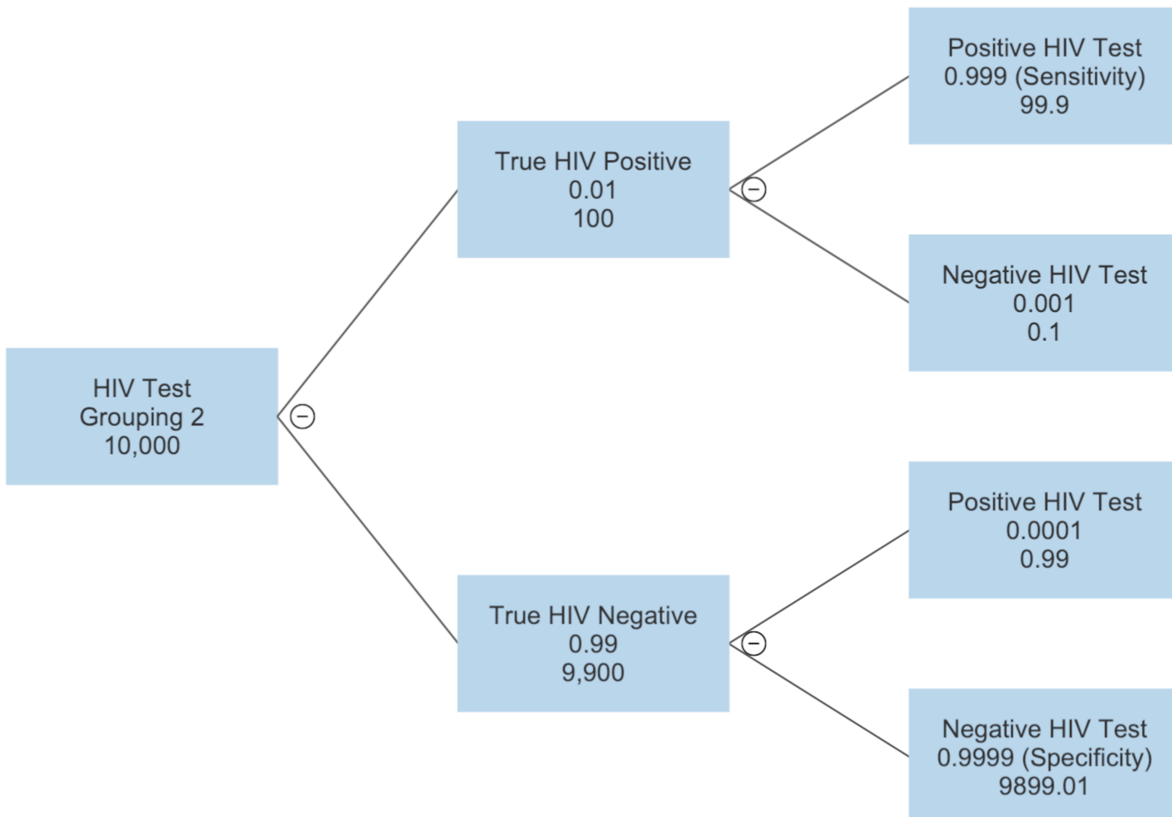

Figure 3: Tree diagram for HIV diagnostic test in high-risk group

X	P(X)
0	0.35
100	0.60
1000	0.05

X	P(X)
0	0.72
30	0.08
20	0.18
50	0.02

Textbook Exercises Chapter 6

6.2

You plan to purchase dental insurance for your three remaining years in school. The insurance makes a one-time payment of $1000 in case of a major dental repair such as an implant, or $100 in case of a minor repair (such as a cavity). If you don't need dental repair over the next 3 years, the insurance expires and you receive no payout. You estimate the chances of requiring a major repair over the next 3 years as 5%, a minor repair at 60%, and no repair at 35%.

a) Why is X = payout of dental insurance a random variable?

The payout of dental insurance is random because there is uncertainty in whether or not a dental repair is needed. We may be able to do things to help prevent cavities (e.g. flossing), but there is a level of randomness involved.

b) Is X discrete or continuous? What are its possible values?

X is a discrete random variable. The sample space S of $X = \{0, 100, 1000\}$.

c) Give the probability distribution of X.

d) How much should the insurance company expect to pay for this policy?

$$E[X] = (0 * 0.35) + (100 * 0.60) + (1000 * 0.05) = 110$$

6.12

You are watching two items posted for sale on Ebay and bid $30 for the first and $20 for the second item. You estimate that you are going to win the first bid with probability 0.1, second with 0.2 and you assume that winning the two bids are independent events. Let X denote the random variable denoting the total amount of money you will spend on the two items.

a) List the sample space of all possible outcomes of winning or losing the two bids.

$$S = \{0, 30, 20, 50\}$$

b) Find the probability of each outcome in the sample space (use the tree diagram).

Answer:

c) Find the probability distribution of X.

Answer:

X	P(X)
0	0.72
30	0.08
20	0.18
50	0.02

d) Find the mean of X.

Answer:

$$E[X] = (0 * .72) + (30 * .08) + (20 * 0.18) + (50 * 0.02) = 7$$

e) Find the standard deviation of X.

Answer:

$$E[X^2] = (0^2 * .72) + (30^2 * .08) + (20^2 * 0.18) + (50^2 * 0.02) = 194$$

$$E[X]^2 = 49$$
$$Var[X] = E[X^2] - E[X]^2 = 194 - 49 = 145$$
$$SD[X] = \sqrt{145} = 12.04159$$

6.39 - Bidding on Ebay

You are bidding on 4 items available on ebay. You think that for each bid, you have a 25% chance of winning it, and the outcomes of the four bids are independent events. Let X denote the number of winning bids out of the 4 items you bid on.

a) Explain why the distribution of X can be modeled by the binomial distribution.

Answer: The probability of winning each item is the same, they are independent, and each bidding for a win can be treated as a Bernoulli trial. X is being modeled as the number of successes out of the 4 Bernoulli trials, which can be modeled with a binomial distribution.

b) Find the probability that you win exactly 2 bids.

Answer:

$$P(X = 2) = \binom{4}{2}(0.25)^2(0.75)^2 = 0.211$$

c) Find the probability that you win 2 bids or fewer.

Answer:

$$P(X \leq 2) = 0.9492188$$

d) Find the probability that you win more than 2 bids.

Answer:

$$P(X > 2) = 1 - P(X \leq 2) = 1 - 0.9492188 = 0.0507812$$

6.40

For each of the following situations, explain whether the binomial distribution applies to X.

a) You are bidding on the first four items available on Ebay. You think that you will win the first bid with probability 25% and the second bid with probability 30%. Let X denote the number of winning bids out of the 4 items you bid on.

Answer: This cannot be modeled by the binomial distribution because the probability of each winning each bid (the events) are not the same.

b) You are bidding on the first four items available on Ebay. Each bid is for $70, and you think there is a 25% chance of winning a bid, with bids being independent events. Let X be the total amount of money you pay for your winning bids.

Answer: X cannot be directly modeled by the binomial distribution because it is the total amount of money payed for the winning bids. The binomial distribution deals with the number of trial successes (winning bids). We could, however, look at the number of winning bids and then multiply each possible outcome by $70 to get the total amount of money paid.