Textbook Exercises

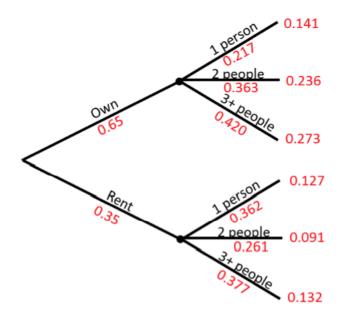
11.56 Housing Units in the US (Graded for Accurateness) According to the 2010 US Census, 65% of housing units in the US are owner-occupied while the other 34% are renter-occupied. The table below shows the probabilities of the number of occupants in a housing unit under each of the two conditions. Create a tree diagram using this information and use it to answer the following questions:

Condition	1	2	3 or more
Owner-occupied	0.217	0.363	0.420
Renter-occupied	0.362	0.261	0.377

- (a) What is the probability that a US housing unit is rented with exactly two occupants?
- (b) What is the probability that a US horsing unit has three or more occupants?
- (c) What is the probability that a unit with one occupant is rented?

Solution

We first create the tree diagram using the information given, and use the multiplication rule to fill in the probabilities at the ends of the branches. For example, for the top branch, the probability of having 1 occupant in an owner-occupied housing unit is $0.65 \cdot 0.217 = 0.141$.



- (a) We see at the end of the branch with rented and 2 occupants that the probability is 0.091.
- (b) There are two branches that include having 3 or more occupants and we use the addition rule to see that the probability of 3 or more occupants is 0.273 + 0.132 = 0.405.

(c) This is a conditional probability (or Bayes? rule). We have:

$$P(\text{rent if 1}) = \frac{P(\text{rent and 1 person})}{P(\text{1person})} = \frac{0.127}{0.141 + 0.127} = \frac{0.127}{0.268} = 0.474$$

If a housing unit has only 1 occupant, the probability that it is rented is 0.474.

11.83 Owner-Occupied Household Size (Graded for Accurateness) The table below gives the probability function for the random variable giving the household size for an owner-occupied housing unit in the US.

X	1	2	3	4	5	6	7
p(x)	0.217	0.363	0.165	0.145	0.067	0.026	0.028

- (a) Verify that the sums of the probabilities is 1 (up to round-off error).
- (b) What is the probability that a unit has only one or two people in it?
- (c) What is the probability that a unit has five or more people in it?
- (d) What is the probability that more than one person lives in a US owner-occupied housing unit?

Solution

- (a) We see that 0.217 + 0.363 + 0.165 + 0.145 + 0.067 + 0.026 + 0.018 = 1.001. This is different from 1 just by round-off error on the individual probabilities.
- (b) We have p(1) + p(2) = 0.217 + 0.363 = 0.580.
- (c) We have p(5) + p(6) + p(7) = 0.067 + 0.026 + 0.018 = 0.111.
- (d) It is easiest to find this probability using the complement rule, since more than 1 occupant is the complement of 1 occupant for this random variable. The answer is 1-p(1) = 1-0.217 = 0.783.

11.85 Average Household Size for Owner-Occupied Units (Graded for Accurateness) The table shown in the previous question gives the probability function for the random variable giving the household size for an owner-occupied housing unit in the US.

- (a) Find the mean household size.
- (b) Find the standard deviation for household size.

Solution

(a) We multiply the values of the random variable by the corresponding probability and add up the results. We have

$$\mu = 1(0.217) + 2(0.363) + 3(0.165) + 4(0.145) + 5(0.067) + 6(0.026) + 7(0.018) = 2.635$$

The average household size for an owner-occupied housing unit in the US is 2.635 people.

(b) To find the standard deviation, we subtract the mean of 2.635 from each value, square the difference, multiply by the probability, and add up the results to find the variance; then take a square root to find the standard deviation.

$$\sigma^2 = (1 - 2.635)^2 \cdot 0.217 + (2 - 2.635)^2 \cdot 0.363 + \dots + (7 - 2.635)^2 \cdot 0.018$$

$$\begin{array}{rcl} &=& 2.03072 \\ \Rightarrow \sigma &=& \sqrt{2.03072} = 1.425 \end{array}$$

11.87 Fruit Fly Lifetimes (Graded for Completeness) Suppose that the probability function shown below reflects the possible lifetimes (in months after emergence) for fruit flies.

X	1	2	3	4	5	6
p(x)	0.30	?	0.20	0.15	0.10	0.05

- (a) What proportion of fruit flies die in their second month?
- (b) What is the probability that a fruit fly lives more than four months?
- (c) What is the mean lifetime for a fruit fly?
- (d) What is the standard deviation of fruit fly lifetimes?

Solution

Let the random variable X measure fruit fly lifetimes (in months).

(a) The probabilities must add to 1, so the proportion of dying in the second month is

$$P(X = 2) = 1 - (0.30 + 0.20 + 0.15 + 0.10 + 0.05) = 1 - 0.80 = 0.20$$

(b)
$$P(X > 4) = P(X = 5) + P(X = 6) = 0.10 + 0.05 = 0.15$$

(c) The mean fruit fly lifetime is

$$\mu = 1(0.30) + 2(0.20) + 3(0.20) + 4(0.15) + 5(0.10) + 6(0.05) = 2.7$$
 months

(d) The standard deviation of fruit fly lifetimes is

$$\sigma = (1 - 2.7) \cdot 0.30 + (2 - 2.7) \cdot 0.20 + \cdots + (6 - 2.7) \cdot 0.05 = 2.31 = 1.52$$
 months

11.95 Getting to the Finish (Graded for Completeness) In a certain board game participants roll a standard six-sided die and need to hit a particular value to get to the finish line exactly. For example, if Carol is three spots from the finish, only a roll of 3 will let her win; anything else and she must wait another turn to roll again. The chance of getting the number she wants on any roll is p = 1/6 and the rolls are independent of each other. We let a random variable X count the number of turn until a player gets the number needed to win. The possible values of X are 1,2,3,... and the probability function for any particular count is given by the formula

$$P(X = k) = p(1 - p)^{k - 1}$$

- (a) Find the probability a player finishes on the third turn.
- (b) Find the probability a player takes more than three turns to finish.

Solution

(a) Using the formula for the probability function with p=1/6 and k=3 we have

$$P(X=3) = \left(\frac{1}{6}\right) \left(1 - \frac{1}{6}\right)^{3-1} = \left(\frac{1}{6}\right) \left(\frac{5}{6}\right)^2 = 0.116$$

(b) The event "more than three turns to finish" or X > 3 includes X = 4, 5, 6, ..., an infinite number of possible outcomes! Fortunately we can use the complement rule.

$$P(X > 3) = 1 - (p(1) + p(2) + p(3))$$

$$= 1 - \left[\left(\frac{1}{6} \right) \left(\frac{5}{6} \right)^0 + \left(\frac{1}{6} \right) \left(\frac{5}{6} \right)^1 + \left(\frac{1}{6} \right) \left(\frac{5}{6} \right)^2 \right]$$

$$= 1 - [0.1667 + 0.1389 + 0.1157]$$

$$= 1 - 0.4213$$

$$= 0.5787$$

11.117 Boys or Girls? (Graded for Completeness) Worldwide, the proportion of babies who are boys is about 0.51. A couple hopes to have three children and we assume that the sex of each child is independent of the others. Let the random variable X represent the number of girls in the three children, so X might be 0, 1, 2, or 3. Give the probability function for each value of X.

Solution

A probability function gives the probability for each possible value of the random variable. This is a binomial random variable with n=3 and p=0.49 (since we are counting the number of girls not boys). The probability of 0 girls is:

$$P(X=0) = {3 \choose 0} (0.49^0)(0.51^3) = 1 \cdot 1 \cdot 0.51^3 = 0.133$$

The probability of 1 girl is:

$$P(X=1) = {3 \choose 1} (0.49^1)(0.51^2) = 3 \cdot (0.49^1)(0.51^2) = 0.382$$

The probability of 2 girls is:

$$P(X=2) = {3 \choose 2} (0.49^2)(0.51^1) = 3 \cdot (0.49^2)(0.51^1) = 0.367$$

The probability of 3 girls is:

$$P(X=3) = {3 \choose 3} (0.49^3)(0.51^0) = 1 \cdot (0.49^3) \cdot 1 = 0.118$$

We can summarize these results with a table for the probability function.

X	0	1	2	3
p(x)	0.133	0.382	0.367	0.118

Notice that the four probabilities add up to 1, as we expect for a probability function.

11.121 Owner-Occupied Housing Units (Graded for Accurateness) In the 2010 US Census, we learn that 65% of all housing units are owner-occupied while the rest are rented. If we take a random sample of 20 housing units, find the probability that:

- (a) Exactly 15 of them are owner-occupied.
- (b) 19 or more of them are owner-occupied.

Solution

If X is the random variable giving the number of owner-occupied units in a random sample of 20 housing units in the US, then X is a binomial random variable with n = 20 and p = 0.65.

(a) To find P(X=15), we first calculate $\binom{20}{15} = \frac{20!}{15!(5!)} = 15,504$. We then find

$$P(X = 15) = {20 \choose 15} (0.65^{15})(0.35^5) = 15,504(0.65^{15})(0.35^5) = 0.1272.$$

(b) We know that $P(X \ge 18) = P(X = 18) + P(X = 19) + P(X = 20)$, and we calculate each of the terms separately and add them up. We have

$$P(X = 18) = {20 \choose 18} (0.65^{18})(0.35^2) = 190(0.65^{18})(0.35^2) = 0.0100$$

$$P(X = 19) = {20 \choose 19} (0.65^{19})(0.35^1) = 20(0.65^{19})(0.35^1) = 0.0020$$

$$P(X = 20) = {20 \choose 20} (0.65^{20})(0.35^0) = 1 \cdot (0.6520) \cdot 1 = 0.0002$$

Then we have

$$P(X \ge 18) = P(X = 18) + P(X = 19) + P(X = 20) = 0.0100 + 0.0020 + 0.0002 = 0.0122$$

- 11.128 Airline Overbooking (Graded for Accurateness) Suppose that past experience shows that about 10% of passengers who are schedule to take a particular flight fail to show up. For this reason, airlines sometimes overbook flights, selling more tickets than they have seats, with the expectation that they will have some no shows. Suppose an airline used a small jet with seating for 30 passengers on a regional route and assume that passengers are independent of each other in whether they show up for the flight. Suppose that the airline consistently sells 32 tickets for every one of these flights.
 - (a) On average, how many passengers will be on each flight?
 - (b) How often will they have enough seats for all of the passengers who show up for the flight?

Solution

Let X measure the number of passengers (out of 32) who show up for a flight. For each passenger we have a 90% chance of showing up, so X is a binomial random variable with n = 32 and p = 0.90.

- (a) The mean number of passengers on each flight is $\mu = np = 32(0.9) = 28.8$ people.
- (b) Everyone gets a seat when $X \leq 30$. To find this probability we use the complement rule (find the chance too many people show up with X = 31 or X = 32, then subtract from one.)

$$P(X \le 30) = 1 - [P(X = 31) + P(X = 32)]$$
$$= 1 - \left[{32 \choose 31} 0.9^{31} 0.1^{1} + {32 \choose 32} 0.9^{32} 0.1^{0} \right]$$

$$= 1 - [32 \cdot 0.9^{31}(0.1) + 1 \cdot 0.9^{32} \cdot 1]$$

$$= 1 - [0.122 + 0.034]$$

$$= 1 - 0.156$$

$$= 0.844$$

Everyone will have a seat on about 84.4% of the flights. The airline will need to deal with overbooked passengers on the other 15.6% of the flights.

Computer Exercises

For each R problem, turn in answers to questions with the written portion of the homework. Send the R code for the problem to Katherine Goode. The answers to questions in the written part should be well written, clear, and organized. The R code should be commented and well formatted.

R problem 1 (Graded for Completeness) Use the data on page 280 from Exercise 4.136 to use R to compute a p-value from the exact probability distribution. Compare with the answer you get from 10,000 simulations of the randomization distribution using R. (Either write new code or reuse code from a previous assignment for the randomization test.)

Solution

In 1980, it was shown that the active ingredient in marijuana outperformed a placebo in reducing nausea in chemotherapy patients. Further experiments have been performed to determine if the drug has other medicinal uses. The experiment which we are interested in done on 55 patients with HIV. The patients were randomly assigned to two groups. One group received cannabis (marijuana) and the other group received a placebo. All of the patients had severe neuropathic pain, and the response variable is whether or not pain was reduced by 30% or more. The following table shows the data from the experiment.

	Pain Reduced	Pain Not Reduced	Total
Cannabis	14	13	27
Placebo	7	21	28
Total	21	34	55

We are interested in determine whether marijuana is more effective than the placebo in relieving pain.

 p_c = proportion of cannabis patients who had their pain reduced by more than 30%

 p_p = proportion of placebo patients who had their pain reduced by more than 30%

Thus, we are interested in testing the following hypotheses.

$$H_0: p_c = p_p \text{ vs } H_A: p_c > p_p$$

Our observed statistics are as follows.

$$p_c = \frac{14}{27} = 0.519$$
 $p_p = \frac{7}{28} = 0.25$

	Pain Reduced	Pain Not Reduced
Cannabis	15	12
Placebo	6	22

In order to determine the p-value, we need to consider the cases where the number of cannabis patients who had reduced pain is greater than 14 since these are the cases, which are more extreme. The following table is one case.

Let X be the number of patients out of a sample of 27 cannabis patients who have reduced pain in this study which has a total of 55 patients of which 21 have reduced pain. Thus, we consider the following probability, which is our p-value.

$$P(14 \le X \le 21) = P(X = 14 \cup X = 15 \cup X = 16 \cup X = 17 \cup X = 18 \cup X = 19 \cup X = 20 \cup X = 21)$$

$$= P(X = 14) + P(X = 15) + P(X = 16) + P(X = 17) + P(X = 18) + P(X = 19)$$

$$+P(X = 20) + P(X = 21)$$

Consider that

$$P(X = 14) = \frac{\text{\# of ways to choose } 27 \text{ cannabis patients out of the } 55 \text{ total so } 14 \text{ have reduced pain}}{\text{total } \# \text{ of ways } 27 \text{ cannabis patients can be chosen}}$$

$$= \frac{\text{(choose } 14 \text{ from } 21 \text{ with reduced pain}) \times \text{(choose other } 13 \text{ from } 34 \text{ with no reduction})}{\text{choose } 27 \text{ from } 55 \text{ total}}$$

$$= \frac{\binom{21}{14}\binom{34}{13}}{\binom{55}{27}}$$

Thus.

$$P(14 \le X \le 21) = P(X = 14 \cup X = 15 \cup X = 16 \cup X = 17 \cup X = 18 \cup X = 19 \cup X = 20 \cup X = 21)$$

$$= P(X = 14) + P(X = 15) + P(X = 16) + P(X = 17) + P(X = 18) + P(X = 19)$$

$$+P(X = 20) + P(X = 21)$$

$$= \frac{\binom{21}{14}\binom{34}{13}}{\binom{55}{27}} + \frac{\binom{21}{15}\binom{34}{12}}{\binom{55}{27}} + \frac{\binom{21}{16}\binom{34}{11}}{\binom{55}{27}} + \cdots + \frac{\binom{21}{20}\binom{7}{7}}{\binom{55}{27}} + \frac{\binom{21}{21}\binom{34}{6}}{\binom{55}{27}}$$

$$= 0.03774$$

This value can be found using all of the following methods in R.

```
 \begin{array}{l} {\rm choose}(21,14)*{\rm choose}(34,13)/{\rm choose}(55,27)+{\rm choose}(21,15)*{\rm choose}(34,12)/{\rm choose}(55,27)+{\rm choose}(21,16)*{\rm choose}(34,11)/{\rm choose}(55,27)+{\rm choose}(21,17)*{\rm choose}(34,10)/{\rm choose}(55,27)+{\rm choose}(21,18)*{\rm choose}(34,9)/{\rm choose}(55,27)+{\rm choose}(21,19)*{\rm choose}(34,8)/{\rm choose}(55,27)+{\rm choose}(21,20)*{\rm choose}(34,7)/{\rm choose}(55,27)+{\rm choose}(21,21)*{\rm choose}(34,6)/{\rm choose}(55,27)+{\rm choose}(21,21)*{\rm choose}(34,6)/{\rm choose}(55,27)+{\rm choose}(34,21)+{\rm choose
```

fisher.test(mat, alternative = "greater")

Now we use R to create a randomization with 10,000 simulations. This is done using the following code.

```
p.hat <- numeric(10000)
p.c.observed.2 <- 14/27
for (i in 1:10000)
{
      c.2 <- sum(sample(c(rep(1,21),rep(0,34)),size=27,replace=FALSE))
      p.hat[i] <- c.2/27
}
pvalue <- sum(p.hat>=p.c.observed)/10000
```

Doing this simulation, we calculate a p-value of 0.037. We note that this value is very similar to the p-value calculated from the exact probability distribution.

R problem 2 (Graded for Accurateness) Consider a hypothesis test $H_0: \mu = 100$ versus $H_A: \mu > 100$ from data where the test statistic \bar{X} is normally distributed with mean μ and standard deviation 5 (so the sample size is large enough for the standard error to be 5).

1. What would the p-value be if $\bar{X} = 108.7$?

Solution

The p-value is calculated as follows.

$$P(\bar{X} \ge 108.7) = 1 - P(\bar{X} \le 108.7)$$

= 0.0409

We used the following code in R to calculate this value.

```
1-pnorm(108.7,100,5)
```

2. What number c would \bar{X} need to exceed for the p-value to be less than 0.05?

Solution

We determine c in the following manner.

$$P(\bar{X} \ge c) = 0.05$$

$$\Leftrightarrow 1 - P(\bar{X} \le c) = 0.05$$

$$\Leftrightarrow 0.95 = P(\bar{X} \le c)$$

$$\Leftrightarrow 108.2243 = c$$

We used the following code in R to calculate this value.

```
qnorm(1-0.05,100,5)
```

3. If the null hypothesis is true, what is the probability that the p-value, as calculated by an area under a normal curve, is less than 0.05?

Solution

First consider that when we calculate a p-value, we always assume the null hypothesis is true. Thus, we first calculate

$$\begin{array}{rcl} P(\bar{X}>a) & < & 0.05 \\ \Leftrightarrow P(\bar{X}$$

under the assumption that $\mu=100$ and $\sigma=5$. Now, we are interested in determining the probability that we would obtain a value that is greater than or equal to a=108.2243, and we are told that the true distribution has $\mu=100$. Thus,

$$P(\bar{X} > 108.2243) = 1 - P(\bar{X} < 108.2243)$$

= 0.05

We used the following code in R.

```
qnorm(0.95,100,5)
1-pnorm(108.2243,100,5)
```

However, we also could have gotten the answer in this manner.

```
1-pnorm(qnorm(0.95,100,5),100,5)
```

4. If the true mean is 104, what is the probability that the p-value is less than 0.05?

Solution

We go through a similar process as in part (c), but this time, the true mean is 104. Thus, when we calculate the probability that we would obtain a value that is greater than or equal to a = 108.2243, we need to use $\mu = 104$, instead of $\mu = 100$. We obtain a probability of 0.199 using the following R code.

```
1-pnorm(qnorm(0.95,100,5),104,5)
```

R problem 3 (Graded for Completeness) A male fruit fly is equally likely to have genotype A or genotype B. If he has genotype A, then in a given cross, all offspring will have red eyes. If he has genotype B, each offspring is equally likely to have red or white eyes, independent of all others. Assume that there are five offspring, all with red eyes.

1. Given all five offspring have red eyes, what is the probability that the fly has genotype A?

Solution

Let

A = genotype A

B = genotype B

R = all 5 offspring have red eyes

Then

$$P(A|R) = \frac{P(A \cap R)}{P(R)}$$

$$= \frac{P(R|A)P(A)}{P(R \cap A) + P(R \cap B)}$$

$$= \frac{P(R|A)P(A)}{P(R|A)P(A) + P(R|B)P(B)}$$

$$= \frac{(1)(\frac{1}{2})}{(1)(\frac{1}{2}) + (\frac{1}{2})^5(\frac{1}{2})}$$

$$= \frac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{64}}$$

$$= 0.969697$$

2. Given all five offspring have red eyes, what is the probability that a sixth offspring will also have red eyes?

Solution

Let

$$R = \text{all 5 offspring have red eyes}$$

$$S = 6$$
th offspring has red eyes

Then

$$P(S|R) = \frac{P(S \cap R)}{P(R)}$$

$$= \frac{P(S \cap R \cap A) + P(S \cap R \cap B)}{P(R \cap A) + P(R \cap B)}$$

$$= \frac{P(S \cap R|A)P(A) + P(S \cap R|B)P(B)}{P(R|A)P(A) + P(R|B)P(B)}$$

$$= \frac{(1)\left(\frac{1}{2}\right) + \left(\frac{1}{2}\right)^{6}\left(\frac{1}{2}\right)}{(1)\left(\frac{1}{2}\right) + \left(\frac{1}{2}\right)^{5}\left(\frac{1}{2}\right)}$$

$$= \frac{\frac{1}{2} + \left(\frac{1}{2}\right)^{7}}{\frac{1}{2} + \left(\frac{1}{2}\right)^{6}}$$

$$= 0.9848485$$

R problem 4 (Graded for Accurateness) In a test to see if a person has ESP, the person identifies a correct shape 33 out of 125 trials. The test is designed so that the number of correct answers should be binomial with p = 0.2 if the null hypothesis of no ESP is true.

1. Use the R function pbinom() to compute a p-value for this hypothesis test.

Solution

We are interested in testing the following hypotheses.

$$H_0: p = 0.2$$

 $H_A: p > 0.2$

We observed $\hat{p} = 33/125 = 0.264$. If we let X = the number of correctly identified shapes, then the p-value that we are interested in is

$$P(33 \le X \le 125) = 0.0502$$

The R code used to compute this value is as follows. Any of these compute the correct answer.

```
pbinom(125, 125, 0.2) - pbinom(32, 125, 0.2)
sum(dbinom(33:125, 125, 0.2))
1-pbinom(32,125,0.2)
```

2. Use R to find a p-value from a randomization distribution (using code from a previous homework or new code). Compare to the previous result.

Solution

In order to find a p-value from a randomization distribution, I used the function created from a previous homework for finding the p-value for a single proportion. The code is shown below.

```
pvalue.p = function(n,x,p0,R,alternative=c("not.equal","less","greater")) {
  alternative = match.arg(alternative)
   p.hat = numeric(R)
   for ( i in 1:R ) {
     p.hat[i] = mean(sample(c(0,1),size=n,replace=TRUE,prob=c(1-p0,p0)))
   p.sample = x/n
   if ( alternative == "not.equal" ) {
     if ( p.sample == p0 ) {
       p.value = 1
     else if ( p.sample < p0 ) {</pre>
       p.value = 2*sum( p.hat <= p.sample ) / R</pre>
     else if ( p.sample > p0 ) {
       p.value = 2*sum( p.hat >= p.sample ) / R
    }
   }
   else if ( alternative == "less" ) {
     p.value = sum( p.hat <= p.sample ) / R</pre>
   else if ( alternative == "greater" ) {
     p.value = sum( p.hat >= p.sample ) / R
   return( p.value )
}
```

For this particular problem, we have $n = 125, x = 33, p_0 = 0.2$, and the alternative is greater than. We choose do 10,000 simulations. We obtain a p-value of 0.0488 using the following code.

We see that this p-value is similar to the one calculated in part (a).

3. Find the mean and standard deviation of the number of correct guesses assuming no ESP. Calculate a p-value by approximating the binomial probability with an area under a normal curve with the same mean and standard deviation. Compare the answer to the first result.

Solution

Since X has a binomial distribution, we have that

$$\mu = np = 125 \cdot 0.2 = 25$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{125 \cdot 0.2(1-0.2)} = \sqrt{20} = 4.472136$$

Now, we approximate the probability

$$P(33 \le X \le 125) = P(X \ge 33)$$

under the assumption that $X \sim N(25, 4.472)$.

$$P(X \ge 33) = 1 - P(X \le 33)$$

= 0.0368

We obtain the p-value using the following R Code

1-pnorm(33,25,4.472)