Problem set 2

## Chapter 3 problems

### Problem 1: Variation on roulette

A casino has opened a roulette table with 48 slots: 22 red, 22 black, and 4 green. The ball is released onto the wheel, where it spins before eventually landing in one of the slots. The slots are all evenly sized and distributed evenly around the wheel.

1. You watch the roulette wheel spin 3 consecutive times and the ball lands on a red slot each time. What is the probability that the ball will land on a red slot on the next spin?
2. You watch the roulette wheel spin 300 consecutive times and the ball lands on a red slot each time. What is the probability that the ball will land on a red slot on the next spin?
3. If you were at a casino and you saw the scenario in part (b) play out, would you be any more likely to bet that the next spin would land in red than if you had seen scenario (a)? Why or why not?

### Problem 2: Dice rolls

If you roll a pair of fair dice, what is the probability of:

1. Getting a sum of 1?
2. Getting a sum of 4?
3. Getting a sum of 12?

### Problem 3: Dice rolls (continued)

If you roll a pair of *unfair* dice, where the probability of rolling 1, 2, 3, 4, 5, and 6 is 0.1, 0.1, 0.2, 0.2, 0.2, and 0.2, respectively (that is, 1s and 2s are half as likely to be rolled as the other numbers), what is the probability of:

1. Getting a sum of 1?
2. Getting a sum of 4?
3. Getting a sum of 12?

### Problem 4: Poverty and language in Durham

In Durham, it is estimated that 17.4% of Durham residents live below the poverty line, 18.8% speak a language other than English (foreign language) at home, and 4.1% fall into both categories.

1. Are living below the poverty line and speaking a foreign language at home disjoint?
2. Draw a Venn diagram summarizing the variables and their associated probabilities.
3. What percent of Durham residents live below the poverty line and only speak English at home?
4. What percent of Durham residents live below the poverty line or speak a foreign language at home?
5. What percent of Durham residents live above the poverty line and only speak English at home?
6. Is the event that a Durham resident lives below the poverty line independent of the event that the person speaks a foreign language at home?

### Problem 5: SES and Smoking

Suppose that 356 people have been polled on their current smoking status (Yes or No) and their socioeconomic status (SES) (High, Middle, or Low), with results shown in the table below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | High | Middle | Low | All |
| Yes | 0.1433 | 0.0618 | 0.1208 | 0.3258 |
| No | 0.4494 | 0.0843 | 0.1404 | 0.6742 |
| Total | 0.5927 | 0.1461 | 0.2612 | 1.0000 |

1. Are being of high SES and currently smoking mutually exclusive?
2. What is the probability that a randomly chosen individual has high SES?
3. What is the probability that a randomly chosen individual has high SES given that they currently smoke?
4. What is the probability that a randomly chosen individual has high SES given that they don’t currently smoke?
5. Do having high SES and smoking appear to be independent?

### Problem 6: Assortative mating reboot

Consider the assortative mating exercise (3.18) in OIS, which says that assortative mating is a nonrandom mating pattern where individuals with similar genotypes and/or phenotypes mate with one another more frequently than what would be expected under a random mating pattern. Suppose researchers revisited this topic and collected data on hair colors of 225 male Duke students and their female partners. The table below summarizes the results. For simplicity, we only include heterosexual relationships in this exercise.

## Female students  
## Male students Blue Brown Green Total  
## Blue 81 24 15 120  
## Brown 17 39 10 66  
## Green 12 9 18 39  
## Total 110 72 43 225

1. What is the probability that a randomly chosen male respondent or his partner has blue eyes?
2. What is the probability that a randomly chosen male respondent with blue eyes has a partner with blue eyes?
3. What is the probability that a randomly chosen male respondent with brown eyes has a partner with blue eyes? What about the probability of a randomly chosen male respondent with green eyes having a partner with blue eyes?
4. Does it appear that the eye colors of male respondents and their partners are independent? Explain your reasoning.

### Problem 7: Covid antibody testing

Artron is working on an antibody test for Covid-19. Suppose that, in a remote region relatively unaffected by the pandemic, only 1% of people have actually had Covid-19. The antibody test is 93% accurate if a person has actually had Covid-19, meaning that the probability of a positive test result when a person has actually had Covid-19 is 0.93. The test is 98% accurate if a person has not had Covid-19. What is the probability that a randomly selected person from this region who tests positive for the Covid-19 antibody by the test actually has had Covid-19?

### Problem 8: Fracking poll

Poll results on the use of hydraulic fracking to extract oil and natural gas resources found that 48% of the respondents supported fracking. Additionally, the poll estimated that of those who did support fracking, 33% had a college degree; on the other hand, 64% of those who did *not* support fracking had a college degree. Suppose we randomly sampled a person who participated in the poll and found that they had a college degree. What is the probability that they support fracking?

## Chapter 4 problems

### Problem 9: Triathlon times (new competitors)

In triathlons, it is common for racers to be placed into age and gender groups. Friends Ronen and Sun both completed the Bull City Triathlon, where Ronen competed in the Men, Ages 30 - 34 group while Sun competed in the Women, Ages 25 - 29 group. Ronen completed the race in 1:24:15 (5055 seconds), while Sun completed the race in 1:30:54 (5454 seconds). Obviously Ronen finished faster, but they are curious about how they did within their respective groups. Can you help them? Here is some information on the performance of their groups:

* The finishing times of the Men, Ages 30 - 34 group has a mean of 4212 seconds with a standard deviation of 609 seconds.
* The finishing times of the Women, Ages 25 - 29 group has a mean of 5289 seconds with a standard deviation of 821 seconds.
* The distributions of finishing times for both groups are approximately Normal.

Remember: a better performance corresponds to a faster finish.

1. Write down the short-hand for these two normal distributions.
2. What are the Z-scores for Ronen’s and Sun’s finishing times? What do these Z-scores tell you?
3. Did Ronen or Sun rank better in their respective groups? Explain your reasoning.
4. What percent of the triathletes did Ronen finish faster than in his group?
5. What percent of the triathletes did Sun finish faster than in her group?
6. If the distributions of finishing times are not nearly normal, would your answers to parts (b) - (e) change? Explain your reasoning.

### Problem 10: Triathlon times (continued)

In the previous problem we saw two distributions for triathlon times: N(; ) for Men, Ages 30 - 34 and N(; ) for the Women, Ages 25 - 29 group. Times are listed in seconds. Use this information to compute each of the following:

1. The cutoff time for the fastest 5% of athletes in the men’s group, i.e. those who took the shortest 5% of time to finish.
2. The cutoff time for the slowest 10% of athletes in the women’s group.

### Problem 11: HSV-2

The CDC estimates that 12% of Americans have HSV-2.

1. Suppose we take a random sample of 100 Americans. Is the use of the binomial distribution appropriate for calculating the probability that exactly 20 out of 100 randomly sampled Americans have HSV-2? Explain.
2. Calculate the probability that exactly 20 out of 100 randomly sampled Americans have HSV-2.
3. What is the probability that exactly 5 out of a new sample of 100 Americans do *not* have HSV-2?
4. What is the probability that at least 1 out of 10 randomly sampled Americans have HSV-2?
5. What is the probability that more than 2 out of 10 randomly sampled Americans have had HSV-2?

### Problem 12: I-40 Speeds

The distribution of passenger vehicle speeds traveling on the Interstate 40 Freeway (I-40) between Durham and Raleigh is nearly normal with a mean of 71.2 miles/hour and a standard deviation of 4.89 miles/hour.

1. What percent of passenger vehicles travel slower than 80 miles/hour?
2. What percent of passenger vehicles travel between 60 and 80 miles/hour?
3. How fast do the fastest 5% of passenger vehicles travel?
4. The speed limit on this stretch of the I-40 is 70 miles/hour. Approximate what percentage of the passenger vehicles travel above the speed limit on this stretch of the I-40.

### Problem 13: DUML Surveys

The Duke Marine Lab campus administration reported that the typical response rate to their alumni surveys is only 14%. If for a particular survey 20,000 alumni are contacted, what is the probability that at least 1,000 will agree to respond?

### Problem 14: Guessing on a quiz

In a multiple choice quiz there are 5 questions and 3 choices for each question (a, b, c). Robyn has not studied for the quiz at all, and they decide to randomly guess the answers. What is the probability that:

1. the first question they get right is the 3rd question?
2. they get exactly 3 or exactly 4 questions right?
3. they get the majority (i.e., 3 or more) of the questions right?