

Chapter 3 Textbook exercises

Solutions to even-numbered questions
Statistics and statistical programming
Northwestern University
MTS 525

Aaron Shaw

September 30, 2020

Contents

3.12 School absences	1
3.22 Exit polling	1
3.28 Birthdays!	1
3.34 Airline baggage fees	2
3.38 Income and gender	3

All exercises taken from the *OpenIntro Statistics* textbook, 4th edition, Chapter 3.

3.12 School absences

- (a) By the addition rule: $P(\text{no missed days}) = 1 - (0.25 + 0.15 + 0.28) = 0.32$
- (b) $P(1 \text{ miss or less}) = P(\text{no misses}) + P(1 \text{ miss}) = 0.32 + 0.25 = 0.57$
- (c) $P(\text{at least 1 miss}) = P(1 \text{ miss}) + P(2 \text{ misses}) + P(\geq 3 \text{ misses}) = 1 - P(\text{no misses}) = 1 - 0.32 = 0.68$
- (d) Assume (foolishly!) that the absences are independent across children. This allows us to use the multiplication rule:
 $P(\text{neither miss any}) = P(\text{no misses}) \times P(\text{no misses}) = 0.32 \times 2 = 0.1024$
- (e) Again, assume that the absences are independent across children and use the multiplication rule:
 $P(\text{both miss some}) = P(\text{at least 1 miss}) \times P(\text{at least 1 miss}) = 0.68 \times 2 = 0.4624$
- (f) Siblings often cohabitate and are therefore likely to get each other sick, so the independence assumption is not sound.

3.22 Exit polling

This one is all about conditional and compound probabilities and could be represented as a tree diagram (if you find those useful).

$$\begin{aligned} P(\text{support}|\text{college}) &= \frac{P(\text{support and college})}{P(\text{college})} \\ &= \frac{0.1961}{0.1961+0.2068} \\ &= 0.49 \end{aligned}$$

3.28 Birthdays!

- (a) Once you have one person's birthday, the probability that the second person has the same birthday is:

$$P(\text{first two share birthday}) = \frac{1}{365} = 0.0027$$

- (b) This one is more challenging! There are many possible approaches, but I find it easiest to think about the probability that none of the three share a birthday in the following way: start with the probability that the first two *don't* share a birthday, followed by the probability that the next person doesn't share a birthday either. This makes it possible to apply the general multiplication rule:

$$\begin{aligned}
 P(\text{at least two share birthday}) &= 1 - P(\text{none of three share birthday}) \\
 &= 1 - P(\text{first two don't share}) \times P(\text{third doesn't share either}) \\
 &= 1 - \left(\frac{364}{365}\right) \times \left(\frac{363}{365}\right) \\
 &= 0.0082
 \end{aligned}$$

3.34 Airline baggage fees

part a First, the average fee per passenger (let's call that \bar{F}) is the sum of the expected values of the fees per passenger at each of the three possible fee levels (determined by number of bags checked). This works out to the sum of the fees per bag (let's write this f_b) times the probability (proportion in the case of a binomial process) of passengers with each number of bags (call that P_b). We can now put that in slightly more formal notation and work out the arithmetic:

$$\begin{aligned}
 \bar{F} &= E(\text{Fee per passenger}) = \sum_{b=0}^2 (f_b \times P_b) \\
 &= \$0(0.54) + \$25(0.34) + \$60(0.12) \\
 &= \$0 + \$8.5 + \$7.2 = \$15.70
 \end{aligned}$$

To calculate the standard deviation of the expected value, we need to find the square root of the variance. To find the variance, we need to find the deviance (difference from the expected value) at each fee level, multiply those deviances by the probability (again, the proportion, in a binomial process) of the respective fee levels, and then sum them up. Here's what that looks like:

Bags	$(F - E(F))^2 =$	Deviance
0	$(0 - 15.70)^2 =$	246.49
1	$(25 - 15.70)^2 =$	86.49
2	$(60 - 15.70)^2 =$	1962.49

Bags	$(F - E(F))^2 \times P(F) =$	Deviance
0	$246.49 \times 0.54 =$	133.10
1	$86.49 \times 0.34 =$	29.41
2	$1962.49 \times 0.12 =$	235.50

I sum that last column of values to find the variance (traditionally notated using the greek letter sigma squared (σ^2)):

$$\sigma_{\bar{F}}^2 = \$133.10 + \$29.41 + \$235.50 = \$398.01$$

And take the square root to find the standard deviation (traditionally notated as sigma (σ)):

$$\sigma_{\bar{F}} = \sqrt{\$398.01} = \$19.95$$

part b To calculate this using the tools introduced in the chapter, we'll need to assume independence between the baggage choices of individual passengers (and this is probably wrong, but maybe not catastrophic for the precision of our estimate? Who knows.):

Once we assume independence between passengers, we can calculate the expected total revenue (let's call that $E(\text{revenue})$) by summing the individual expected revenue over the 120 passengers. We can calculate the corresponding standard deviation of the expected total revenue by summing the individual variances

and then taking the square root of that sum. Plug and chug using the values we calculated in Part a of this exercise to find the answers:

$$\begin{aligned} E(\text{revenue}) &= 120 \times \$15.70 = && \$1,884 \\ \sigma_{E(\text{revenue})}^2 &= 120 \times \$398.01 = && \$47,761.20 \\ \sigma_{E(\text{revenue})} &= \sqrt{\$47,761.20} = && \$218.54 \end{aligned}$$

3.38 Income and gender

- (a) The distribution is right skewed, with a median somewhere around \$35-\$50,000. There's a long tail out to the right (high positive values).
- (b) By the addition rule:

$$P(\text{Income} < \$50k) = 2.2 + 4.7 + 15.5 + 18.3 + 21.2 = 62.2\%$$

- (c) If we assume that income and gender are independent then we can use the general multiplication rule to work out an answer based on compound probability:

$$P(\text{Income} < \$50k \text{ and female}) = P(\text{Income} < \$50k) \times P(\text{female}) = 0.622 \times 0.41 = 0.255$$

- (d) If the variables income and gender were independent (unrelated) then we might expect the actual proportion of women with incomes less than \$50k to equal the total sample proportion. The actual proportion of women with incomes less than \$50k (71.8%) turned out to be a lot higher than than the sample proportion (62.2% from the table). Shockingly, it seems that the assumption that income and gender are independent (unrelated) may not be valid in this data.