Review for Exam 1

I picked one problem from each general area of the homework:

- p43, #29: contingency table problem
- (not in book): comparing groups
- p149, #19: normal probabilities, plus an extra part
- p301, #28: simulation
- p382, #32: basic probability
- p406, #25: probability (hint: draw a picture)
- p407, #33: conditional probability (hint: draw a picture)
- p428, #18: probability model, i.e. expected value and standard deviation
- p448, #20: probability and binomial model

p43, #29

(a)

We want a fraction of the number of Asian applicants to all applicants: $P(Asian) = 292/1755$.

(b)

Of only those students accepted (931), 110 were Asian: $P(Asian|accepted) = 110/931$.

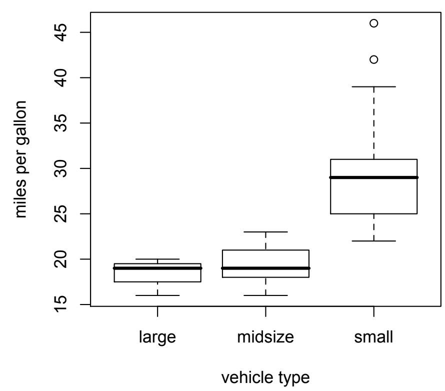
(c)

There were 292 Asian applicants, and 110 of them were accepted: $P(accepted|Asian) = 110/292$.

(d)

Of all students (1755), 931 were accepted: $P(accepted) = 931/1755$.

cars example (not in book): compare the mileage across vehicle type.



Topics we want to touch on are shape, center, spread, and any unusual characteristics. Below is a very thorough comparison.

It appears that the typical large and midsize vehicles have about the same mileage and have a center around the medians at 18 (19?) mpg whereas small cars typically do much better with a median of about 28 mpg[1]. Variation in mileage seems appears to be larger in smaller cars, which have an IQR of about 7 mpg while large and midsize cars have smaller variation with IQRs of about 2 to 3, respectively[2].

All distributions are roughly symmetric, with perhaps slight skew towards the high end (right skew) for small vehicles and slight skew to the low end (left skew) for large vehicles. There are only two potentially unusual characteristics, which are two possible outliers in the small group at about 41 and 46 mpg.

Note that we cannot determine how many modes there are, which we would want to discuss if we had histograms.

[1]Notice that I am not just listing off statistics but *comparing* the center of the groups.

[2]Again, while I use statistics in my description, it is important to compare across groups.

p149, #19

The weights of the cattle follow $N(1152, 84)$.

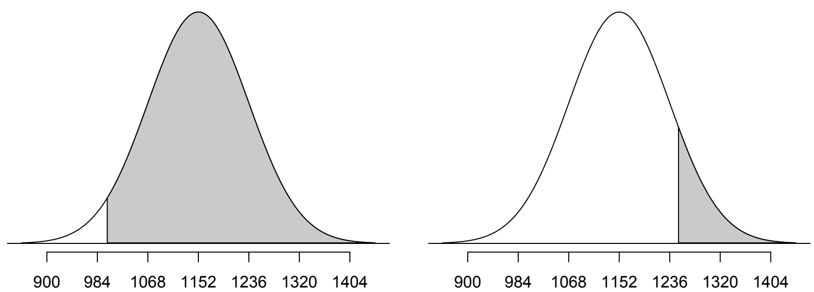
(a)

$Z_{1000} = \frac{1000-1152}{84} = -1.81$, i.e. a steer weighting 1000 lbs would be 1.81 standard deviations below the mean.

(b)

To determine which is more unusual, we look at the Z score of each: $Z_{1250} = 1.17$. Since the Z score for a steer at 1000 lbs is more standard deviations from the mean than a steer weighing 1250 lbs and we are working with a normal model, the 1000 lb steer is more unusual.

(c, extra) How often would you expect to see a steer that weights 1000 lbs or more? What about 1250 lbs or more?



Looking in the normal probability table we get 0.035 and 0.878, however, the table always gives us the lower tails. To get the upper tails, we take 1 minus each value (separately) to get 0.965 and 0.122.

p301, #28. Take the setup and do a simulation of 10 people.

To setup the simulation, we must assign some random numbers to be type O blood and the rest to be not type O blood. Since 44% should be type O, we could use two digits per simulation. We could assign

00-43 type O (O)

44-99 not type O (N)

(There are other alternatives to this setup, e.g. [00-55] is not type O and [56-99] is type O.)

I use the following random numbers: 22678 50247 39401 11042. Then I get O (22), N (67), N (85), O (02), N (47), O (39), O (40), O (11), O (10), O (42).

p382, #32

(a)

In this example, we have 45% type O, 40% type A, 11% type B, and the rest (100% - 45% - 40% - 11% = 4%) type AB. Being type A or B are two disjoint outcomes, so we can just find

$$P(A \text{ or } B) = P(A) + P(B) = 0.4 + 0.11 = 0.51$$

The probability of not type O is one minus the probability of type O: 0.55.

(b)

As long as they are not coming in as a family to donate blood, it seems reasonable to assume the four donors are independent. I use superscripts to denote donor ^{1}, donor ^{2}, and so on.

$$\begin{aligned}
P(^1O,^2O,^3O,^4O) &= P(^1O) * P(^2O) * P(^3O) * P(^4O) = (0.45)^4 = 0.041 \\
P(\text{no one is type AB}) &= P(^1\text{not AB},^2\text{not AB},^3\text{not AB},^4\text{not AB}) \\
&= P(^1\text{not AB}) * P(^2\text{not AB}) * P(^3\text{not AB}) * P(^4\text{not AB}) \\
&= (0.96)^4 = 0.849 \\
P(\text{not all type A}) &= 1 - P(\text{all type A}) \\
&= 1 - P(^1A,^2A,^3A,^4A) \\
&= 1 - P(^1A) * P(^2A) * P(^3A) * P(^4A) \\
&= 1 - (0.4)^4 = 0.974 \\
P(\text{at least one type B}) &= 1 - P(\text{no type B}) \\
&= 1 - P(^1\text{not B},^2\text{not B},^3\text{not B},^4\text{not B}) \\
&= 1 - P(^1\text{not B}) * P(^2\text{not B}) * P(^3\text{not B}) * P(^4\text{not B}) \\
&= 1 - (0.89)^4 = 0.373
\end{aligned}$$

p406, #25

You should make a Venn diagram for this problem. I would make one for this answer key, however, I don't have any (easy) way to make it presently so I will not.

(a)

We are given $P(\text{campylobacter}) = 0.81$, $P(\text{salmonella}) = 0.15$, and $P(\text{both}) = 0.13$. If you make a Venn diagram (you should), you will get the following disjoint probabilities

$$\begin{aligned} P(\text{both camp. and salm.}) &= 0.13 \\ P(\text{camp., no salm.}) &= 0.81 - 0.13 = 0.68 \\ P(\text{salm., no camp.}) &= 0.15 - 0.13 = 0.02 \\ P(\text{no contamination}) &= 1 - 0.13 - 0.68 - 0.02 = 0.17 \end{aligned}$$

(b)

NO. It is possible for contamination with both bacteria at the same time (with probability 0.13).

(c)

We can check the definition of independence:

$$\begin{aligned} P(both) &\stackrel{?}{=} P(camp) * P(salm) \\ 0.13 &\stackrel{?}{=} 0.81 * 0.15 \\ 0.13 &\neq 0.1215 \end{aligned}$$

Not independent (but it is close).

p407, #33

I will denote the "first flight on time" with `flYes` and if the flight isn't on time with `flNo`. If the luggage makes it, I denote that with `lugYes` and if it does not make it I use `lugNo`.

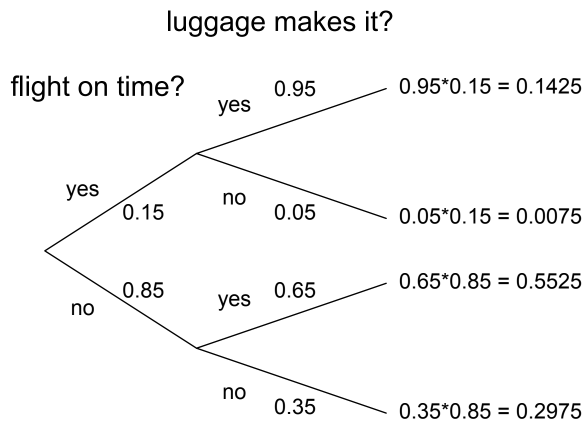
We are given $P(flYes) = 0.15$, $P(lugYes|flYes) = 0.95$, and $P(lugYes|flNo) = 0.65$. Since we see that we can split things up first by flight and then by whether the luggage makes it (or not), we should realize that we might need to make a tree diagram at some point for this problem.

(a)

No, the flight and luggage are absolutely **not** independent. Whether the flight is on time or not influences the probability of whether the luggage will make it. (If they were independent, the flight being on time would have no influence on this probability.)

(b)

Tree diagram time. We want $P(lugYes)$.



Then $P(lugYes)$ is the sum of the two (disjoint) cases where the luggage makes it: $0.1425 + 0.5525 = 0.695$.

p428, #18

We want to find the expected value and standard deviation. For these sorts of problems, we want to build a table like the following:

X	0	1	2	3	4	5	sum (when relevant)
$P(X)$	0.05	0.25	0.35	0.15	0.15	0.05	
$X * P(X)$	0.05	0.25	0.7	0.45	0.6	0.25	$EX = 2.25$
$Dev = X - EX$	-2.25	-1.25	-0.25	0.75	1.75	2.75	
$Dev^2 = (X - EX)^2$	5.063	1.563	0.063	0.563	3.063	7.563	
$Dev^2 * P(X)$	0.253	0.391	0.022	0.084	0.459	0.378	$\sigma^2 = 1.588$

The expected value was 2.25 red lights (part a). The standard deviation is $\sigma = \sqrt{1.588} = 1.26$ red lights (part b).

p448, #20

This problem we should identify as having $n = 6$ trials (arrows shot), although this won't be relevant for (a-c). I use superscript numbers to denote each arrow and `Y` to denote a bull's-eye and `N` to denote NOT a bull's-eye.

Throughout, we assume independence between shots holds. We also have $P(success) = 0.8$.

(a)

$$P(\text{first bull's-eye on third arrow}) = P(^1N,^2N,^3Y) = P(N) * P(N) * P(Y) = 0.2 * 0.2 * 0.8 = 0.032$$

Here we have said nothing about the last three arrows so we need not include them in our probability.

(b)

$$P(\text{miss at least once}) = 1 - P(\text{makes it every time}) = 1 - 0.8^6 = 0.738$$

(c)

Notice that these are two disjoint events. (Your first bull's-eye cannot be on the 4th AND the 5th shot.)

$$\begin{aligned} P(\text{first Y on 4th or 5th}) &= P(\text{first Y on 4th}) + P(\text{first Y on 5th}) \\ &= P(NNNY) + P(NNNNY) \\ &= 0.0064 + 0.00128 = 0.00768 \end{aligned}$$

(d)

Now we start dealing with binomial problems. We have six trials ($n = 6$) and we are looking for four bull's-eyes, which we call a success ($x = 4$), and we have a probability of an individual success as $p = 0.8$. Since we are looking for *exactly* four successes, this is a binomial probability:

$$\begin{aligned} P(\text{exactly 4 successes in 6 trials}) &= \binom{6}{4} (0.8)^4 (1-0.8)^{6-4} \\ &= \frac{6!}{4!(6-4)!} (0.8)^4 (1-0.8)^{6-4} \\ &= \frac{6*5*4*3*2*1}{(4*3*2*1)*(2*1)} (0.8)^4 (1-0.8)^{6-4} \\ &= \frac{6*5}{2*1} (0.8)^4 (1-0.8)^{6-4} \\ &= 0.24576 \end{aligned}$$

(e)

The probability of getting *at least* four bull's-eyes is NOT a binomial problem, although we can make it into a binomial problem by breaking up the cases of *at least 4* into 4 OR

5 OR 6 bull's-eyes:

$$\begin{aligned} P(\text{at least 4 successes in 6 trials}) &= P(\text{4 or 5 or 6 successes in 6 trials}) \\ &= P(4) + P(5) + P(6) \end{aligned}$$

where $P(4)$ is P(exactly four successes in 6 trials), etc. Each of these three probabilities are binomial. We've already found the first in part (d): 0.246. We can find the second by using $x = 5$ instead of $x = 4$, and we can find the last by using $x = 6$ (although you should be able to easily do the last one without the binomial formula).

The final result should be 0.901; you will need to find $P(5) = 0.393$ and $P(6) = 0.262$ to get this answer.

(f)

She gets at most 4 bull's-eyes. This again, is not explicitly binomial but we can make it binomial. We know *at most 4* is the same as 0 OR 1 OR 2 OR 3 OR 4. Alternatively, I would advise using the complement to reduce your computations, as below:

$$P(\text{at most 4}) = 1 - P(\text{at least 5}) = 1 - P(\text{5 or 6}) = 1 - [P(5) + P(6)]$$

We already found $P(5)$ (probability of exactly 5 bull's-eyes in 6 shots) in part (e), and same for $P(6)$, so we can now compute our final result:

$$1 - [P(5) + P(6)] = 1 - P(5) - P(6) = 1 - 0.393 - 0.262 = 0.345$$