Statistical Foundations of Data Science
Example Problems

The content of this assignment is based on the second lecture in the course titled "How to lie with statistics".

1. A survey of six lion-prides reveals that they have 8,14, 16,20,22 and 30 members.
   (a) How many lions are in the average pride? [3] $\frac{8+14+16+20+22+30}{6} = 18.333$
   (b) What would be the average value if we asked each lion how many lions are in their pride? [7] $\frac{8^2+14^2+16^2+20^2+22^2+30^2}{110} = \frac{2300}{110} = 20.9090909$

2. A mathematician, a computer scientist and a statistician discuss on which day of the week they were born. They decide to call the day of the week they were born on there week-birthday. So for example if you were born on a Tuesday your week-birthday would be Tuesday. Assuming that people are equally likely to be born on any day of the week and that these week-birth-dates are independent of each other find:
   (a) The probability that at least two of out protagonists share a week-birthday. [4]
   $1 - \frac{7}{7}\frac{6}{7}\frac{5}{7} = \frac{19}{49} = 0.387755$
   (b) The probability that all three share a week-birthday. [2]
   $\frac{1}{49} = 0.02040816326$ Fix the mathematician's birthday. There is then a one in seven chance each that the computer scientist and statistician share it.
   (c) The probability that exactly two share a week-birthday (and the third does not) [2]
   Using the previous two question $\frac{19}{49} - \frac{1}{49} = \frac{18}{49} = 0.36734693877$
   (d) The probability that the Statistician and Computer Scientist share a week-birthday but that the mathematician is on a different day. [2]
   By symmetry it's a third of our answer in part c so $\frac{6}{49} = 0.12244897959$

3. Monty Hall decides to shake up his game show. He now has four doors behind two of which are cars, one car is red and the other is blue and behind the other two are goats. The doors are labelled A,B,C and D. Our heroine chooses door A. Monty as usual opens a door behind which is a goat this time he
   (a) Find the probability of winning either car if we do not switch. [2]
   $\frac{1}{2} = 0.5$
   (b) Find the probability of winning either car if we do switch. [3]
   $\frac{3}{4} = 0.75$ Here the total probability should add up to two. This is distributed over the other two doors.
   (c) Find the probability of winning the red car if we do not switch. [2]
   By symmetry this should be half our answer in part a $\frac{1}{4} = 0.25$
   (d) Find the probability of winning the red car if we do switch. [3]
   By symmetry this should be half our answer in part b $\frac{3}{8} = 0.375$

4. (a) Six computers randomly and uniformly choose integer numbers between one and six inclusive. Compute the probability that they all select different numbers [3]

$$\frac{6}{6}\cdot\frac{5}{6}\cdot\frac{4}{6}\cdot\frac{3}{6}\cdot\frac{2}{6}\cdot\frac{1}{6} = \frac{6!}{6^6} = \frac{120}{46656} = 0.00257201646$$

(b) Compute the probability if the numbers are now selected between one and ten (there should remain six computers) [3]

$$\frac{10}{10}\cdot\frac{9}{10}\cdot\frac{8}{10}\cdot\frac{7}{10}\cdot\frac{6}{10}\cdot\frac{5}{10} = 0.1512$$

(c) Compute the probability if the numbers are now selected between one and ten and we there are now ten computers choosing [3]

$$\frac{10}{10}\cdot\frac{9}{10}\dots\cdot\frac{2}{10}\cdot\frac{1}{10} = \frac{10!}{10^{10}} = 0.00036288$$

(d) What do your answers to the first and third sub-question suggest about the probability of no ties when we have n computers and the numbers are selected from between 1 and n as n tends to infinity. [1]
It suggests (and it turns out to be true) that the probability of n players all choosing different numbers decreases as n increases. Also that the probability tends to zero (which is again true).

5. In the town of Ezissaib, there are two kinds of businesses. The thee large corporate businesses each employ one hundered people and the twenty small shops each employ five people. Everyone in the town is employed by exactly one business.

(a) What proportion of population work in small businesses? [3]
One hundred people work in the twenty small shops and three hundred work in the three large corporations. Therefore the proportion working in the small businesses is: $\frac{100}{400} = 0.25$

(b) If we were to ask everyone in the town how many people work in the business that they're employed in. What would the average of the answers we'd get if everyone answered honestly? [4]
One hundred people work in the twenty small shops and three hundred work in the three large corporations. Therefore the proportion working in the small businesses is: $0.25\times 5 + 0.75\times 100 = 76.25$

(c) If we asked each business how many people work in their business and averaged those answers, what answer would we get? [3]
$\frac{3}{23}\times 100 + \frac{20}{23}\times 5 = 17.3913043478$

6. Prove $\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}$ algebraically. [10]

$$
\begin{aligned}
\binom{n}{k} + \binom{n}{k+1} &= \\
\frac{n!}{k!(n-k)!} + \frac{n!}{(k+1)!(n-k-1)!} &= \\
\frac{n!}{k!(n-k-1)!}[\frac{1}{n-k} + \frac{1}{k+1}] &= \\
\frac{n!}{k!(n-k-1)!}[\frac{k+1+n-k}{(n-k)(k+1)}] &= \\
\frac{(n+1)!}{(k+1)!(n-k)!} &= \\
\binom{n+1}{k+1}
\end{aligned}
$$

7. (a) 8 students write a statistics exam. The exam is set up so that no students can possibly tie and a gold, silver and bronze medal are awarded to the first, second and third place winners respectively. How many ways can these medals be given out? [3]

$$8 \times 7 \times 6 = 336$$

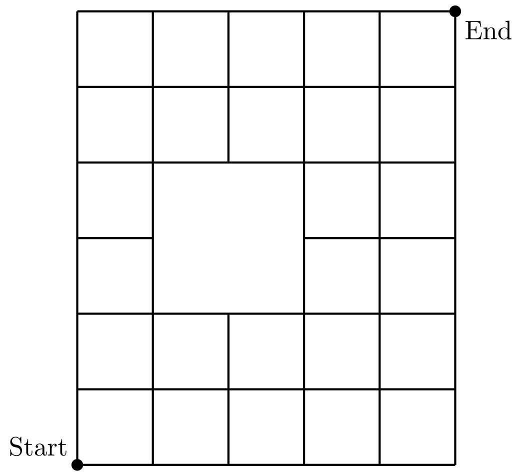
(b) The next year another 8 students take the course and once again ties are impossible. However the budget for medals grows so that now one gold, two silver and three bronzes are awarded. How many ways can the medals be distributed? [4]

$$
\begin{aligned}
8 \times \binom{7}{2} \times \binom{5}{3} &= \\
8 \times 21 \times 10 &= \\
1680
\end{aligned}
$$

(c) The following year we have yet another 8 students. The medals are now inscribed with a placing (first for the gold, second and third for the silvers and fourth, fifth and sixth for the three bronzes). How many ways can these medals be distributed?

$$8P6 = 8 \times 7 \times .... \times 3 = 20160$$

8. How many ways can one go from the start to the end in the picture below if you are only allowed to take right and upwards steps? [10]



Two solutions: The first is to draw in the lines missing from the central square and count $\binom{11}{5}$ paths and then subtract the $\binom{5}{2}\binom{6}{3}$ path that go through the draw in vertex. This gives a solutions of

$$\begin{aligned}\binom{11}{5} - \binom{5}{2} \times \binom{6}{3} &= \\ 462 - 10 \times 20 &= \\ 262\end{aligned}$$

However this solution requires seeing a trick. The higher work but more straight forward solution is to compute recursively. Naturally we get the same answer.