QSCI 381 Homework #3 - Vinsensius

Number 1

In this question we will look at scenarios using permutations and combinations. In each case show your working before presenting the final answer You are working as an insect taxonomist in a laboratory. You have 11 specimens whose label tags have unfortunately been mixed up. Each label tag contains information on species, order (Hymenoptera, Coleoptera, Diptera, Lepidoptera) and year collected. The tags are:

1. Dendroctonus pseudotsugae – Coleoptera - 1998
2. Dendroctonus pseudotsugae – Coleoptera - 1997
3. Choristoneura fumiferana – Lepidoptera - 2001
4. Scolytus ventralis – Coleoptera - 2011
5. Medetera aldrichii – Diptera - 2002
6. Medetera aldrichii – Diptera - 1999
7. Coeloides vancouverensis – Hymenoptera - 2015
8. Temnoschiela chlorodia – Coleoptera - 2001
9. Choristoneura biennis – Lepidoptera - 1996
10. Glypta fumiferana – Hymenoptera – 2007
11. Glypta fumiferana – Hymenoptera - 2005

As you have been newly hired, you have no knowledge of what tag belongs to which insect, and so you assign the tags to the insects at random.

1a

How many different ways are there of assigning the tags to the specimens? (2 pts)

Answer:

There are 11! ways of assigning the tags to the specimen. The answer would be $11! = 39916800$.

1b

What is the probability that the specimens are all labeled with the correct tag? (2 pts)
Answer:

There is only 1 way to label the specimen correctly and there are 11! to label the specimens, i.e. the sample space is 11!.

Thus

$$\text{P(label tag correctly)} = \frac{1}{39916800}$$

1c

What is the probability that the first three specimens are labeled correctly? (3 pts)
Answer:

$$\begin{aligned}\text{P(first three specimens with correct label)} &= \frac{1}{11} * \frac{1}{10} * \frac{1}{9} \\ &= \frac{1}{990}\end{aligned}$$

1d

By considering the number of permutations of orders (Coleoptera, Diptera, Hymenoptera, Lepidoptera), what is the probability that every specimen is assigned a label showing the correct order? (4 pts)

Answer:

There are 4 Coleoptera, 2 Diptera , 3 Hymenoptera and 2 Lepidoptera orders. Thus, the number of permutation based on orders is

$$\text{Number of ways to rearrange the specimen based on order} = \frac{11!}{4!3!2!2!} = 69300$$

There is only 1 correct way to arrange the specimens based on orders

Thus,

$$\text{P(correct way to arrange specimen based on order)} = \frac{1}{69300}$$

1e

Six of the specimens are contained within a large display cabinet. Of the 11 labels, you pick 6 at random and place them in the information section next to the cabinet, which

doesn't label the individual specimens, but just lists the specimens contained within the cabinet. What is the probability that you correctly pick the tags belonging to the specimens contained within the cabinet? (4 pts)

Answer:

$$\text{Number of ways of getting 6 tags of 11 tags} = \frac{11!}{5!6!} = 462$$

There is only 1 way of getting the order of tags correctly. As a result the probability is

$$\text{P(Labelling 6 specimen with right tag)} = \frac{1}{462}$$

1f

What is the probability that your selection of six labels contains all the Hymenoptera labels, and any other three labels? (4 pts) **Answer:**

There are total 3 Hymenoptera label and 8 labels of other orders. So, the number of ways of getting 3 Hymenoptera label out of 3 total Hymenoptera label is 1 way. Number of ways of getting the 3 other label out of 8 label total is $\frac{8!}{3!5!}$, which is 56 ways. Thus total of ways of getting 3 Hymenoptera labels and 3 other label is $56 * 1$ or 56 ways considering they are independent events.

The sample space would be the number of ways choosing 6 labels out of 11 tags randomly, which is $\frac{11!}{5!6!}$ or 462 ways.

Thus, the probability is

$$\text{P(3 Hymenoptera label and 3 other label)} = \frac{56}{462} = \frac{4}{33}$$

Number 2

In a health study, 25 participants are asked to record the number of days over the previous three weeks (21 days) in which they exercised for more than one hour. The following data shows the number of days reported by the 25 participants Read the following data into R:

exdays<-c(14,9,17,11,17,12,13,12,7,5,18,6,14,12,17,13,16,20,1,15,7,2,4,21,8)

2a

Given the description of this observational study, what distribution would you use to describe these data and why? Describe the parameters of this distribution and how they relate to this experiment. (4 pts)

Answer:

The distribution of the data that i would use is binomial distribution. The number of trial is 21 days and each trial there are only 2 independent events, whether the person has exercised more than an hour or not. The parameter n is 21 days and p is the probability of a person exercise for a single day, which can be calculated based on the mean of the data.

2b

What is the mean and variance of the collected data? Round your answers to 2 decimal places (2 pts)

Answer: The mean of the data is 11.64 days and the variance is 30.57 days.

2c

Using the mean you calculated in 2b, what is the mean probability of exercising on any particular day? Round your answer to 3 decimal places (2 pts)

Answer: The mean of probability is $\frac{11.64}{21}$, which is 0.554.

2d

Using your answer in 2c, what would be the expected number of exercise days per week, and per month (30 days)? Round your answer to 2 decimal places (4 pts)

Answer: The expected number of exercise per week is 3.88(0.554*7) days. The expected number of exercise per month is 16.62(0.551*30) days.

2e

Given the assumed distribution, use your answer in 2c to determine the probability that an individual will have 5 days of exercise in a 14 day period. Round your answer to 3 decimal places (4 pts)

Answer: Using R *dbinom* function with parameter n = 14, x=5 and prob = 0.554(from (c)).

$$\text{P(exercise 5 days in 14 days period)} = 0.0730$$

2f

What is the probability an individual will exercise less than 11 days in a 14-day period? Round your answer to 3 decimal places and show your work (4 pts)

Answer: The calculation is done using R *dbinom* function with x is [11,14], number of trials(n) is 14 and probability value based on part(c).

$$\begin{aligned}\text{P(exercise less than 11 days in 14 days period)} &= 1 - \text{P(exercise} \geq \text{11 days)} \\ &= 0.933\end{aligned}$$

Number 3

In this example we will use the hermit dataset

hermit< − read.csv(file = "hermit.csv", header = T)

This dataset shows experimental results exploring how the pH of water affects the presence of a shell in male and female hermit crabs. Each row is a result from a single hermit crab and contains information on whether the had a shell (yes, no), its gender (male, female), and the pH of the water.

3a

Calculate the separate probabilities of having a shell for hermit crabs in water > 7 pH (basic water) and for ≤ 7 pH (acidic water) irrespective of gender. Round your answers to 3 decimal places (4 pts)

Answer:

$$\begin{aligned} \text{P(shell} \mid \text{basic water)} &= \frac{36}{57} \\ \text{P(shell} \mid \text{acid water)} &= \frac{19}{56} \end{aligned}$$

3b

Using your answer from (3a), if you have two bodies of water each containing 75 hermit crabs, one which is acidic (≤ 7 pH), and another which is basic (> 7 pH), what would be the mean and standard deviation of the number of hermit crabs with shells for each body of water? Round your answer to 2 decimal places (4 pts).

Answer: We will assume the data to have binomial distribution because for each body of water we are looking at two event, the hermit crab that has shell and not having shell. We will be using mean and sd based on binomial distribution to calculate.

	Shell and basic	shell and acid
Mean	47.37	25.45
Standard Dev.	4.18	4.1

3c

Using your answer from (3a), if you had two bodies of water each containing 30 hermit crabs, one which was acidic (≤7 pH), and another which basic (> 7 pH), what would be the probability that at least 15 of the hermit crabs have shells in the acidic water, and

separately for the basic water. Round your answer to 3 decimal places and show your working (4 pts)

Answer: The calculation is done using R *dbinom* function with x is [15,30], number of trials(n) is 30 and probability value based on part(a)

$$P(x \geq 15 \mid \text{basic water }) = 0.952$$
$$P(x \geq 15 \mid \text{acid water }) = 0.050$$