

ISDS Individual Assignment 1 Solutions

The submission deadline for this assignment is **12pm Monday 23rd October**. Submission can be done through the module Ultra page. I **strongly recommend** submitting at least a few hours ahead of the deadline, in case of technical issues.

Each of the four questions carries similar but not identical weight.

The submission portal for this assignment will ask you to either choose an option, enter a value, enter text, or upload a document, depending on the sub-question being answered. I **strongly recommend** clicking on the submission portal as soon as is convenient, so you know what each sub-question requires you to submit.

Question 1.1:

Multiple choice questions. [\[1 mark\]](#) per correct answer

1. A short e-survey is released, asking the following question: “How many cups of tea do you drink a day?”. The available answers were “0”, “1”, “2”, and “3 or more”.

What type of data is the e-survey collecting?

- (a) Nominal
- (b) [Ordinal](#)
- (c) Discrete
- (d) Continuous

2. Data is collected relating to the continent of origin of students at Durham. The data is to be shown graphically to the University Council, to help them understand where students come from most often, and least often.

Council will want to compare numbers between continents - they are not interested in the proportion each continent of origin makes up of the whole of the student body. Which of the following graphs would be the best method for displaying this data?

- (a) [Bar chart](#)
- (b) Pie chart
- (c) Histogram
- (d) Stem and leaf diagram

3. In a recent survey, 46% of the British population said they preferred dogs to cats. In another recent survey, 12% of the British population listed jazz as one of their favourite musical genres.

Assume that whether a person prefers dogs to cats is independent of whether they consider jazz one of their favourite musical genres. What probability, expressed as a percentage, should we give to the event that a British person prefers dogs to cats **and** considers jazz one of their favourite musical genres.

- (a) 58%
- (b) 5.52%
- (c) 55.2%
- (d) 5.80%

The two important ideas here are i) the proportion of people in the population who have a given quality is equivalent to the probability that a **specific** person from that population has the given quality, and ii) we are allowed to assume these qualities are independent. We can therefore use the multiplicative rule - multiplying the probabilities of each separate quality together to get the probability someone has both qualities

4. Consider an outcome space $\Omega = \{1, 8, 15, 35, 69, 732, 983\}$. Let $A = \{1, 8, 69, 983\}$ and let $B = \{1, 8, 15, 35, 69\}$.

Assuming a uniform probability distribution on Ω , which of these is the probability that A AND B occur?

- (a) $\frac{3}{7}$
- (b) $\frac{4}{7}$
- (c) $\frac{5}{7}$
- (d) $\frac{6}{7}$

There are four outcomes in A , and five in B , but only three outcomes in both A **and** B . The probability of getting an outcome that is in A **and** in B is therefore $\frac{3}{7}$.

5. A random variable U is defined with the probability density function below:

$$f(u) = \begin{cases} \frac{u}{2} & 0 \leq u \leq 2 \\ 0 & \text{otherwise} \end{cases}.$$

Which of the following is the value of $P(U > 0.30)$?

- (a) 0.09775
- (b) 0.03

- (c) 0.70
- (d) 0.9775

We have

$$\begin{aligned}P(U > 0.30) &= \int_{0.3}^{\infty} f(u) du \\&= \int_{0.3}^2 \frac{u}{2} du\end{aligned}$$

using the fact that the integral of 0 is always 0, so we don't need to worry about the part of the integral where $f(u) = 0$.

Next,

$$\begin{aligned}\int_{0.3}^2 \frac{u}{2} du &= \left[\frac{u^2}{4} \right]_{0.3}^2 \\&= \frac{4}{4} - \frac{0.09}{4} \\&= 0.9775.\end{aligned}$$

Question 1.2:

Using the `ToothGrowth` data set in R:

1. Find the modal tooth length across the entire data set.
26.4 [1 mark]
2. Find the mean tooth length for guinea pigs who were given their vitamins via orange juice.
20.66 [1 mark]
3. Create a diagram in R of two box-and-whisker plots, one showing tooth length for guinea pigs given their vitamins via asorbic acid, and one showing tooth length for guinea pigs given their vitamins via orange juice. The box-and-whisker plots should be adjacent to each other, using the same axis, to enable easy comparison.

30

35

[2 marks]

4. Using the box-and-whisker plots created in Question 1.2.3, comment on which of the two vitamin delivery approaches is more effective in promoting tooth growth. Justify your answer.

OJ is both a more reliable approach to vitamin dissemination (smaller IQR) and with a higher median. This would suggest OJ is the superior option. On the other hand, the maximum length was achieved via VC.
[2 marks] for similar comments, or other valid comments.

Question 1.3:

A company that produces party poppers claims on their website that only 0.6% of their party poppers will fail to go off when the string is pulled.

1. The company sells party poppers in boxes of 200. What is the expected number of party poppers in a box which will fail to go off?

$$0.006 \times 200 = 1.2$$

[1 mark]

2. I decide I want to represent the number of party poppers which fail to go off in a box using a random variable, X .
 - (a) Which distribution would be the most appropriate to use for X ? Give both the name of the distribution, and the value of the parameters assuming the company's claim on their website is correct.

$$X \sim \text{Bin}(200, 0.006)$$

[1 mark]

- (b) What assumptions regarding the party poppers would need to hold in order to justify that choice of distribution?

Assumptions: a party popper goes off or it doesn't, each party popper goes off or doesn't independently, each party popper has same probability of not going off, the number of poppers in a box is indeed 200

[2 marks] for any three of these

3. I order a box for myself, but due to an administrative error, the box arrives containing not 200 party poppers, but 2.

- (a) Define the distribution for Y , the random variable representing the number of party poppers which fail to go off in my box of 2, assuming the claim on the company's website is correct.

$$Y \sim \text{Bin}(2, 0.006)$$

[1 mark]

- (b) Find $E(Y)$ and $\text{Var}(Y)$, under the assumptions needed for Question 1.3.2 b), and under the assumption that the claim on the company's website is correct. NOTE: You will need to scan and upload your working.

We have

$$\begin{aligned} E(Y) &= 0 \times P(Y = 0) + 1 \times P(Y = 1) + 2 \times P(Y = 2) \\ &= \frac{2!}{1!1!} \times 0.006 \times (1 - 0.006) + 2 \times \frac{2!}{2!0!} \times 0.006^2 \\ &= 0.0120 \\ \text{Var}(X) &= (0 - 0.012)^2 \times P(Y = 0) + (1 - 0.012)^2 \times P(Y = 1) \\ &\quad + (2 - 0.012)^2 \times P(Y = 2) \\ &= (-0.012)^2 \times \frac{2!}{0!2!} \times (1 - 0.006)^2 \\ &\quad + 0.988^2 \times \frac{2!}{1!1!} \times 0.006 \times (1 - 0.006) + 1.988^2 \times \frac{2!}{2!0!} \times 0.006^2 \\ &= 0.0119 \end{aligned}$$

[3 marks]

Question 1.4:

Let $X \sim \text{Pois}(0.4)$, and let $Y \sim \text{Pois}(\lambda)$. Find:

1. $P(X = 4)$.

We have

$$\begin{aligned} P(X = 4) &= \frac{e^{-0.4} 0.4^4}{4!} \\ &= 0.001 \end{aligned}$$

[1 mark]

2. $P(X < 3)$. NOTE: You will need to scan and upload your working.

We have

$$\begin{aligned}P(X < 3) &= P(X = 0) + P(X = 1) + P(X = 2) \\&= \frac{e^{-0.4}0.4^0}{0!} + \frac{e^{-0.4}0.4^1}{1!} + \frac{e^{-0.4}0.4^2}{2!} \\&= 0.992\end{aligned}$$

[2 marks]

3. An algebraic expression for $P(Y = 4|Y \geq 3)$ NOTE: You will need to scan and upload your working.

We have

$$\begin{aligned}P(Y = 4|Y \geq 3) &= \frac{P(Y = 4 \text{ AND } Y \geq 3)}{P(Y \geq 3)} \\&= \frac{P(Y = 4)}{P(Y \geq 3)} \\&= \frac{P(Y = 4)}{1 - P(Y < 3)} \\&= \frac{\frac{e^{-\lambda}\lambda^4}{4!}}{1 - \left(\frac{e^{-\lambda}\lambda^0}{0!} + \frac{e^{-\lambda}\lambda^1}{1!} + \frac{e^{-\lambda}\lambda^2}{2!}\right)} \\&= \frac{\lambda^4}{12(2e^\lambda - 2 - 2\lambda - \lambda^2)}\end{aligned}$$

[3 marks]