Name:____ **Date:**____________________

PSTAT 5A: Discussion Worksheet 01

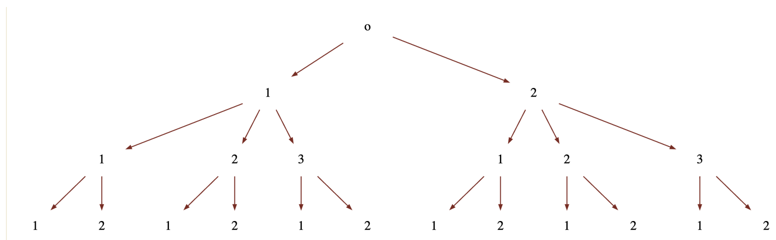
Summer Session A 2023, with *Ethan P. Marzban*

Welcome to the first PSTAT 5A Discussion Section! We encourage you to solve the following problems in groups. Statistics and Data Science are not meant to be lonely fields- we have quite a bit we can learn from each other!

1. A random number generator picks a number from the set $\{1, 2\}$ at random, then picks another number from the set $\{1, 2, 3\}$ at random, and finally picks a third number from the set $\{1, 2\}$ at random. The number selected at each stage is recorded.

a) Use a tree diagram to specify the outcome space of this experiment.

Solution:



b) Are we justified in using the classical approach to probability? Why or why not?

Solution: Yes, becuase the three numbers are selected *at random*.

c) Use the classical approach to probability to compute the probabilities of the following events (being sure to use proper notation!):

(i) E = "the first number selected is the number '1'"

Solution: The outcome space contains 12 elements. Of these 12 outcomes, there are only 6 in which the first number selected is 1:

$$E = \{(1, 1, 1),\ (1, 1, 2),\ (1, 2, 1),\ (1, 2, 2),\ (1, 3, 1),\ (1, 3, 2)\}$$

Hence, by the classical appraoch to probability,

$$\mathbb{P}(E) = \frac{6}{12} = \frac{1}{2} = 50\%$$

(ii) F = "the second number selected is the number '2'"

> **Solution:** Of the 12 outcomes in Ω, there are only 4 in which the second number selected is 1:
>
> $$E = \{(1, 2, 1),\ (1, 2, 2),\ (2, 2, 1),\ (2, 2, 2)\}$$
>
> Hence, by the classical appraoch to probability,
>
> $$\mathbb{P}(F) = \frac{4}{12} = \frac{1}{3} = 33.\overline{3}\%$$

(iii) G = "either the first number selected is the number '1' or the second number selected is the number '2' (or both)"

> **Solution:** Note that the event G is equivalent to $E \cup F$, where E and F are defined as in parts (i) and (ii) above. As such, we can apply the **addition rule** to compute
>
> $$\mathbb{P}(G) = \mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F)$$
>
> We computed $\mathbb{P}(E)$ and $\mathbb{P}(F)$ in the previous parts; as such, we only need to compute $\mathbb{P}(E \cap F)$. Note that the event $E \cap F$ is equivalent to "the first number selected was '1' and the second number selected was '2'"; there are only 2 outcomes in Ω that satisfy this:
>
> $$E \cap F = \{(1, 2, 1),\ (1, 2, 2)\}$$
>
> meaning, by the classical approach to probability,
>
> $$\mathbb{P}(E \cap F) = \frac{2}{12} = \frac{1}{6}$$
>
> and so
>
> $$\begin{aligned} \mathbb{P}(G) &= \mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F) \\ &= \frac{1}{2} + \frac{1}{3} - \frac{1}{6} = \frac{2}{3} = 66.\overline{6}\% \end{aligned}$$

d) Compute the probability that the sum of the last two numbers selected is strictly greater than the first number. **Hint:** Remember the complement rule!

Solution: Admittedly, this part requires a bit of thought. Following the hint, we let A denote the event "the sum of the last two numbers is strictly greater than the first number" and examine

$$A^{\complement} = \text{“the sum of the last two numbers is less than or equal to the first number”}$$

The smallest sum we can obtain from the last two numbers is 2 (i.e. when the last two numbers selected were both 1), which reveals to us that there is only one outcome in $A^{\complement}$:

$$A^{\complement} = \{(2, 1, 1)\}$$

Hence, by the classical approach to probability,

$$\mathbb{P}(A^{\complement}) = \frac{1}{12}$$

and so by the **complement rule** we have

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^{\complement}) = 1 - \frac{1}{12} = \frac{11}{12}$$

2. For each of the following variables, classify them as either discrete, continuous, ordinal, or nominal, and use this to determine what type of visualization is best to plot the distribution of observations collected on each.

a) x = the place/ranking (first, second, third, etc.) of athletes at the end of a marathon

Solution: This is an example of an ordinal variable. It is categorical since the addition of any two elements does not have any interpretive meaning (e.g. “first place” plus “second place” does not equal “third place”), and it is ordinal since there is a natural ordering to the possible values (“first place” is better than “second place”, which is better than “third place”, etc.)

b) y = the number of children in various families residing in the city of Santa Barbara

Solution: This is an example of a discrete variable. It is numerical since the addition of any two elements does have interpretive meaning (e.g. “one child” plus “two children” equals “three children”), and it is discrete since the set of possible values ($\{0, 1, 2, \cdots\}$) has jumps.

c) z = the species to which 100 different plants, selected from *Leadbetter* Beach, belong.

> **Solution:** This is an example of an nominal variable. It is categorical since the addition of any two elements does not have any interpretive meaning, and it is ordinal since there is no natural ordering to the possible values.

3. Consider a list of numbers $X = \{x_i\}_{i=1}^n$ and another list of numbers $Y = \{a \cdot x_i\}_{i=1}^n$, where a is a fixed constant. In other words, the elements of Y are found by taking the elements of X and multiplying by a. Show that $\overline{y} = a \cdot \overline{x}$.

Solution:

$$\begin{aligned}
\overline{y} &= \frac{1}{n}\sum_{i=1}^{n} y_i && \text{[Definition of mean]} \\
&= \frac{1}{n}\sum_{i=1}^{n}(a \cdot x_i) && \text{[Definition of } y_i\text{]} \\
&= \frac{1}{n}(a \cdot x_1 + a \cdot x_2 + \cdots + a \cdot x_n) && \text{[Definition of Sigma Notation]} \\
&= \frac{1}{n} \cdot a \cdot (x_1 + x_2 + \cdots + x_n) && \text{[Factorization]} \\
&= a \cdot \left[\frac{1}{n}(x_1 + x_2 + \cdots + x_n)\right] && \text{[Commutativity of multiplication]} \\
&= a \cdot \overline{x} && \text{[Definition of mean]}
\end{aligned}$$

