ESI 4606: Analytics I - Foundations of Data Science
Homework 2
Due: September 21st (11:00AM), 2022

Problem 1 (2 points)

Consider randomly selecting a student at USF, and let A be the event that the selected student has a Visa card and B be the analogous event for MasterCard. Suppose that $\Pr(A) = 0.6$ and $\Pr(B) = 0.4$

(a) Could it be the case that $\Pr(A \cap B) = 0.5$? Why or why not?

Solution:

No, this is not possible. Since event $A \cup B$ is contained within event B, it must be the case that $P(A \cap B) <= P(B)$. However, $.5 > .4$.

<u>We cannot assume event A and event B are independent here. So we can't get $P(A \cap B) = P(A)P(B)$</u>

(b) From now on, suppose that $\Pr(A \cap B) = 0.3$. What is the probability that the selected student has at least one of these two types of cards?

Solution:

By the addition rule: $P(A \cup B) = .6 + .4 - .3 = .7$.

(c) What is the probability that the selected student has neither type of card?

Solution:

$P(\textit{neither A nor B}) = P(A' \cap B') = P((A \cap B)') = 1 - P(A \cup B) = 1 - .7 = .3$.

(d) Calculate the probability that the selected student has exactly one of the two types of cards.

Solution:

From a Venn diagram, we see that the probability of interest is $P(\textit{exactly one}) = P(\textit{at least one}) - P(\textit{both}) = P(A \cup B) - P(A \cap B) = .7 - .3 = .4$.

Problem 2 (2 points)

A customer service center has 5 telephone lines. Let X denote the number of lines in use at a specified time. The probability mass function (pmf), $f(x)$, for X, is shown in Table 1.

Table 1: Table summary of the probability mass function

x	0	1	2	3	4	5
$f(x)$	0.10	0.15	0.20	0.25	0.20	0.10

(a) What is the corresponding cumulative distribution function (cdf), $F(x)$?
Solution:

$$F(0) = P(X \leq 0) = p(0) = .10$$
$$F(1) = P(X \leq 1) = p(0) + p(1) = .25$$
$$F(2) = P(X \leq 2) = p(0) + p(1) + p(2) = .45$$
$$F(3) = .70$$
$$F(4) = .90$$
$$F(5) = 1.00$$

The complete cdf of X is:

$$F(X) = \begin{cases} .00 & X < 0 \\ .10 & 0 \leq X < 1 \\ .25 & 1 \leq X < 2 \\ .45 & 2 \leq X < 3 \\ .70 & 3 \leq X < 4 \\ .90 & 4 \leq X < 5 \\ 1.00 & X \geq 5 \end{cases}$$

(b) Let event A={at most three lines are in use}, compute $\Pr(A)$.
Solution:
$P(A) = P(X \leq 3) = F(3) = .70$

(c) Let event B={fewer than two lines are in use}, compute $\Pr(B)$.
Solution:

$P(B) = P(X < 2) = P(X \leq 1) = F(1) = .25$

(d) Compute the mean and the variance of X, i.e., E(X) and Var(X).
Solution:

$$\begin{aligned}
E(X) &= \sum_{i=1}^{n} x_i f(x_i) \\
&= 0 \times .10 + 1 \times .15 + 2 \times .20 + 3 \times .25 + 4 \times .20 + 5 \times .10 \\
&= 2.60 \\
Var(X) &= \sum_{i=1}^{n} x_i^2 f(x_i) - [E(X)]^2 \\
&= 0^2 \times .10 + 1^2 \times .15 + 2^2 \times .20 + 3^2 \times .25 + 4^2 \times .20 + 5^2 \times .10 - 2.60^2 \\
&= 2.14
\end{aligned}$$

Problem 3 (1 point)

The lifetime of a certain brand of lightbulb has an exponential distribution with a mean of 800 hours. **Use R** to answer the following questions. [Hint: Using help() to learn R functions (e.g., dexp,pexp,qexp,rexp) related to the exponential distribution.]

(a) Find the probability that a randomly selected lightbulb of this kind lasts 700 to 900 hours.

```
# Function pexp(x, rate) in which rate=1/mean, computes the cdf F(X=x)
# part (a) is to compute P(700<=X<=900)=P(X<=900)-P(X<700)
# For continuous variable, it equals to F(900)-F(700)
> pexp(900, 1/800)-pexp(700, 1/800)
[1] 0.09220955
```

(b) Find the probability that a randomly selected lightbulb of this kind lasts longer than 850 hours.

```
# part (b) is to compute P(X>850)=1-P(X<=850)=1-F(850)
> 1-pexp(850, 1/800)
[1] 0.3455908
```

(c) Find the 80th percentile of the lifetime of this kind of lightbulb.

```
# Function qexp(x, rate) in which rate=1/mean, computes the quantile
> qexp(0.8, 1/800)
[1] 1287.55
```

Note: **To get full points, include R codes in the appendix sections**