ISQA 8160 Exam I

Brian Detweiler

July 19th, 2016

Assigned Problems

1.) Laura Naples, Manager of Heritage Inn, periodically collects and tabulates information about a sample of the hotel's overnight guests. This information aids her in planning and scheduling decisions she must make. The table below lists data on ten randomly selected hotel registrants, collected as the registrants checked out. The data listed for each registrant are: number of people in the group; birth date of person registering; shuttle service used, yes or no; total telephone charges incurred; and reason for stay, business or personal.

Name of Registrant	People In Group	Birth Date (mm/dd/yy)	Shuttle Used	Telephone Charges	Reason For Stay
Adam Sandler	1	05/07/59	yes	$0.00	personal
Mica Pepper	4	11/23/48	no	12.46	business
Claude Shepler	2	04/30/73	no	1.20	business
Amy Rodriquez	2	12/16/71	no	2.90	business
Tony DiMarco	1	05/09/39	yes	0.00	personal
Amy Franklin	3	09/14/69	yes	4.65	business
Tammy Roberts	2	04/22/66	no	9.35	personal
Ed Blackstone	5	10/28/54	yes	2.10	personal
Mary Silverman	1	11/12/49	no	1.85	business
Todd Atherton	2	01/30/62	no	5.80	business

a. How many elements are there in the data set?

There are 10 elements in the data set.

b. How many variables are there in the data set?

There are 5 variables in the data set.

c. How many observations are there in the data set?
There are 10 observations in the data set.

d. What are the observations for the second element listed?
The second element contains the following observations:

People in Group: 4
Birth Date (mm/dd/yy): 11/23/48
Shuttle Used: no
Telephone Charges: 12.46
Reason for Stay: business

e. What is the total number of measurements in the data set?
There are $10 * 5 = 50$ measurements in the data set.

f. Which variables are quantitative?
The following variables are quantitative:
People in Group
Birth Date (mm/dd/yy)
Telephone Charges

g. Which variables are qualitative?
The following variables are qualitative:

Shuttle Used
Reason For Stay

h. What is the scale of measurement for each of the variables?

People in Group: Interval Scale (cannot have 0 people in a group), with $X \in \{1, 2, \ldots\}$
Birth Date (mm/dd/yy): Ordinal Scale, with $X \in$ {all dates}
Shuttle Used: Nominal with $X \in$ {yes, no}
Telephone Charges: Ratio, with $X \in \mathbb{R}$
Reason For Stay: Nominal with $X \in$ {business, personal}

i. Does the data set represent cross-sectional or times series data?

This data is cross-sectional.

j. Does the data set represent an experimental or an observational study?

This data set is observational.

■

2.) Missy Walters owns a mail-order business specializing in baby clothes. She is considering offering her customers a discount on shipping charges based on the dollar-amount of the mail order. Before Missy decides the discount policy, she needs a better understanding of the dollar-amount distribution of the mail orders she receives. Missy had an assistant randomly select 50 recent orders and record the value, to the nearest dollar, of each order as shown below.

136	281	226	123	178	445	231	389	196	175
211	162	212	241	182	290	434	167	246	338
194	242	368	258	323	196	183	209	198	212
277	348	173	409	264	237	490	222	472	248
231	154	166	214	311	141	159	362	189	260

We can turn the data into order statistics by sorting it. This will help us find various quantiles.

123	162	178	196	212	231	246	277	338	409
136	166	182	196	212	231	248	281	348	434
141	167	183	198	214	237	258	290	362	445
154	173	189	209	222	241	260	311	368	472
159	175	194	211	226	242	264	323	389	490

a. Determine the mean, median, and mode for this data set.

$$\begin{aligned} \overline{x} &= \frac{\sum_i^n x_i}{n} \\ &= 251.46 \\ Q_{0.50} &= 228.5 \\ x_{\text{mode}} &= 231 \end{aligned}$$

b. Determine the 80th percentile.

$$\begin{aligned} Q_{0.80} &= \text{Order Statistic}_{\lceil 0.80*50 \rceil} \\ &= \frac{\text{Order Statistic}_{40} + \text{Order Statistic}_{41}}{2} \\ &= \frac{323 + 338}{2} \\ &= 330.5 \end{aligned}$$

c. Determine the first quartile.

$$\begin{aligned} Q_{0.25} &= \text{Order Statistic}_{\lceil 0.25*50 \rceil} \\ &= \text{Order Statistic}_{\lceil 12.5 \rceil} \\ &= \text{Order Statistic}_{13} \\ &= 183 \end{aligned}$$

d. Determine the range and interquartile range.

$$\begin{aligned}
\text{Range} &= max(X) - min(X) \\
&= 490 - 123 \\
&= 367 \\
Q_{0.75} &= \text{Order Statistic}_{\lceil 0.75*50 \rceil} \\
&= \text{Order Statistic}_{\lceil 37.5 \rceil} \\
&= \text{Order Statistic}_{38} \\
&= 290 \\
\text{Interquartile Range} &= Q_{0.75} - Q_{0.25} \\
&= 290 - 183 \\
&= 107
\end{aligned}$$

e. Determine the sample variance, sample standard deviation, and coefficient of variation.

$$\begin{aligned}
S^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \\
&= \frac{1}{49}\sum_{i=1}^{4} 9(x_i - 251.46)^2 \\
&= 8388.049388 \\
S &= \sqrt{S^2} \\
&= \sqrt{8388.049388} \\
&= 91.58629476 \\
c_s &= \frac{s}{\overline{x}} \\
&= \frac{91.58629476}{251.46} \\
&= 0.3642181451
\end{aligned}$$

f. Determine the z-scores for the minimum and maximum values in the data set.

$$\begin{aligned}
z &= \frac{x - \overline{x}}{s} \\
z_{\text{min}} &= \frac{123 - 251.46}{91.58629476} \\
&= -1.402611606 \\
z_{\text{max}} &= \frac{490 - 251.46}{91.58629476} \\
&= 2.604538164
\end{aligned}$$

■

3.) Over the past 200 days:

Number of Houses Sold	Number of Days
0	60
1	80
2	40
3	16
4	4

a. How many sample points are there?
There are 5 sample points.

b. Assign probabilities to the sample points and show their values.

$$\begin{aligned} P(X=0) &= \frac{60}{200} = \frac{3}{10} = 0.3 \\ P(X=1) &= \frac{80}{200} = \frac{2}{5} = 0.4 \\ P(X=2) &= \frac{40}{200} = \frac{1}{5} = 0.2 \\ P(X=3) &= \frac{16}{200} = \frac{2}{25} = 0.08 \\ P(X=4) &= \frac{1}{50} = 0.02 \end{aligned}$$

c. What is the probability that the agency will not sell any houses in a given day?

$$P(X=0) = 0.3$$

d. What is the probability of selling at least 2 houses?

$$\begin{aligned} P(X \geq 2) &= P(X=2) + P(X=3) + P(X=4) + P(X>4) \\ &= 0.2 + 0.08 + 0.02 + 0 \\ &= 0.3 \end{aligned}$$

e. What is the probability of selling 1 or 2 houses?

$$\begin{aligned} P(X=1) + P(X=2) &= 0.4 + 0.2 \\ &= 0.6 \end{aligned}$$

f. What is the probability of selling less than 3 houses?

$$\begin{aligned} 1 - P(X \geq 3) &= 1 - P(X=3) + P(X=4) + P(X>4) \\ &= 1 - 0.08 + 0.02 + 0 \\ &= 0.9 \end{aligned}$$

■

4.) A very short quiz has one multiple-choice question with five possible choices (a, b, c, d, e) and one true or false question. Assume you are taking the quiz but do not have any idea what the correct answer is to either question, but you mark an answer anyway.

a. What is the probability that you have given the correct answer to both questions?

$$\begin{aligned} P(X = \text{correct}, Y = \text{correct}) &= P(X = \text{correct})P(Y = \text{correct}) \\ &= \frac{1}{5} \cdot \frac{1}{2} \\ &= 0.1 \end{aligned}$$

b. What is the probability that only one of the two answers is correct?

$$\begin{aligned} P(\text{Only X or Y correct}) &= P(X = \text{correct})P(Y = \text{incorrect}) + P(X = \text{incorrect})P(Y = \text{correct}) \\ &= \frac{1}{5} \cdot \frac{1}{2} + \frac{4}{5} \cdot \frac{1}{2} \\ &= 0.5 \end{aligned}$$

c. What is the probability that neither answer is correct?

$$\begin{aligned} P(X = \text{incorrect}, Y = \text{incorrect}) &= P(X = \text{incorrect})P(Y = \text{incorrect}) \\ &= \frac{4}{5} \cdot \frac{1}{2} \\ &= 0.4 \end{aligned}$$

d. What is the probability that only your answer to the multiple-choice question is correct?

$$\begin{aligned} P(X = \text{correct}, Y = \text{incorrect}) &= P(X = \text{correct})P(Y = \text{incorrect}) \\ &= \frac{1}{5} \cdot \frac{1}{2} \\ &= 0.1 \end{aligned}$$

e. What is the probability that you have only answered the true or false question correctly?

$$\begin{aligned} P(X = \text{incorrect}, Y = \text{correct}) &= P(X = \text{incorrect})P(Y = \text{correct}) \\ &= \frac{4}{5} \cdot \frac{1}{2} \\ &= 0.4 \end{aligned}$$

■

5.) A government agency has 6,000 employees. The employees were asked whether they preferred a four-day work week (10 hours per day), a five-day work week (8 hours per day), or flexible hours. You are given information on the employees' responses broken down by gender.

	Male	Female	Total
Four days	300	600	900
Five days	1,200	1,500	2,700
Flexible	300	2,100	2,400
Total	1,800	4,200	6,000

a. What is the probability that a randomly selected employee is a man and is in favor of a four-day work week?

$$\begin{aligned} P(X = \text{man}, Y = 4 \text{ day}) &= \frac{300}{6000} \\ &= \frac{1}{20} \end{aligned}$$

b. What is the probability that a randomly selected employee is female?

$$\begin{aligned} P(X = \text{female}) &= \frac{4200}{6000} \\ &= \frac{7}{10} \end{aligned}$$

c. A randomly selected employee turns out to be female. Compute the probability that she is in favor of flexible hours.

$$\begin{aligned} P(\text{flexible}|\text{female}) &= \frac{P(\text{flexible}, \text{female})}{P(\text{female})} \\ &= \frac{\frac{2100}{6000}}{\frac{4200}{6000}} \\ &= \frac{\frac{7}{20}}{\frac{7}{10}} \\ &= \frac{\frac{7}{20}}{\frac{7}{10}} \\ &= 0.5 \end{aligned}$$

d. What percentage of employees is in favor of a five-day work week?

$$\begin{aligned} P(\text{five days}) &= \frac{2700}{6000} \\ &= \frac{9}{20} \end{aligned}$$

e. Given that a person is in favor of flexible time, what is the probability that the person is female?

$$
\begin{aligned}
P(\text{female}|\text{flexible}) &= \frac{P(\text{female}, \text{flexible})}{P(\text{flexible})} \\
&= \frac{\frac{2100}{6000}}{\frac{2400}{6000}} \\
&= \frac{\frac{7}{20}}{\frac{2}{5}} \\
&= \frac{7}{8}
\end{aligned}
$$

f. What percentage of employees is male and in favor of a five-day work week?

$$
\begin{aligned}
P(\text{male}, \text{five days}) &= \frac{1200}{6000} \\
&= \frac{1}{5}
\end{aligned}
$$

■

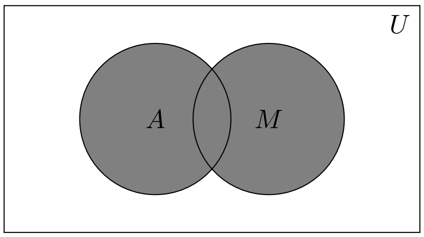
6.) Assume you have applied for two scholarships, a Merit scholarship (M) and an Athletic scholarship (A). The probability that you receive an Athletic scholarship is 0.18. The probability of receiving both scholarships is 0.11. The probability of getting at least one of the scholarships is 0.3.

a. What is the probability that you will receive a Merit scholarship?

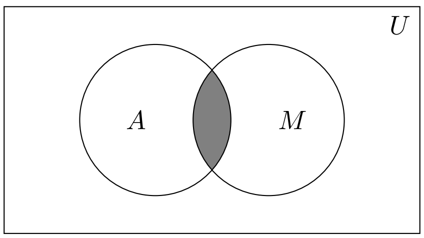
We are given:

$$\begin{aligned} P(A) &= 0.18 \\ P(A \cap M) &= 0.11 \\ P(A \cup M) &= 0.30 \end{aligned}$$

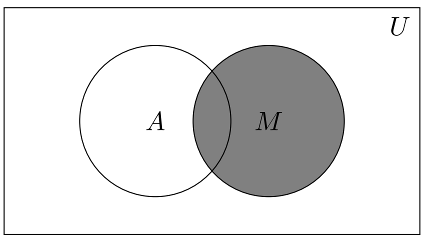
Using Venn Diagrams, we can see that $P(A \cup M) = 0.30$ is drawn as



And likewise, $P(A \cap M) = 0.11$ is drawn as



In order to find $P(M)$, we need to find $P(A \cup M) - P(A) + P(A \cap M)$.



This results in $P(M) = 0.30 - 0.18 + 0.11 = 0.23$.

b. Are events A and M mutually exclusive? Why or why not? Explain.
No. Mutually exclusive events cannot occur at the same time. If they were mutually exclusive, we would have

$$P(A \cap M) = 0$$

But instead we have

$$P(A \cap M) = 0.11 > 0$$

c. Are the two events, A and M, independent? Explain, using probabilities.
No. If they were independent, the following would be true:

$$\begin{aligned} P(A \cap M) &= P(A) \cdot P(M) \\ &= 0.18 \cdot 0.23 \\ &= 0.0414 \neq 0.11 \end{aligned}$$

d. What is the probability of receiving the Athletic scholarship given that you have been awarded the Merit scholarship?

$$\begin{aligned} P(A|M) &= \frac{P(A \cap M)}{P(M)} \\ &= \frac{0.11}{0.23} \\ &= 0.478260869565 \end{aligned}$$

e. What is the probability of receiving the Merit scholarship given that you have been awarded the Athletic scholarship?

$$\begin{aligned} P(M|A) &= \frac{P(M \cap A)}{P(A)} \\ &= \frac{0.11}{0.18} \\ &= 0.6\overline{11} \end{aligned}$$

■

7.) The probability distribution of the daily demand for a product is shown below.

Demand	Probability
0	0.05
1	0.10
2	0.15
3	0.35
4	0.20
5	0.10
6	0.05

a. What is the expected number of units demanded per day?

$$\begin{aligned}
E[X] &= \sum_{i=1}^{n} x_i \cdot f(x_i) \\
&= \sum_{i=1}^{7} x_i \cdot f(x_i) \\
&= 0(0.05) + 1(0.10) + 2(0.15) + 3(0.35) + 4(0.20) + 5(0.10) + 6(0.05) \\
&= 0 + 0.1 + 0.3 + 1.05 + 0.8 + 0.5 + 0.3 \\
&= 3.05 \text{ units per day}
\end{aligned}$$

b. Determine the variance and the standard deviation.

$$\begin{aligned}
Var(X) &= E[X^2] - \big(E[X]\big)^2 \\
&= \sum_{i=1}^{n} x_i^2 f(x_i) - \left(\sum_{i=1}^{n} x_i f(x_i)\right)^2 \\
&= \sum_{i=1}^{7} x_i^2 f(x_i) - 9.3025 \\
&= 0(0.05) + 1(0.10) + 4(0.15) + 9(0.35) + 16(0.20) + 25(0.10) + 36(0.05) - 9.3025 \\
&= 0 + 0.1 + 0.6 + 3.15 + 3.2 + 2.5 + 1.8 - 9.3025 \\
&= 11.35 - 9.3025 \\
&= 2.0475
\end{aligned}$$

The standard deviation of X is simply the square root of the variance, which is just 1.430908802.

■

8.) Twenty-five percent of all resumes received by a corporation for a management position are from females. Fifteen resumes will be received tomorrow.

a. Define the random variable in words for this experiment.

We can use the random variable X to describe the event of a proportion of resumes received by males and females in one day.

Let X be a binomial random variable, such that $p = 0.25$ is the probability that a resume received was submitted by a female, and $q = 1 - p = 0.75$ is the probabilty that a resume was submitted by a male.

We know that 15 resumes will be submitted tomorrow, so we can construct the PMF for this particular situation:

$$\begin{aligned} f(x) &= \binom{n}{x} p^x \cdot q^{n-x} \\ &= \binom{15}{x} (0.25)^x \cdot (0.75)^{15-x} \end{aligned}$$

It follows that the CDF is given by

$$\begin{aligned} F(x) &= \sum_{i=1}^{\lfloor x \rfloor} \binom{n}{x} p^x \cdot q^{n-x} \\ &= \sum_{i=1}^{\lfloor x \rfloor} \binom{15}{i} (0.25)^i \cdot (0.75)^{15-i} \end{aligned}$$

b. What is the probability that exactly 5 of the resumes will be from females?

Now we can refer back to our PMF.

$$\begin{aligned} P(X = 5) = f(5) &= \binom{15}{5} (0.25)^5 \cdot (0.75)^{15-5} \\ &= 3003 \cdot (0.0009765625) \cdot (0.0563135147095) \\ &= 0.165145981126 \end{aligned}$$

c. What is the probability that fewer than 3 of the resumes will be from females?

Here, we can use the CDF.

$$\begin{aligned} P(x \leq 2) = F(2) &= \sum_{i=1}^{\lfloor 2 \rfloor} \binom{15}{i} (0.25)^i \cdot (0.75)^{15-i} \\ &= \binom{15}{0} (0.25)^0 \cdot (0.75)^{15} + \binom{15}{1} (0.25)^1 \cdot (0.75)^{14} + \binom{15}{2} (0.25)^2 \cdot (0.75)^{13} \\ &= 0.0133634610102 + 0.0668173050507 + 0.155907045119 \\ &= 0.23608781118 \end{aligned}$$

d. What is the expected number of resumes from women?

$$
\begin{aligned}
E[X] &= \sum_{i=1}^{n} x_i f(x) \\
&= \sum_{i=1}^{15} x_i \cdot f(x_i) \\
&\cdots \text{ [see Excel for calculations]} \\
&= 3.75
\end{aligned}
$$

e. What is the variance of the number of resumes from women?

$$
\begin{aligned}
Var(X) &= E[X^2] - \big(E[X]\big)^2 \\
&= \sum_{i=1}^{n} x_i^2 f(x) - (3.75)^2 \\
&\cdots \text{ [see Excel for calculations]} \\
&= 2.8125
\end{aligned}
$$

■

9.) A retailer of electronic equipment received six HDTVs from the manufacturer. Three of the HDTVs were damaged in the shipment. The retailer sold two HDTVs to two customers.

a. Can a binomial formula be used for the solution of the above problem?

No. Our first clue is that we are told each of the values up front, rather than the probabilities. If we know the total number of objects N, and the number of sub-objects K, then we are likely to be asked a question about sampling these objects without replacement (the retailer is selling the TVs, thus not putting them back), which then becomes a job for the *hypergeometric* distribution.

b. What kind of probability distribution does the above satisfy, and is there a function for solving such problems?

This can indeed be modeled by the *hypergeometric* distribution, whose PDF is given as

$$f(k;n) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

Substituting in our values, we have

$$\begin{aligned} f(k;2) &= \frac{\binom{3}{k}\binom{6-3}{2-k}}{\binom{6}{2}} \\ &= \frac{\binom{3}{k}\binom{3}{2-k}}{15} \end{aligned}$$

c. What is the probability that both customers received damaged HDTVs?

$$\begin{aligned} f(2;2) &= \frac{\binom{3}{2}\binom{3}{2-2}}{15} \\ &= \frac{\binom{3}{2}\binom{3}{0}}{15} \\ &= \frac{3}{15} \\ &= \frac{1}{5} \end{aligned}$$

d. What is the probability that one of the two customers received a defective HDTV?

$$
\begin{aligned}
f(1;2) &= \frac{\binom{3}{1}\binom{3}{2-1}}{15} \\
&= \frac{\binom{3}{1}\binom{3}{1}}{15} \\
&= \frac{9}{15} \\
&= \frac{3}{5}
\end{aligned}
$$

■

10.) The average number of calls received by a switchboard in a 30-minute period is 15.

a. Define the random variable in words for this experiment.

The random variable X can be defined as the number of calls received in a 30-minute time period, such that $\lambda = 15$. This can be modeled as a *Poisson* distribution with the following PDF:

$$
\begin{aligned}
f(k;\lambda) &= \frac{\lambda^k \cdot e^{-\lambda}}{k!} \\
f(k;15) &= \frac{15^k \cdot e^{-15}}{k!}
\end{aligned}
$$

It follows that the CDF is given as:

$$
\begin{aligned}
F(k;\lambda) &= e^{-\lambda} \sum_{i=0}^{\lfloor k \rfloor} \frac{\lambda^i}{i!} \\
F(k;15) &= e^{-15} \sum_{i=0}^{\lfloor k \rfloor} \frac{15^i}{i!}
\end{aligned}
$$

b. What is the probability that between 10:00 and 10:30 the switchboard will receive exactly 10 calls?

$$
\begin{aligned}
P(X = 10) &= f(10;15) \\
&= \frac{15^{10} \cdot e^{-15}}{10!} \\
&= 0.0486107508298
\end{aligned}
$$

c. What is the probability that between 10:00 and 10:30 the switchboard will receive more than 9 calls but fewer than 15 calls?

$$
\begin{aligned}
P(9 < X < 15) &= F(14;15) - F(9;15) \\
&\cdots \text{ [see Excel for calculations]} \\
&= 0.3958000482
\end{aligned}
$$

d. What is the probability that between 10:00 and 11:00 the switchboard will receive fewer than 7 calls?

Note that this spans two time periods, so we must double our λ. Given that $2\lambda = 30$, we can expect our probability to be a pretty small number.

$$\begin{aligned}
P(X \leq 6|2\lambda = 30) &= e^{-30} \sum_{i=0}^{\lfloor 6 \rfloor} \frac{30^i}{i!} \\
&\cdots \text{ [see Excel for calculations]} \\
&= 0.00000011731942
\end{aligned}$$

■