Edhesive AP Statistics **Unit 1 – Solutions**

**Multiple Choice:** Choose the best answer choice for the following problems.

1. Martina’s class conducted an experiment using a large number of seeds, recording the time in days that it took each seed to sprout after planting. Some of the seeds were planted in an indoor planter that was kept in the classroom, and some were planted outside. These data were used to generate the boxplots shown below.



Which of the following statements is true?

* 1. The mean time to sprout for seeds that were planted inside is approximately 20 days.
  2. The first quartile of time to sprout for seeds that were planted outside is approximately 15 days.
  3. The range in time to sprout for seeds that were planted inside is approximately 40 days.
  4. Approximately half of the seeds planted inside took longer to sprout than any seed that was planted outside.
  5. The interquartile range of the time to sprout for seeds that were planted inside is approximately 10 days.

According to the boxplots, the maximum time to sprout for seeds planted outside was 20 days, and the median time to sprout for seeds planted inside was 20 days. Since approximately half of any dataset is larger than its median, approximately half of the seeds planted inside took longer than 20 days to sprout.

1. A call center used the median instead of the mean when it advertised the amount of time that its representatives spend on the average phone call. When the call center’s data on phone call time is displayed in a histogram, it can be seen that the distribution of call times is very right-skewed. Which of the following explains why the call center would report the median instead of the mean in this situation?
   1. The mean is much smaller than the median when the data is very right-skewed.
   2. The median is exactly halfway between the minimum call time and the maximum call time.
   3. The median is not affected by the skewness of data, while the mean can be strongly affected.
   4. The median is always the preferred statistic when the underlying data is a measure of time.
   5. The mean cannot be calculated for skewed data.

Right-skewed data has a distribution with a long tail on the right, and these values that are much larger than normal can inflate the mean of the dataset so that it does not very well represent the “average” data point. The median, in contrast, is not affected by such skewness or by outliers.

1. The music teachers at a high school surveyed students in the marching band, asking them how long they have been playing their instrument. The results are reported by the students’ instrument type in the table below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | <1 year | 1 to 3 years | 3 to 5 years | >5 years | Total |
| Brass | 4 | 15 | 10 | 2 | 31 |
| Percussion | 6 | 12 | 3 | 3 | 24 |
| Woodwind | 3 | 11 | 14 | 4 | 32 |
| Total | 13 | 38 | 27 | 9 | 87 |

The observed number of woodwind players who have played their instrument for less than 1 year was 3. What would be the expected number of woodwind players in this sample who have played their instrument for less than 1 year, if the length of time playing an instrument were independent from instrument type?



If A and B are independent, then *P*(A and B) = *P*(A) · *P*(B). So, under the assumption that length of time playing an instrument is independent from instrument type, one would calculate *P*([woodwind] and [<1 year]) = *P*([woodwind]) · *P*([<1 year]). From the collected data, *P*([woodwind]) ≈ and *P*([<1 year]) ≈ , so that the probability of a student both playing a woodwind instrument and having played it for less than one year is ≈ 0.055.

In the current sample, 87 students were surveyed, so the expected *number* of students with this characteristic = =

1. The Weekender Movie Theatre randomly sampled 25 of its movie showings, recording the number of ticket purchases for each showing. The histogram below displays the data collected by the movie theatre.



In which interval is the median number of ticket purchases located?

* 1. 20 tickets to less than 40 tickets
  2. 40 tickets to less than 60 tickets
  3. 60 tickets to less than 80 tickets
  4. 80 tickets to less than 100 tickets
  5. 100 tickets to less than 120 tickets

There were 10 movie showings that had ticket sales of less than 40, and 12 showings that had ticket sales of 60 or above. The remaining 3 showings had ticket sales in the 40 to 60 interval, and the median is the largest of these three.

1. A local charity called 100 of their previous supporters at home in December to solicit end-of-year donations. A manager at the charity was curious if there was a relationship between whether the supporters answered the phone and whether they made an end-of-year donation. The data collected by the manager is displayed in the table below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Made an End-of-Year Donation | |  |
|  |  | Yes | No | Total |
| Answered  Phone Call | Yes | 30 | 10 | 40 |
| No | 25 | 35 | 60 |
|  | Total | 55 | 45 | 100 |

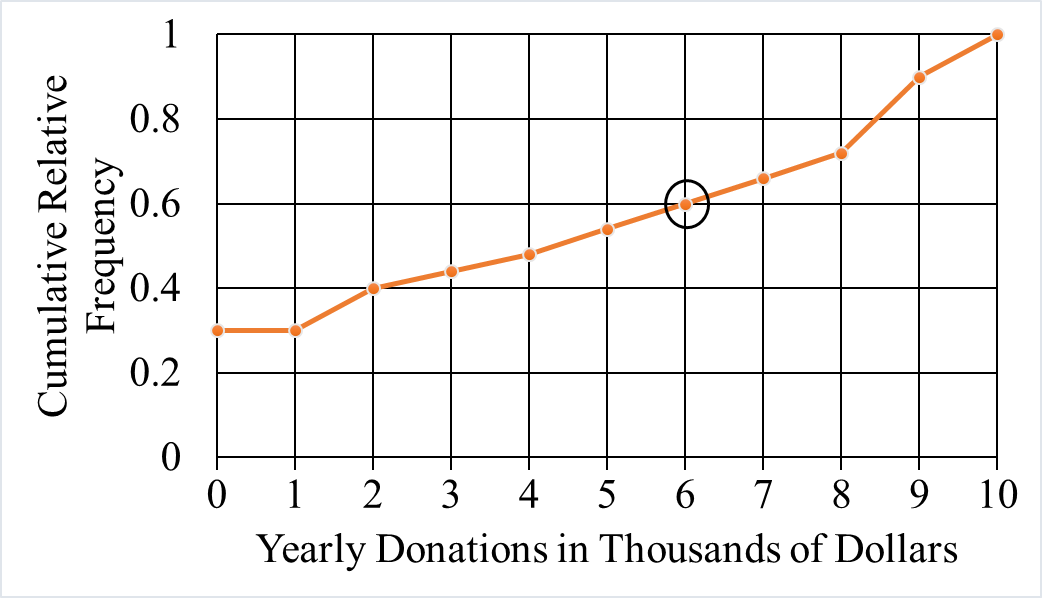
Which of the following statements is true for these 100 supporters?

* 1. They were more likely not to answer the phone call than to answer it.
  2. They were more likely not to make a donation than to make one.
  3. There were more supporters who answered the phone call and made a donation than supporters who did not answer the phone call and did not make a donation.
  4. Answering the phone call but not making a donation was more likely than not answering the phone call but making a donation.
  5. Not answering the phone call but making a donation was more likely than not answering the phone call and not making a donation.

According to the table, there were a total of 40 supporters who answered the phone call and 60 supporters who did not answer the phone call. Thus, not answering the phone call was the more likely outcome.

**Free Response – Solutions**

1. A large university keeps track of the yearly donations given by alumni. The figure below presents the cumulative relative frequency plot of the past year’s donations (in thousands of dollars) from the alumni.



* 1. In the context of the question, what information is conveyed by the circled point on the graph?

The circled point approximates that 60% of the alumni donated $6,000 or less.

* 1. What proportion of alumni donated between 2 and 6 thousand dollars during the past year?

The proportion of alumni who donated $6000 or less is 0.6 and $2000 or less is 0.4. The approximate proportion of alumni who donated between 2 and 6 thousand dollars is 0.6 – 0.4 = 0.2, or 20%.

* 1. Between donations of 0 and 1 thousand dollars, the cumulative frequency plot is flat. Interpret this information in the context of the question.

There are no alumni who donated between $0 and $1000.

* 1. A selection of artisan cheeses will be sent by the university to the alumni who submitted donations in the top 60 percent of donations. What is the minimum amount an alumnus would have to donate to receive the selection of artisan cheeses from the University?

The top 60% of donations would occur at the 40th percentile of the distribution, which represents $2000. Any alumni who donates at least $2000 will be in the top 60%.