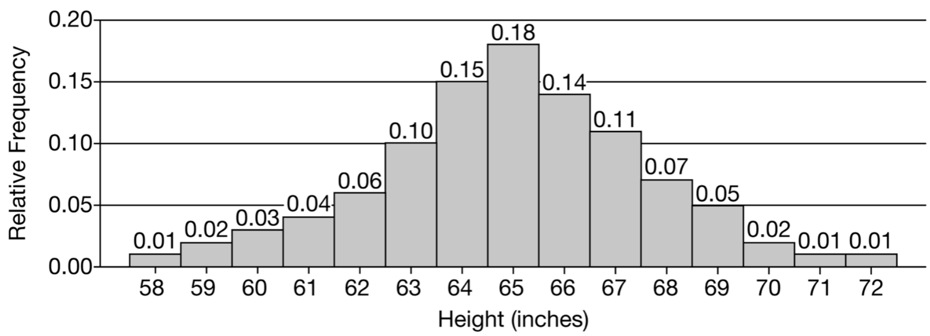
Deep Thoughts

September 9, 2021

1. The following histogram shows the relative frequencies of the heights, recorded to the nearest inch, of a population of women. The mean of the population is 64.97 inches, and the standard deviation is 2.66 inches.



(a) Based on the histogram, what is the probability that the selected woman will have a height of at least 67 inches? Show your work.

> **Solution:** The probability is equal to the sum of the relative frequencies of the heights from 67 to 72.
>
> $$0.11 + 0.07 + 0.05 + 0.02 + 0.01 + 0.01 = 0.27$$
>
> Essentially correct (E) if the response correctly sums the relative frequencies of the values from 67 to 72 and shows the work.
>
> Partially correct (P) if the response provides the correct probability with no work;

(b) What is the area of the bar corresponding to a height of 67 inches in the graph, and what does the area represent in terms of probability?

> **Solution:**
>
> **Essentially correct (E)** if the response satisfies the following two components: Indicates the bar is a rectangle with base 1, height 0.11, and area 0.11. Indicates the area of 0.11 is equal to the probability that a randomly selected woman has a height of 67 inches.
>
> Partially correct (P) if the response satisfies only one of the two components.
>
> Incorrect (I) if the response does not satisfy the criteria for E or P.

(c) The histogram displays a discrete probability model for height. However, height is often considered a continuous variable that follows a normal model. Consider a normal model that uses the mean and standard deviation of the population of women as its parameters.

i. Use the normal model and the relationship between area and relative frequency to find the probability that the randomly selected woman will have a height of at least 67 inches. Show your work.

Solution: For a normal model with mean 64.97 inches and standard deviation 2.66 inches, the z-score for 67 inches is $z = \frac{67-64.97}{2.66} \approx 0.76$. Using area as probability, the area under the standard normal curve to the right of 0.76 is approximately 0.2236.

ii. Does your answer in part (c)i match your answer in question 1? If not, give a reason for why the answers might be different.

Solution: The probability of 0.2236 is a little less than the answer in part (a) of 0.27 because there is less area in the tail of the normal curve than in the bars of the histogram.
Essentially correct (E) if the response satisfies the following four components.

- Finds the correct z Alt text: z-score
- Finds the correct area
- Recognizes the two probabilities are slightly different
- Provides a reasonable explanation of why they might be different.

(d) Let the random variable H represent the height of a woman in the population. $P(H < 60)$ represents the probability of randomly selecting a woman with height less than 60 inches. Based on the information given, the probability can be found using either the discrete model or the normal model.

i Give an example of a probability of H that can be found using the discrete model but not the normal model. Explain why.

Solution: The probability that the selected woman has a height equal to 63 inches, $P(H = 63)$, is an example of a probability that can be found with the discrete model as the area (or height) of the bar that corresponds to 63 inches. However, there is no area above a single number on the normal model.

ii Give an example of a probability of H that can be found using the normal model but not the discrete model. Explain why.

Solution: The probability that the selected woman has a height greater than 63.5 inches, $P(H > 63.5)$, is an example of a probability that can be found with the normal model by standardizing 63.5. However, it is not clear on the discrete model what the relative frequency of fractional heights might be.

2. A small coffee shop sells freshly squeezed juices in a refrigerated unit with slots where juice is displayed. These slots are called facings. The manager of the coffee shop suspects that the distribution of juice sales is different than the distribution of facings for each type of juice, so the manager records the sales of each juice over a two-week period. The proportion of facings and the sales for each type of juice are shown in the tables.

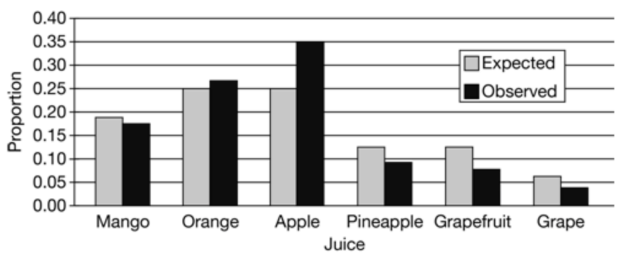
Juice	Mango	Orange	Apple	Pineapple	Grapefruit	Grape
Proportion of Facings	0.1875	0.250	0.250	0.125	0.125	0.0625

Juice	Mango	Orange	Apple	Pineapple	Grapefruit	Grape
Observed Number of Sales	23	35	46	12	10	5

(a) Construct a single bar chart that contains both the expected proportion of sales based on the proportion of facings and the observed proportion of sales for each type of juice.

Solution:

Part (a):



Scoring

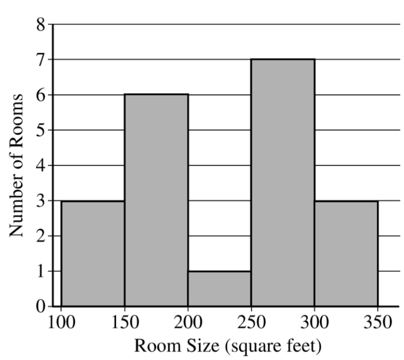
Part (a) is scored as follows:

Essentially correct (E) if the response correctly constructs AND labels a bar chart that contains both the expected and observed proportions

Partially correct (P) if the response fails to accurately construct a bar chart or uses frequencies instead of proportions in the plot

Incorrect (I) if the response does not satisfy the criteria for E or P

3. The sizes, in square feet, of the 20 rooms in a student residence hall at a certain university are summarized in the following histogram.



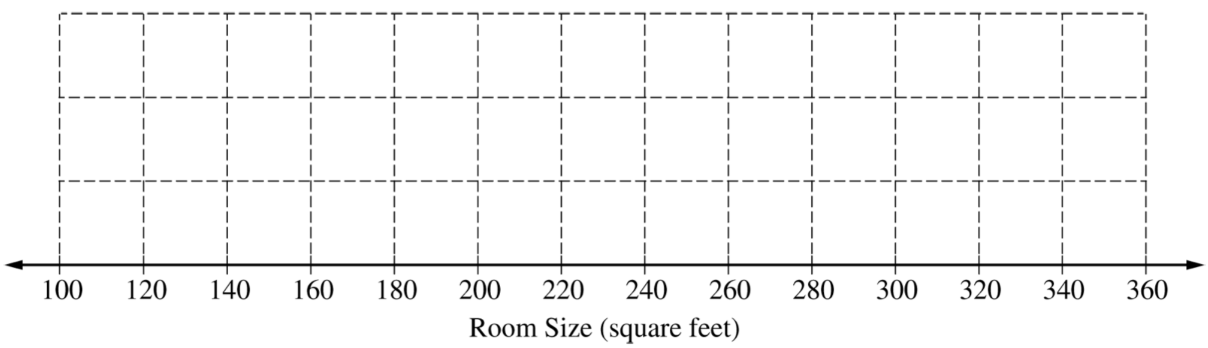
(a) Based on the histogram, write a few sentences describing the distribution of room size in the residence hall.

Solution: The distribution of the sample of room sizes is bimodal and roughly symmetric with most room sizes falling into two clusters: 100 to 200 square feet and 250 to 350 square feet. The center of the distribution is between 200 and 300 square feet. The range of the distribution is between 150 and 250 square feet. There are no apparent outliers.

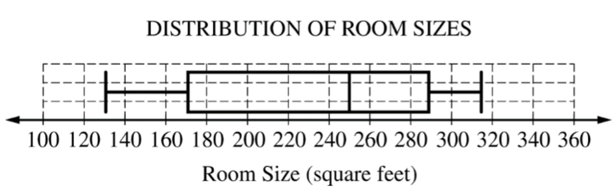
(b) Summary statistics for the sizes are given in the following table.

Mean	Standard Deviation	Min	Q1	Median	Q3	Max
231.4	68.12	134	174	253.5	292	315

Determine whether there are potential outliers in the data. Then use the following grid to sketch a boxplot of room size.



Solution: The interquartile range is $IQR = 292 - 174 = 118$ square feet. There are no potential outliers because the minimum room size of 134 square feet does not fall below $Q_1 - 1.5(IQR) = -3$ square feet, and the maximum room size of 315 square feet does not exceed $Q_3 + 1.5(IQR) = 469$ square feet.



(c) What characteristic of the shape of the distribution of room size is apparent from the histogram but not from the boxplot?

Solution: The histogram clearly shows the bimodal nature of the distribution of room sizes, but this is not apparent in the boxplot.