

Case 1

Estuardo is interested in how people in California give ratings to the taste of ranch pizza. He collected 20 people in Los Angeles and made a frequency table on the frequency of pizza ratings (1 through 5).

Ratings	Frequency (Freq.)	Cumulative Freq.	Relative Freq.	Cumulative Relative Freq.
1	3	3	$\frac{3}{3+2+5+6+4} = \frac{3}{20} = \frac{15}{100} = 0.15 = 15\%$	15% (=0.15)
2	2	3 + 2 = 5	$\frac{2}{3+2+5+6+4} = \frac{2}{20} = \frac{10}{100} = 0.1 = 10\%$	15% + 10% = 25% (=0.25)
3	5	3 + 2 + 5 = 10	$\frac{5}{3+2+5+6+4} = \frac{5}{20} = \frac{25}{100} = 0.25 = 25\%$	15% + 10% + 25% = 50% (=0.5)
4	6	3 + 2 + 5 + 6 = 16	$\frac{6}{3+2+5+6+4} = \frac{6}{20} = \frac{30}{100} = 0.3 = 30\%$	15% + 10% + 25% + 30% = 80% (=0.8)
5	4	3 + 2 + 5 + 6 + 4 = 20	$\frac{4}{3+2+5+6+4} = \frac{4}{20} = \frac{20}{100} = 0.2 = 20\%$	15% + 10% + 25% + 30% + 20% = 100% (=1)

- Can you fill in the blanks of cumulative freq., relative freq., and cumulative relative freq.?

Done.

- What is the cumulative relative freq. for giving 2 stars or lower?

15% + 10% = 25%. 25% is equivalent to 0.25.

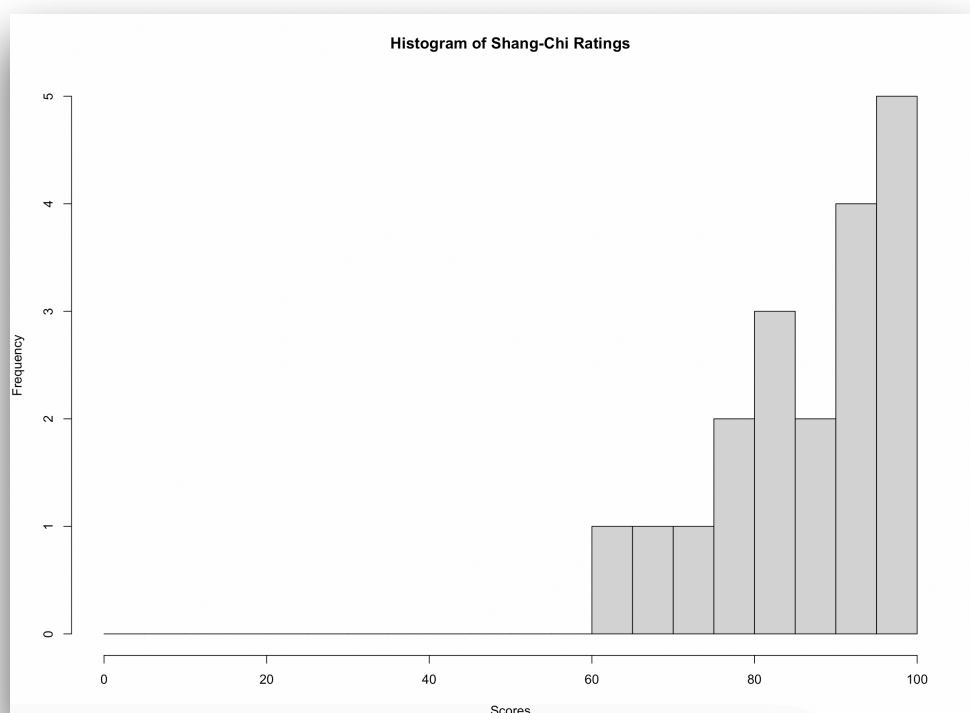
Case 2

Ari has been working as a data scientist at Regal. Since Shang-Chi of the Legend of the Ten Rings is on the screen, she wants to know how people rate the movie. She made a questionnaire on which score you want to give to the movie. The score ranges from 0 to 100 on a 5-point scale. So far, she has collected responses from 19 people, which is summarized in the following table:

Scores	65	70	75	80	85	90	95	100
Frequency	1	1	1	2	3	2	4	5

- How does the distribution look like? Normally distributed, negatively skewed, or positively skewed?

The distribution is negatively skewed. See the following histogram.



- **What are the mean, the median, and the mode of the sample? Feel free to use a calculator, if necessary.**

(1) Sample mean =

$$\begin{aligned}\bar{X} &= \frac{\sum_{i=1}^{19} X_i}{19} = \frac{X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8 + X_9 + X_{10} + X_{11} + X_{12} + X_{13} + X_{14} + X_{15} + X_{16} + X_{17} + X_{18} + X_{19}}{19} \\ &= \frac{65 + 70 + 75 + 80 + 80 + 85 + 85 + 85 + 90 + 90 + 95 + 95 + 95 + 95 + 100 + 100 + 100 + 100 + 100}{19} \\ &= \frac{65 \times 1 + 70 \times 1 + 75 \times 1 + 80 \times 2 + 85 \times 3 + 90 \times 2 + 95 \times 4 + 100 \times 5}{19} = \frac{1685}{19} = 88.68\end{aligned}$$

Please note that I rounded off to the second decimal place.

(2) Sample median = $\frac{19 + 1}{2}$ th value, when values are arranged in an ascending order = 10th value = 90

(3) Sample mode = 100

- **Can you calculate the population mean?**

No, we cannot. This is because we do not have population data.

- **Imagine Ari receives one response from a person who is very disappointed with the movie. If that person gives a score of 0, how do the mean, the median, and the mode change?**

The score of 0 is an outlier (i.e., extreme value). Since the mean is the most vulnerable to the existence of an

outlier, the mean will change. Depending on the actual $\frac{n + 1}{2}$ th value in the sample, the median might change or not. The mode will not change because a score of 100 is still the most frequent. Let's calculate the mean, the median, and the mode.

(1) Sample mean =

$$\begin{aligned}\bar{X} &= \frac{\sum_{i=1}^{20} X_i}{20} = \frac{X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8 + X_9 + X_{10} + X_{11} + X_{12} + X_{13} + X_{14} + X_{15} + X_{16} + X_{17} + X_{18} + X_{19} + X_{20}}{20} \\ &= \frac{0 + 65 + 70 + 75 + 80 + 80 + 85 + 85 + 85 + 90 + 90 + 95 + 95 + 95 + 95 + 100 + 100 + 100 + 100 + 100}{20} \\ &= \frac{0 \times 1 + 65 \times 1 + 70 \times 1 + 75 \times 1 + 80 \times 2 + 85 \times 3 + 90 \times 2 + 95 \times 4 + 100 \times 5}{20} = \frac{1685}{20} = 84.25\end{aligned}$$

We can see that the mean has decreased by about 4.43.

(2) Sample median = $\frac{20 + 1}{2}$ th value, when values are arranged in an ascending order = 10.5th value = the mean between the 10th value and the 11th value = the 10th value is 90, and the 11th value is also 90 = the mean of these two values is $\frac{90 + 90}{2}$, which is 90. Thus, the sample median is 90. We can see that the

median value has not changed.

(3) Sample mode = 100