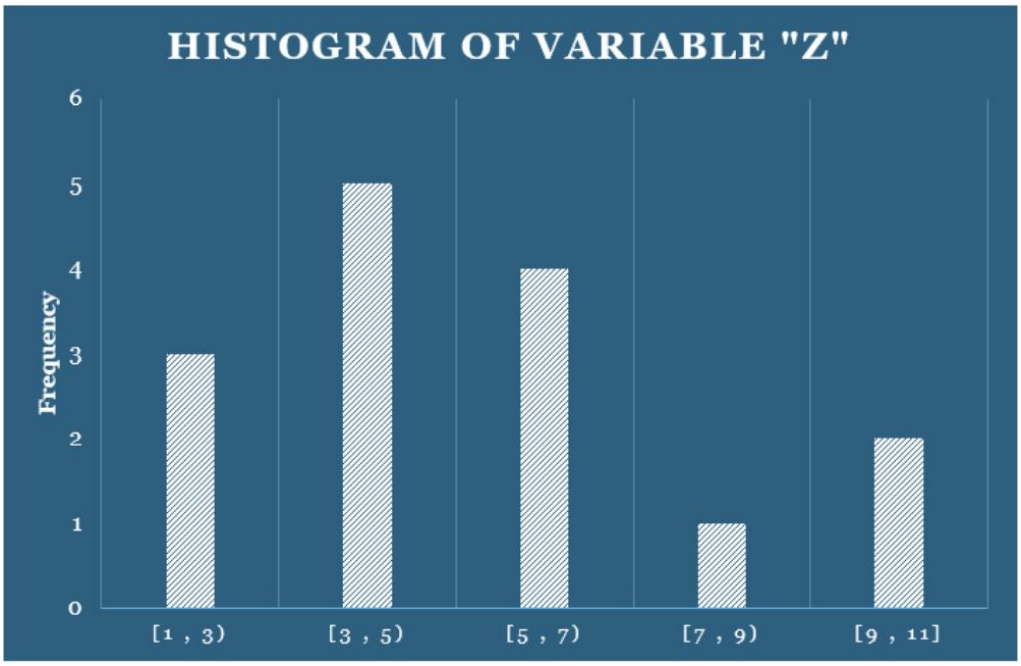
POL 201.30 – Introduction to Statistical Methods in Political Science
Online Quiz #2 – Solution Key

Total Quiz Score:
- **8 points.**

Notes:
- For "**Select All That Apply**" questions:
  - The "✅"symbol next to a bolded option indicates a *correct* selection.
  - The "☒" symbol next to an option indicates an *incorrect* selection.

Question 1 (6 points)



Pay close attention to the figure and all its elements, including the text below. Note that $[a, b]$ represents a closed interval, that is, both a and b are included. In contrast, $[a, b)$ is a half-open interval, that is, a is included but b is not.

From the following options, **select all that apply**:

(***Note**: Incorrect/false selections will discount partial credit from this question)*

Question 1. options:
- ✅ The data set that was used to generate the histogram corresponds to the sample: {2, 3, 1, 5, 4, 2, 6, 11, 3, 9, 6, 4, 8, 3, 6} (*+1 if selected, -1 if NOT selected*)
- ☒ The bins' width is 2.5 (*+1 if NOT selected, -1 if selected*)
- ✅ The range of Z is 10 (*+1 if selected, -1 if NOT selected*)
- ☒ The histogram shows a symmetric distribution (*+1 if NOT selected, -1 if selected*)
- ✅ Z's sample standard deviation is 2.8 (*rounded to the first decimal*) (*+1 if selected, -1 if NOT selected*)
- ☒ The data set that was used to generate the histogram corresponds to the sample: {1, 3, 6, 10, 4, 2, 6, 3, 9, 7, 8, 6, 3, 6, 1, 10, 11, 4} (*+1 if NOT selected, -1 if selected*)

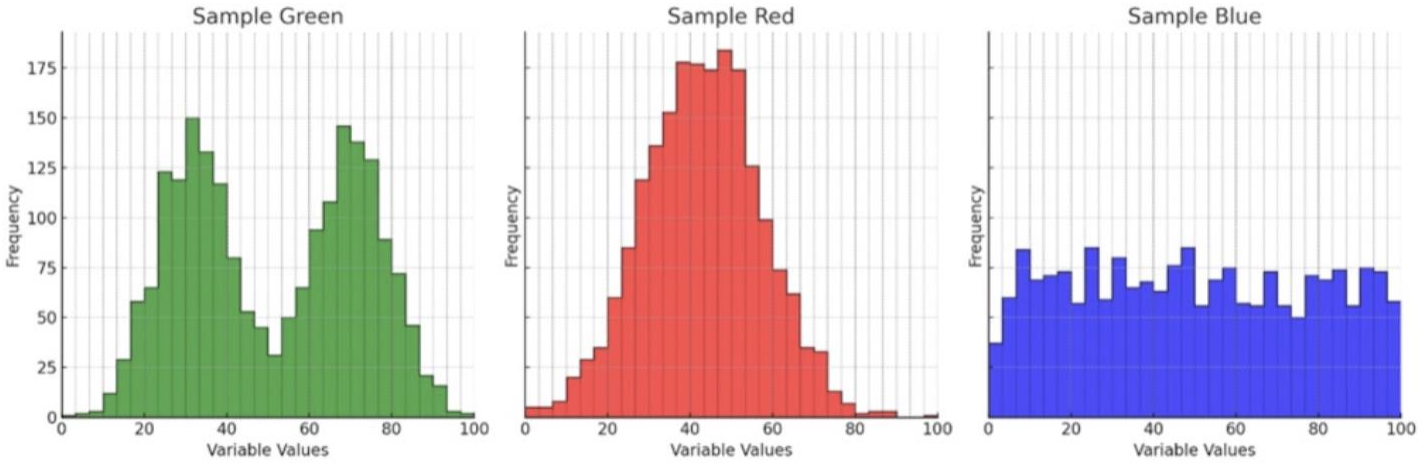
Justification:

- First option: **Correct**. The data set {2, 3, 1, 5, 4, 2, 6, 11, 3, 9, 6, 4, 8, 3, 6} can be used to generate the histogram above. The main component to understand is the difference between "[]" and "()". "[]" means that the number closest to the bracket is included within the bin and "()" means the number closest to the parenthesis is not included in the bin. For example, the first bin [1,3) would have a frequency of 3 because from the sample, {2,1,2} would be included. The next bins would be represented as follows: [3,5) would have a frequency of 5 and have {3,4,3,4,3}, [5,7) would have a frequency of 4 and have {5,6,6,6}, [7,9) would have a frequency of 1 and have {8}, and lastly, [9,11] would have a frequency of 2 and have {11,9}.
- Second option: **Incorrect.** The bins' width is not 2.5. A histogram with a bin width of 2.5 would have bins that look like [1, 3.5), [3.5, 6), and so on. The width of the bins in this histogram is 2, since we can compute it as: (*Upper Limit – Lower Limit*).
- Third option: **Correct.** The range of a histogram can be calculated by subtracting the smallest or minimum value from the largest or maximum value. In this histogram, the highest value that can be found is 11 and the lowest being 1. The range formula would be: *Range = Maximum value - Minimum Value*. So the range of this histogram would be 11 - 1 = 10. So, the range is 10.
- Fourth option: **Incorrect.** *A symmetric distribution is a distribution in which the shape on the left and right sides of the center are mirror images of each other, meaning that the frequencies change at the same rate and direction moving away from the center in both directions*. In contrast, a skewed distribution has an asymmetrical shape, where one tail extends further than the other. This histogram does not exhibit symmetry because the frequencies do not decrease evenly on both sides of the center. Instead, the values decrease only to the right while remaining relatively high on the left, indicating that the distribution is skewed to the right rather than symmetrical.
- Fifth option: **Correct.** To determine the sample standard deviation of Z, a few mathematical equations are necessary. Below is a google sheet displaying the proper way to find the sample standard deviation. The work below displays that the sample standard deviation is in fact 2.8.

		xi	xi - x̄	(xi - x̄)^2
Using the formula for Sample Standard Deviation:		2	-2.867	8.218
		3	-1.867	3.484
S = √(1/n-1)∑(xi - x̄)^2		1	-3.867	14.951
		5	0.133	0.018
And the data set we found from answer choice 1:		4	-0.867	0.751
		2	-2.867	8.218
{2,3,1,5,4,2,6,11,3,9,6,4,8,3,6}		6	1.133	1.284
		11	6.133	37.618
We can find the sample standard deviation		3	-1.867	3.484
		9	4.133	17.084
		6	1.133	1.284
	n = 15	4	-0.867	0.751
	n - 1 =	8	3.133	9.818
	15 - 1 = 14	3	-1.867	3.484
		6	1.133	1.284
	∑xi =	73	∑(xi - x̄)^2 =	111.733
	x̄ = (∑xi) / n	4.867	s^2= (111.733) / (n-1)	7.981
			s = √7.981	2.825

- Fifth option: **Incorrect**. The data set {1, 3, 6, 10, 4, 2, 6, 3, 9, 7, 8, 6, 3, 6, 1, 10, 11, 4} can't be used to generate the histogram above. The data set does correctly generate the correct frequency for [1, 3), [3, 5), and [5, 7), *it does not create the correct frequency for [7, 9) and [9, 11].* The frequency for [7, 9) is 2 and has {7,8} included. In the histogram above, the frequency for this bin is only 1, this data set has 2. In addition, [9, 11] has a frequency of 4 and has {10, 9, 10, 11} included. The frequency for this bin is 2 in the histogram, while this data set has a frequency of 4.

Question 2 (2 points)



Pay close attention to the figure and all its elements, including text. Consider the included histograms for a numerical variable collected from three different samples (green, red, and blue).

From the following options, **select all that apply**:

(***Note**: Incorrect/false selections will discount partial credit from this question)*

Question 2 options:

- ☑ The three distributions have roughly similar medians. (*+0.5 if selected, -0.5 if NOT selected)*
- ☒ The blue sample distribution is bimodal (*-0.5 if selected, +0.5 if NOT selected)*
- ☑ The red sample distribution shows less spread around its center than the blue distribution (*+0.5 if selected, -0.5 if NOT selected)*
- ☒ For the green sample distribution, the median is also the most likely observation. (*+0.5 if selected, -0.5 if NOT selected)*

Justification:

- First option: **Correct.** *The median is defined as the value where approximately 50% of the distribution lies to the left and 50% to the right*. The distributions of the green, red, and blue histograms all have a similar median, centered around a variable value of 50. The green sample has two peaks with a valley between them at the center or median of the histogram, which is around 50. The red sample has a single uniform peak, reaching its maximum height when the variable value is about 50, making the median 50. Lastly, the blue distribution has a uniform frequency throughout the variable values. Since it is uniform, the center or median is located directly between the minimum of 0 and the maximum of 100, making the median 50.
- Second option: **Incorrect.** The blue sample is not a bimodal distribution because a bimodal distribution has two distinct peaks. The blue sample does not have two distinct peaks; it is a uniform distribution where there are no peaks, and all relative frequencies are about the same.
- Third option: **Correct.** The red distribution shows less spread around its center than the blue distribution. The red distribution mainly ranges from about 15 to 80, while the blue distribution is uniform throughout, ranging from 0 to 100. The red sample is distributed much more tightly around its center than the blue sample.
- Fourth option: **Incorrect.** The mode is the measure of central tendency that represents the most likely observation, as it accumulates the highest frequency. In discrete distributions, it is the most frequent value, and in continuous distributions, it is the value within a closed interval that accumulates the highest relative likelihood. The median value for the green sample is about the same as in the other distributions, around 50. However, since the green sample is bimodal, its most likely observations have variable values of about 30 and 70, corresponding to the two highest peaks on the histogram. At the median of 50, the frequency is lower.