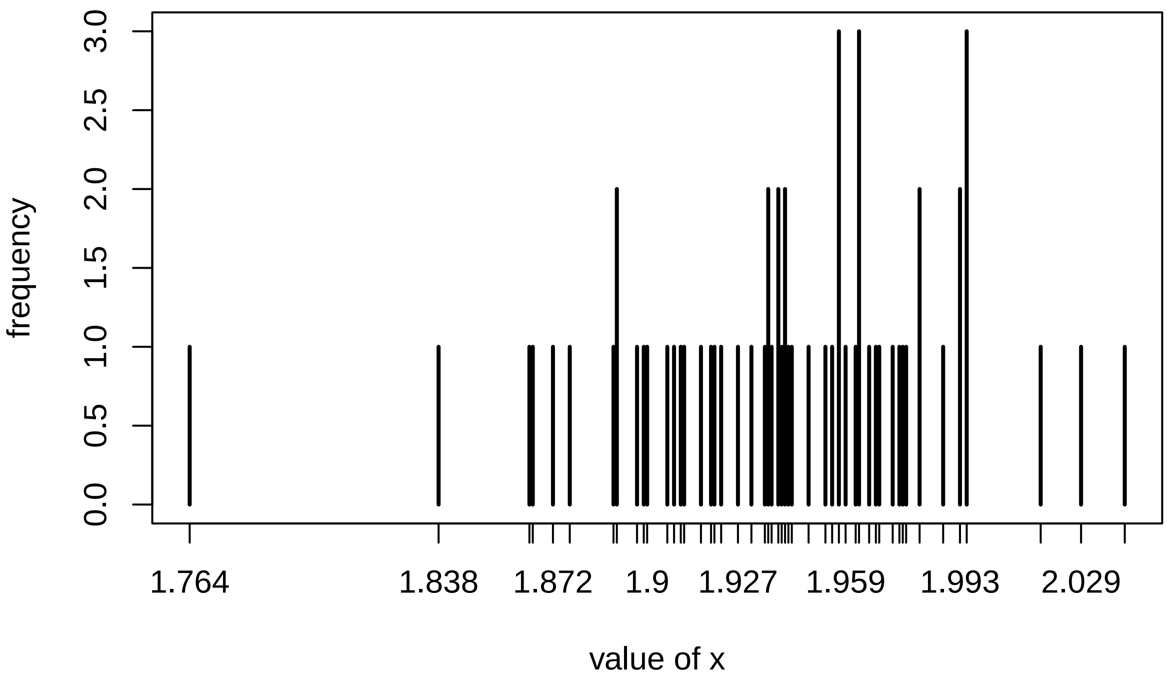
Suggested Answers for Distribution Module

The Normal Distribution: Copper/Sulfur Ratio in Copper Sulfide

Q1. A histogram shows the frequency for a range of possible outcomes, called a bin. The bin on the far-left, for example, shows that one of the 62 trials has a value for x in the range 1.760–1.775. Why do we need to bin the data to create a useful histogram?

Answer. For a continuous distribution, most of the individual measurements have unique values. A plot of the frequency of individual results, therefore, is likely to have many vertical lines with heights of one and a few vertical lines with heights of two or three. For the data in this exercise, 41 of the 62 values of x are unique, six appear twice, and three appear three times. Of the 279 possible values between the minimum value of 1.764 and the maximum value of 2.042, 229 (or 82%) have a frequency of zero. As the plot below suggests, the shape of the distribution is difficult to see.



Q2. The data in this experiment follows a normal distribution described by the equation

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

where $p(x)$ is the probability of a given outcome, μ is the mean (average), and σ is the standard deviation. The dashed line overlaid on the histogram shows the normal distribution for the mean and standard deviation set by the sliders. Use the sliders to explore how the mean and the standard deviation affect the normal distribution's position, width, and overall shape. Explain how your observations are consistent with the equation for a normal distribution.

Answer. Adjusting the slider for the mean shifts the position of the normal distribution along the x-axis without altering the distribution's height along the y-axis or its breadth along the x-axis. Selecting a smaller value for the standard deviation increases the distribution's height and narrows its breadth; selecting a larger value for the standard deviation, decreases the distribution's height and increases its width.

Q3. Adjust the mean and the standard deviation to provide the best fit of the normal distribution equation to the histogram. When you are satisfied with your fit, download the data and calculate the experimental mean and standard deviation, Compare these values to your estimated values from fitting the equation to the data.

Answer. Responses will vary, but a mean in the range 1.92 to 1.96 and a standard deviation in the range of 0.035 to 0.050 are reasonable approximations. The original data has a mean of 1.94 and a standard deviation of 0.047.

Q4. A normal distribution assumes all samples come from a population with a single mean and standard deviation. Suppose our 62 samples came from two populations, one where sulfur was used in excess and one where it was not. Describe the shape of the resulting histogram.

Answer. Because each population is characterized by its own mean and standard deviation, we expect the histogram to be bimodal, with two peaks each centered on its population's mean value.

The Poisson Distribution: Emission of Beta Particles from ^{40}K

Q1. Although a barplot and a histogram look similar, they are not interchangeable. A histogram is useful for continuous data and a barplot is useful for discrete data. Why are the results of beta emission considered discrete instead of continuous?

Answer. A continuous distribution is one in which all results are possible, limited only by the equipment used to make measurements. For a discrete distribution, the possible results are limited by the nature of the data. For the data in this experiment, the number of emitted particles is restricted to integers values, which makes the distribution discrete instead of continuous.

Q2. The data in this experiment follow a Poisson distribution in which the probability of a particular outcome, $p(x)$, in a fixed period of time is given by the equation

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

where λ is the outcome's average rate and where the exclamation point indicates the factorial. For the Cu_xS data we drew the theoretical distribution as a smooth line, but here we show only a set of points connected by line segments. Explain why we can display the normal distribution as a smooth line, but cannot do so with a Poisson distribution.

Answer. For a continuous distribution, every possible position along x-axis has a corresponding value along the y-axis; thus, the resulting plot consists of many points connected by short line segments that we see as a continuous smooth curve. For a discrete distribution, there are a limited number of possible values along the x-axis, which gives rise to a small number of individual points. We can choose to connect the points using line segments to aid in our visualization of the distribution's shape, however, the line segments provide no additional information about the distribution.

Q3. How does the shape of the Poisson distribution change as you adjust the value of λ from 0 to 13?

Answer. As the value of λ increases, the curve for the Poisson distribution shifts to the right and becomes more symmetrical in its shape, approaching the shape of a normal distribution when λ is sufficiently large. For this data, values of λ greater than 5.5 result in a distribution that appears normal.

Q4. Use the slider to determine the value of λ for this data. Use your value of λ to calculate the probability that a single beta particle is emitted, *p(x = 1)*. Explain how to convert this probability into a number of events and compare this result to the data in the figure.

Answer. Responses answers will vary, but a value for λ between 5 and 5.5 is reasonable. Using $\lambda = 5$ gives the probability as

$$p(x = 1) = \frac{5^1 \times e^{-5}}{1!} = 0.0337$$

With 365 separate trials, we expect a single beta particle to be emitted $365 \times 0.0337 = 12$ times. From the figure, the actual number is 10.

Q5. The mean for a Poisson distribution is equal to λ and its standard deviation is equal to $\sqrt{\lambda}$. Report the mean and standard deviation for this data using your value for λ and then download the data and check your result by calculating the actual mean and standard deviation.

Answer. Using $\lambda = 5$ gives an expected mean of 5 and an expected standard deviation of $\sqrt{5}$, or 2.24. The original data has a mean of 5.14 and a standard deviation of 2.29.

The Uniform Distribution: Certification of Class A 10-mL Pipettes

Q1. Is this data continuous or discrete? How do you know?

Answer. The distribution is continuous as all values within the limits of 9.98 mL and 10.02 mL is possible.

Q2. This distribution is described as uniform. What does it mean to be uniform? What will happen to the shape of this distribution and the counts in each bin if you increase or decrease the number of bins? Use the slider to check your prediction.

Answer. A uniform distribution is one in which all possible results are equally likely. The resulting histogram, therefore, should consist of bins of equal height. For 10 bins we expect each bin to contain 100 pipets; for this data set, individual bins contain between 93 and 105 pipets. If we decrease the number of bins, then we expect to see more pipets per bin and to see bins of nearly equal height; if we increase the number of bins, then we expect to see fewer pipets per bin and to see bins of nearly equal height.

Q3. For a uniform distribution, the mean is $\frac{a+b}{2}$ and the standard deviation is $\sqrt{\frac{1}{12}(a-b)^2}$ where a and b are, respectively, the distribution's largest and smallest values. Calculate the expected mean and standard deviation for Class A 10-mL pipets and then download the data and check your result by calculating the actual mean and standard deviation.

Answer. Using the limits of $a = 10.02$ mL and of $b = 9.98$ mL gives an expected mean of $\frac{10.02+9.98}{2}$, or 10.00 mL, and an expected standard deviation of $\sqrt{\frac{1}{12}(10.02-9,98)^2}$, or 0.0115 mL. The original data has a mean of 10.00(02) and a standard deviation of 0.0115(087), where the numbers in parentheses shows extra significant figures.

The Binomial Distribution: ^{13}C Atoms in Cholesterol

Q1. Is this data continuous or discrete? How do you know?

Answer. This is an example of discrete data because the number of atoms of ^{13}C in any molecule must be a positive integer.

Q2. The data in this experiment follow a binomial distribution in which the probability of a particular outcome, $p(x)$, is given by the equation

$$p(x) = \frac{n!}{x!(n-x)!} \times p^x \times (1-p)^{n-x}$$

where n is the number of carbons atoms in cholesterol (also called the size), x is the number of ^{13}C atoms in a molecule, and p is the probability that a randomly selected carbon atom is ^{13}C. Do some research to determine (a) the number of carbon atoms in cholesterol and (b) the probability that an atom of carbon is ^{13}C. Adjust the sliders to these values and comment on the fit of the equation to the data.

Answer. The formula for cholesterol is $C_{27}H_{46}O$, which makes $n = 27$. The isotopic abundance for ^{13}C is 1.1%, which makes $p = 0.011$. Adjusting the sliders to these values provides an excellent fit to the data.

Q3. The mean for a binomial distribution is equal to $n \times p$ and its standard deviation is equal to $\sqrt{n \times p \times (1-p)}$. Report the mean and the standard deviation for this data using your values for n and p, and then download the data and then check your result by calculating the actual mean and standard deviation.

Answer. The expected mean is 27×0.011, or 0.297 and the expected standard deviation is $\sqrt{27 \times 0.011 \times (1-0.011)}$, or 0.542. The original data has a mean of 0.32 and a standard deviation of 0.584.