Histogram with normal distribution

This laboratory discusses the display bases of display, probability, normal distribution, and z scores. The following packages are required for this laboratory: SFSMISC Auto Psych TidyVerse We remind you that histograms are used to view continuous data. The histograms are not used to view the categorical data. Instead, a bar chart is recommended for categorical data. The GEOM_HIST () function creates Histograms in R using GGPlot views. The following is an example of creating a variable stage histogram inside the DS data set. GgPlot (DS, AES (ETÃ)) + geom_histogram () ## `stat_bin ()` `using bins = from 30 '. Choose a better relationship with `binwidth`. The histogram displays the frequency of the EtÃ for certain boxes. Alternatively, the density of age can be displayed instead of frequency making a slight change in the view. Use the aesthetic mapping within the GEOM_HISTOGRAM () function and the X setting as VARIABLE ETÃ and Y as .. GgPlot (DS) + GEOM_HISTOGRAM (AES (X = ETÃ, Y = .. DensitÃ ..)) ## `stat_bin ()` `using bins = 30`. Choose a better relationship with `binwidth`. The shape of the plot is the same for frequency and density histograms; However, the Y axis measures different units. The area associated with the largest value of the Y axis suggest that a higher percentage of interviewees are susceptible to provide an age within the centuries on the axis of the Ascisses. Data are organized in intervals, known as bins, to compose the X axis. The number of bins is a potentially controversial topic; However, a good recommendation is to set the bin number of (SQRT N), where N is the number of observations. To change the number of containers, use bins = N inside the GEOM_HISTOGRAM () function. The square root of N for the current data set is a little more than 50, then set the bins to be 50. GGPlot (DS, AES (ETÃ)) + (Geom_Histogram bins = 50) Use of various functions with the Histogram function, the display is improved with more significant information. These functions can help: xlab (Ã ¢ x-axis labelÃ ¢) sets the label for the x axis x (y labelÃ ¢) sets the label for the y ggtitle axis (Ã ¢ titleÃ ¢) coord_cartesian (ylim = C (min: max), XLIM = C (min: max)) Sets the limits of the XE Y axis. The following is an excellent example of a cure data histogram. GgPlot (DS, AES (ETÃ)) + GEOM_HISTOGRAM (bins = 50) + XLAB ("Age") + YLAB ("frequency") + ggtitle ("Age histogram") + coord_cartesian (ylim = C (0175), XLIM = C (15.95)) + theme_light () # Sets the theme. There are a lot to choose from. The approximate data from normal distribution can define probability. Using R, normal distribution A CurveÃ ¢ bell can be projected on a histogram. Given an average and standard identified deviation and a density histogram, the Stat_Function () function can project a normal distribution as follows. Specify fun = Dnorm. ggplot (DS, AES (ETÃ)) + GEOM_HISTOGRAM (AES (X = ETÃ, Y = .. DensitÃ ..), Bins = 50) + STAT_FUNCTION (FUN = DNORM, ARGS = List (Media = Media (DS $ ETÃ) , SD = SD (DS $ ETÃ)), Color = "Red") Comparing the Histogram texture for the normal distribution curve generated can be difficult. The GEOM_DENSITY () function can draw a line using the density data for the ages alongside the projected line of what would seem normal distribution as the average data and deviation. The two forms can therefore visually compare to interpret whether the ages can be approximated by normal distribution. ggplot (DS, AES (ETÃ)) + GEOM_HISTOGRAM (AES (X = ETÃ, Y = .. Densità ..), Bins = 50) + STAT_FUNCTION (FUN = DNORM, ARGS = List (Media = Media $ ETÃ), SD = SD (DS $ ETÃ)), Color = "Red") + Geom_Density (Color = "Blue") The culmination of the histogram, curve, and the density line improved by adding limits and Labels to the X-axis, defining a plurality of containers, and the title of the chart. Including filling colors and outline for the histogram can also make more readable: ggplot (DS, AES (ETÃ)) + GEOM_HISTOGRAM (AES (X = ETÃ, Y = .. Densità ..), Binoni = 50, FILL = "# d3d3d3", color = "black") + stat_function (fun = dnorm, args = list (media = media (ds $ eth), sd = sd (ds $ etã)), eth)), + GEOM_DENSITY (Color = "Blue") + GGTITLE ("ETÃ" histogram) + XLAB ("AGE") + YLAB ("DENSITY") + theme_BW () + Coord_Cartesiano (XLIM = C (15.95), YLIM = C (0.0.04)) supports a series of distributions; However, for the purposes of these laboratories we will mainly focus on normal and binomial distributions. Displays the guide documentation (distributions) to explore the distributions supported by R. Guide (distributions) The following R functions are applicable to normal distribution: DNORM () PNorm () QNORM () RNORM () The DNORM () function provides the Height of a probability distribution function to a given X value: Dnorm (X, Mean = $ Mu $, Standard Deviation = $ SIGMA $). Note: The value returned by the DNORM () function is not the probability associated with the occurrence of the value X! The default average and the standard deviation for the DNORM () function is 0 and 1, respectively. The following example finds the height of the distribution function of the probability a (x = 2) with (mu = 4) and (sigma = 1.5). DNORM (2, Mean = 4, SD = 1.5) ## [1] 0.10934 The DNORM () function used in combination with the agent variable from the DS data set can find the height of the probability distribution function. In the following example, the DNORM () function will find the height of the probability distribution function for 65. Similar to previous examples, there is a topic to ignore the Na and missing values. DNORM (65, Mean = Media (DS $ AGE, NA.RM = T), SD = SD (DS $ AGE, NA.RM = T)) ## [1] 0.02662361 DNORM () Returns the height of the probability Distribution function as 0.027. Note: This is a random value and, alone, is not significant. The DNORM () function returns the relevant probability, which can lead to a probability; However, to understand this value also requires an explanation of the calculation. For continuous data, the probability of a single value is small (near zero), so instead the approach should be to find the probability that a value occurs within a specified interval. The probability associated with a value that occurs within a specified interval is the same as the area of the probability distribution function between the two points. In calculation this is defined how to find the integral of the probability distribution function. [INT _ {- x_1} {f_x (t)} {f_x (t)} {f_x (t), dx}] The above formula is the cumulative distribution function for two points, (x_1) and (x_2). In this case, (x_1) is defined as the lower limit and (x_2) is defined as the upper limit. R includes the integrated calculus function .xy () to return the odds. Integrate.xy (density (DS $ AGE) $ X, density (DS $ AGE) $ Y, 65.66) ## [1] 0.02917993 The probability associated with an age between 65 and 66 in the age variable is .029 (about ...) 3% of possibilities). Similarly, the PNORM () function calculates the odds associated with a given X value. The default function for the PNORM () function is the cumulative distribution function with a lower limit of (INFTY) and a limit Top of X. The following example calculates the probability associated with a value of age between (- infty) and 5, given (mu = 6) and (sigma = 6). PNORM (5, 6 = average, sd = 2) ## [1] The, 3085.375 thousand PNorm () calculates the probability of observing a value between INFTY and 5 as 0.31. The following example uses the PNorm () function with the DS data set to find the probability that a responder is 65 or less. PNORM (65, Mean = Media (DS $ AGRA, NA.RM = T), SD = SD (DS $ ETÃ €, na.rm = t)) ## [1] 0.6277983 to calculate the probability associated with an age 65 years old or larger, the minor.tail = false argument will look at the upper queue (right side of the probability distribution function). This is the same as Between 1 and the previously calculated lower queue probability. PNORM (65, Media = Media (DS $ ETÃ, NA.RM = T), SD = SD (DS $ ETÃ, NA.RM = T), Lower.tail = false) ## [1] IL, 3722,017 Mila QNorm () The function is the reverse function of the PNorm () function. A probability, media and standard deviation, the QNorm () function returns a value X from the probability distribution probability The following example finds the upper X value of the X of the probability distribution function associated with the probability or area under the curve, 0.3 data (MU = 5) and (sigma = 1) . ## [1] 4.475599 The following calculates the higher limit from the ages variable in the DS data set to demonstrate further associated with a probability of 40%. That is, the QNorm () function calculates the age that 40% of respondents is equal to or not. QNORM (.4, Mean = Media (DS $ ETÃ, NA.RM = T), SD = SD (DS $ ETÃ, NA.RM = T)) ## [1] 56.7677 Finally, the RNORM () function will generate values Random that follow a normal data distribution A number of points (N), provided (MU), and (sigma). The following example calculates 200 random values provided (MU = 6) and (sigma = 2). Random values are stored in the RVALUES object. RVALUES â €