Brian Kang

HW Set #1

|  |
| --- |
|  |
|  | # hw\_lect2\_1 |
|  |  |
|  | # if survial and Zelig (depend on survival) packages are not installed |
|  | # install.packages(c("survival", "Zelig")) |
|  | # library(survival) |
|  | library(Zelig) |
|  |  |
|  | # I will be using the PErisk function's barb2, prsexp2, prscorr2, gdpw2 |
|  | # variables for my 2 categorical/discrete and 2 continuous variables |
|  | # 2 categorical/discrete: barb2 and gdpw2 |
|  | # 2 continuous: prsexp2 and prscorr2 |
|  |  |
|  | # part a) |
|  | data("PErisk") |
|  | dat <- PErisk |
|  | # subset the function to extract the columns (or variables) I want |
|  | modifiedDat <- dat[,c("barb2", "prsexp2", "prscorr2", "gdpw2")] |
|  | # print the data in indicated format |
|  | print(modifiedDat) |
|  | barb2 prsexp2 prscorr2 gdpw2  Argentina -0.7207754 1 3 9.690170  Australia -6.9077550 5 4 10.304840  Austria -4.9103370 5 4 10.100940  Bangladesh 0.7759748 1 0 8.379768  Belgium -4.6173440 5 4 10.250120  Bolivia -2.4614400 0 0 8.583543  Botswana -1.2448680 4 3 8.777710  Brazil -0.4570337 4 3 9.375601  Burma 1.6043430 3 1 7.096721  Cameroon -4.2290650 3 1 8.120886  Canada -6.9077550 5 5 10.410180  Chile -1.5427610 3 2 9.261224  Colombia -2.0578210 3 2 9.191973  Congo-Kinshasa -2.3232880 1 0 7.095064  Costa Rica -5.0900030 3 4 9.167329  Cote d'Ivoire -4.2290650 4 2 8.228711  Denmark -6.9077550 5 5 10.106510  Dominican Republic -2.3788620 2 2 8.899731  Ecuador -1.8453370 3 2 9.117786  Finland -6.9077550 5 5 10.123670  Gambia, The -1.5433320 4 2 7.501082  Ghana -1.0115170 2 1 7.597396  Greece -2.0732370 3 3 9.701494  Hungary -0.9041942 4 3 9.351840  India -2.1051040 4 2 7.970049  Indonesia -2.1002320 3 0 8.392310  Iran 2.3374250 0 2 9.368114  Ireland -6.9077550 5 4 9.891465  Israel -2.3199960 4 4 10.067770  Italy -6.9077550 4 3 10.260780  Japan -6.9077550 5 4 9.892022  Kenya -2.3276050 2 2 7.619724  Korea, South -2.6557950 4 1 9.422787  Malawi -1.4694240 3 3 7.029973  Malaysia -3.9279490 4 3 9.178953  Mexico -1.6579350 2 2 9.661735  Morocco -3.1569580 3 1 8.780480  New Zealand -6.9077550 5 5 10.176260  Nigeria 0.3001068 1 1 7.687080  Norway -6.9077550 5 5 10.298330  Papua New Guinea -2.6361580 4 2 8.126518  Paraguay -0.9707628 3 0 8.727616  Philippines -2.9647760 1 1 8.384804  Poland 1.3170210 3 3 9.052400  Portugal -2.4596250 4 3 9.444543  Sierra Leone 1.4064060 3 1 7.759614  Singapore -4.8485160 5 5 9.882724  South Africa -2.1758200 3 4 9.191871  Spain -6.9077550 5 3 10.047330  Sri Lanka -1.8643430 2 2 8.627661  Sweden -6.9077550 4 5 10.224340  Switzerland -6.9077550 5 5 10.341100  Syria 1.7251660 1 1 9.664151  Thailand -6.9077550 3 2 8.548692  Togo -4.2290650 4 1 7.331715  Tunisia -2.5853990 2 2 9.047586  Turkey -2.6732430 3 2 8.978912  United Kingdom -6.9077550 5 5 10.127270  Uruguay -2.1277750 2 2 9.414342  Venezuela 0.4288450 3 2 9.848820  Zambia 0.9658105 3 1 7.726213  Zimbabwe -0.6403214 3 2 7.965893 |
|  | # part b) |
|  | # makes 2x2 grid to display the 4 histograms |
|  | par(mfrow = c(2,2)) |
|  | # plot histograms for each of the 4 variables |
|  | # continuous variables will be assigned bins ("breaks") |
|  | # discrete variables will be plotted |
|  | hist(dat[,3], breaks = seq(-7, 3, by = 1)) |
|  | hist(dat[,4]) |
|  | hist(dat[,5]) |
|  | hist(dat[,6], breaks = 15) |
|  |  |
|  |  |
|  | # ----------------------------------------------------------------------- |
|  |  |
|  | # hw\_lect2\_2 |
|  |  |
|  | # part a) |
|  | # a temp variable containing the each levels' percent |
|  | temp <- c(0.41, 0.27, 0.22, 0.02, 0.06, 0.02) |
|  | # contains the counts for each level of the rv |
|  | # the round() command was used to round the decimal values |
|  | counts <- round(temp\*51) |
|  |  |
|  | # part b) |
|  | # plot(c(5:0), counts, type="h") |
|  | # rep(c(5,4), c(21,14)) generates 21 5s and 14 4s |
|  | # generate and contains rv data using the counts |
|  | histDat = rep(c(5,4,3,2,1,0), counts) |
|  |  |
|  | # part C) |
|  | # to see the histogram with visible bars |
|  | # hist(histDat, breaks = 20) |
|  | # to see histogram that looks like the one from plot() |
|  | # (to make the boxes basically look like lines) |
|  | hist(histDat, breaks = 36000) |
|  |  |
|  | # ----------------------------------------------------------------------- |

**The graph above is NOT the one from plot()**

**Answer to #1.7a is at under of #1.16**

|  |  |
| --- | --- |
|  |  |
|  | # 1.16 |
|  | idtDat <- c(28.1,31.2,13.7,46.0,25.8,16.8,34.8,62.3, |
|  | 28.0,17.9,19.5,21.1,31.9,28.9,60.1,23.7, |
|  | 18.6,21.4,26.6,26.2,32.0,43.5,17.4,38.8, |
|  | 30.6,55.6,25.5,52.1,21.0,22.3,15.5,36.3, |
|  | 19.1,38.4,72.8,48.9,21.4,20.7,57.3,40.9) |
|  | # create histogram with class length of 10 |
|  | hist(idtDat, breaks = seq(10, 80, by = 10)) |
|  | # could have just done |
|  | # hist(idtDat) |
|  |  |
|  | # create histogram of data log\_10(data) with class length of 0.1 |
|  | idtModDat = log10(idtDat) |
|  | hist(idtModDat, breaks = seq(1.1, 2.0, by = 0.1)) |
|  | # did not do 1.9, just for extra clarity |
|  |  |
|  | # The logarithmic transformation yielded a set of numbers with |
|  | # "nicer" looking data. The new histogram is more bell shaped and |
|  | # the mode (1.4) is close to both median (1.448) and mean (1.47), |
|  | # although it seems to be possibly bimodal. On the other hand, |
|  | # the original data is right skewed (the "tail" is on the right end) |
|  | # and seemingly unimodal. |

#1.7a

\*(Numbers in parentheses are frequencies)

\***# bolded numbers at end of decimals repeat**

0: iiiiiii (7)\*, rel. freq. = 7/60 = 0.11**6\***

1: iiiiiiiiiiiii (12), rel. freq. = 12/60 = 0.2

2: iiiiiiiiiiiii (13), rel. freq. = 13/60 = 0.21**6**

3: iiiiiiiiiiiiii (14), rel. freq. = 14/60 = 0.2**3**

4: iiiiii (6), rel. freq. = 6/60 = 0.1

5: iii (3), rel. freq. = 3/60 = 0.05

6: iii (3), rel. freq. = 3/60 = 0.05

7. I (1), rel. freq. = 1/60 = 0.01**6**

8. I (1), rel. freq. = 1/60 = 0.01**6**

- sum of all frequencies = 60 and sum of all relative frequencies = 1