

EDS 6333 - Probability and Statistics

Name : Rahul Krishna Gunneri

PSID : 2210264

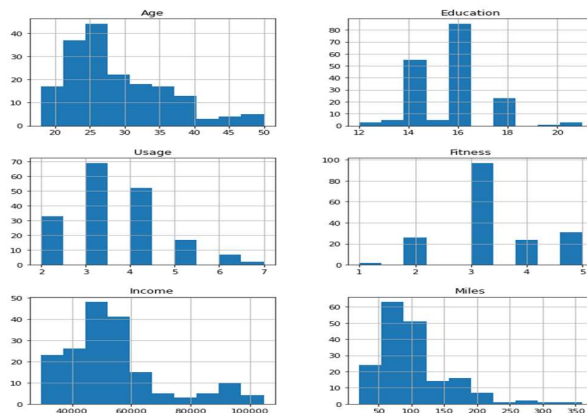
Introduction :

CardioGood aim is to determine variations in product lines by consumer profile to improve their sales. We will assess both categorical and numerical variables. Male or female, single or in a relationship, and self-rated level of fitness are categorical variables (1 to 5). Age (years), education (years), annual family income (\$), the average number of times the client plans to use the treadmill per week, and the average number of miles the customer expects to run/walk are variables that are categorised as numeric.

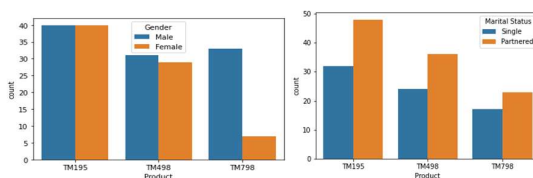
- Analysis of basic statistics about each feature, such as count, min, max, and mean

	Age	Education	Usage	Fitness	Income	Miles
count	180.000000	180.000000	180.000000	180.000000	180.000000	180.000000
mean	28.788889	15.572222	3.455556	3.311111	53719.577778	103.175556
std	6.943498	1.617055	1.084797	0.958869	16506.684226	51.864911
min	18.000000	12.000000	2.000000	1.000000	29562.000000	21.200000
25%	24.000000	14.000000	3.000000	3.000000	44058.750000	65.800000
50%	26.000000	16.000000	3.000000	3.000000	50596.500000	94.000000
75%	33.000000	16.000000	4.000000	4.000000	58668.000000	114.600000
max	50.000000	21.000000	7.000000	5.000000	104581.000000	360.000000

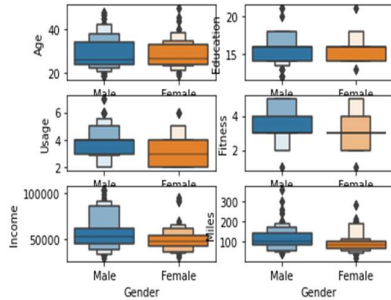
- Plotting the frequency tables for all the numerical characteristics of a customer.



- Frequency tables for categorical characteristics of a customer.



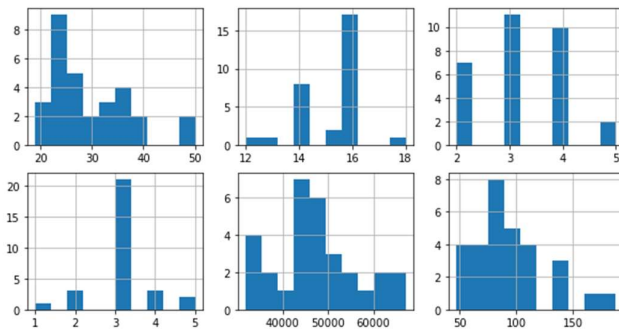
- We are determining outliers by using boxplots and from the below plots we can know that Fitness, Education and Miles have outliers.



- We are taking three distinct products TM195, TM498, and TM798 data into dataframes for product comparison.
- In order to determine the population mean and variance, we iterate the samples several times because we are aware that we cannot draw conclusions from sample results.
- We are iterating 75 times, randomly selected bootstrapped samples using 30 samples from these datasets.
- Population variance is similar to the sample variance when divided by $(n)^{1/2}$.

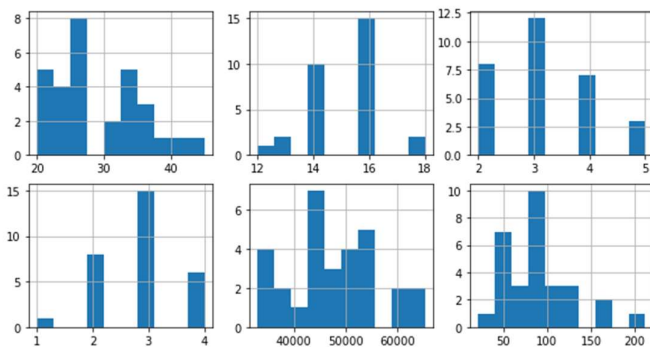
TM195:

- Frequency tables for Product TM195



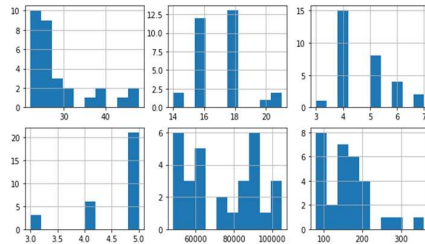
TM498:

- Frequency tables for Product TM498



TM798 :

- Frequency tables for Product TM798



Confidence Intervals :

Calculated confidence intervals for confidence levels 90%,92%,95%,98%,99% for products TM195,TM498 and TM798.

TM195	90%	92%	95%	98%	99%
Age	(26.28,30.84)	(26.13,31.0)	(25.82,31.31)	(25.26,31.86)	(24.86,32.26)
Education	(14.61,15.38)	(14.59,15.40)	(14.54,15.45)	(14.46,15.55)	(14.38,15.61)
Usage	(2.96,3.43)	(2.94,3.45)	(2.91,3.48)	(2.85,3.54)	(2.81,3.58)
Fitness	(2.70,3.02)	(2.69,3.03)	(2.67,3.05)	(2.63,3.09)	(2.61,3.12)
Income	(44041.37, 49799.02)	(43846.34, 49994.05)	(43454.97,50385.42)	(42748.80,51091.59)	(42250.06,51590.33)
Miles	(75.96,90.09)	(75.49,90.57)	(74.53,91.53)	(72.79,93.26)	(71.57,94.49)
TM498	90%	92%	95%	98%	99%
Age	(26.89,30.50)	(26.77,30.62)	(26.53,30.86)	(26.08,31.31)	(25.77,31.62)
Education	(14.87,15.59)	(14.84,15.61)	(14.79,15.66)	(14.70,15.75)	(14.64,15.81)
Usage	(2.99,3.40)	(2.97,3.42)	(2.95,3.44)	(2.90,3.49)	(2.86,3.53)
Fitness	(2.75,3.04)	(2.74,3.05)	(2.72,3.07)	(2.68,3.11)	(2.65,3.14)
Income	(46589.44,50889.35)	(46444.79,51035.0)	(46151.51,51327.28)	(45624.13,51854.66)	(45251.67,52227.12)
Miles	(81.75,95.62)	(81.28,96.09)	(80.33,97.03)	(78.63,98.73)	(77.43,99.93)
TM798	90%	92%	95%	98%	99%
Age	(26.95,31.44)	(26.80,31.59)	(26.50,31.89)	(25.95,32.44)	(25.56,32.83)
Education	(16.81,17.85)	(16.78,17.88)	(16.71,17.95)	(16.58,18.08)	(16.49,18.17)
Usage	(4.48,5.11)	(4.45,5.14)	(4.41,5.18)	(4.33,5.26)	(4.28,5.31)
Fitness	(4.42,4.48)	(4.41,4.85)	(4.38,4.88)	(4.33,4.93)	(4.29,4.96)
Income	(71670.15,82777.77)	(71293.90,83154.02)	(70538.88,83909.05)	(69176.54,85271.38)	(68214.37,86233.55)
Miles	(150.39,190.0)	(149.05,191.34)	(146.36,194.03)	(141.50,198.89)	(138.07,202.32)

Hypothesis Testing :

- For hypothesis testing we are considering two cases
case1-
H0: TM195-TM498>=0;
H1: TM195-TM498<0
case2-
H0: TM498-TM798>=0;
H1: TM498-TM798<0
Case3-
H0: TM195-TM798>=0;
H1: TM195-TM798<0
- For the 'Age' characteristic of Hypothesis Test 1, the null hypothesis was accepted in case 1 and rejected in case 2, however we are investigating case 3 because we cannot draw any conclusions from these two situations. Product TM195 is being used by younger individuals, whereas Product TM498 is being used by older people and Product TM798 is being used by middle-aged people.

- The second hypothesis test is for "Education." In both instances, the null hypothesis is rejected, indicating that individuals with lower levels of education use ProductTM195, those with higher levels use ProductTM798 and the remaining individuals use ProductTM498.
- Third Hypothesis This test is for "Usage." Because we cannot draw a conclusion from these two situations, we are contemplating case3, which was rejected. The null hypothesis is accepted for case1 and rejected for case2. People with little usage tend to favor the TM195 product, those with medium usage tend to favor the TM798 product, and those with high usage tend to favor the TM498 product.
- The fourth hypothesis test is for fitness. Individuals who are in bad shape are using ProductTM195, people who are moderately fit are using ProductTM498 and people who are in good shape are using TM798. The null hypothesis is rejected in both circumstances.
- The fifth hypothesis test, "Income," is for. People with lower incomes tend to favor TM195, whereas those with higher incomes tend to use TM798 and the remaining population tends to use TM498. The null hypothesis is rejected in both circumstances.
- The sixth hypothesis test is for "Miles." Customer who expects to walk more are using TM798, customer who expects to walk less are using TM195, and rest of the people are using TM498 Product, rejecting null hypothesis in both circumstances.

Conclusion :

Treadmill TM195 is heavily used by both genders equally, as are all other customer demographics, with the exception of younger customers who use it less, as do those who are less fit and have lower income. Treadmill TM498 is the least used, and TM798 is primarily used by men who are in excellent physical condition, have higher incomes, and plan to walk more miles. Overall, based on our observations, we may draw the conclusion that buyers of all types tend to favor the TM195 product.