Dalton Shaver

STAT 3010

September 19, 2021

**Data Analysis of Statistics Class**

**1)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name** | **Label** | **General Type** | **Specific Type** | **Measurement Units** |
| Gender | Male or Female | Categorical | Nominal | N/A |
| Tattoo | Does the student have a tattoo? (Yes or No) | Categorical | Nominal | N/A |
| CDs | Categorized number of the student's estimate of how many music CDs he/she owns (BelowAvg = 0 to 40, Average = 41 to 80, AboveAvg = more than 80 | Categorical | Ordinal | N/A |
| Height | Self-reported height in inches | Quantitative | Continuous | Inches |
| EarPrc | Number of ear piercings | Quantitative | Discrete | Ear Piercings |

Table 1: List of Variables in the Penn State Data

This dataset involves three categorical and two quantitative attributes for each sample of 205 students in a statistics class. The table above displays relevant information such as the label, specific variable type, and measurement units.

**2)**

Table 2: Descriptive Statistics for Height and Ear Piercings

|  |  |  |
| --- | --- | --- |
|  | **Height** | **EarPrc** |
| Mean | 67.03 | 2.76 |
| Median | 67.00 | 2.00 |
| Standard Deviation | 3.79 | 2.42 |
| Minimum | 60.00 | 0.00 |
| Q1 | 64.00 | 0.00 |
| Q3 | 70.00 | 4.00 |
| Maximum | 77.00 | 13.00 |
| Range | 17.00 | 13.00 |
| IQR | 6.00 | 4.00 |
| Count | 205.00 | 205.00 |

Table 2 displays the descriptive statistics of both quantitative variables, height and ear piercings. For height among students, the mean and median were extremely close. We can assume the distribution of this variable is symmetrical. The mean would be the best representation of central tendency in this situation. The standard deviation would better represent the distribution of height because the mean and median are nearly identical, therefore we can assume there are no extreme outliers. When examining the number of ear piercings, there is a sizeable difference between the mean and median calculations. The median is a better fit to represent the data as there are possible extreme outliers. Additionally, the IQR would be the best representation of dispersion because any outliers could cause the standard deviation to be significantly affected.

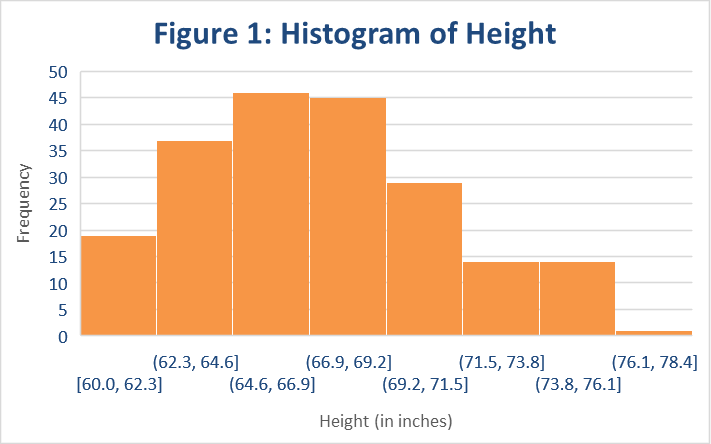


Figure 1 shows the distribution of student height is unimodal and symmetrical. The mean, 67.03 inches, is the best representation of central tendency. The standard deviation, 3.79 inches, is the appropriate measure of dispersion.

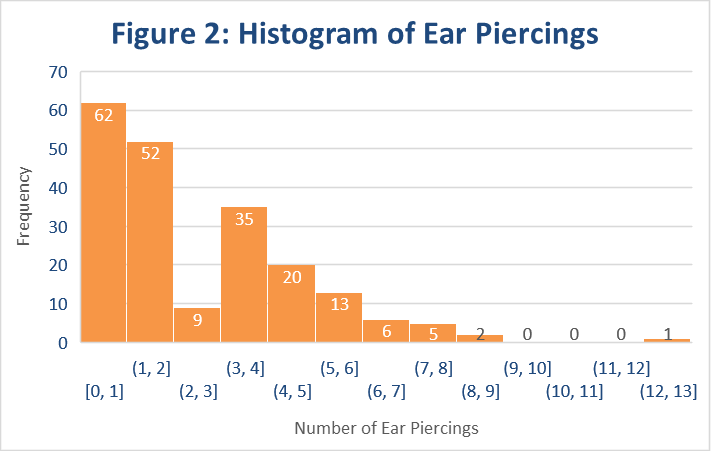


Figure 2 shows the distribution of student ear piercings is unimodal and significantly right-skewed. The graph also shows there is an outlier, representing a singular student with 13 ear piercings. Therefore, the median (2) and Interquartile Range (4) are the appropriate methods of measurement to determine central tendency and dispersion.

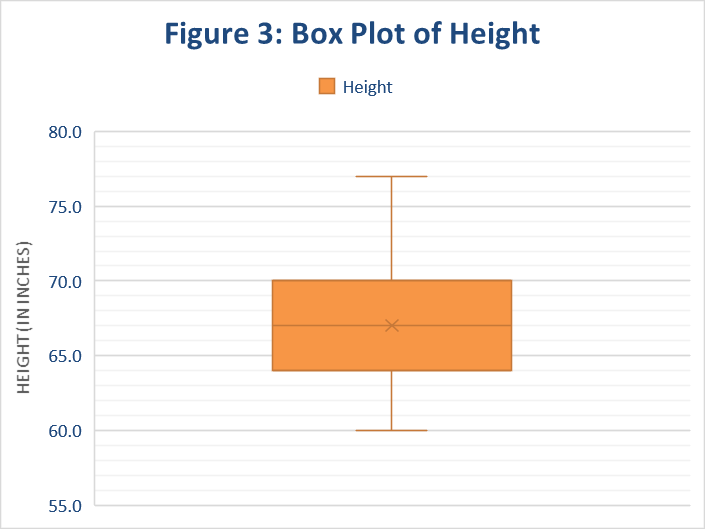


Figure 3 shows a box plot of students' heights. Both the mean and median are overlapping, meaning the two values are nearly the same. There are no outliers and no visible skewness regarding the minimum and maximum values. This results in an unimodal, symmetrical distribution for the variable Height.

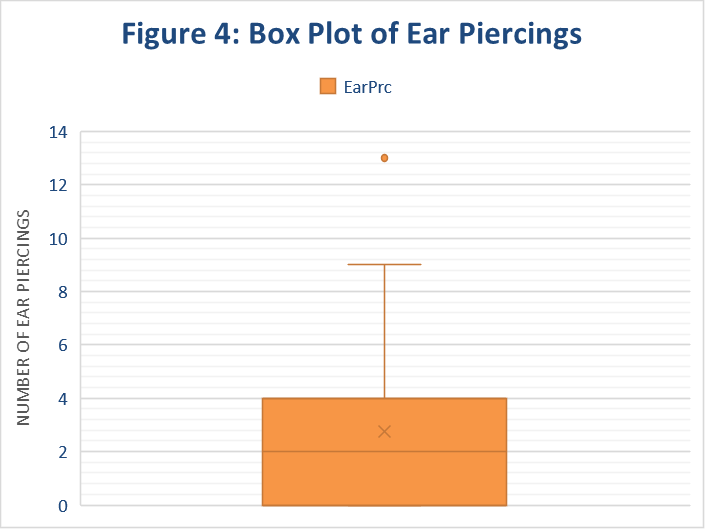


Figure 4 reveals a right-skewed graph of the number of students' ear piercings. It shows the median (2 ear piercings) is less than the mean (2.76 ear piercings). The graph also shows Quartile 1 is equivalent to the minimum, while there is a maximum value. Additionally, there is an outlier present in the distribution (13 ear piercings). This results in an unimodal, right-skewed distribution for the variable EarPrc.

Table 3: Frequency Table for Gender

|  |  |  |
| --- | --- | --- |
| **Gender** | **Frequency** | **Percent** |
| Female | 137 | 67% |
| Male | 68 | 33% |
| **Grand Total** | **205** | **100%** |

Table 3 shows approximately two thirds (67%) of the statistics class were female students, therefore the mode for the variable Gender is female.

Table 4: Frequency Table for Tattoo

|  |  |  |
| --- | --- | --- |
| **Tattoos** | **Frequency** | **Percent** |
| No | 174 | 85% |
| Yes | 31 | 15% |
| **Grand Total** | **205** | **100%** |

Table 4 shows far more students did not have a tattoo (85%). This means the mode for the variable Tattoo is No (does not have a tattoo).

Table 5: Frequency Table for CDs

|  |  |  |
| --- | --- | --- |
| **CDs** | **Frequency** | **Percent** |
| BelowAvg | 75 | 37% |
| Average | 70 | 34% |
| AboveAvg | 60 | 29% |
| **Grand Total** | **205** | **100%** |

Table 5 displays the frequency counts for students that owned either a below average, average, or above average number of CDs. The number of students owning a below average amount of CDs is slightly higher (37%), thus the mode of the variable CDs is BelowAvg.

**3)**

Table 6: Stratified Analysis of Height and EarPrc by Tattoo Status

|  |  |  |
| --- | --- | --- |
| **Tattoo** | **Mean of Height** | **Mean of EarPrc** |
| No | 66.86 | 2.55 |
| Yes | 67.97 | 3.94 |
| **Grand Total** | **67.03** | **2.76** |

Table 6 shows a stratified analysis of the quantitative variables Ear Piercings, Height, and Tattoo. The most significant result of the calculations is the mean number of piercings relative to tattoo status. There is a 1.39 ear piercings difference between the students with no tattoo and those with one or more tattoos. The mean height presents a slight difference between students with and without tattoos. Students with tattoos have a mean height of 67.97 inches, while students lacking tattoos have a mean height of 66.86 inches, a difference of only 1.11 inches. We see a much larger difference in the number of ear piercings than height relative to tattoo status.

**4)**

Table 7: Stratified Analysis of Ear Piercings by CDs

|  |  |  |
| --- | --- | --- |
| **CDs** | **Sum of EarPrc** | **Mean of EarPrc** |
| BelowAvg | 228.00 | 3.04 |
| Average | 201.00 | 2.87 |
| AboveAvg | 137.00 | 2.28 |
| **Grand Total** | **566.00** | **2.76** |

Table 7 displays a stratified analysis of the variables Ear Piercings and CDs. The sum is the appropriate calculation to analyze how many ear piercings are present within each group of CDs. The table shows the CDs group BelowAvg has a total of 228 ear piercings, while group Average has 201, and group AboveAvg has 137. These values are more useful because they show total count, rather than then just the mean of values in each group.

**5)**

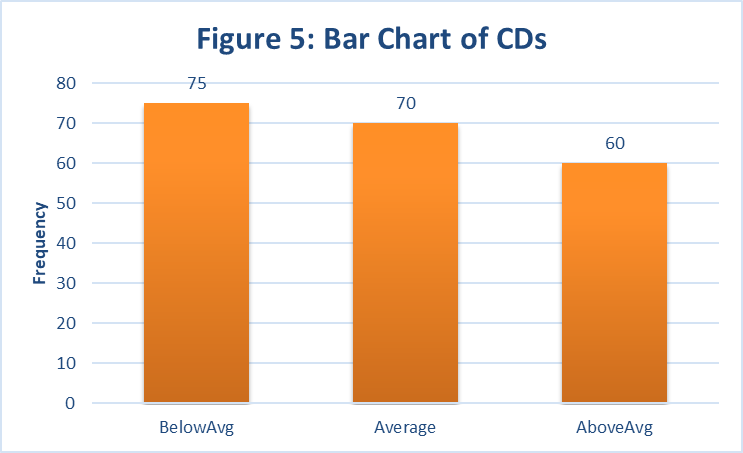
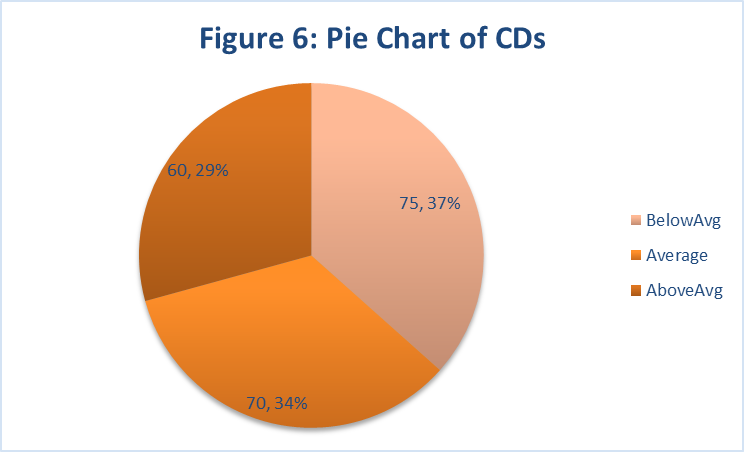


Figure 5 shows a bar chart of CDs groups while Figure 6 shows a pie chart of the same variable. The bar chart displayed in Figure 5 is a far better visualization of the variable. This is because it is difficult to see the slight difference of each pie slice, and it also does not intuitively show the order of the variable (BelowAvg, Average, AboveAvg).



**6)**

Table 8: Contingency Table for Tattoo status by CDs

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **.** | **CDs** |  |  |  |
| **Tattoo** | **BelowAvg** | **Average** | **AboveAvg** | **Grand Total** |
| No | 67 | 58 | 49 | 174 |
| Yes | 8 | 12 | 11 | 31 |
| **Grand Total** | **75** | **70** | **60** | **205** |

Table 8 is a contingency table that shows the distribution of variables Tattoo and CDs. The table points out that a majority of students do not have a tattoo, therefore there is a larger distribution of CDs groups for that tattoo status. Firstly, there is 67 students with no tattoo, that own a below average number of CDs. This is more than the 58 students with an average amount of CDs, and significantly more that then 49 students possessing an above average amount.

Table 9: Percentage of row for Tattoo status by CDs

|  |  |  |  |
| --- | --- | --- | --- |
| **.** | **Tattoo** |  |  |
| **CDs** | **No** | **Yes** | **Grand Total** |
| BelowAvg | 89.33% | 10.67% | 100.00% |
| Average | 82.86% | 17.14% | 100.00% |
| AboveAvg | 81.67% | 18.33% | 100.00% |
| **Grand Total** | **84.88%** | **15.12%** | **100.00%** |

In Table 9, the percentage of rows is displayed for the variables Tattoo and CDs. The percentages allow a closer analysis of CDs distribution among students with tattoos. It shows that 18.33% of students with tattoos have an above average number of CDs, compared to only 10.67% owning a below average amount.

Table 10: Percentage of column for Tattoo status by CDs

|  |  |  |  |
| --- | --- | --- | --- |
| **.** | **Tattoo** |  |  |
| **CDs** | **No** | **Yes** | **Grand Total** |
| BelowAvg | 38.51% | 25.81% | 36.59% |
| Average | 33.33% | 38.71% | 34.15% |
| AboveAvg | 28.16% | 35.48% | 29.27% |
| **Grand Total** | **100.00%** | **100.00%** | **100.00%** |

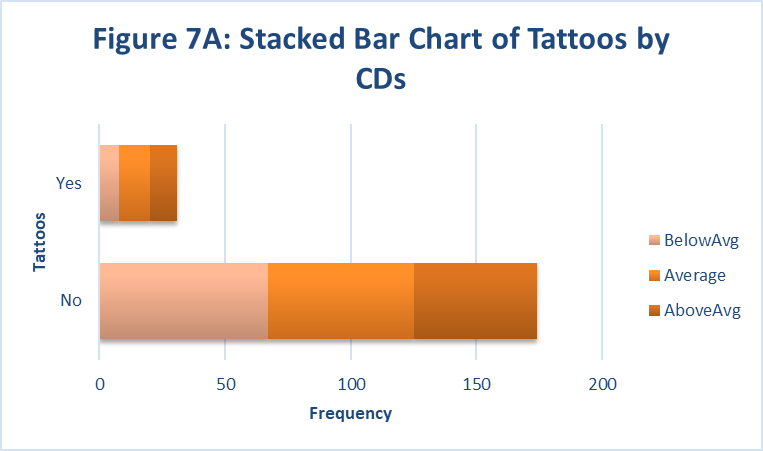
For Table 10, the percentage of columns are displayed for the previous variables. According to the table, it is apparent that students with no tattoos are more likely to have a below average number of CDs (38.51%) than an above average (28.16%). For students with tattoos, they are more likely to have an above average amount of CDs (35.48%) than a below average amount (25.81%).

Table 11: Percentage of total for Tattoo status by CDs

|  |  |  |  |
| --- | --- | --- | --- |
| **.** | **Tattoo** |  |  |
| **CDs** | **No** | **Yes** | **Grand Total** |
| BelowAvg | 32.68% | 3.90% | 36.59% |
| Average | 28.29% | 5.85% | 34.15% |
| AboveAvg | 23.90% | 5.37% | 29.27% |
| **Grand Total** | **84.88%** | **15.12%** | **100.00%** |

Table 11 shows the total percentages of all students in the classroom. Only 3.90% of the students had tattoos and a below average number of CDs, while 32.68% of students in the class had no tattoos and owned a below average amount of CDs.

**7)**



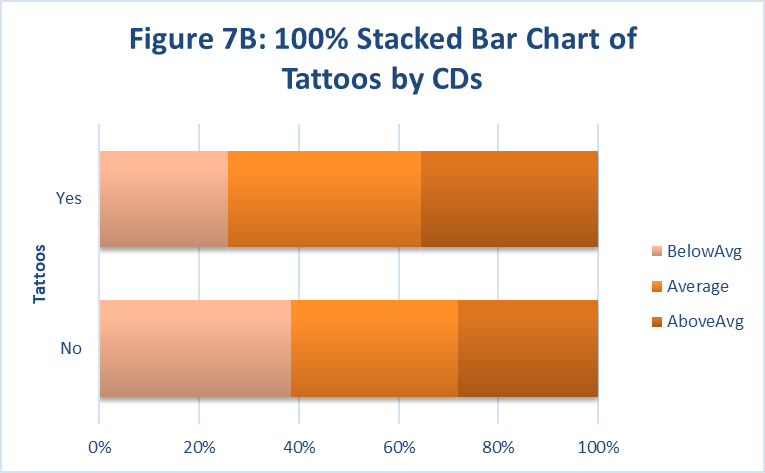


Figure 7A shows a stacked bar chart of the variables Tattoo and CDs. Figure 7B is the representation of the same variables, but shown as a 100% stacked bar chart. The chart 7B is a better representation of the relationship between tattoo status and CDs possession because you can compare the distribution of the CDs groups much more easily than in 7A.

**8)**

Table 12: Random Sample (n=35)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Gender** | **Tattoo** | **CDs** | **Height** | **EarPrc** | **Random** |
| Female | Yes | Average | 62.5 | 7 | 0.00723468 |
| Male | No | BelowAvg | 69.0 | 0 | 0.00731810 |
| Female | No | AboveAvg | 66.0 | 4 | 0.01050671 |
| Female | Yes | Average | 64.0 | 4 | 0.02169663 |
| Male | No | AboveAvg | 70.0 | 0 | 0.02401736 |
| Male | No | AboveAvg | 74.0 | 0 | 0.02539078 |
| Female | No | BelowAvg | 67.0 | 2 | 0.02683147 |
| Female | No | BelowAvg | 70.0 | 4 | 0.03000857 |
| Male | No | AboveAvg | 73.0 | 0 | 0.03589335 |
| Male | No | AboveAvg | 68.0 | 0 | 0.04756544 |
| Male | No | BelowAvg | 76.0 | 0 | 0.05038095 |
| Male | No | Average | 71.0 | 0 | 0.05572796 |
| Female | No | AboveAvg | 63.0 | 4 | 0.05634299 |
| Male | Yes | Average | 72.0 | 2 | 0.05679727 |
| Female | No | Average | 64.0 | 2 | 0.05707654 |
| Female | No | BelowAvg | 62.5 | 6 | 0.06238325 |
| Female | No | Average | 66.0 | 0 | 0.06479609 |
| Female | No | BelowAvg | 67.0 | 5 | 0.06797105 |
| Female | No | Average | 67.0 | 4 | 0.06801928 |
| Female | No | Average | 60.0 | 2 | 0.07233910 |
| Female | No | BelowAvg | 64.0 | 5 | 0.07940061 |
| Female | No | BelowAvg | 65.0 | 2 | 0.08168229 |
| Female | No | Average | 64.0 | 4 | 0.08533762 |
| Female | No | BelowAvg | 66.0 | 2 | 0.08560321 |
| Male | No | AboveAvg | 67.0 | 0 | 0.09395020 |
| Female | No | BelowAvg | 63.0 | 7 | 0.09470112 |
| Male | No | Average | 72.0 | 0 | 0.09524781 |
| Male | No | BelowAvg | 70.0 | 0 | 0.09775457 |
| Female | No | Average | 68.0 | 2 | 0.10030257 |
| Female | No | AboveAvg | 63.0 | 4 | 0.11140619 |
| Male | No | BelowAvg | 74.0 | 0 | 0.11304642 |
| Female | No | Average | 64.0 | 3 | 0.11978178 |
| Female | No | BelowAvg | 66.0 | 2 | 0.12156340 |
| Female | No | Average | 67.0 | 2 | 0.12632488 |
| Female | No | BelowAvg | 64.5 | 5 | 0.12715498 |

Table 13: Descriptive Statistics for the Random Sample (n=35)

|  |  |  |
| --- | --- | --- |
|  | **Height** | **EarPrc** |
| Sample Mean | 67.13 | 2.40 |
| Median | 67.00 | 2.00 |
| Sample SD | 3.82 | 2.17 |
| Population SD | 3.79 | 2.42 |
| Minimum | 60.00 | 0.00 |
| Q1 | 64.00 | 0.00 |
| Q3 | 70.00 | 4.00 |
| Maximum | 76.00 | 7.00 |
| Range | 16.00 | 7.00 |
| IQR | 6.00 | 4.00 |
| Count | 35.00 | 35.00 |

Table 13 shows the descriptive statistics of the variables Height and EarPrc for the sample of 35 students. The mean of the sample is 67.13 inches, while in Table 2 it shows the population mean is 67.03 inches. Similarly, the sample median is 67 inches, while the population median is also exactly 67 inches.

**9)**

Table 14: 90%, 95%, 99% Confidence Intervals for Height

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Sample Mean** | **Confidence Level** | **a** | **z\*** | **Standard Deviation** | **Margin of Error** | **Lower Bound** | **Upper Bound** | **Confidence Interval** |
| 67.1 | 90% | 0.10 | 1.645 | 3.8 | 1.1 | 66.1 | 68.2 | (66.1, 68.2) |
| 67.1 | 95% | 0.05 | 1.96 | 3.8 | 1.3 | 65.9 | 68.4 | (65.9, 68.4) |
| 67.1 | 99% | 0.01 | 2.575 | 3.8 | 1.7 | 65.5 | 68.8 | (65.5, 68.8) |

Table 14 displays the confidence intervals for Height (in inches). Based on the random sample of 35 students, we can be 95% confident that the average height (or true mean) of students in the selected statistics class was between 65.9 and 68.4 inches.