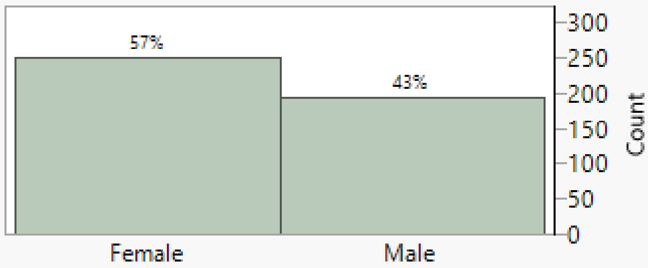
Gender



Frequencies

Level	Count	Prob
Female	249	0.56720
Male	190	0.43280
Total	439	1.00000

N Missing 0
2 Levels

A number of different variables were analyzed to get an understanding of a university student so Brainstorm Media can effectively market to them. Firstly, Brainstorm Media wanted to find out how many females are taking introductory statistics classes at the University of Tennessee, to see if they need to advertise differently to focused more on females. However, it appears that while gender is somewhat equal when it comes to this specific sample, there are more females than males taking this class, with about 56.7% of the class identifying as female, while only about 43.2% identifying as male. This is a very small difference of only 59 more females than males in this sample but it does suggest that females are the majority of the class, based on those percentages. This gives insight that Brainstorm Media doesn’t need to create more advertisements that are more specifically targeted towards either gender, females or males, since even with the slight difference, they are still closely equal.

Confidence Intervals

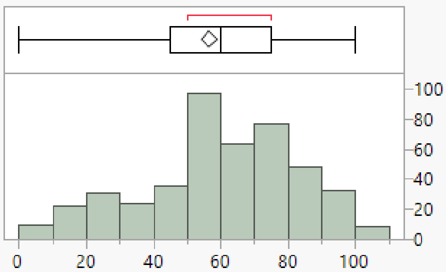
Level	Count	Prob	Lower CI	Upper CI	1-Alpha
Female	249	0.56720	0.524762	0.608671	0.925
Male	190	0.43280	0.391329	0.475238	0.925
Total	439				

Note: Computed using score confidence intervals.

Brainstorm Media then wanted to create a 92.5% confidence interval for the variable gender using female as the desired response of interest to see what proportion of females were in introductory statistics classes at the University of Tennessee Knoxville, to see further evidence if they needed to advertise these classes more specifically towards females. First, the data has to be checked to see if it passes the 3 conditions for a confidence interval for sample proportions. The first condition is randomization, and the data does pass this condition since it was a random sample of 439 students. This condition simply makes sure that the data is independent and by having randomly selected data, it reduces the amount of bias in the sample. The second condition is the 10% condition, which the data would pass as long as there are more than 4,390 students among all of the introductory statistics classes taught at University of Tennessee Knoxville. This condition also checks independency of the data. The final condition is the success/failure condition, which just checks that the sample size is large enough to create a confidence interval by making sure there are at least 10 successes and 10 failures. We can check this by simply looking at the count of males and females, and since females are considered our successes in this context, we know we have 249 successes and 190 failures. Therefore, the data passed the third condition since both are greater than 10. Since all of the conditions are met, we can interpret the data. We are 92.5% confident that the true proportion of students who identify as female in an introductory statistics class at the University of Tennessee Knoxville is contained in the interval 52.5% and 60.9%.

Since Brainstorm Media is trying to figure out if their advertising needs to market these introductory statistics classes more specifically to students who identify as female, it could be beneficial to see if the gender of the students in the introductory statistics classes at the University of Tennessee Knoxville is around 50% for both genders or if one gender is more of the majority. If females are more of the majority, then Brainstorm Media doesn't need to start creating more advertisements more targeted towards females for these classes, but if males are more of the majority, then that information could be used to better understand the target market of their advertisements and make adjustments to the ads. We would be testing the initial hypothesis that the true proportion of students in the introductory statistics classes at the University of Tennessee Knoxville who identify as female is 50% against the alternative hypothesis that the true proportion of students in the introductory statistics classes who identify as female is not 50%. Since the interval is from 52.5% and 60.9% of students identifying as female, we would reject the null since 50% is not in that interval. This means that there is sufficient evidence to conclude that the true proportion of students in the introductory statistics classes at the University of Tennessee Knoxville who identify as female is not 50%.

PercMoreIntelligentThan



Quantiles

100.0%	maximum	100
99.5%		100
97.5%		99
90.0%		85
75.0%	quartile	75
50.0%	median	60
25.0%	quartile	45
10.0%		20
2.5%		10
0.5%		0
0.0%	minimum	0

Summary Statistics

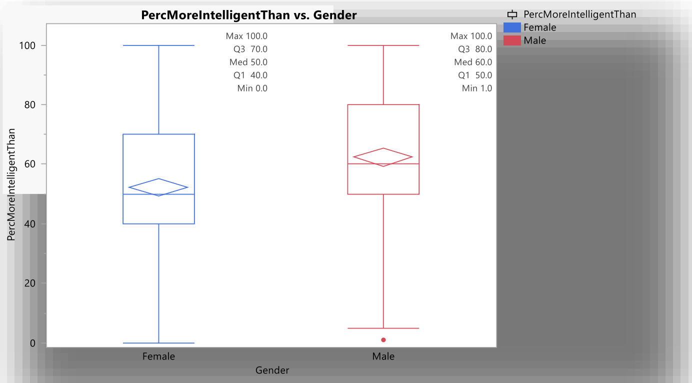
Mean	56.680182
Std Dev	23.644879
Std Err Mean	1.1285085
Upper 95% Mean	58.898147
Lower 95% Mean	54.462217
N	439

The next variable that Brainstorm Media wanted to look at was what percent of people an individual believed they were smarter than, since that could give insight as to what typical students in an introductory statistics class at the University of Tennessee Knoxville believe about themselves compared to others when it comes to their intelligence. This variable is important because it may help us understand why less of the population of students at the University of Tennessee Knoxville are taking these introductory statistics classes, depending on what percent of people in their grade level and gender the majority of current students believe they are more intelligent than. Based on the data, the histogram created is skewed to the left with most of the data starting after 50%, which means that more of the students in the sample thought that they were more intelligent than 50% or more of the people in their same grade level and gender. Since the data is skewed, it would be safe to assume that we should use the median since it would not be affected by those students who thought of themselves as only more intelligent than 50% or less of people in their same grade level and gender. Those students on the left side of the histogram, may be taking the class to build their intelligence but the majority is further right. Based on the median, about 50% of the students surveyed believe they are more intelligent than 60 percent of people in their same grade level and gender. Furthermore, based on the IQR of the histogram, the middle 50% of students believed they were between more intelligent than 45 to 75 percent of people in their same grade level and gender., which leads to a range of about 35% for the middle 50% of the students surveyed. This seems like most believe they are more intelligent than half of their peers that are in their same grade level and gender which could be associated with future students reasoning that they may not be smart enough for the class if they know how current students feel intelligence wise when related to their peers.

Confidence Intervals

Parameter	Estimate	Lower CI	Upper CI	1-Alpha
Mean	56.68018	54.66612	58.69425	0.925
Std Dev	23.64488	22.30692	25.16199	0.925

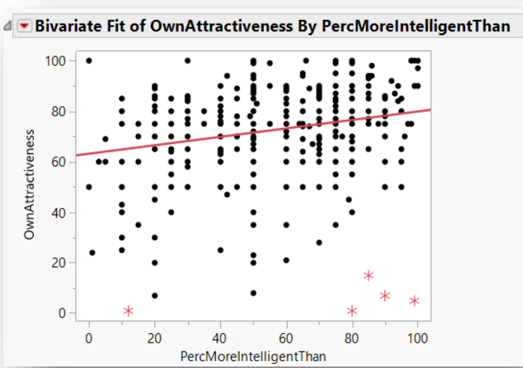
Brainstorm Media then wanted to create a 92.5% confidence interval to see if the true average percentage of people in their same grade level and gender that students thought they were more intelligent than, was close to what the histogram predicted as the average which was 56.68%. In order to do a confidence interval for population mean, there are three conditions that need to be checked. The first condition is the randomization condition, which checks that the data is independent. This is done by making sure the sample was a simple random sample of the population which it was so the data passes the first condition. The second condition, which is the 10% condition, also checks independency of the data by making sure that the sample size of 439 is no larger than 10% of the population. This would be met as long as the population of students taking the introductory statistics class at the University of Tennessee Knoxville is over 4,390. The final condition is the nearly normal condition, which checks that the data from the population is unimodal and symmetric while also checking that the sample size is appropriate in order to use the t-distribution. While the data for the percentage of people a student felt they were more intelligent than, was skewed, we are assuming that the condition is met since it was unimodal and a sample size of 439 means it is safe to use the t-model since the data is not extremely skewed. Therefore, it is safe to create a confidence interval and interpret what it means. We are 92.5% confident that the true population average for percentage of people in their same grade level and gender is contained within the interval 54.67% and 58.69%.



Means and Std Deviations

Level	Number	Mean	Std Dev
Female	249	46.25502	24.631728
Male	190	61.394211	24.880634

While the confidence interval examine the percentage for all students, analysis was done to see if there was a relationship between male and female (Gender) and the percentage of people in their same grade level and gender. Based on the side-by-side boxplot, it appears that there is a relationship between the two variables. While it appears the overall shape of the two groups are essentially the same, female (0,100) and male (1, 100), there is a difference of 15.14 percentage between the average percentage for females 46.26% and average percentage for male 63.39%. Males seem to have a much higher view, on average, that they are more intelligent than people in their same grade level and gender. A statistical test could further determine whether this difference of the two groups is by random chance or if this relation is statistically significant.



Linear Fit

OwnAttractiveness = 63.182383 + 0.1671553*PercMoreIntelligentThan

Summary of Fit

RSquare	0.048679
RSquare Adj	0.046502
Root Mean Square Error	17.44126
Mean of Response	72.71366
Observations (or Sum Wgts)	439

Lack Of Fit

Analysis of Variance

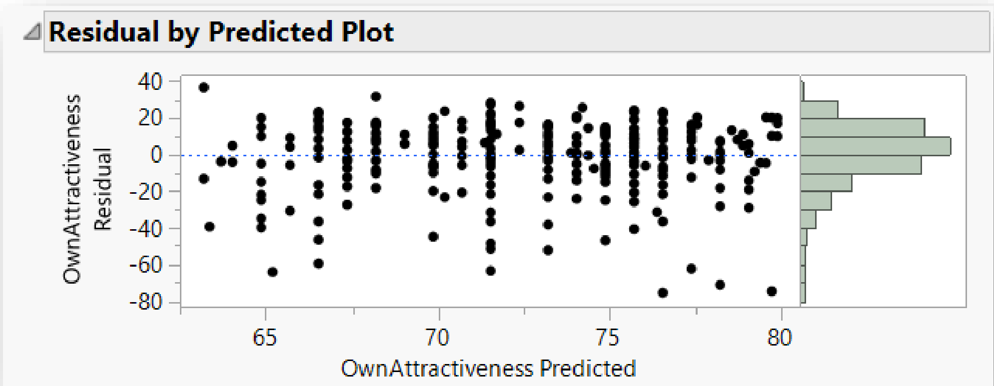
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	6802.29	6802.29	22.3614
Error	437	132934.32	304.20	Prob > F
C. Total	438	139736.61		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	63.182383	2.180716	28.97	<.0001*
PercMoreIntelligentThan	0.1671553	0.035348	4.73	<.0001*

The scatterplot above shows the relationship between the percentage of people in their same grade and gender that a student feels they are more intelligent than and their ranking on their own attractiveness. Brainstorm Media wanted to see if these two variables had any possible relationship in order to further elaborate on the decrease in enrollment in the introductory statistics classes at the University of Tennessee Knoxville. By looking at the scatterplot, we can see a slight positive direction, since it appears that as the percentage of people that a student feels they are more intelligent than increases, there is a larger majority of data points higher up on the scale of rankings for their own attractiveness. The form of the scatterplot is quite weak however, since the points bunch up close together in the top right corner of the scatter plot, with very minimal points below 40 on the own attractiveness scale. This could be explained by the observation earlier on in the histogram, where more than 50% of the students felt they were more intelligent than 50% of the people in their same gender and level, and since the direction is positive, that makes lower rankings of own attractiveness possible outliers since they go against the trend. There are about 8 points that stick out in the scatterplot, specifically those that rest below 20 on the ranking of own attractiveness, since they are a distance away from the majority and do not follow the trend. One unusual feature of the data is the point located around 99% of people the student feels more intelligent than but only with a ranking of 7 or 8 for their own attractiveness. This point stands out since the rest of points around that percentage are higher up, typically above 80 on the ranking of their own attractiveness. However, the several outliers could just be students who are shy about their appearance and may have other lurking reasons why they don’t feel as attractive but we will be keeping that data in the scatterplot. Even though we are keeping these outliers in the scatterplot, we understand that outliers can do anything to the measurement of association between these two variables.

Next, a regression was performed on the data to figure out if the percentage of people a student believes they are more intelligent than can be used to predict their own ranking of attractiveness. Before the regression could be done, there are three conditions that need to be met before we even make the residual plot. The first condition is that there are two quantitative variables. This condition is met since both the percentage of people an individual believes they are smarter than and their own attractiveness ranking are quantitative variables. The second condition is the straight enough condition, which the data does not pass. The scatterplot is too clumped together above the line while it spreads out a lot below the line. The third condition is the no outliers condition, which the scatter plot also does not pass since there are several outliers near the bottom of the graph that are a distance away from the majority. One point we would define as an outlier is the point previously mentioned at a percentage of 99% with only a ranking of 7 or 8, since it is the furthest from the majority for that percentage. Therefore, the data only passes one of the three initial conditions to check but we will still create the residual to see if the fourth condition is met or not.



The fourth condition is based on the residual, and it is the does the plot thicken? condition. This is checked by looking at the residual and seeing if the data fans out at any point or if it is consistent throughout the plot. Based on this residual plot, we would say that the data doesn't pass the condition, since the data is more spread out at the beginning and then clump together near the predicted own attractiveness ranking of 75. This could be due to the fact that the initial best fit line was closer to the top of the scatterplot, where a lot of the data was located, so the residual between predicted ranking of own attractiveness for higher percentages of people a student thought they were more intelligent than, had less of a variety of residuals the larger the percentage went. Since the initial data for percentage of people a student felt that they were more intelligent than did lean on the upper 50% of the scale, the scatterplot may have passed these conditions if there was more data on the lower end of percentages to give more accuracy. If we did decide to remove the lower set of outliers in the scatterplot or simply focus on the upper 50% of data, we may be able to create a more accurate regression line and residual plot for these two variables.

After checking the four conditions, we then want to see whether or not the relationship between these two variables is due to random chance. We can check this by comparing the p-value of the slope to the threshold of 0.05. Since the p-value for this set of data is less than 0.0001 which is below 0.05, we can conclude that the relationship between the percentage of people a student feels that they are more intelligent than and their personal ranking of their own attractiveness is not due to random chance. Since the relationship is not due to random chance, we can use this model to predict student's rankings of their own attractiveness based on what percentage of people they believe they are more intelligent than. For every additional increase in the percentage of people a student believes they are more intelligent than, we would expect their ranking of their own attractiveness to increase by 0.167 points, on average. For a student who believes they are more intelligent than 0% of people in their same grade and gender, our model predicts that they would have a ranking of 63.18 on their own attractiveness. We can also use the model to look at how much of the variation in one variable is explained by the other variable. Therefore, 4.87% of the variation in a student's individual ranking of their own attractiveness is associated with the variation in percentage of people a student believes that they are more intelligent than.