EFFECT OF COLLEGE TYPE
PUBLIC VS PRIVATE

Hypothesis: There is not a difference in yearly cost of attendance, yearly debt acquired, household income, and retention rate after their first year between public and private colleges?

All analyses done through a t-test for the difference of means.

Introduction:

Trying to decide between a public and private college can be a very daunting task for high school students — many ask, "Is it worth it to attend a private college?" The following dataset, which includes responses from nearly 7,500 colleges, was obtained online from the Department of Education's College Scorecard. We looked at data summarizing cost, family income, debt, and retention rate for many of these universities. In the original dataset, private schools were split into two categories: for-profit and not for-profit. We decided to combine both into one variable.

Conclusions:

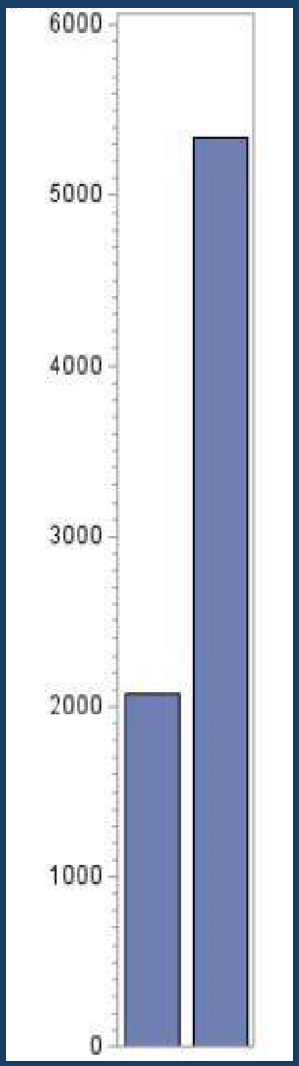
	Sample Size	Skewness	Median	Conclusion
Cost	4,055	Right	$20,450	Reject
Debt	6,311	Right	$8,749	Reject
Income	7,130	Right	$22,937	Reject
Retention	2,316	Left	0.7196	Reject

From our analysis, we concluded that public schools are a better "bang for your buck" on average because they cost less, leave students with less debt, and have higher retention rates. It is also important to note that students who attended public schools came from families with higher incomes than those students who attended private schools.

Note: In figures, 1 refers to public colleges and 2 refers to private colleges

NUMBER OF DATA POINTS

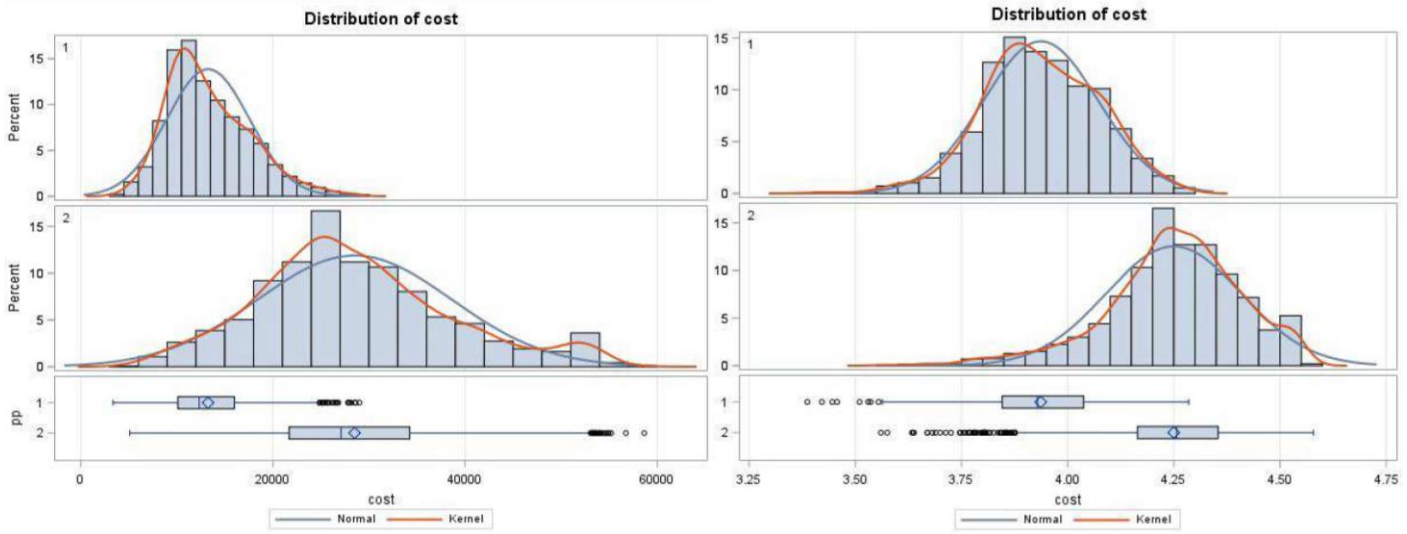
PUBLIC (1) VS PRIVATE (2)



By: Mitch Tamashunas, Brooke Arnold, and Brandon Kuhn

Data from College Scorecard 2010 to 2011

H_O: The cost of public and private colleges is the same
H_A: The cost of public and private colleges is different



The left distributions are the untransformed data. It can be seen that they are both skewed to the right. To make the distributions more normal, we used logarithim transformations. It is important to note that we can use log units when analyzing the data because log units are monotonic, which means that when transformed back to the original data, the sample mean for public schools will still be less than that of private schools. This resulted in the distributions on the right, which are much more normal. Even though there are outliers in both distributions, this can be expected because the sample size is in the thousands. Looking at the boxplots, it can be seen that the mean and median for each distribution are almost equal.

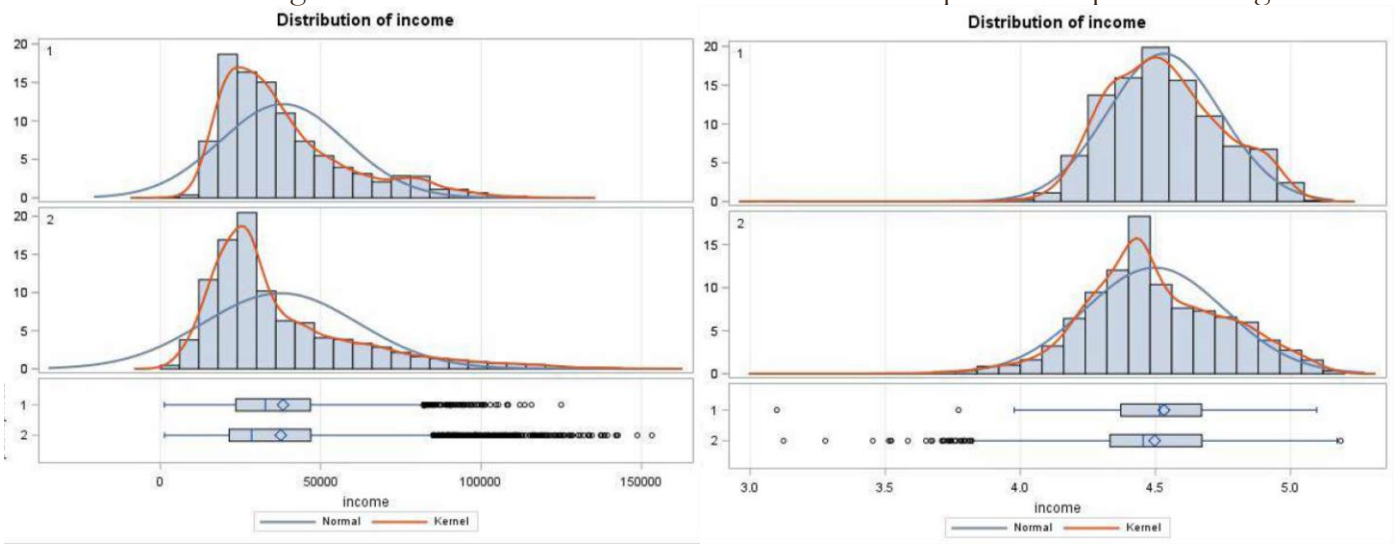
publicprivate	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1		3.9367	3.9301	3.9433	0.1360	0.1315	0.1408
2		4.2497	4.2434	4.2561	0.1591	0.1547	0.1637
Diff (1-2)	Pooled	-0.3131	-0.3225	-0.3036	0.1501	0.1469	0.1534
Diff (1-2)	Satterthwaite	-0.3131	-0.3222	-0.3039			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	4053	-58.01	<.0001
Satterthwaite	Unequal	3492.2	-65.94	<.0001

The mean cost for public schools is 3.9367 log units and the mean price for private schools is 4.2497 log units. Using a t-test for difference of means, we can be 95% confident that the true mean difference between public and private colleges lies between -0.3222 and -0.3039 log units. Looking at the second table, we can see that the p-value for this test is less than 0.0001. Using either of the previous two results, at a confidence level of $\alpha = 0.05$ we can reject the null hypothesis and conclude there is a difference between the average cost of public and private schools. Also, looking back at the confidence interval, we can infer that private schools on average cost more than public schools.

Note: In figures, 1 refers to public colleges and 2 refers to private colleges

H_O: The average household income is the same for students in public and private colleges
H_A: The average household income is different for students in public and private colleges



The left distributions are the untransformed data. It can be seen that they are both skewed to the right. To make the distributions more normal, we used logarithim transformations. It is important to note that we can use log units when analyzing the data because log units are monotonic, which means that when transformed back to the original data, the sample mean for public schools will still be less than that of private schools. This resulted in the distributions on the right, which are much more normal. Even though there are outliers in both distributions, the number was dramatically decreased and some can always be expected because the sample size is in the thousands. Looking at the boxplots, it can be seen that the mean and median for each distribution are almost equal.

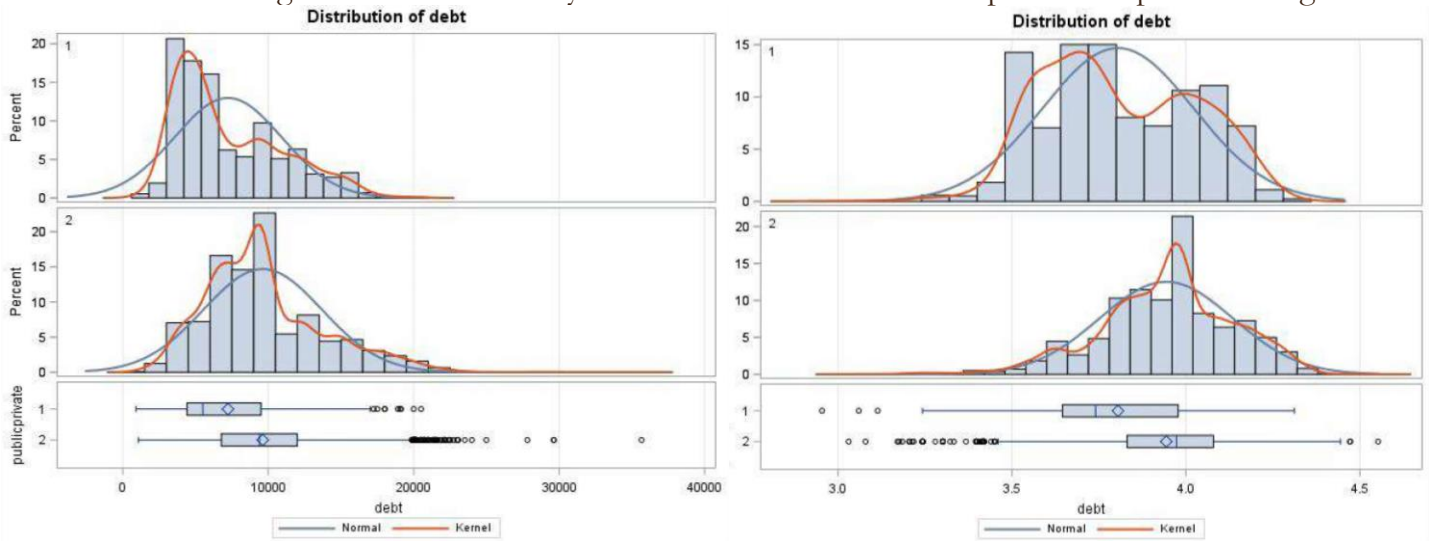
publicprivate	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1		4.5314	4.5224	4.5404	0.2093	0.2031	0.2158
2		4.4967	4.4896	4.5039	0.2592	0.2542	0.2643
Diff (1-2)	Pooled	0.0347	0.0221	0.0472	0.2458	0.2418	0.2499
Diff (1-2)	Satterthwaite	0.0347	0.0232	0.0462			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	7128	5.40	<.0001
Satterthwaite	Unequal	4696.7	5.90	<.0001

The mean household income for public schools is 4.5314 log units and the mean household income for private schools is 4.4967 log units. Using a t-test for difference of means, we can be 95% confident that the true mean difference between public and private colleges lies between 0.0232 and 0.0462 log units. Looking at the second table, we can see that the p-value for this test is less than 0.0001. Using either of the previous two results, at a confidence level of $\alpha = 0.05$ we can reject the null hypothesis and conclude there is a difference between the household income of students in public and private schools. Also, looking back at the confidence interval, we can infer that families of students at private schools on average earn less than those at public schools.

Note: In figures, 1 refers to public colleges and 2 refers to private colleges

H_O: The average debt after the first year is the same for students in public and private colleges
H_A: The average debt after the first year is different for students in public and private colleges



The left distributions are the untransformed data. It can be seen that they are both skewed to the right with a strange shape. To make the distributions more normal, we used logarithim transformations. This resulted in the distributions on the right, which are more normal. The distributions still show an abnormal shape, however, which must be kept in mind when drawing conclusions. Even though there are outliers in both distributions, the number was dramatically decreased, and some can always be expected because the sample size is in the thousands.

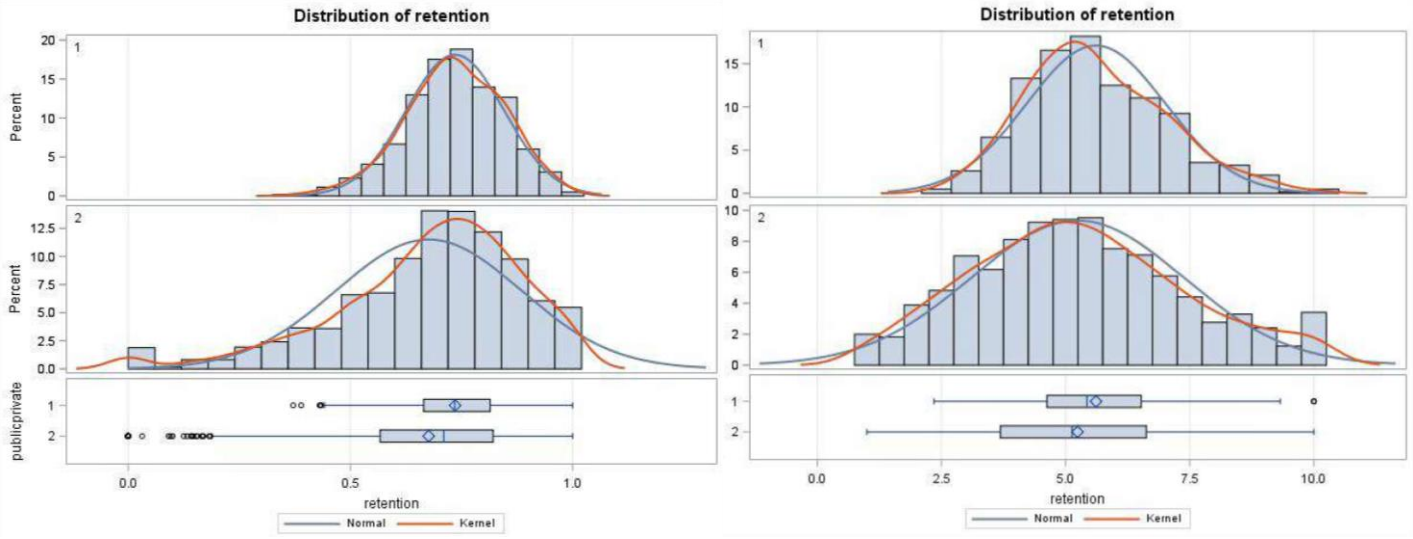
publicprivate	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1		3.8047	3.7943	3.8150	0.2174	0.2104	0.2250
2		3.9435	3.9379	3.9490	0.1912	0.1874	0.1952
Diff (1-2)	Pooled	-0.1388	-0.1498	-0.1277	0.1986	0.1952	0.2022
Diff (1-2)	Satterthwaite	-0.1388	-0.1505	-0.1271			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	6309	-21.22	<.0001
Satterthwaite	Unequal	3334.4	-22.21	<.0001

The mean debt acquired after the first year for public schools is 3.8047 log units and the mean debt acquired after the first year for private schools is 3.9435 log units. Using a t-test for difference of means, we can be 95% confident that the true mean difference between public and private colleges lies between -0.1505 and -0.1271 log units. Looking at the second table, we can see that the p-value for this test is less than 0.0001. Using either of the previous two results, at a confidence level of $\alpha = 0.05$ we can reject the null hypothesis and conclude there is a difference between the mean debt acquired after the first year for students in public and private schools. Also, looking back at the confidence interval, we can infer that students who attend public schools on average acquire less debt after the first year than those who attend private schools. As mentioned, we must question whether these conclusions are correct because of the shape of the distributions. Looking back at our conclusion in our previous two tests helps enforce the conclusion in this test. That is because if public colleges cost less and the students' families earn more, it logically follows that the students will have less debt.

Note: In figures, 1 refers to public colleges and 2 refers to private colleges

H_O: The retention rate of public and private colleges is the same
H_A: The retention rate of public and private colleges is different



The left distributions are the untransformed data. It can be seen that the distribution for private schools is skewed to the left. To make the distributions more normal, we used exponential transformations. It is important to note that we can use exponential units when analyzing the data because exponential units are monotonic, which means that when transformed back to the original data the sample mean for public schools will still be more than that of private schools. This resulted in the distributions on the right, which are more normal and have only one outlier. The transformation caused the distribution for public schools to become slightly skewed to the right, but when comparing the left two distributions to the right two, the right two are overall much more symmetric. Looking at the boxplots, it can be seen that the mean and median for each distribution are almost equal.

publicprivate	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1		5.6114	5.5007	5.7221	1.3991	1.3251	1.4820
2		5.2394	5.1379	5.3409	2.1345	2.0651	2.2088
Diff (1-2)	Pooled	0.3720	0.1907	0.5533	1.9661	1.9110	2.0244
Diff (1-2)	Satterthwaite	0.3720	0.2219	0.5221			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	2314	4.02	<.0001
Satterthwaite	Unequal	1661.9	4.86	<.0001

The mean retention rate for public schools is 5.6114 exponential units and the mean retention rate for private schools is 5.2394 exponential units. Using a t-test for difference of means, we can be 95% confident that the true mean difference between public and private colleges lies between 0.2219 and 0.5221 exponential units. Looking at the second table, we can see that the p-value for this test is less than 0.0001. Using either of the previous two results, at a confidence level of $\alpha = 0.05$ we can reject the null hypothesis and conclude there is a difference between the retention rate of public and private schools. Also, looking back at the confidence interval, we can infer that public schools on average have higher retention rates than private schools.

Note: In figures, 1 refers to public colleges and 2 refers to private colleges

List of each person's contributions:

All parts of the project were done together in the library, so all members contributed equally to final project.

Note: In figures, 1 refers to public colleges and 2 refers to private colleges