Analysis of the Strength of the Relationship Between IQ Scores and GPA

Math HL Internal Assessment

Personal Code: hsn921

Introduction and Rationale **2**
Aim 2
Hypothesis 2
Initial Results 3
Correlation **4**
Regression Model 4
Standard Error of Estimate 5
Pearson Correlation Coefficient 6
Test of Independence **7**
Q-Q Plot 7
Shapiro Wilk Test 9
Chi-Squared Test for Independence 10
Evaluation **13**
Extensions **16**
Bibliography **16**
Appendix **17**

.

Introduction and Rationale

For me and students all across America, school is a competition. Get the highest SAT score, enroll in the hardest possible classes, become valedictorian, etc. There has always been a stigma that a high-Grade Point Average (GPA) is a consequence of naturally higher intelligence; students at the top of the class rankings worked hard, but they also have higher natural ability. However, I don't believe this: I don’t think that kids who happen to have a high GPA are necessarily any smarter than anyone else. School should not be a game that assigns winners and losers in an arbitrary point system. At my current school, there has been much discussion on whether class rank should be abolished, and with it the traditions of valedictorian. Our class rank system is based on student GPA. The purpose of this investigation is to either confirm or refute the claim that there is a strong link between GPA and intelligence. I plan to measure “intelligence,” with an Intelligence Quotient (IQ) test. IQ Tests have been used for more than a century, initially developed as a test for struggling school children, they have been used in a variety of different applications[1]. While an IQ test is not an absolute measure, it does provide a decent baseline. Showing that there is no direct correlation between GPA and “intelligence,” would help to disprove the stigma that has been surrounding class rank. I am very interested in finding out whether the students at the top of the rankings, can actually be deemed to be any smarter. The outcome of this experiment could also be used to argue a case for the abolishment of class rank.

Aim

I intend to collect a sample of students to conduct my research. I will ask each of these students for their GPAs. I will then ask each student to take an IQ test. This IQ test will be an online test, that is around 20 minutes in length. At the end of the test, a score will be formulated along with the questions they got right and wrong. They will be prompted to respond to questions regarding their opinions on the validation of GPA and IQ values. These results will then be tabulated in a table for analysis.

Hypothesis

I will conclude that intelligence level is a good predictor of a GPA if the Pearson Correlation Coefficient (r-value), is greater than 0.7 or less than -0.7.

$$|r| > 0.7$$

This value would suggest that the coefficient of determination (r^2) value is around 0.49, signifying that around half of the data sets are closely related to each other. This coefficient of determination value also implies that the linear line of best fit accurately predicts about half the data.

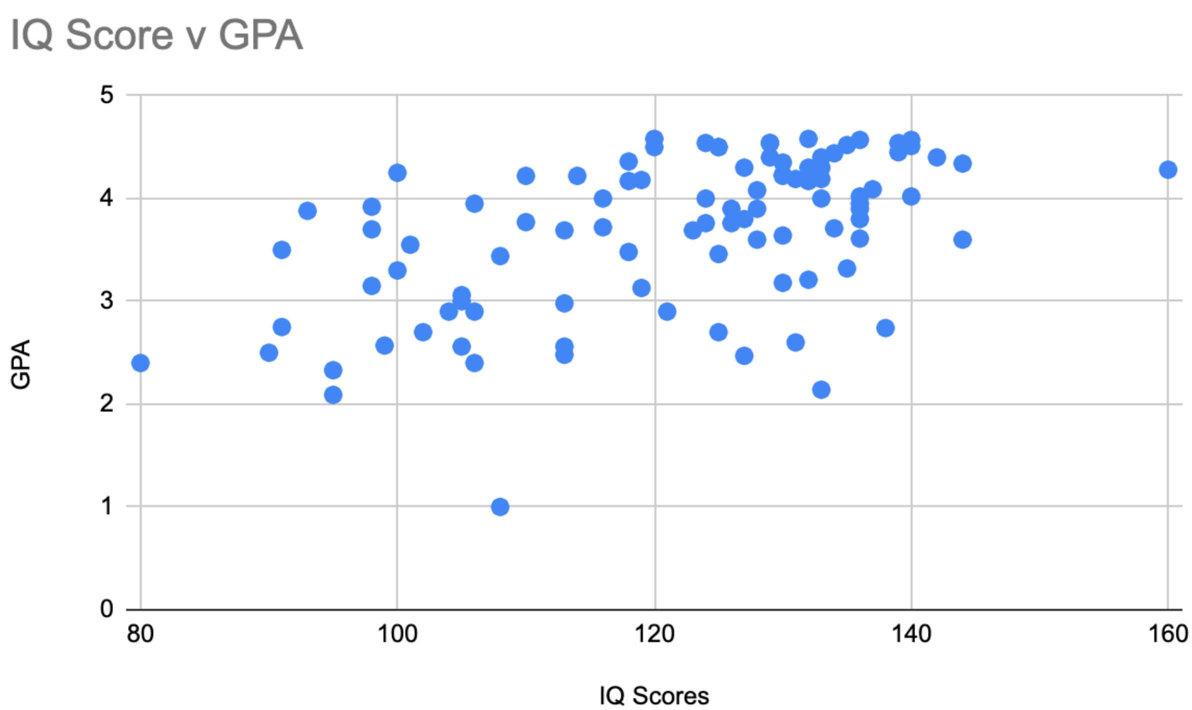
[1] Dombrowski, Stefan 2020

.

Therefore I will define my hypothesizes as follows:
Null Hypothesis H_0: There is no strong correlation between a student's IQ score and their IQ
Alternate Hypothesis H_1: There is a strong correlation between a student's IQ score and their IQ

Initial Results

From the survey, I was able to accumulate 99 pairs of IQ scores and GPAs from students. The data is shown in the scatter plot below:



The full data table can be found in the appendix*

Each data point represents a single student. Because of the context of the experiment, I have chosen GPAs to be the dependent variable. Upon immediate observation, the strength of the correlation cannot easily be determined. This suggests that there may not be a strong link between an IQ score and a student's subsequent GPA. In determining the association between the two variables, a Least Squares Regression Line can be calculated.

.

Correlation

Regression Model

The least squares regression line will tabulate an equation in the form $y = mx + b$. I will utilize the following equations[2] to find the unknown *m* and *b* values:

$$m = \frac{N\Sigma(xy) - \Sigma x \Sigma y}{N\Sigma(x^2) - (\Sigma x)^2} \qquad b = \frac{\Sigma y - m\,\Sigma x}{N}$$

Where
- N is the total amount of data points
- X is the independent variable: IQ Score
- Y is the dependent variable: GPA value

The following shows sample computations for 10 of the students in my sample.

IQ Score	GPA	x^2	XY
128	3.60	16384	460.8
137	4.09	18769	560.33
100	4.25	10000	425
133	4.00	17689	532
142	4.04	20164	624.8
120	4.58	14400	549.6
134	4.44	17956	594.96
124	4.54	15376	562.96
135	4.52	18225	610.2
133	4.40	17689	585.2

Here are the totals for all of the students in the sample.

N	ΣX	ΣY	ΣX^2	ΣXY
99	12035	364.76	1485933	44927.95

[2] Math is Fun

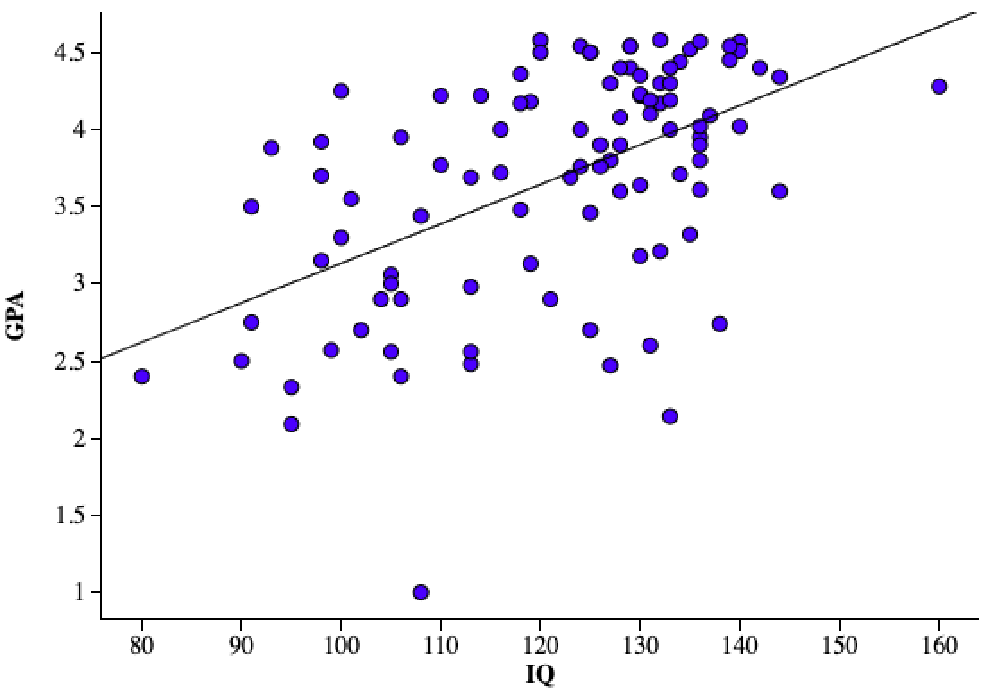
.

Values are then substituted in from the above table:

$$m = \frac{(99)(44927.95)-(12035)(364.76)}{(99)(1485933)-(12035)^2} = 0.0256$$

$$b = \frac{364.76-(0.0256)(12035)}{99} = 0.574$$

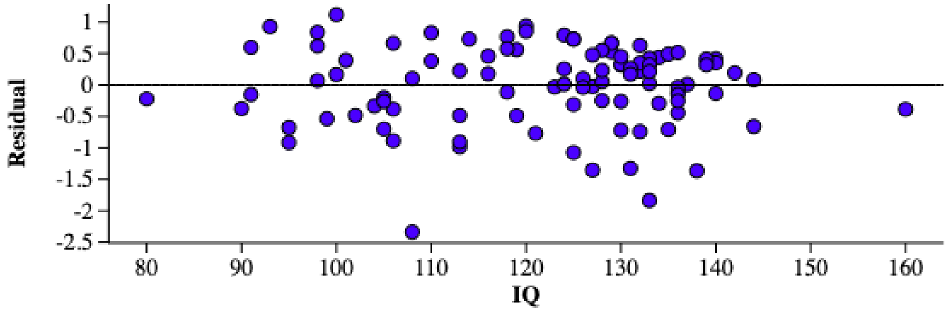
This results in the least-squares regression line of $\hat{y} = 0.0256x + 0.574$. This can be interpreted as, for every point increase in a student's IQ, their GPA is predicted to increase by 0.0256 points. This signifies a positive association between the two variables.



.

Standard Error of Estimate

The residual plot shows how far each actual data point is from the value predicted by the least-squares regression line.



As shown above, there isn't a strong pattern in the residual plot which may be indicative of the data being normally distributed. There are some data points with rather large residuals, implying large uncertainty in the prediction accuracy of the line of best fit. I wanted to see how far student data point was from the line of best fit, so I used the following equation to find the standard error of estimate:

Standard Error of Estimate = $\sqrt{\frac{\Sigma(Residual_i)^2}{n-2}}$

N	$\Sigma Resiudal^2$
99	39.6889

$$\sqrt{\frac{39.6889}{97}} = 0.640$$

[3]Any given data point within this distribution is on average ± 0.640 points away from the line of best fit. A line of best fit predicting a GPA with a margin of error of 0.640 points, is moderately large - a GPA of 4.0 and 3.36 are quite different. This suggests that a linear

[3] The denominator value of n-2 is utlised because of the 2 degrees generated by the 2 parameters in a scatterplot. Thus in order to compensate for residual bias, the denominator is set to n-2. (Stats.Stack Exchange)

.

relationship may not best represent the data or that there is not a strong correlation between the two variables.

Pearson Correlation Coefficient

Finding the Pearson correlation coefficient provides a numerical statistic displaying the strength of the correlation. This measure was the initial test I stated earlier in my hypothesis. I will use the following equations to calculate the correlation coefficient (r-value):

$$r = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_i (x_i - \overline{x})^2} \sqrt{\sum_i (y_i - \overline{y})^2}}$$

$$\overline{x} = \frac{\Sigma x}{N} \quad \overline{y} = \frac{\Sigma y}{N}$$

Using excel, I calculated the following values from my data set:

N	$\overline{x}$	$\overline{y}$	$\Sigma(X_I - \overline{X})(Y_I - \overline{Y})$	$\Sigma(X_I - \overline{X})^2$	$\Sigma(Y_I - \overline{Y})^2$
99	121.57	3.68	585.661	22890.323	54.673

After substituting in values I can solve for r:

$$r = \frac{585.661}{\sqrt{(22890.323)(54.673)}} = 0.5325$$

$$r^2 = 0.284$$

The resulting Pearson Correlation Coefficient reveals a distribution that is moderately positively correlated. The subsequent Coefficient of Determination is 0.284, signifying that the current regression model accurately accounts for 28% of the data values[4]. Because the Correlation Coefficient is relatively low, I will reject the previously stated Alternate Hypothesis (there is a strong correlation between IQ scores and GPA).

Test of Independence

After determining that there was not a particularly strong linear correlation between GPA and IQ scores, I aimed to find out whether the two variables were related or not. To determine

[4] https://blog.minitab.com/en/adventures-in-statistics-2/regression-analysis-how-do-i-interpret-r-squared-and-assess-the-goodness-of-fit

.

this, I decided to apply a test of independence to the distribution. This would reveal if there were any strong trends. Before I could assign a test, I had to first figure out what kind of distribution each variable was. Normally distributed variables can be assigned to a Matched T-Test. If the variables are not normally distributed, an alternate test will be used.

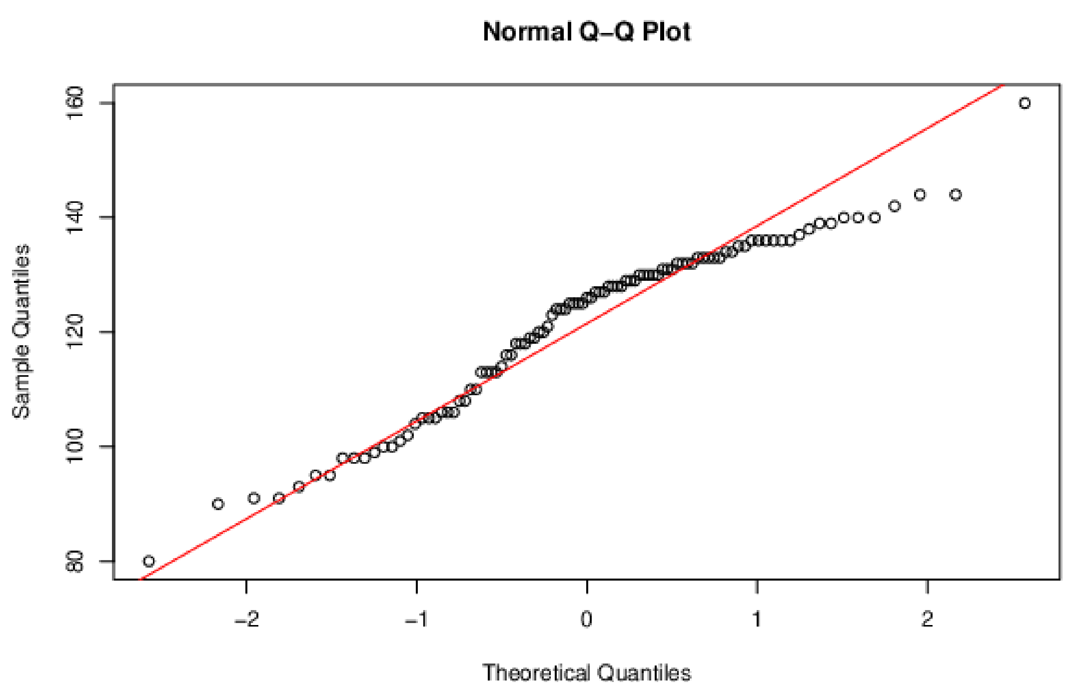
Q-Q Plot

A Quantile-quantile Plot (Q-Q Plot), is a visual tool used to determine the similarity of two different distributions. A Normal Q-Q Plot compares data from a sample distribution against a theoretical normal distribution. Plotting both IQ Scores and GPA on a QQ Plot will help to reveal what kind of distributions these variables follow. On a Q-Q Plot, the x-axis represents each data point's Z score while the y axis shows the actual numerical data point. The Z score for each data point was calculated with the following equation:

$$Z = \frac{x_i - \bar{x}}{\sigma}$$

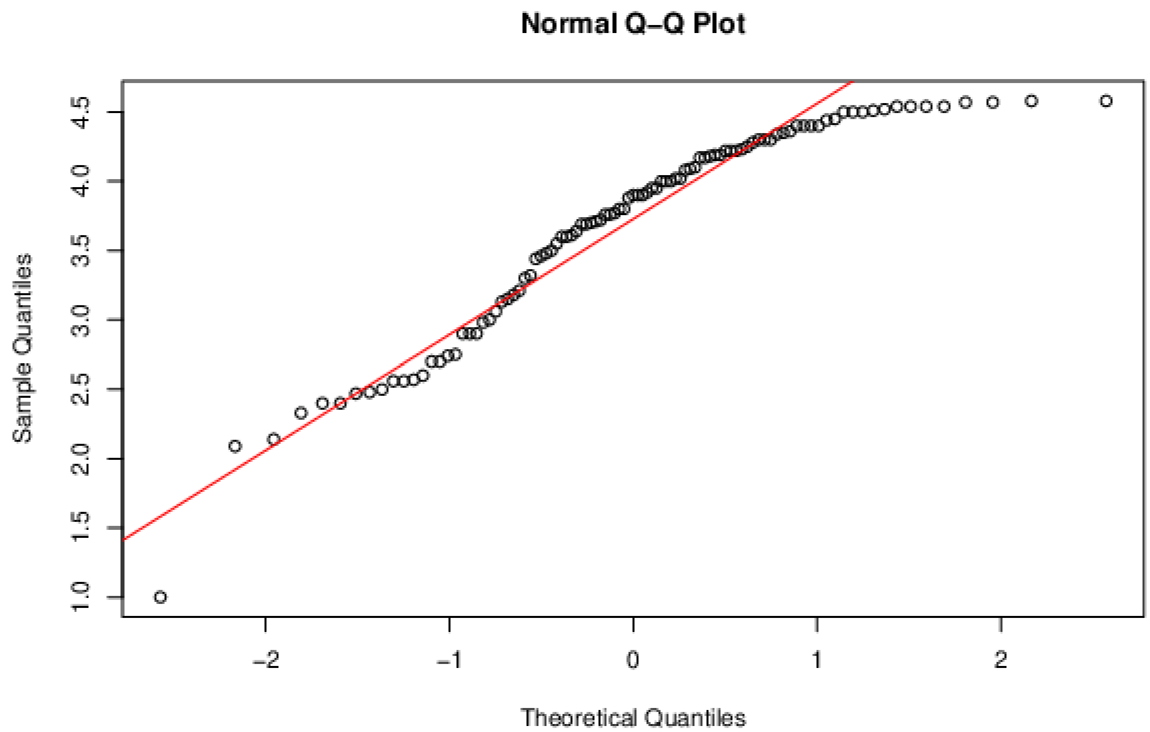
If the variable follows a normal distribution, the plot should resemble a linear line. If the plot varies strongly from linearity, then the variable likely is in the form of another sort of distribution.

IQ scores



.

<u>GPAs</u>



[5]

Both plots show data sets that have large outlier skews and are not representative of a normal distribution. This would suggest that a T-Test is not applicable. Before making this decision, another test will be applied that provides more quantitative analysis.

Shapiro Wilk Test

Just like Q-Q plots, the Shapiro Wilk Test is a method for evaluating how well a sample data set compares to a normal distribution. The test yields a test statistic, W, which on a range between 0 to 1 explains how closely the data set fits normality. The calculated W value is compared to a Wvalue from the Shapiro Wilk probability table.

For the purposes of this investigation, the test is being evaluated with 99 n terms and an α level of 0.05. The α level denotes what percentage confidence the test has incorrectly predicted, whether the distribution achieves normality or not. Because I am using an α level of 0.05, the test has a probability of 0.95 of correctly interpreting the distribution. This is the most

[5] Wessa

.

commonly used α level[6]. This α level, in combination with the number of terms in the sample, 99, is used to find the corresponding W_{table} value. This is the minimum value that the W_{calc} can be to pass the test for normality. Below is the equation to find W_{calc}, or the test statistic:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2},$$ [7]

The a_I value represents the coefficient for the corresponding sample value and is derived from the Shapiro Wilk Coefficient table. The sample values are arranged from smallest to largest, so x_1 is the smallest value in the data set. If $W_{calc} < W_{table}$ then the assumption of normality must be rejected. Because the IQ and GPA sample sets have the same amount of terms, and the same α level, they have the same critical point for the assumption of normality. This value from the probability table is $W = 0.9744$. Due to the large number of calculations required, I used excel for computations.

	W_{calc}	Conclusion
IQ	0.9509	$W_{calc} < W_{table}$
GPA	0.9095	$W_{calc} < W_{table}$

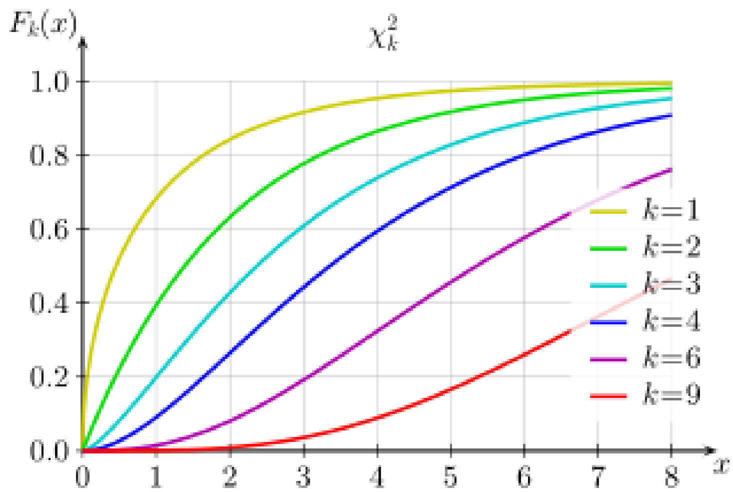
In both distributions, the W_{calc} value is less than the critical value of 0.9744. Therefore the assumption of normality for each distribution must be rejected. The Matched Pairs T-Series test requires the distributions to be normal, therefore an alternate test for independence must be found.

Chi-Squared Test for Independence

Because neither variable follows a normal distribution, I will utilize the Chi-Squared Test for Independence. Like the Shapiro Wilk test, a test statistic is compared to a critical value and then a conclusion is drawn. The critical values for the test, are obtained from the Chi-Squared distribution:

[6] Science Direct
[7] Science Direct

.



[8]

X^2k is the notation for the Chi-Squared distribution. *K* represents the degrees of freedom being applied. Degrees of freedom is the number of independent values or quantities that can be assigned to a statistical distribution[9]. The y-axis, $F_k(x)$, represents the confidence applied to the test. Just like the Shapiro Wilk, an α level of 0.05 represents a test with a 5% risk of drawing an incorrect conclusion. The x-axis represents critical value probabilities. In order to determine what critical value should be used, the degrees of freedom for the test must first be found. The equation to determine what degree of freedom is being used is:

$$df = (r - 1)(c - 1)$$

The variables *r* and *c,* represent the number of rows and columns respectively from a contingency table. A contingency table displays the frequencies for particular combinations of multiple discrete variables. I used the median of the GPA and IQ data sets in order to create “high” and “low” categories. High GPAs and IQs refer to the values in the top 50th percentile of the data set, while low GPAs and IQs refer to the values in the bottom 50th percentile. These are the categories used in the following table:

	High GPA	Low GPA	**Totals**
High IQ			
Low IQ			
Totals			

Upon seeing that the table has two rows of combinations and two columns (the totals rows and columns are not included), it is evident that the degree of freedom is 1. According to

[8] Wikipedia - Chi Squared Distriubtion
[9] Google Dictionary

.

the Chi-Squared distribution, the critical value obtained from 1 degree of freedom with an α level of 0.05, is 3.84. I will now establish my hypothesis:

Null Hypothesis H_0: IQ scores and accumulated GPA values have an independent relationship
Alternate Hypothesis H_1: IQ scores and accumulated GPA values do not have an independent relationship

If the test statistic is less than the critical value of 3.84 then the null hypothesis is accepted and the two variables are deemed to have an independent relationship. Finding the expected count for each combination precedes the computation of the test statistic. The expected count is the frequency that would be expected in a cell, on average, if the variables are independent[10]. These will be the two formulas used:

Expected Count

$$\Sigma_{ij} = \frac{(r_i)(c_j)}{N}$$

Test Statistic

$$\sum_{i, j=1}^{n} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Where

- *i* is the number of rows
- *j* is the number of columns
- *N* is the total sample size
- *n* is the number of combinations of rows and columns
- O is the observed count in a combination cell
- E is the expected count in a combination cell

Essentially the test statistic is the summation of the difference between the observed and expected values in each cell squared over each cell's expected value. I can now input the tabulated values into my contingency table:

---

10 https://support.minitab.com/en-us/minitab/19/help-and-how-to/statistics/tables/how-to/chi-square-test-for-association/interpret-the-results/all-statistics/

.

Observed Value Table	High GPA	Low GPA	**Totals**
High IQ	35	15	50
Low IQ	15	34	49
Totals	50	49	99

Based on the expected count formula, *E* for each cell is found by multiplying the corresponding row and cell totals and then dividing that value over the sample total.

Expected Value Table	High GPA	Low GPA	**Totals**
High IQ	25.25	24.75	50
Low IQ	24.75	24.25	49
Totals	50	49	99

Now that both the observed and expected count values have been determined, the test statistic can be solved.

	High GPA	Low GPA
High IQ	$(O - E)$: 9.75 $(O - E)^2$: 95.0625 $\frac{(O-E)^2}{E}$: 3.76	$(O - E)$: - 9.75 $(O - E)^2$: 95.0625 $\frac{(O-E)^2}{E}$: 3.84
Low IQ	$(O - E)$: - 9.75 $(O - E)^2$: 95.0625 $\frac{(O-E)^2}{E}$: 3.84	$(O - E)$: 9.75 $(O - E)^2$: 95.0625 $\frac{(O-E)^2}{E}$: 3.92

Test Statistic = 3.76 + 3.84 + 3.84 + 3.92
= 15.36

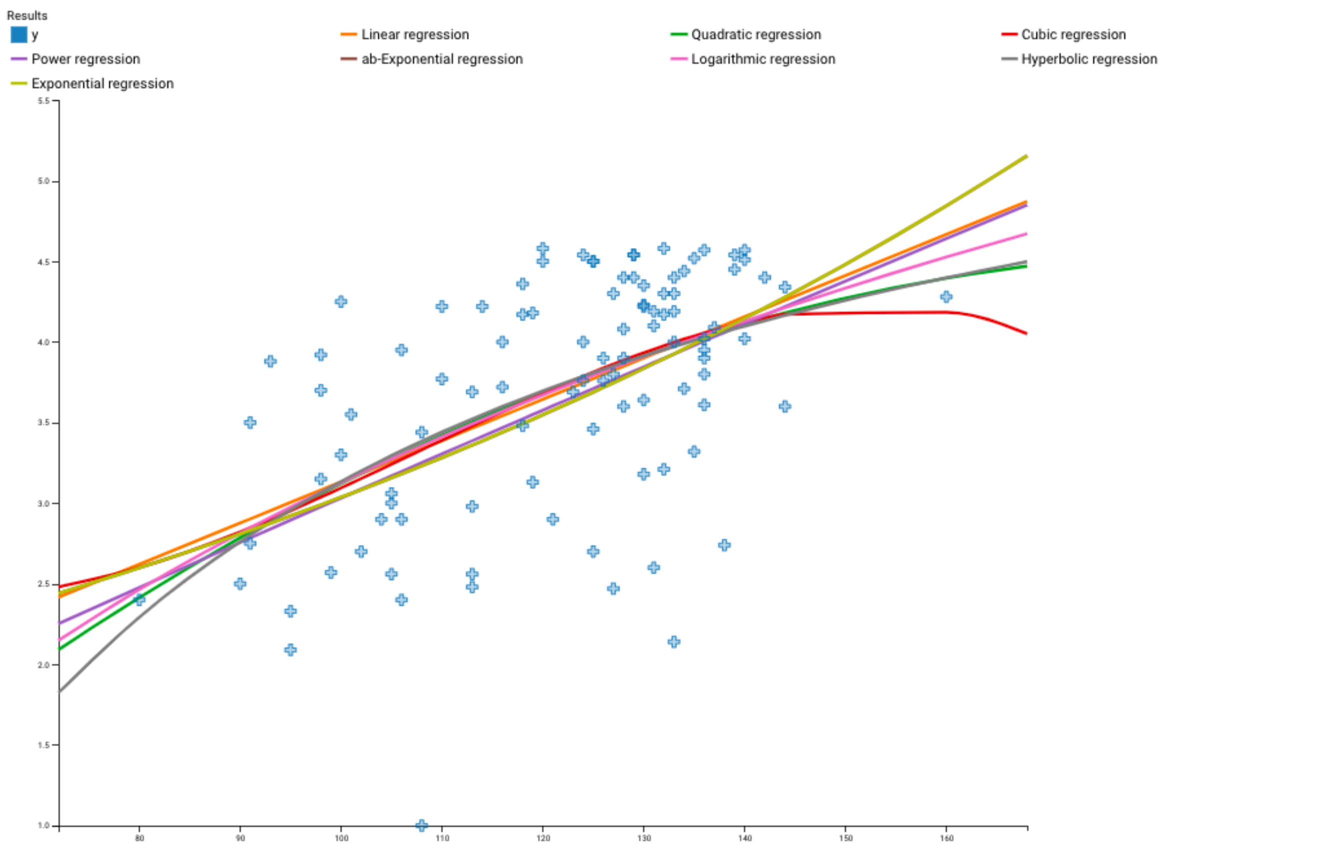
The test statistic is greater than the critical value of 3.84. Thus, the null hypothesis is rejected, and it is concluded that there is a dependent relationship between IQ scores and GPAs. According to the contingency tables, a high or low IQ can be a decent predictor of whether the resulting GPA will be high or low.

.

Evaluation

I found it quite interesting that the data did not have a strong correlation value but had a relatively high test statistic, suggesting that the variables were related but did not correlate highly. This made me wonder if the correlation value could have been stronger if another regression type was used. I used an online calculator to model my data using 8 different regression types: linear, quadratic, cubic, power, ab exponential, logarithmic, hyperbolic, and exponential. The following shows a scatter plot of the original data against the various regression types.

Function Approximation with Regression Analysis



[11]

---

[11] Planet Calc

.

Regression Approximation Table

	Correlation Coefficient	Coefficient of Correlation	Relative Error %	Regression
Linear	0.524	0.274	16.5725	$\hat{y} = 0.0256x + 0.5741$
Quadratic	0.528	0.278	16.3963	$\hat{y} = -0.0002x^2 + 0.067x - 1.824$
Cubic	0.530	0.280	16.4071	$\hat{y} = -0.00001x^3 + 0.002x^2 - 0.13x + 6.2$
Power	0.515	0.265	16.4203	$\hat{y} = 0.0468x^{0.9055}$
AB exponential	0.507	0.258	16.5639	$\hat{y} = 1.3942 * 1.0078^x$
Logarithmic	0.5264	0.271	16.4607	$\hat{y} = -10.5803 + 2.9767 * lnx$
Hyperbolic	0.5256	0.276	16.4697	$\hat{y}$ =6.5038 - 336.8653/x
Exponential	0.507	0.258	16.5639	$\hat{y} = e^{0.3323+0.0078x}$

The analysis shows that no particular regression is significantly stronger than the others. This suggests that an IQ score is only a moderately strong predictor of a GPA. Nonetheless, a positive association remains, and typically stronger IQ scores result in stronger GPAs. All these regressions support the finding of the Chi-Squared Independence Test and maintain that there is a relationship. IQ scores and GPAs seem to serve well together at describing generalities, instead of making accurate, precise predictions.

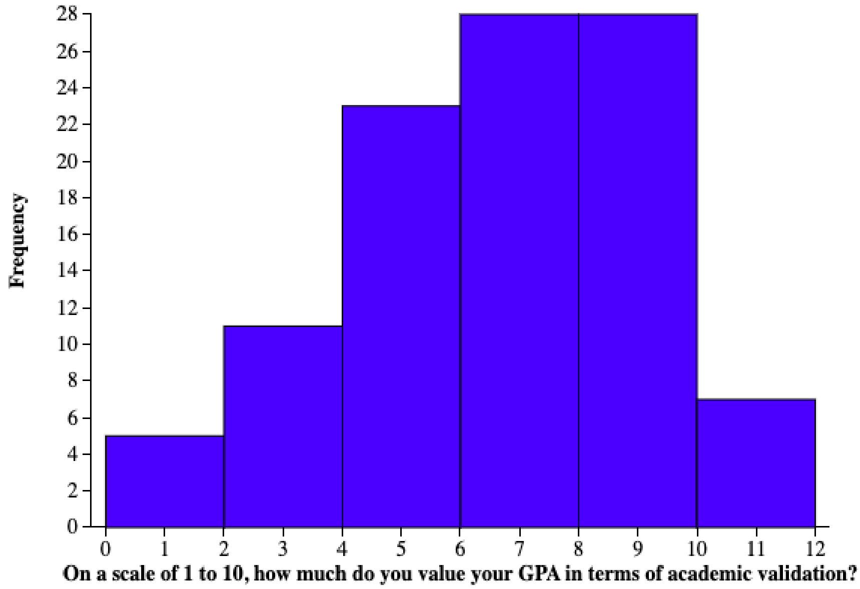
Overall, the relationships found by these methods were reasonable. The lack of a strong correlation and the strength of the relative error can be explained by the many confounding variables surrounding GPAs. Socioeconomic status, family stability, and grade inflation can all affect a student's GPA[12]. These factors may have contributed towards the acceptance of the firstly stated null hypothesis: a low correlation value will be found. However, the strength of the Chi-Squared Test Statistic supports some of the original stigma surrounding GPAs. High scores from IQ tests can quite possibly be the resultant of the same factors that affect GPAs. Students' personal opinions on the tests also could have affected the results. Students who thought more highly of IQ tests may have been more likely to give more effort in the test than a kid who did not have a strong opinion on the test.

According to a survey I included in this investigation, students generally held GPA and IQ in high regard - especially the former. The high value of GPA by students may be the

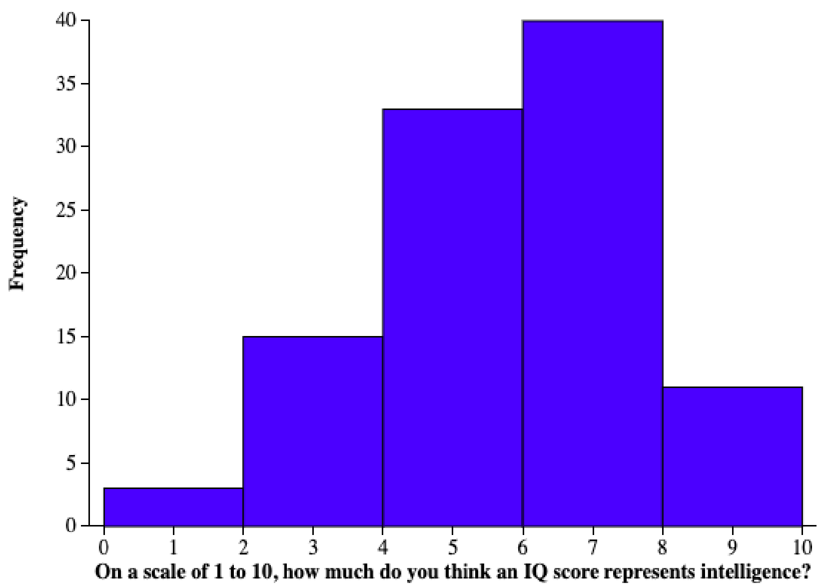
[12]The Classroom

.

resultant of the type of student included in this survey, or it may be because of the familiarity students have with GPA.



[13]



Pertaining to a question such as "Is GPA still a useful measure in school?", the results of this exploration are somewhat inconclusive, as the answer can change depending on the perspective the question is being evaluated on. Overall the study shows that while GPA can be incredibly hard to predict and is probably determined more by work rate than intelligence, trends still do exist within a student population. Also, I am pleased with the method used during the study. Through this exploration, I have been able to expand on my knowledge of normal and

[13] Because of the existence of data points with the value "10" within the top histogram, the range appears to be 0 - 12 even though the survey is conducted on a 1 - 10 scale

.

non-normal distributions. I found it quite interesting to be able to learn and practice the use of independence testing. This investigation helped me to break through the intimidating allure that accompanies distributional analysis.

Extensions

- The data set was not originally compatible with the Chi-Squared test due to its continuous nature. A future experiment may contain categorical data that would not require additional grouping
- More data values may have yielded a stronger correlation value
- Adding in students from other schools would have diversified the data and maybe have produced more interesting findings
- Comparing GPA to other factors (i.e family income) could potentially yield a different type of distribution (i.e normal)
- While additional tests weren't used due to the strength of the Shapiro Wilk Test, additional tests such as a Kolmogorov–Smirnov Test could have been used instead
- This test included mainly International Baccalaureate and AP students, and therefore may have been impacted by undercoverage bias; a test that could test for an entire grade may have yielded a different distribution

.

Works Cited

Awana, Momi. “Factors That Affect GPA.” *The Classroom | Empowering Students in Their College Journey*, 5 Nov. 2021, https://www.theclassroom.com/factors-affect-gpa-5500.html.

Editor, Minitab Blog. “Regression Analysis: How Do I Interpret R-Squared and Assess the Goodness-of-Fit?” *Minitab Blog*, https://blog.minitab.com/en/adventures-in-statistics-2/regression-analysis-how-do-i-interpret-r-squared-and-assess-the-goodness-of-fit.

“Function Approximation with Regression Analysis.” *Planet Calc*, https://planetcalc.com/5992/.

“Interpret All Statistics for Chi-Square Test for Association.” *Minitab*, https://support.minitab.com/en-us/minitab/19/help-and-how-to/statistics/tables/how-to/chi-square-test-for-association/interpret-the-results/all-statistics/.

King, Andrew P. *Statistics for Biomedical Engineers and Scientists*. Academic Press, 2019.

“Least Squares Regression.” *Math Is Fun*, https://www.mathsisfun.com/data/least-squares-regression.html.

Ramachandran, Kandethody M., and Chris P. Tsokos. *Mathematical Statistics with Applications in R*. Academic Press, 2021.

“The Dark History of IQ Tests - Stefan C. Dombrowski.” *TED*, TED-Ed, https://ed.ted.com/lessons/the-dark-history-of-iq-tests-stefan-c-dombrowski.

“Why Divide by N-2 for Residual Standard Errors.” *Cross Validated*, Feb. 2017, https://stats.stackexchange.com/questions/201352/why-divide-by-n-2-for-residual-standard-errors#.

.

Appendix

Raw Data Tables:

Record your score from the IQ Test	Record your GPA	On a scale of 1 to 10, how much do you think an IQ score represents intelligence?	On a scale of 1 to 10, how much do you value your GPA in terms of academic validation?
128	3.6	6	3
137	4.09	3	10
100	4.25	5	7
133	4	6	7
142	4.4	8	8
120	4.58	7	8
134	4.44	7	10
124	4.54	7	8
135	4.52	7	8
133	4.4	5	9
140	4.57	5	4
140	4.51	7	4
120	4.5	5	2
136	4.57	5	6
139	4.54	1	5
125	4.5	7	7
139	4.45	9	10
129	4.54	6	6
132	4.58	7	6
129	4.4	3	10
129	4.54	8	2
106	3.95	6	2
125	4.5	8	8
106	2.4	5	8
95	2.09	4	4
123	3.69	5	4
104	2.9	3	6
95	2.33	6	8

.

91	3.5	5	8
110	3.77	6	7
121	2.9	6	8
80	2.4	5	5
127	3.8	6	3
108	1	8	4
99	2.57	3	1
113	2.48	2	1
113	3.69	3	6
90	2.5	8	8
133	2.14	7	5
124	4	7	6
98	3.15	5	7
126	3.9	6	6
130	3.64	5	4
160	4.28	6	9
93	3.88	3	2
125	3.46	6	4
136	3.95	5	9
113	2.98	8	5
106	2.9	6	8
116	4	5	8
101	3.55	7	5
105	3.06	5	5
128	3.9	4	7
98	3.7	6	7
100	3.3	7	5
130	3.18	7	7
105	2.56	1	5
135	3.32	9	6
116	3.72	5	5
132	4.17	4	7
132	4.3	5	4
128	4.08	3	5
114	4.22	3	9

.

127	4.3	5	1
119	4.18	6	5
136	3.61	5	8
138	2.74	3	7
118	3.48	3	6
118	4.36	6	3
127	2.47	4	10
118	4.17	7	9
124	3.76	2	6
133	4.3	6	8
136	4.02	5	2
130	4.22	7	8
136	3.9	5	7
130	4.23	7	7
98	3.92	5	9
144	3.6	6	3
140	4.02	6	7
110	4.22	4	9
144	4.34	9	7
126	3.76	7	9
91	2.75	5	8
119	3.13	7	7
131	4.19	3	8
133	4.19	8	7
131	2.6	3	9
125	2.7	7	1
105	3	5	5
132	3.21	5	5
134	3.71	5	7
130	4.35	6	9
108	3.44	4	1
136	3.8	5	2
113	2.56	7	5
102	2.7	2	2
131	4.1	7	6

.

128	4.4	1	10
124	3.4	8	4
122	4.11	5	10

.