A Defence of Statistics through Chi Squared Tests

Jonathan Peniket

April 2026

1 Introduction

Statistics are by far the least liked section of A-Level Mathematics, and for good reason. Meaningless tasks about biased spinners and Venn diagram notation become incredibly tedious and pointless, especially when you could be learning a new integration technique or discovering how to predict the motion of flying objects. However, this is due to the lack of application the Statistics module has at A-Level- the extent of which is learning about the cloud coverage in a small town in Scotland. This essay will present an example of why I think statistics are immensely over-hated, through the application of my favourite topic in statistics: Chi-Squared tests of independence. I will apply a Chi-Squared test to groundbreaking, Nobel prize winning Geopolitical-Economic research. Daron Acemoglu, James Robinson, and Simon Johnson's work on 'How Nations Fail' provided a new perspective on why some nations who faced brutal colonization and war have flourished in the modern world, and why others have 'failed'. One of their main reasons for this is the strength of 'inclusive institutions'. To simplify, this means that nations where the government and government agencies are not being exploited are more likely to succeed in stopping their economies and people from being exploited. In this essay, I will test this, using a Chi-Squared test of independence to see if statistics representing the strength of institutions are independent of those representing a nation's economic success. Whilst the bulk of this essay will be a representation of the power of Chi-Squared tests, it's true purpose is to display the beauty of statistics and make an argument in their favour.

2 Compiling the Data to create the Observed Frequencies

2.1 Deciding on measures

2.1.1 Institution Strength

To carry out the test, a measure was needed to define the strength of a nation's 'inclusive institutions'. I selected 'Rule of Law' (WJP Rule of Law index 2025), 'Democracy Index' (Polity 5 Democracy Index), and 'Corruption Perception' (Transparency International Corruption Perception Index 2023) based on the economists' examples given in their work.

2.1.2 Success of a Nation

Those three figures represent the strength of a country's institutions, and therefore are chosen to test against the economic 'success' of a nation. To measure this success, I returned to the original economists to define it, and, therefore chose gross national income per capita (GNI per capita) (UN Human Development Report 2021), and GDP per capita (Our World in Data GDP report).

2.2 Preparing the Data for a test

For the type of Chi-Squared test I will be performing, the data entries need two figures attached to them. In this case, that will mean that each country must only be defined by their success and the strength of their institutions. That means the figures which I just selected are no good unless combined to form single statistics measuring the success of the nation and the strength of that nation's institutions.

2.2.1 Cleaning the data

To create effective statistics to apply to the Chi-squared test, I needed to clean the complicated and over-sized datasets available to the public online. For this, I utilized the Python library 'Pandas' to remove unwanted data and merge the different specific statistics into one large dataset. This included removing unnecessary columns and converting statistics into the right data types, usually strings (text) to floats (numbers).

2.2.2 Normalizing the differing statistics against each other

Firstly, I focused on the 'Institution Strength' data. In the Rule of Law dataset, I converted the decimal (0-1) ratings to scores out of 10, then proceeded similarly onto the other 'Institution Strength' data, converting all scores to a score out of 10 to one decimal place. For example, with the figures for democracy, which were ranging from -10 to 10, I added 10, and divided by two. I totalled the GNI per capita and GDP per capita to create an arbitrary rating for success, later to be converted into a grade.

2.2.3 Defining the 'Institution Score'

To create an 'Institution score', I took the average of the three ratings out of 10. I again used Pandas to sort these into 'bins' of class width 1. This meant that in the dataset, instead of a specific Institution score, each nation was defined by an interval in which their institution score is situated. This is necessary for the Chi-Squared test, as each possibility should contain a reasonably high frequency, which it would not if the institution scores were specific values, rather than intervals.

2.2.4 Choosing the grade boundaries for the success grade

To group the success scores, I statistically sorted them into grades, rather than separating them into bins, due to their variability and lack of a clean, simple upper limit. To assign each of these grades, I calculated every nation's z-scores using this formula:

$$z = \frac{x - \mu}{\sigma}$$

Where X is $GNI + GDP$. The grade boundaries are defined as such:

$$A : z \geq 1$$

$$B : 0.5 \leq z < 1$$

$$C : -0.5 \leq z < 0.5$$

$$D : -1 \leq z < -0.5$$

$$E : z < -1$$

The z-score represents how much further or closer an entry is from the mean than the average distance from the mean. For example, a z-score of 0 would mean that the data point is exactly equal to the mean, whereas a z-score of 1 shows that the data point is exactly one standard deviation above the mean. In this situation, where a score above the mean is desirable, the higher the z-score, the better, and therefore the higher z-scores are designated to better grades of success.

2.3 Entering the data into a Contingency table

To complete a Chi-Squared test on the independence of two variables, a contingency table is needed. This is a table with one variable as the rows (in this case, the success scores), and another variable as the columns (in this case, the institution strength score). The cells (boxes) of the table are filled in by frequency, for example, a cell under column ' $8 - 9$ ' and in row 'Grade B' would contain the number of countries who simultaneously have an institution score of between 8 and 9, and a success grade of a B. Through Pandas, I merged my success grades and institution score bins to one dataset, and then collapsed it into a two-way contingency table. The resulting table was as such:

	Score Class								
Grade	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	All
A	0	1	2	0	1	1	11	3	19
B	0	0	0	0	0	8	2	0	10
C	0	3	4	11	18	4	1	0	41
D	6	5	12	18	5	0	0	0	46
All	6	9	18	29	24	13	14	3	116

3 Performing the Chi-Squared Test

3.1 Defining needed figures

3.1.1 Finding the table of Expected frequencies

For a contingency table, the expected frequency of any one cell is:

$$\frac{row\ total \times column\ total}{grand\ total}$$

Calculating this using the python package NumPy returned this table as the table of expected results:

	Score Class								
Grade	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	Total
A	0.98	1.47	2.95	4.75	3.93	2.13	2.29	0.49	19
B	0.52	0.78	1.55	2.50	2.07	1.12	1.21	0.26	10
C	2.12	3.18	6.36	10.25	8.48	4.59	4.95	1.06	41
D	2.38	3.57	7.14	11.50	9.52	5.16	5.55	1.19	46
Total	6	9	18	29	24	13	14	3	116

However, some of these expected values are below 5, so will not provide an accurate Chi-Squared test, due to the nature of the Chi-Squared distribution only being accurate for values above 5. To fix this, we must combine certain rows or columns that are adjacent, to ensure that every cell has an expected frequency above 5. First, I merged the score classes into 2-5, 5-7, 7-10, to create this table:

Grade	Low (2–5)	Mid (5–7)	High (7–10)	Total
A	5.40	8.68	4.91	19
B	2.85	4.57	2.59	10
C	11.66	18.73	10.60	41
D	13.09	21.02	11.90	46
Total	33	53	30	116

Then after combining rows A, B and C:

Grade	Low (2–5)	Mid (5–7)	High (7–10)	Total
A+B+C	19.91	31.98	18.1	70
D	13.09	21.02	11.90	46
Total	33	53	30	116

3.1.2 Combining our observed frequency table

Combining the observed frequency table in the same way results in this:

Grade	Low (2–5)	Mid (5–7)	High (7–10)	Total
A+B+C	10	30	30	70
D	23	23	0	46
Total	33	53	30	116

3.1.3 Degrees of Freedom

Degrees of freedom are the number of entries that can freely vary without altering the mean or standard deviation. For example, if you have 5 numbers that have a mean of 10, 4 of those could be picked at random to be literally any number, however, once those 4 have been picked, there is only one possible value of the 5th. In Chi-Squared tests, degrees of freedom define the distribution that is selected out of the 'Chi-Squared Family of Distributions', and so we must calculate it here. Degrees of Freedom are represented by the Ancient Greek letter nu (ν). In the case of a two-way frequency table, $\nu = (r-1)(c-1)$ where c is the number of columns and r is the number of rows. As our table has 2 rows and 3 columns, $\nu = (2-1)(3-1) = 1 \times 2 = 2 \therefore \nu = 2$. This means we will be testing our 'Goodness of fit' value against the χ^2_2 distribution.

3.2 Calculating the Goodness of Fit value

Goodness of fit is represented by χ^2, and is calculated by this formula:

$$\sum \frac{O_x^2}{E_x} - N$$

Where O_x is the observed frequency, E_x is the expected frequency, and N is the grand total. This table shows all the O_x values and their corresponding E_x values.

O_x	E_x
10	19.91
23	13.09
30	31.98
23	21.02
30	18.1
0	11.9

Table 1: Observed and Expected Frequencies

This results in:

$$\chi^2 = 142.87 - 116 = 26.87$$

To use this in a test, the critical value of the χ_2^2 distribution must be found. For this test, the level of significance will be 1%, so the value of $\chi_2^2(0.01)$ must be found. This value is 9.210.

3.3 Summary: the full test

H_0: A nation's economic success is independent of the strength of it's inclusive institutions.
H_1: A nation's economic success is not independent of the strength of it's inclusive institutions.

$\nu = 2 \Rightarrow$ to reject H_0, the goodness of fit value must exceed $\chi_2^2(0.01) \approx 9.210$

$$\chi^2 = \sum \frac{O_x^2}{E_x} - N \approx 26.87$$

$26.87 > 9.210$
$\therefore$ Reject H_0
There is sufficient evidence to suggest that a nation's economic success is not independent of the strength of it's inclusive institutions.

4 Conclusions

4.1 Conclusion

In simple terms, the Chi Squared test showed that, if economic success and institutions were independent, it is less than 1% likely that the statistics I collected could have been what they are. The intention of this project was to show that statistics are an incredibly valuable, powerful section of maths, that can show things about the real world. Whilst the conclusions of this test are quite easy to assume, these tests can be used for much more bizarre and seemingly illogical situations. That is why Chi Squared tests of independence are my favourite topic in maths, as they can reveal links between seemingly separate things, bridging the gap between discussion and science. As someone hugely interested in politics, economics, philosophy, history, and many other social disciplines, statistics and especially chi squared tests are the best way to gain a fresh, evidential, and powerful perspective on matters which are usually kept to words and theories.

4.2 A final argument for statistics

I sincerely hope this Chi-Squared test on modern, exciting, impactful economic work has displayed that statistics is so much more than calculating the probability that a random boy called Harry will roll a 2 on a dice. Statistics penetrates every area of study, from English to Physics, and from Philosophy to Sports science, and is therefore the most impactful area of maths on the majority of people's lives. It can provide the only truly, fully, empirical evidence of arguments, it can make markets fail or boom and make people live or die. This is just one example of the power that they hold: the possibilities are endless.