

Comparison of Perceived University Experience Satisfaction Between Mature and Younger Students

Abstract

A study was conducted to investigate the potential difference in perceived university experience satisfaction between mature university students (aged 45 to 62) and younger students (aged 17 to 20). Based on a reliable test, numeric satisfaction scores were assigned to a sample of 10 individuals from each of these two subsections of the student population. Both of these samples were then analysed and compared to identify whether there was a significant difference between them. Grouped frequency, skewness and kurtosis analysis were conducted on both samples to confirm that they fit the normal distribution, with both skewness and kurtosis being within the requisite value of ± 2 for both samples. A Shapiro-Wilk normality test was carried out, giving a W value of 0.96, corresponding to a p value of 0.81. This enabled confirmation of the auxiliary hypothesis that the data could be accurately described as following the normal distribution pattern.

A null hypothesis was developed: that there was no significant difference between the two groups. The alternative hypothesis was, therefore, that there was a significant difference between the two groups. A Student's T-test was conducted, finding $t_{\text{obt}} = 4.26$, allowing the alternative hypothesis to be confirmed at the $p = 0.1$, $p = 0.02$ and the $p = 0.01$ significance levels. The data was also examined to identify potential outlier values and other possible future methods of analysis were discussed. Ultimately, it was found that, in comparison to younger students, mature students can reliably be said to exhibit a different degree of satisfaction with the university experience than younger students.

Introduction

Before engaging in any form of scientific discourse on the relationships between different variables and their potential causes, it is first essential to reaffirm the key truth behind all experimentation: correlation is not equal to causation. It would be extremely easy, after a quick look at the two clearly differentiable student data sets and jump to the assumption that the simple fact of being a mature student in some way raises the satisfaction level above that of a younger student. In the same way, a scientist with no knowledge of the effects of smoking might assume that people with yellower (tobacco stained) teeth were inherently more susceptible to lung cancer.

It was in order to avoid such erroneous logical missteps that Bradford Hill first came up with what are now known as *Hill's criteria for causation*, in 1965 [1]. These criteria limit the possibility of false conclusions from data which, although correlated, does not possess a direct causal link. Of these, the consistency, coherence and analogy criteria, most directly concern the relatibility of an experimental hypothesis to pre-established evidence and findings. In deference to these criteria, some of the experimental history and context behind this study must be at least briefly examined.

A 2007 paper by Ramsay, Jones and Barker identified the importance of adjustment and support factors, which vary extremely between institutions, on the relative satisfaction levels of young and mature students [2]. In addition, there are other factors to be taken into account, such as the difference in relative socio-economic class between young and mature students [3]. Also of relevance is the evidence of a significant increase in the non-continuation rates of mature as opposed to young students, clearly evidenced in HESA performance indicators [4].

None of these aspects can be ignored when formulating conclusions on the subject of age-dependent student satisfaction. However, causation issues and related research only represent one side of the coin when it comes to confirming or denying the experimental hypothesis. Even more crucial is the question of whether or not any claimed relationship between variables is truly a valid representative of correlation.

It would be very possible for this document to have rejected the null hypothesis, on the basis of a t-test and yet come to an invalid conclusion. This is because the t-test is, along with all other parametric tests, based on a set of assumptions about the nature of the data-set it is acting upon, which allow it to make a stronger, more significant statement about the correlation within the data [5]. The key basis of these assumptions is that the data set obeys the normal distribution pattern. Also known as the Gaussian distribution, first observed by Carl Friedrich Gauss in 1809, it is the pattern to which the measurements of almost all natural stochastic processes converge, including human properties such as height distributions, as noticed by Quetelet in 1835 [6], [7].

Due to it being a fundamental assumption of all parametric tests, with the chance of invalid conclusions in the case of a non-normal distribution, it was understood to be of key importance to first confirm that the student data sets did indeed conform to this pattern. This confirmation may be obtained in multiple ways, even simply through direct visual observation but also by frequentist approaches, such as the Shapiro-Wilk test [8].

Methodology

A grouped frequency tally chart was created for each data set. This allowed an initial understanding and impression of the distribution pattern to be built up. It was also an essential step towards building up a histogram for each data set. These histograms could then be contrasted with a plot of the approximate normal distribution, which was given a matching mean and standard deviation to the data set it was compared with.

The direct qualitative observation of the “goodness of fit” of the data set to the Gaussian distribution line can be thought of as a graphical method of determining whether the data set could be described as normally distributed and hence whether or not the application of parametric tests would be justified.

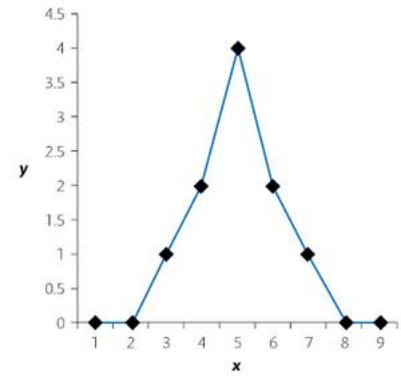
More quantitative methods were then applied, first through comparison of the mean and median, which in a Gaussian distribution would be exactly representative of the 50th percentile. The level of difference between these two measures of average represents the level of distortion, or skewness, of

the distribution pattern. The skewness was then measured explicitly, using the excel descriptive statistics feature, along with the kurtosis value, which essentially represents the “peakedness” of the distribution (see Figure 1) [9]. The cut-off point for both the skewness and kurtosis values, beyond which the distribution may no longer be called Gaussian, was taken as ± 2 , in accordance with accepted convention [10] [11].

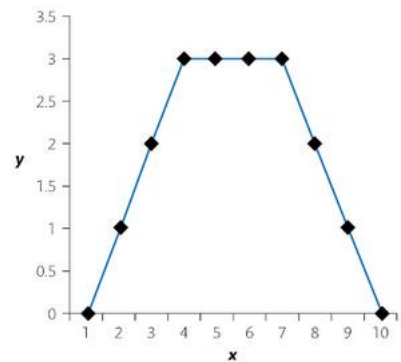
On the basis of this analysis, an auxiliary hypothesis was then formed: that the data was normally distributed. The confirmation or rejection of this hypothesis was then based upon a frequentist approach, using the Shapiro-Wilk test. Other possible options, such as Pearson’s chi-squared normality test or the Kolmogorov–Smirnov test, would provide alternative methods for assessing how closely the real data set matched with the values that would be expected if it perfectly followed a normal distribution pattern. However, the Shapiro-Wilk test is known to be very sensitive and superior in general when it comes to small data samples, as in this case [10].

It should also be noted that, for small sample sizes, no test can claim to provide a fool-proof guarantee of positive evidence for a normally distributed data set. Rather, each test simply reveals whether the data set is *not* normally distributed [11]. It is for this reason that the data is analysed from multiple different perspectives: qualitative analysis; skewness and kurtosis and the Shapiro-Wilk test, shown in Equation 1.

$$W = \frac{(\sum_{i=1}^n a_i(x_{(n+1-i)} - x_i))^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$



Leptokurtic distribution



Platykurtic distribution

Figure 1: Distributions with a higher kurtosis value are referred to as leptokurtic, whilst low kurtosis distributions are platykurtic [9].

Equation 1 [12]

Here a_i represent the values in a Shapiro-Wilk table for the number of elements in the data-set, x -subscript represents each value by its position in the ordered database and W is the Shapiro-Wilk value which approaches 1 as the data converges towards the normal distribution.

Having confirmed a normal distribution pattern, a null hypothesis was formed, stating that there was no significant difference between the two groups. It was decided to use a two tailed t test, so as to minimise the chance of type 1 error (which is generally regarded as more serious than type 2 error). Hence, the alternative hypothesis was that there was indeed a significant difference between the two groups, which would identify a correlation between age perceived satisfaction and student age.

These hypotheses were tested at the 0.1, 0.02 and 0.01 significance levels, which, if all passed, would indicate a less than 1% probability of the null hypothesis being true. On the other hand, if all were failed, this would indicate a probability of greater than 10% that any apparent differences in observed results was a matter of chance, and the random fluctuations when taking a sample from a population.

Lastly, potential outliers were examined using the Z score, i.e. by rephrasing the suspected outlying scores in units of standard deviation (σ) from the mean value. It is a well-known rule of thumb that outlying values can be identified as any point which lies outside of $\pm 3\sigma$ from the mean, as this would imply it lay outside of 99.7% of the data within the normal distribution.

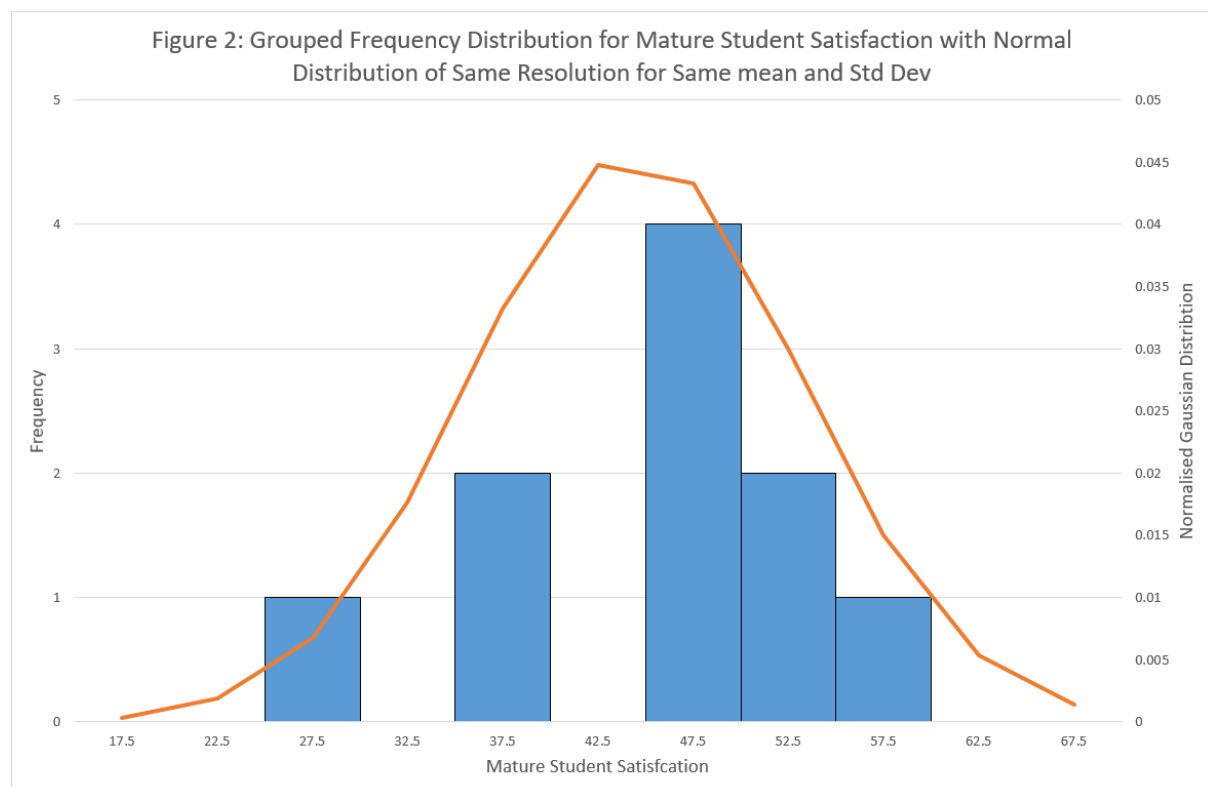
Results

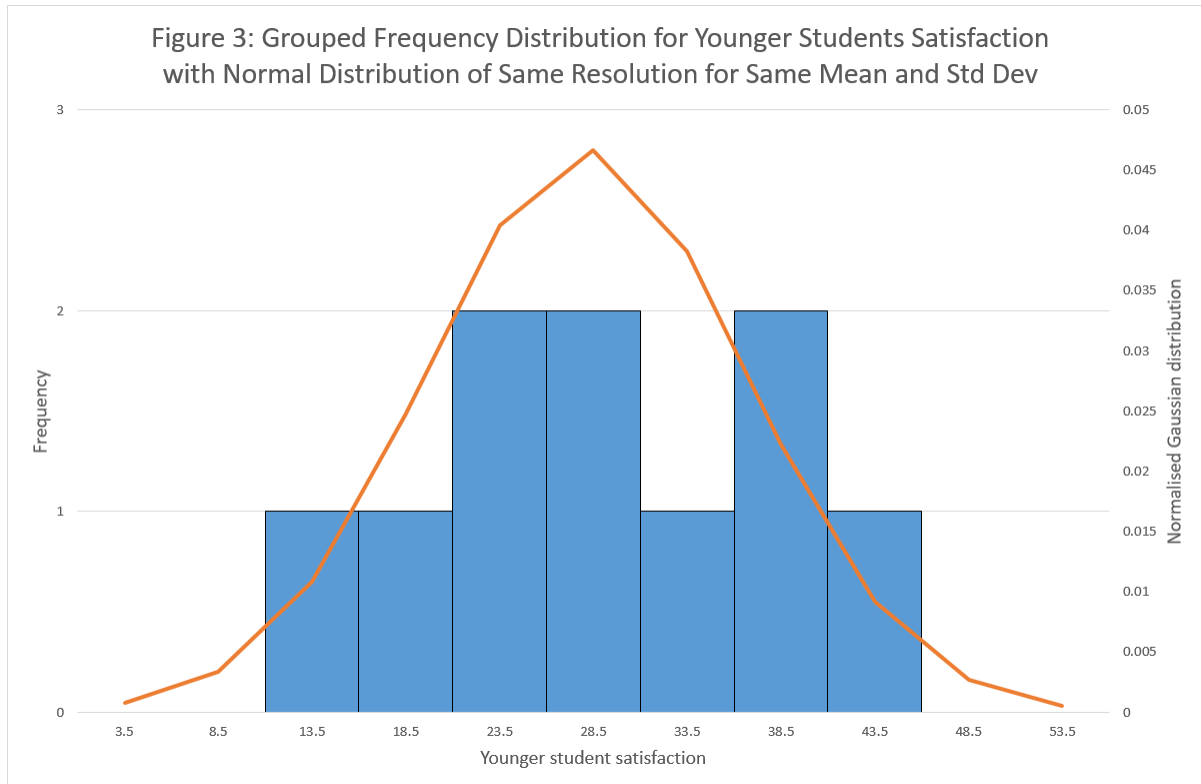
The initial grouped frequency tallies immediately revealed the apparently Gaussian nature of the two distributions (See Table 1).

Table 1 – Grouped frequency distribution for both datasets

Mature Students			Younger Students		
25-29		1	11-15		1
30-34			16-20		1
35-39		2	21-25		2
40-44		0	26-30		2
45-49		4	31-35		1
50-54		2	35-40		2
55-59		1	41-45		1

The data for each frequency table was then plotted as a histogram, with the equivalent Gaussian curve (constructed using the same number of values) included so as to allow easier qualitative comparison (see Figure 2 and Figure 3).





As can be seen quite clearly, despite the extremely low sample size, which makes the distribution appear extremely low resolution, the data set histograms already appear to have a reasonable fit beneath the normal distribution bell curve.

In order to generate this bell-curve it was necessary to calculate the mean and standard deviation for each data set. These were calculated simultaneously with the 50th percentile (median), skewness and kurtosis values using the excel 'descriptive statistics' feature as shown in table 2.

Table 2: Descriptive statistics for both datasets			
Mature Students		Younger Students	
Mean	44.5	Mean	28.1
Standard Error	2.745704	Standard Error	2.701645589
Median	46	Median	26.5
Mode	46	Mode	#N/A
Standard Deviation	8.682678	Standard Deviation	8.543353492
Sample Variance	75.38889	Sample Variance	72.98888889
Kurtosis	1.979848	Kurtosis	1.181902204
Skewness	-1.2629	Skewness	0.053883295
Range	30	Range	26
Minimum	25	Minimum	15
Maximum	55	Maximum	41
Sum	445	Sum	281
Count	10	Count	10

As can be seen both the skewness and kurtosis factors are within the requisite range of ± 2.0 in order for our distribution to be classified as normal. Furthermore, the difference between mean and median for both samples, while indicating some skewness, is not large in comparison to the mean or standard deviation of either data set.

Performing the Shapiro-Wilk test on the young student sample provides a W value of 0.96. As this falls between the p values of 0.9 and 0.5 on the Shapiro-Wilk p value table, we can by linear approximation using the two adjacent W values, estimate the p value this corresponds to as 0.81. Hence, we may accept our auxiliary hypothesis that the data is normally distributed.

Upon running Student's t-test, a value of $t_{\text{obt}} = 4.26$ was generated.

The t_{crit} values for the 0.1, 0.02 and 0.01 significance levels respectively were: 1.73, 2.55 and 2.88 respectively. As t_{obt} was significantly higher than all of these, we can reject the null hypothesis at the 0.01 significance level, accepting the alternative hypothesis that there is a significant difference between the two groups of students.

The most deviated value across both data sets was the lowest student satisfaction of the mature student group. Converted to a Z value in standard deviation units it has a magnitude of 2.25σ , which is less than the 3σ necessary to declare it an outlier. As such, no outliers both data sets may be declared free of outlying values.

Discussion

Ultimately, we were able to confirm that our data was normally distributed, with a Shapiro-Wilk p-value of 0.81. We were then able to use parametric analysis in the form of a Student's T-test, to reject the null hypothesis, that there was no significant difference between the satisfaction of the two types of student, at the 0.01 level. Hence we may claim that there is in fact a significant difference in satisfaction.

However, there are many aspects of this experiment left unexplored. Firstly, a repeated experiment with significantly larger data sets would be extremely beneficial, not only would this reduce random error but also, as mentioned, the normal nature of a distribution cannot easily be verified for small sample sizes. A related logical expansion might be to increase the number of ranges of students, perhaps in to analyse in subgroups of 3 or 4 years, then compare the multiple, more finely graduated age groups of students against each other using the ANOVA test. Alternatively, although technically challenging, it would be most illustrative to repeat the satisfaction test on students who re-entered higher education later in life. This would enable one to use a paired t-test, as it would be possible to compare the same group of students to their younger selves, after a long period of time. Lastly, although in this case it was possible to reject the null hypothesis using a two tailed test, a one tailed test, although more prone to type one error, might have made the significance of the difference clearer, whilst also enabling the directional statement "Mature students are more satisfied than younger students".

Works Cited

- [1] A. B. Hill, "The Environment and Disease: Association or Causation?," *Proceedings of the Royal Society of Medicine*, vol. 58, no. 5, pp. 295-300, 1965.
- [2] S. Ramsay, E. Jones and M. Barker, "Relationship between adjustment and support types: Young and mature-aged local and international first year university students," *Higher Education*, vol. 54, no. 2, pp. 247-266, 2007.
- [3] D. McVitty and K. Morris, "Never Too Late To Learn: Mature students in higher education," National Union of Students, 2012.
- [4] HESA, "Non-continuation: UK Performance Indicators 2016/17," Higher Education Statistics Agency, 8 March 2018. [Online]. Available: <https://www.hesa.ac.uk/news/08-03-2018/non-continuation-tables>. [Accessed 04 12 2018].
- [5] W. S. (. Gosset, "The Probable Error of a Mean," *Biometrika*, vol. 6, no. 1, pp. 1-25, 1908.
- [6] C. F. Gauss, Theory of the motion of the heavenly bodies moving about the sun in conic sections: a translation of Gauss's "Theoria motus." With an appendix, Boston: Little, Brown and Company, 1857.
- [7] A. Quetelet, Sur l'homme et le développement de ses facultés, ou Essai de physique sociale, Paris, 1835.
- [8] S. S. Shapiro and M. B. Wilk, "An Analysis of Variance Test for Normality (Complete Samples)," *Biometrika*, vol. 52, no. 3, pp. 591-611, 1965.
- [9] D. Holmes, P. Moody, D. Dine and L. Trueman, Research methods for the biosciences, Oxford: Oxford University Press, 2017.
- [10] L. Trueman, "Checking that data is normally distributed using Excel," Oxford University Press, 17 November 2016. [Online]. Available: https://www.youtube.com/watch?time_continue=60&v=EG8AF2B_dps. [Accessed 02 12 2018].
- [11] D. George and P. Mallery, SPSS for Windows Step by Step: A Simple Guide and Reference 17.0 Update, 10 ed., Boston: Pearson, 2010.
- [12] N. Razali and Y. B. Wah, "Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests," *Journal of Statistical Modeling and Analytics*, vol. 2, no. 1, pp. 21-33, 2011.
- [13] A. Sen and M. Srivastava, "5.2.2 Tests For Normality," in *Regression Analysis: Theory, Methos and Applications*, New York, Springer, 1990, p. 105.