

Beyond Traditional Quantitative Research Methods: Integrating G*Power for Sample Size Determination and WebPower for Multivariate Normality Assessment

*Melangkaui Kaedah Penyelidikan Kuantitatif Tradisional: Mengintegrasikan G*Power untuk Penentuan Saiz Sampel dan WebPower untuk Penilaian Kenormalan Multivariat*

Teh Suhaila Tajuddin¹, Izlawanie Muhammad²

¹ Selangor Islamic University, Faculty of Economics, Accountancy and Management, UIS, Malaysia

² Universiti Sains Islam Malaysia, Faculty of Economics and Muamalat, USIM, Malaysia

Email: tehsuhaila@uis.edu.my, izlawanie@usim.edu.my

ABSTRACT

Determining an adequate sample size and assessing statistical assumptions are fundamental to methodological rigour in quantitative research. Despite their importance, social science studies may still rely on conventional rules of thumb rather than evidence-based procedures. This paper demonstrates an integrated methodological workflow using G*Power for a priori sample size determination and WebPower for post-collection assessment of multivariate normality, illustrated through a tax morale study involving Malaysian salaried taxpayers. Based on five direct predictors specified in the structural model, the G*Power analysis indicated a minimum sample size of 92 respondents. After data screening, 319 usable responses were available for analysis, substantially exceeding this power-based benchmark. Multivariate normality was subsequently assessed using WebPower through Mardia's multivariate skewness and kurtosis statistics. The results indicated significant multivariate skewness ($\beta = 9.46$, $p < .01$) and multivariate kurtosis ($\beta = 107.44$, $p < .01$), demonstrating that the observed data departed from multivariate normality. These distributional findings provide additional methodological support for the use of variance-based PLS-SEM, while recognising that the choice of estimation approach should also reflect the study's analytical objectives and model characteristics. By integrating a priori power analysis with post-collection distributional assessment, this paper offers a practical and transparent framework for strengthening the planning, justification, and reporting of quantitative research in taxation and the broader social sciences.

Keywords: G*Power analysis; sample size; WebPower; quantitative approach; normality

ABSTRAK

Penentuan saiz sampel yang mencukupi dan penilaian andaian statistik merupakan asas kepada keteguhan metodologi dalam penyelidikan kuantitatif. Walaupun kedua-duanya penting, kajian sains sosial masih boleh bergantung pada peraturan praktikal konvensional berbanding prosedur statistik berasaskan bukti. Kajian ini menunjukkan aliran kerja metodologi bersepadu dengan menggunakan G*Power untuk penentuan saiz sampel secara a priori dan WebPower untuk penilaian kenormalan multivariat selepas pengumpulan data, berasaskan kajian moral cukai yang melibatkan pembayar cukai



bergaji di Malaysia. Berdasarkan lima pemboleh ubah peramal yang mempunyai hubungan langsung dalam model struktur, analisis G*Power menunjukkan saiz sampel minimum sebanyak 92 responden. Selepas proses penyaringan data, sebanyak 319 respons yang boleh digunakan tersedia untuk analisis, jauh melebihi penanda aras berasaskan kuasa statistik tersebut. Kenormalan multivariat kemudiannya dinilai menggunakan WebPower melalui statistik kepencongan dan kurtosis multivariat Mardia. Hasil analisis menunjukkan kepencongan multivariat ($\beta = 9.46$, $p < .01$) dan kurtosis multivariat ($\beta = 107.44$, $p < .01$) yang signifikan, menunjukkan bahawa data yang diperhatikan menyimpang daripada kenormalan multivariat. Ciri taburan ini memberikan sokongan metodologi tambahan terhadap penggunaan PLS-SEM berasaskan varians, dengan mengambil kira bahawa pemilihan pendekatan penganggaran juga perlu selaras dengan objektif analisis dan ciri model kajian. Melalui integrasi analisis kuasa secara a priori dengan penilaian taburan selepas pengumpulan data, kajian ini menawarkan kerangka kerja yang praktikal dan telus untuk memperkukuh perancangan, justifikasi, dan pelaporan penyelidikan kuantitatif dalam bidang percukaian dan sains sosial secara lebih meluas.

Kata kunci: Analisis G*Power; saiz sampel; WebPower; pendekatan kuantitatif; kenormalan

INTRODUCTION

Sample size determination and the assessment of statistical assumptions are foundational to the validity and reliability of quantitative research instruments. An inadequately justified sample size may result in insufficient statistical power to detect meaningful effects, whereas an unnecessarily large sample may lead to inefficient use of research resources without corresponding analytical benefits (Faul et al., 2009). Determining an appropriate sample size through statistical power analysis, however, requires researchers to consider several interrelated parameters and can be technically demanding (Kang, 2021). Consequently, many social science researchers continue to rely on conventional rules of thumb rather than evidence-based statistical procedures grounded in power analysis (Memon et al., 2020). Such practices may weaken the methodological justification of quantitative studies, particularly when complex multivariate techniques are employed.

Two methodological considerations are particularly important in addressing this concern. The first is the a priori determination of a minimum adequate sample size capable of detecting anticipated effects at an acceptable level of statistical power. Rather than determining sample size primarily on the basis of convenience, precedent, or general rules of thumb, power analysis enables the required sample to be justified using explicit statistical parameters. The second consideration concerns the assessment of the distributional characteristics of the observed data, particularly multivariate normality, before model estimation. Evaluating these characteristics provides important diagnostic evidence for determining the suitability of alternative estimation approaches. Together, sample size determination and assumption assessment provide a stronger methodological foundation for subsequent statistical analysis.

This paper addresses these considerations through the complementary application of G*Power and the WebPower online platform. G*Power is employed at the research-design stage to conduct a priori power analysis and establish an empirically justified minimum sample size based on the predictors specified in the research model. Following data collection, WebPower is used to assess multivariate normality through Mardia's multivariate skewness and kurtosis statistics. Rather than treating non-normality as the sole criterion for selecting an estimation technique, the results provide diagnostic evidence that can be considered alongside the research objectives and model characteristics when determining an appropriate analytical approach. The integration of G*Power and WebPower therefore establishes a systematic methodological workflow that links sample planning, distributional assessment, and analytical decision-making through statistical evidence.

The tax morale context provides an appropriate empirical setting for demonstrating this methodological workflow. Tax morale generally refers to individuals' intrinsic motivation to fulfil their tax obligations (Alm & Torgler, 2004; Emiljan, 2022; Tajuddin, 2025; Torgler, 2007). Research in this area frequently relies on survey-based measures and multi-item constructs to examine the psychological, social, and institutional factors associated with taxpayers' attitudes and behaviour (Shaharuddin, 2017; Hamid, 2013). PLS-SEM has also been applied in behavioural tax research involving complex relationships among latent constructs (Pacaldo et al., 2020; Hayat et al., 2022). Although PLS-SEM is comparatively flexible regarding sample size and distributional assumptions, its use does not eliminate

the need to justify sample adequacy or examine the distributional characteristics of the data (Hair et al., 2019). Accordingly, the tax morale dataset used in this study provides a practical context for demonstrating how these methodological decisions can be supported systematically.

Against this background, this study has three primary objectives. First, it demonstrates the application of G*Power for a priori sample size determination and statistical power analysis. Second, it illustrates the use of WebPower to assess multivariate normality through Mardia's multivariate skewness and kurtosis statistics and demonstrates how the resulting evidence can inform the selection of an appropriate estimation approach. Third, it integrates these procedures within the broader methodological framework of sample size adequacy and assumption assessment, emphasising that both should be incorporated systematically into quantitative research design rather than treated merely as post hoc statistical checks. In doing so, the study provides a practical and transparent methodological workflow for strengthening the planning, justification, and reporting of quantitative research in taxation and the broader social sciences. The remainder of the paper is organised as follows. The literature review examines sample size determination, statistical power analysis, multivariate normality assessment, and their relevance to PLS-SEM research. The methodology section describes the sampling procedures and the applications of G*Power and WebPower, followed by the presentation and discussion of the empirical results. The final section concludes the paper by highlighting its methodological implications, limitations, and directions for future research.

LITERATURE REVIEW: THEORETICAL FOUNDATIONS OF POWER ANALYSIS

Sample Size Determination in Quantitative Social Science Research

Sample size determination is a consequential component of quantitative research design because it directly affects a study's ability to detect meaningful effects and estimate relationships with adequate precision. A sample that is too small may produce insufficient statistical power, whereas an unnecessarily large sample may consume additional resources without proportionate analytical benefit (Faul et al., 2009). Nevertheless, sample sizes in social science research have often been determined using heuristic practices, such as round-number targets, population percentages, or sample sizes adopted from earlier studies. Memon et al. (2020) caution that such practices lack an explicit statistical basis and may result in either underpowered studies or unnecessary data collection. The issue becomes particularly important in multivariate research, including SEM, where sample-size requirements are influenced by the structure of the model and the magnitude of the relationships to be detected.

Statistical Power Analysis and G*Power

Statistical power analysis provides an evidence-based alternative to rule-of-thumb sample size determination by linking four interrelated parameters: sample size, effect size, significance level (α), and statistical power ($1 - \beta$) (Faul et al., 2009). G*Power is widely used for this purpose because it is freely available, supports a broad range of statistical tests, and does not require specialised programming knowledge (Kang, 2021). In an a priori power analysis, researchers specify the expected effect size, significance level, desired statistical power, and number of predictors before data collection, and G*Power estimates the minimum sample size required under those assumptions. In PLS-SEM research, the analysis can be based on the maximum number of direct predictors associated with an endogenous construct in the structural model (Hair et al., 2019). This approach provides a transparent, model-informed benchmark for sample-size adequacy rather than a generic threshold applied irrespective of the study design.

Sample Size Adequacy in PLS-SEM

The relationship between sample size and PLS-SEM requires careful interpretation because the method's ability to accommodate relatively small samples is sometimes mistaken for an absence of sample-size requirements. Earlier PLS-SEM applications frequently relied on the '10-times rule', which recommends a minimum sample based on the largest number of structural paths directed at an endogenous construct (Barclay et al., 1995). Although simple to apply, this rule does not explicitly account for effect size, significance level, desired statistical power, or other features that influence the ability to detect structural relationships. More recent methodological guidance therefore recommends power-based or PLS-SEM-

specific approaches for establishing minimum sample-size requirements (Hair et al., 2019; Kock & Hadaya, 2018).

A power-based minimum should also be distinguished from the sample ultimately achieved in a study. Meeting the minimum threshold indicates that the specified effect can be detected with the planned level of statistical power under the assumptions entered into the analysis; it does not imply that additional observations have no value. Larger samples can improve estimation precision and the stability of bootstrapped standard errors and confidence intervals in PLS-SEM (Hair et al., 2019). At the same time, a larger sample does not, by itself, establish population representativeness, which depends on the sampling design. This distinction between statistical adequacy and sampling representativeness is maintained throughout the present study.

In addition to generic power analysis using G*Power, PLS-SEM-specific procedures have been proposed to estimate minimum sample size. Kock and Hadaya (2018), for example, developed the inverse square root and gamma-exponential methods, which incorporate the anticipated path coefficient, desired statistical power, and significance level while reflecting the behaviour of PLS-SEM path estimates. Their simulation evidence indicates that the inverse square root method can approximate more computationally intensive Monte Carlo results while remaining relatively straightforward to apply. These approaches provide useful complementary evidence when researchers seek to evaluate sample-size adequacy specifically within a PLS-SEM framework.

Accordingly, G*Power should be viewed as a transparent power-based benchmark rather than as a universal sample-size solution for PLS-SEM. Hair et al. (2022) recommend considering the assumptions underlying the selected procedure and, where appropriate, cross-checking the resulting estimate against PLS-SEM-specific approaches. Such triangulation can strengthen the methodological justification for the minimum sample size without displacing the role of a priori power analysis in research planning. This study therefore uses G*Power as its primary planning tool while identifying PLS-SEM-specific cross-validation as an avenue for future methodological extension.

Multivariate Normality Assessment and the WebPower Platform

Assumption assessment represents a second important component of methodological rigour in SEM-based research. Conventional maximum-likelihood estimation in CB-SEM is sensitive to departures from multivariate normality, although robust estimation procedures are available for non-normal data. Mardia's (1970) multivariate skewness and kurtosis statistics provide established measures for assessing the joint distributional characteristics of multiple variables rather than examining each variable in isolation. The WebPower platform offers an accessible, non-programming-based implementation of these statistics, enabling researchers to incorporate formal multivariate normality assessment into the post-collection diagnostic stage of quantitative research.

Variance-Based SEM as a Response to Non-Normal Data

PLS-SEM is comparatively flexible with respect to distributional assumptions and does not require multivariate normality for parameter estimation (Hair et al., 2019). This characteristic can be advantageous in behavioural survey research, where responses measured using rating scales may depart from normality. Nevertheless, non-normality should not be treated as an automatic or exclusive justification for choosing PLS-SEM over CB-SEM. The analytical approach should also reflect the research objectives, model characteristics, and intended emphasis on explanation or prediction. In this study, the normality assessment is therefore treated as diagnostic evidence that complements, rather than independently determines, the choice of PLS-SEM.

Sample Size, Assumption Testing, and Tax Morale Research

Tax morale, broadly understood as the intrinsic motivation to comply with tax obligations beyond formal enforcement, is commonly investigated through survey-based models involving psychological, social, and institutional constructs. Such models can involve multiple latent variables and structural relationships, making sample-size justification and distributional assessment important elements of methodological reporting. Although PLS-SEM has been applied in taxation research (Hair et al., 2019), explicit integration of a priori power-based sample-size justification with formal multivariate normality assessment remains less visible in applied studies. The present paper addresses this methodological need

by demonstrating a sequential workflow in which G*Power supports research planning and WebPower supports post-collection distributional assessment.

METHODOLOGY

Population and Sampling Procedures

The target population comprised salaried individual taxpayers in Malaysia. To structure the sampling process systematically, the study adapted the five sequential stages described by Sekaran and Bougie (2016): defining the target population, identifying the sampling frame, selecting the sampling design, determining the required sample size, and implementing the sampling process. Steps I to IV describe the design and planning decisions, while Step V reports the implementation and resulting sample. Because the study employed non-probability purposive sampling, the emphasis is on sample adequacy and geographical coverage rather than probability-based representativeness.

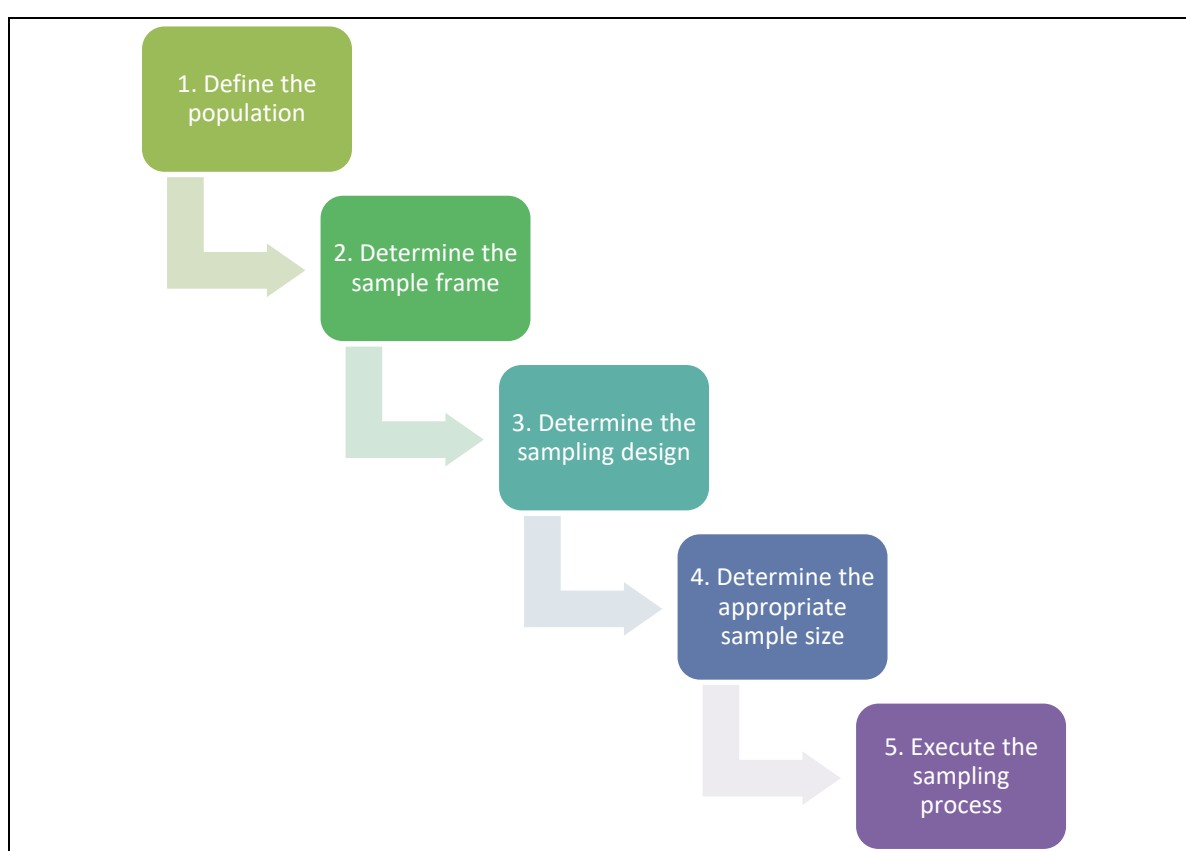


Figure 1: Research Sampling Process

Source: Sekaran & Bougie (2016)

Step I: Defining the Target Population

The first step was to define the target population clearly and establish the eligibility characteristics relevant to the research objectives. The target population comprised registered salaried individual taxpayers from Johor, Kedah, Kelantan, Melaka, Negeri Sembilan, Pahang, Perak, Perlis, Pulau Pinang, Sabah, Sarawak, Selangor, Terengganu, the Federal Territory of Kuala Lumpur, and the Federal Territory of Putrajaya. Defining these boundaries ensured that respondent selection was aligned consistently with the substantive focus of the tax morale study.

The Salaried Group (SG) refers to individual taxpayers earning employment income from public or private-sector employment. Population statistics for the Year of Assessment (YA) 2020 were obtained from the Inland Revenue Board of Malaysia (IRBM); however, individual taxpayer identities were not disclosed because of confidentiality requirements. Consistent with the study's focus, eligible respondents were required to receive employment income, hold a registered income tax file, and have prior experience submitting and paying income tax. These criteria ensured that respondents had direct experience with the tax system and could provide informed responses concerning tax morale. The focus

on individual taxpayers is also consistent with the OECD's (2019) demand for more country-level evidence on individual tax morale.

Table 1 presents the population of registered individual taxpayers for YA 2020 across the Malaysian states and Federal Territories included in the IRBM data. Registered individual taxpayers were classified into the Salaried Group (SG) and the Others Group (OG). Of the 9,305,218 registered individual taxpayers reported for the year ended 31 December 2020 (IRBM, 2021), 7,242,578 were classified as SG taxpayers. Because the study focuses specifically on salaried taxpayers, OG taxpayers were excluded from the target population.

Table 1: Registered SG and OG Taxpayers for Year of Assessment 2020

No	State	Salaried Group (SG)	Others Group (OG)	Total
1	Johor	715,532	332,440	1,047,972
2	Melaka	215,324	60,425	275,749
3	Negeri Sembilan	205,524	76,209	281,733
4	Perak	350,890	215,022	565,912
5	Kelantan	133,277	66,479	199,756
6	Kedah	219,885	79,347	299,232
7	Pulau Pinang	639,310	163,184	802,494
8	Pahang	201,252	108,935	310,187
9	Wilayah Persekutuan Kuala Lumpur	1,399,366	235,313	1,634,679
10	Wilayah Persekutuan Putrajaya Putrajaya	986,573	194,032	1,180,605
11	Terengganu	123,702	57,581	181,283
12	Selangor	1,192,711	225,651	1,418,362
13	Perlis	39,234	13,319	52,553
14	Sabah	340,642	115,665	456,307
15	Sarawak	479,356	119,038	598,394
Grand Total		7,242,578	2,062,640	9,305,218

Source: IRBM (2021)

These population statistics provide the contextual basis for defining the target population and planning geographical coverage. In the present study, IRBM data were used to describe the distribution of salaried taxpayers across Malaysia and to establish state-level allocation benchmarks. Because respondent selection was purposive rather than probability-based, the population proportions are not interpreted as evidence of statistical representativeness. Instead, they provide a systematic basis for broadening the geographical coverage of eligible respondents.

Step II: Determining the Sample Frame

A sampling frame ordinarily identifies the individual elements from which a probability sample can be drawn. In this study, the target population comprised eligible salaried taxpayers registered with IRBM who had prior experience submitting and paying income tax. However, IRBM did not provide an identifiable list of taxpayers because such information is confidential. Consequently, a probability-based sampling frame was unavailable. This constraint informed the decision to adopt a non-probability purposive sampling design in the subsequent stage.

Step III: Determining the Sampling Method

The choice between probability and non-probability sampling depends on both the research purpose and access to the target population. Under probability sampling, each population element has a known probability of selection, which requires an appropriate sampling frame. In non-probability sampling, selection probabilities are unknown (Memon et al., 2017). Saunders et al. (2020) note that probability sampling requires access to a sufficiently complete list of the target population. Because IRBM could not disclose the taxpayer list, the conditions necessary for probability sampling were not met. The study therefore adopted non-probability purposive sampling to recruit respondents who satisfied the predefined eligibility criteria.

Non-Probability Purposive Sampling Technique

Purposive sampling is a non-probability technique in which respondents are selected because they possess characteristics or experiences relevant to the research question (Abdullah, 2019). This approach was appropriate because the study required participants with direct experience of Malaysia's individual income tax system. Rather than seeking statistical representativeness through random selection, the sampling strategy prioritised eligibility and substantive relevance to the study objectives.

Figure 2 illustrates the purposive sampling approach adopted in the study. The eligibility criteria required respondents to receive employment income, hold a registered income tax file, and have prior experience submitting and paying income tax. These criteria were established before recruitment and applied consistently to focus data collection on taxpayers able to provide informed responses about tax morale, attitudes, and compliance-related behaviour. The approach also reflected the practical limitation that the confidential taxpayer population could not be accessed through a conventional random-sampling frame (Obilor, 2023).

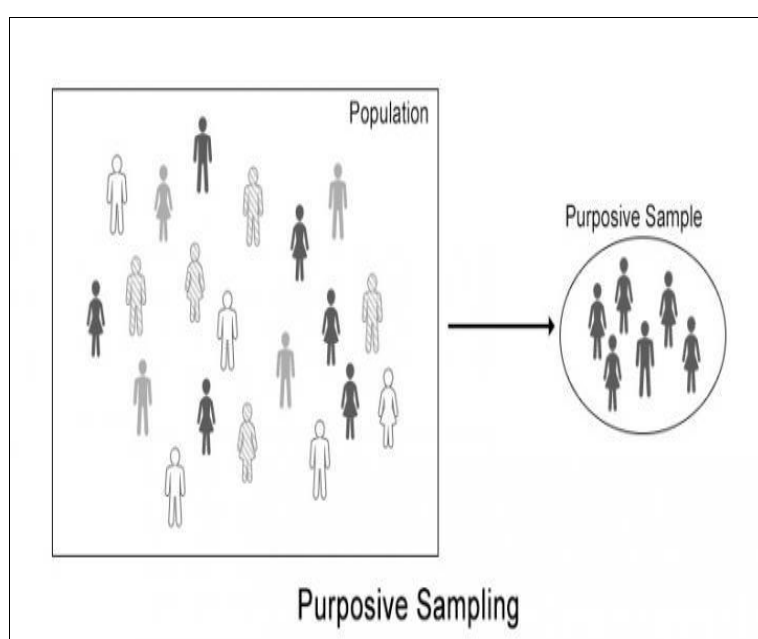


Figure 2: Purposive Sampling

Source: Obilor (2023)

Step IV: Determining the Sample Size

Determining an appropriate sample size was the fourth stage of the sampling process. In quantitative research, an insufficient sample may reduce statistical power, whereas collecting substantially more observations than required may increase resource demands without proportionate analytical benefit (Memon et al., 2020). Sample-size determination should therefore be linked to the intended statistical analysis and the magnitude of the relationships the study seeks to detect.

Because the study employed non-probability purposive sampling, probability-based formulas designed to estimate population parameters were not used as the primary basis for determining the analytical sample requirement. Instead, an a priori statistical power analysis was conducted using G*Power to estimate the minimum number of observations required for the structural model under specified assumptions concerning effect size, significance level, desired statistical power, and number of direct predictors (Faul et al., 2009).

G*Power analysis

G*Power 3.1.9.7 was used to conduct the a priori power analysis. The analysis was specified using the F-test family and Linear multiple regression: Fixed model, R^2 deviation from zero. The a priori option was selected to enable the required sample size could be estimated from the specified effect size, significance level, desired power, and number of predictors. This procedure provided an explicit statistical basis for the minimum sample requirement before the empirical analysis was undertaken. R^2

Figure 3 presents the G*Power specification used in the study. Based on the SmartPLS structural model, five predictor constructs were specified as having direct relationships with the endogenous construct, tax morale. There are five direct predictors (Social Norm > Tax morale, Tax Fairness > Tax Morale, Trust in Tax Authorities > Tax Morale, Social Norm > Tax Fairness > Tax Morale and Social Norm > Trust in Tax Authorities > Tax Morale) used in this study. Accordingly, five predictors were entered into G*Power. With the parameters specified in the analysis, the resulting minimum sample-size requirement was 92 respondents. This estimate represents a power-based analytical benchmark rather than a claim of population representativeness. As emphasised by Hair et al. (2017) and Memon et al. (2020), sample-size decisions should be considered in relation to model complexity, statistical power, and anticipated effect magnitude.

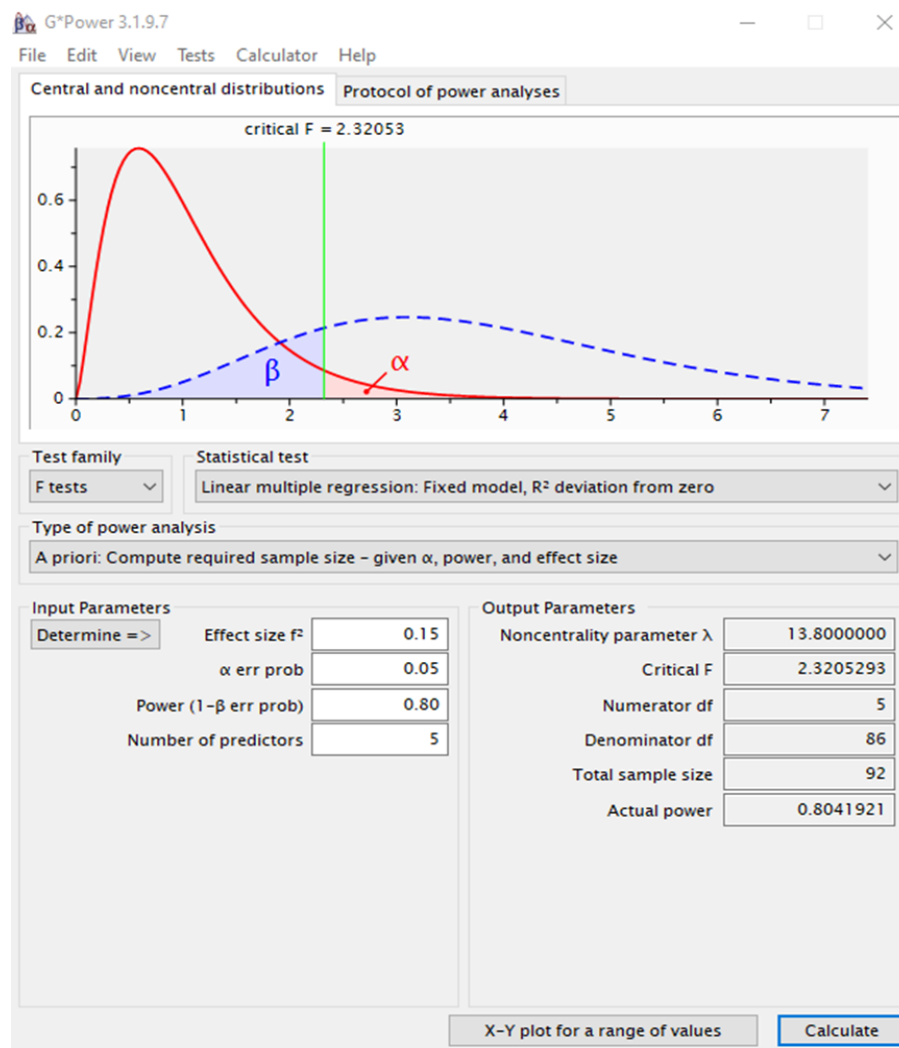


Figure 3: G*Power Analysis

The G*Power result of 92 observations was therefore adopted as the minimum analytical sample-size benchmark for the study. This power-based approach was preferred to the traditional 10-times rule because it explicitly incorporates the assumptions underlying the ability to detect structural relationships. Although the 10-times rule offers a simple heuristic, it has been criticised for potentially underestimating sample requirements, particularly when effects are relatively small (Kock & Hadaya, 2018; Hair et al., 2022).

Table 2 presents the distribution of registered salaried taxpayers across the Malaysian states and Federal Territories represented in the population data. The G*Power benchmark of 92 observations was allocated proportionately across these geographical units to provide a structured guide for recruitment. Because the study used purposive sampling, these state-level figures should be interpreted as allocation

benchmarks designed to broaden geographical coverage rather than as formal probability strata or guarantees of representativeness.

Table 2: Population and Sample Size of Registered SG Taxpayers for YA 2020 by States in Malaysia

No	State	YA 2020		Minimum Sample Size
		SG	%	
1	Johor	715,532	10.00	9
2	Melaka	215,324	2.73	2
3	Negeri Sembilan	205,524	2.73	2
4	Perak	350,890	4.55	4
5	Kelantan	133,277	1.82	2
6	Kedah	219,885	2.73	2
7	Pulau Pinang	639,310	9.09	8
8	Pahang	201,252	2.73	2
9	Wilayah Persekutuan Kuala Lumpur	1,399,366	19.09	17
10	Wilayah Persekutuan Putrajaya	986,573	13.64	15
11	Terengganu	123,702	1.82	2
12	Selangor	1,192,711	16.36	15
13	Perlis	39,234	1.82	2
14	Sabah	340,642	4.55	4
15	Sarawak	479,356	6.36	6
Grand Total		7,242,578	100%	92

Source: IRBM (2021)

RESULTS AND DISCUSSION

Step V reports the implementation of the sampling process and the resulting dataset, followed by the distributional diagnostics and the assessment of potential data-collection mode differences.

Step V: Execute The Sampling Process

Table 3 presents the responses collected from eligible salaried taxpayers across the geographical areas included in the study. A total of 321 responses were obtained, with the largest numbers coming from Johor (63), the Federal Territory of Kuala Lumpur (59), and Selangor (55). The smallest numbers were recorded in Perlis (3), Sabah (6), and Sarawak (6). Responses collected through paper and electronic questionnaires were consolidated into a common dataset for screening and subsequent data analysis processes and procedures.

Table 3: Actual Sample Size of Registered Salaried Group of Taxpayers by States

No	State	YA 2020		Actual Samples Collected	%
		Registered SG Taxpayers	%		
1	Johor	715,532	10.00	63	19.63
2	Melaka	215,324	2.73	12	3.74
3	Negeri Sembilan	205,524	2.73	8	2.49
4	Perak	350,890	4.55	10	3.12
5	Kelantan	133,277	1.82	20	6.23
6	Kedah Darul Aman	219,885	2.73	12	3.74
7	Pulau Pinang	639,310	9.09	17	5.30
8	Pahang	201,252	2.73	11	3.43
9	Federal Territory of Kuala Lumpur	1,399,366	19.09	59	18.38
10	Federal Territory of Putrajaya	986,573	13.64	31	9.66
11	Terengganu	123,702	1.82	7	2.18
12	Selangor	1,192,711	16.36	55	17.13
13	Perlis	39,234	1.82	3	0.93
14	Sabah	340,642	4.55	6	1.87
15	Sarawak	479,356	6.36	7	2.18
Grand Total		7,242,578	100%	321	100%

Source: IRBM (2021)

The study focused on salaried individual taxpayers in Malaysia to examine perceptions of the moral, social, and psychological factors associated with tax morale. Eligibility required respondents to be salaried individuals, registered taxpayers with IRBM, and to have prior experience submitting and paying income tax. These criteria ensured that participants had direct experience with the self-assessment system and the tax obligations relevant to the study (Hidayat, 2022). The recruitment process covered respondents across Malaysia to broaden geographical coverage of the target population. Of the 321 responses collected, two incomplete cases were excluded, leaving 319 usable responses for the normality, descriptive, and measurement analyses.

Tests for Data Normality

Multivariate normality was assessed because distributional characteristics provide useful diagnostic evidence for selecting and justifying an estimation approach. Skewness reflects asymmetry in a distribution, whereas kurtosis describes aspects of tail weight and peakedness (Balakrishnan & Scarpa, 2012; Cain et al., 2017). Mardia's (1970) multivariate skewness and kurtosis statistics extend these concepts to the joint distribution of multiple variables and are widely used to evaluate multivariate normality. Accordingly, the study used the WebPower platform to calculate Mardia's statistics for the observed dataset.

Univariate normality was also examined in SPSS using the Kolmogorov-Smirnov test with Lilliefors correction and the Shapiro-Wilk test. These tests were used as supplementary diagnostics rather than substitutes for the multivariate assessment. A statistically significant result ($p < .05$) indicates evidence against the null hypothesis of normality. Previous methodological comparisons have generally found the Shapiro-Wilk test to have stronger power than the Kolmogorov-Smirnov test for detecting departures from normality (Mohd Razali & Wah, 2011).

Table 4 presents Mardia's multivariate skewness and kurtosis statistics generated using WebPower, with the original output reproduced in Appendix 1. A CSV dataset containing 319 usable responses across eight variables was uploaded to the WebPower platform. The results indicated significant multivariate skewness ($\beta = 9.46$, $p < .01$) and multivariate kurtosis ($\beta = 107.44$, $p < .01$), providing evidence that the observed data departed from multivariate normality. These findings provide additional methodological support for the use of PLS-SEM, which does not require multivariate normality for parameter estimation. Nevertheless, the analytical choice is justified in conjunction with the study's objectives and model characteristics rather than on non-normality alone.

Table 4: Output of Skewness and Kurtosis Calculation

Table 4. Output of Skewness and Kurtosis Calculation						
Sample size:	319					
Number of variables:	8					
Univariate skewness and kurtosis						
	Skewness	SE_skew	Z_skew	Kurtosis	SE_kurt	Z_kurt
DN	-0.083	0.137	-0.608	-0.264	0.272	-0.970
IN	-0.180	0.137	-1.318	0.164	0.272	0.602
LS	-0.656	0.137	-4.805	1.057	0.272	3.883
PN	-0.013	0.137	-0.094	-0.486	0.272	-1.784
SN	-0.174	0.137	-1.273	-0.245	0.272	-0.899
TF	-0.521	0.137	-3.819	0.743	0.272	2.730
TM	-0.343	0.137	-2.513	0.330	0.272	1.211
TTA	-0.667	0.137	-4.888	1.188	0.272	4.363
Mardia's multivariate skewness and kurtosis						
	b	z	p-value			
Skewness	9.460815	502.99998	0			
Kurtosis	107.437410	19.37084	0			

The supplementary univariate results are reported in Table 5. The Shapiro-Wilk test was statistically significant for all eight variables ($p < .05$), indicating departures from univariate normality. The Kolmogorov-Smirnov test was significant for seven variables, whereas the result for tax morale was not significant ($p = .080$). Considered together with Mardia's multivariate statistics, these diagnostics consistently indicate that the observed dataset does not conform to normality. This evidence complements the broader methodological rationale for employing PLS-SEM.

Table 5: Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
DN	.123	319	.000	.979	319	.000
IN	.134	319	.000	.974	319	.000
LS	.132	319	.000	.949	319	.000
PN	.057	319	.013	.983	319	.001
SN	.077	319	.000	.983	319	.001
TF	.097	319	.000	.967	319	.000
TM	.047	319	.080	.974	319	.000
TTA	.093	319	.000	.943	319	.000

Note: a = Lilliefors Significance Correction

Mann-Whitney U Test

Because the normality diagnostics indicated departures from normality, the non-parametric Mann-Whitney U test was used to examine whether response distributions differed between the two independent data-collection groups: electronic and paper questionnaires (MacFarland & Yates, 2016). Table 6 reports the descriptive statistics, while Tables 7 and 8 present the rank-based and inferential results. The observed mean differences were small, and the Mann-Whitney U tests detected no statistically significant differences between the two modes at the .05 level. The two sets of responses were therefore combined for subsequent multivariate analysis.

Table 6: Group Statistics

Variables	Mode	N	Mean	Mean difference*	Standard Deviation	Standard Error Mean
Tax Morale (TM)	Electronic	91	3.8264	0.0852	.68845	.07217
	Paper	228	3.7412		.72844	.04824
Descriptive Norms (DN)	Electronic	91	3.9890	0.0329	1.24494	.13051
	Paper	228	3.9561		1.38361	.09163
Subjective Norms (SN)	Electronic	91	4.2000	-0.0608	1.23216	.12917
	Paper	228	4.2608		1.40442	.09321
Personal Norms (PN)	Electronic	91	4.1703	0.0179	1.44860	.15185
	Paper	228	4.1524		1.47709	.09782
Injunctive Norms (IN)	Electronic	91	4.1346	-0.0112	1.12442	.11787
	Paper	228	4.1458		1.25576	.08316
Trust in Tax Authorities (TTA)	Electronic	91	5.2418	-0.0075	.97381	.10208
	Paper	228	5.2493		1.09593	.07258
Tax Fairness (TF)	Electronic	91	4.8984	-0.1981	1.11740	.11714
	Paper	228	5.0965		1.11508	.07385
Life Satisfaction (LS)	Electronic	91	5.1923	-0.1257	1.05480	.11057
	Paper	228	5.3180		1.01870	.06747

*Mean difference = Mean electronic questionnaire – Mean paper questionnaire

Table 7 presents the Mann-Whitney U ranks for the electronic and paper questionnaire groups. Tax morale (N1 = 91, N2 = 228) is used to illustrate the interpretation. The electronic group had a mean rank of 166.75 compared with 157.30 for the paper group; however, the difference was not statistically significant ($U = 9759.500$, $Z = -0.830$, $p = .406$). Thus, the analysis provided insufficient evidence that the distribution of tax morale responses differed by data-collection mode.

Table 7: Mann-Whitney U Test with Ranks

Variables	Mode	N	Mean Rank	Sum of Ranks
Tax Morale (TM)	Electronic	91	166.75	15174.50
	Paper	228	157.30	35865.50
	Total	319		
Descriptive Norms (DN)	Electronic	91	160.02	14561.50
	Paper	228	159.99	36478.50
	Total	319		
Subjective Norms (SN)	Electronic	91	155.93	14189.50
	Paper	228	161.63	36850.50
	Total	319		
Personal Norms (PN)	Electronic	91	156.87	14275.00
	Paper	228	161.25	36765.00
	Total	319		
Injunctive Norms (IN)	Electronic	91	160.36	14593.00
	Paper	228	159.86	36447.00
	Total	319		
Trust in Tax Authorities (TTA)	Electronic	91	157.80	14360.00
	Paper	228	160.88	36680.00
	Total	319		
Tax Fairness (TF)	Electronic	91	148.01	13469.00
	Paper	228	164.79	37751.00
	Total	319		
Life Satisfaction (LS)	Electronic	91	148.74	13535.50
	Paper	228	164.49	37504.50
	Total	319		

Table 8 summarises the Mann-Whitney U statistics for all study variables. Group 1 comprised respondents who completed the electronic questionnaire (N1 = 91), while Group 2 comprised those who completed the paper questionnaire (N2 = 228). All two-tailed p-values exceeded .05, indicating that no statistically significant mode differences were detected across the study variables. These results support pooling the electronic and paper responses in the subsequent analyses.

Table 8: Mann-Whitney Test Statistics*

Variables	Mann-Whitney U	Wilcoxon W	Z	Asymp. Sig. (2-tailed)
Tax Morale (TM)	9759.500	35865.500	-.830	.406
Descriptive Norms (DN)	10372.500	10372.500	-.002	.998
Subjective Norms (SN)	10003.500	10003.500	-.500	.617
Personal Norms (PN)	10089.000	10089.000	-.384	.701
Injunctive Norms (IN)	10341.000	10341.000	-.045	.964
Trust in Tax Authorities (TTA)	10174.000	14360.000	-.271	.786
Tax Fairness (TF)	9283.000	13469.000	-1.477	.140
Life Satisfaction (LS)	9349.500	13535.500	-1.396	.163

*Group Variable: Type

IMPLICATIONS FOR DATA ANALYSIS

The findings have several methodological implications for quantitative research employing Structural Equation Modelling. First, the significant Mardia statistics demonstrate that the observed data depart from multivariate normality. This finding provides additional methodological support for the use of variance-based PLS-SEM, which does not require multivariate normality for parameter estimation. Importantly, non-normality should not be treated as the sole criterion for selecting PLS-SEM over CB-SEM. The choice of SEM approach should also be aligned with the study's analytical objectives, model characteristics, and intended emphasis on explaining variance or prediction. In the present study, the distributional evidence therefore complements, rather than independently determines, the methodological rationale for PLS-SEM.

Second, the G*Power analysis demonstrates the value of determining sample-size requirements a priori rather than relying on conventional rules of thumb. Based on five direct predictors and the parameters specified in the analysis, G*Power produced a minimum requirement of 92 observations, while 319 usable responses were ultimately available for analysis. The achieved sample therefore substantially exceeded the power-based benchmark. In PLS-SEM, observations beyond the minimum requirement can contribute to greater estimation precision and more stable bootstrapped standard errors and confidence intervals. However, exceeding a statistical power threshold should not be interpreted as establishing population representativeness, particularly under a non-probability purposive sampling design.

Third, the study demonstrates the complementary roles of G*Power and WebPower at different stages of the research process. G*Power supports a priori planning by linking the minimum sample requirement to explicitly specified assumptions concerning effect size, significance level, statistical power, and the number of direct predictors. WebPower complements this process after data collection by assessing the joint distributional characteristics of the observed variables through Mardia's statistics. Their combined use creates a transparent sequence from research planning to data diagnostics and analytical justification, while reducing reliance on unsupported rules of thumb such as the traditional 10-times rule (Kock & Hadaya, 2018).

Finally, the findings underscore the need to distinguish among sample-size adequacy, distributional assessment, and sampling representativeness. Statistical power addresses whether the planned sample is sufficient to detect effects under specified assumptions; normality assessment describes the distributional characteristics of the observed data; and representativeness depends on how respondents are selected from the target population. These concepts are related but not interchangeable. Accordingly, the geographical allocation used in this study should be interpreted as a strategy for broadening coverage rather than as evidence of probability-based representativeness. Explicit reporting of these distinctions, together with the G*Power parameters, usable sample size, WebPower results, and analytical rationale, strengthens transparency and reproducibility in quantitative research.

CONCLUSION

This study demonstrates an integrated, evidence-based workflow for strengthening two important methodological decisions in quantitative research: a priori sample-size determination and post-collection assessment of multivariate normality. Using a Malaysian tax morale study as an empirical illustration, G*Power was applied to establish a statistically justified minimum sample-size benchmark, while WebPower was subsequently used to assess the multivariate distributional characteristics of the collected data through Mardia's skewness and kurtosis statistics. Based on five direct predictors and the parameters specified in G*Power, the minimum required sample was 92 observations, whereas 319 usable responses were available for analysis. The WebPower results indicated significant departures from multivariate normality. These findings provide additional methodological support for the subsequent application of variance-based PLS-SEM, considered together with the study's analytical objectives and model characteristics.

The principal contribution of the study lies in demonstrating how G*Power and WebPower can be incorporated systematically at different stages of the research process rather than treating either tool as a substitute for broader methodological judgement. G*Power strengthens research planning by making the assumptions underlying minimum sample-size determination explicit, while WebPower provides an accessible means of evaluating multivariate distributional characteristics after data collection. Together, they encourage methodological decisions that are documented, empirically informed, and reproducible. The findings should nevertheless be interpreted within clear boundaries. The G*Power result is a power-based benchmark that depends on the statistical model and input parameters specified by the researcher, not a universal sample-size requirement for PLS-SEM. Similarly, the use of purposive sampling means that broad geographical coverage should not be equated with probability-based representativeness. Finally, multivariate non-normality provides relevant diagnostic evidence but does not, by itself, determine the choice between PLS-SEM and alternative SEM approaches.

Future research can extend this workflow by comparing G*Power estimates with PLS-SEM-specific minimum-sample-size procedures, including the inverse square root and gamma-exponential methods, as well as simulation-based power analysis. Replication using probability samples, longitudinal or panel designs, and other behavioural research contexts would help assess the transferability of the proposed approach. Researchers may also compare PLS-SEM with CB-SEM procedures that employ robust corrections for non-normality under comparable empirical conditions. Overall, methodological rigour is strengthened when sample-size determination, distributional assessment, sampling design, and analytical technique selection are treated as interconnected but conceptually distinct decisions. Integrating G*Power and WebPower within this broader decision-making process provides a practical and transparent framework for improving the planning, justification, and reporting of quantitative studies in taxation and the wider social sciences.

REFERENCES

- Abdullah, H. (2019). *Research Methods and Report Writing*. UPNM Press.
- Ahad, N. A., Yin, T. S., Othman, A. R., & Yaacob, C. R. (2011). Sensitivity of normality tests to non-normal data. *Sains Malaysiana*, 40(6), 637–641.
- Alm, J., & Torgler, B. (2004). Estimating the determinants of tax morale. In *Proceedings. Annual conference on taxation and minutes of the annual meeting of the National Tax Association* (Vol. 97, pp. 269–274). National Tax Association.
- Balakrishnan, N., & Scarpa, B. (2012). Multivariate measures of skewness for the skew-normal distribution. *Journal of Multivariate Analysis*, 104(1), 73–87. <https://doi.org/10.1016/j.jmva.2011.06.017>.
- Barclay, D., Higgins, C., & Thompson, R. (1995). The partial least squares (PLS) approach to casual modeling: personal computer adoption and use as an Illustration.
- Cain, M. K., Zhang, Z., & Yuan, K.-H. (2017). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence, and estimation. *Behavior Research Methods*, 49(5), 1716–1735. <https://doi.org/10.3758/s13428-016-0814-1>
- Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). John Wiley & Sons.
- Creswell, J. W., & Creswell, J. D. (2023). *Research design: Qualitative, quantitative, and mixed methods approaches* (6th ed.). SAGE Publications.
- Emiljan, K. (2022). Evaluating tax morale in Albania. *Euro-Balkan Law and Economics Review*, 1, 87–101.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BF03193146>
- Groves, R. M., Fowler Jr., F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). John Wiley & Sons.
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). *A primer on partial least squares structural equation modeling (PLS-SEM)* (2nd ed.). SAGE Publications.
- Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, 31(1), 2–24. <https://doi.org/10.1108/EBR-11-2018-0203>
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)* (3rd ed.). SAGE Publications.
- Hamid, S. (2013). *Tax compliance behavior of tax agents: A comparative study of Malaysia and New Zealand* [Unpublished doctoral dissertation]. University of Canterbury.
- Hayat, N., Salameh, A. A., Mamun, A. A., Helmi Ali, M., & Makhbul, Z. K. M. (2022). Tax compliance behavior among Malaysian taxpayers: A dual-stage PLS-SEM and ANN analysis. *SAGE Open*, 12(3). <https://doi.org/10.1177/21582440221127190>
- Hidayat, N. (2022). *Islamic Religiosity and Tax morale: The Case of Self-employed Taxpayers in Indonesia*. Unpublished PhD Thesis. The University of Western Australia.
- Inland Revenue Board of Malaysia (2021). *Ibu Pejabat Lembaga Hasil Dalam Negeri Malaysia*.

- Kang, H. (2021). Sample size determination and power analysis using the G*Power software. *Journal of Educational Evaluation for Health Professions*, 18, 17.
- Kock, N., & Hadaya, P. (2018). Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods. *Information Systems Journal*, 28(1), 227–261.
- MacFarland, T. W., & Yates, J. M. (2016). Mann-Whitney U test. In *Introduction to nonparametric statistics for the biological sciences using R* (pp. 103–132). Springer. https://doi.org/10.1007/978-3-319-30634-6_4
- Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3), 519–530. <https://doi.org/10.1093/biomet/57.3.519>
- Memon, M. A., Ting, H., Ramayah, T., Chuah, F., & Cheah, J.-H. (2017). A review of the methodological misconceptions and guidelines related to the application of structural equation modeling: A Malaysian scenario. *Journal of Applied Structural Equation Modeling*, 1(1), i–xiii.
- Memon, M. A., Ting, H., Cheah, J.-H., Thurasamy, R., Chuah, F., & Cham, T. H. (2020). Sample size for survey research: Review and recommendations. *Journal of Applied Structural Equation Modeling*, 4(2), i–xx. [https://doi.org/10.47263/JASEM.4\(2\)01](https://doi.org/10.47263/JASEM.4(2)01)
- Mohd Razali, N., & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, 2(1), 21–33.
- Obilor, E. I. (2023). Convenience and Purposive Sampling Techniques: Are they the Same? *International Journal of Innovative Social & Science Education Research*, 11(1), 1–7.
- Organization for Economic Co-operation and Development (OECD). (2019). Tax Morale: What Drives People and Businesses to Pay Tax?. OECD Publishing, Paris.
- Pacaldo, R. S., & Ferrer, R. C. (2020). Determinants of tax morale using structural equation model (SEM). *DLSU Business & Economics Review*, 29(2), 40–57.
- Sarstedt, M., Bengart, P., Shaltoni, A. M., & Lehmann, S. (2018). The Use of Sampling Methods in Advertising Research: A Gap Between Theory and Practice. *International Journal of Advertising*, 37(4), 650–663. <https://doi.org/10.1080/02650487.2017.1348329>.
- Saunders, M.N.K., Lewis, P., & Thornhill, A. (2020). *Research Methods for Business Students*. Eight Edition. Pearson Education.
- Sekaran, U., & Bougie, R. (2016). *Research methods for business: A skill-building approach* (7th ed.). John Wiley & Sons.
- Shaharuddin, N. S. (2017). *Sole proprietorship and tax compliance intention: A theory of planned behavior approach* [Unpublished doctoral dissertation]. Universiti Kebangsaan Malaysia.
- Tajuddin, T. S. (2025). *The influence of social norms, trust, fairness and life satisfaction on tax morale among salaried individual taxpayers in Malaysia* [Unpublished doctoral dissertation]. Universiti Sains Islam Malaysia.
- Torgler, B. (2007). *Tax compliance and tax morale: A theoretical and empirical analysis*. Edward Elgar Publishing.
- Wulandari, D., Sutrisno, S., & Nirwana, M. B. (2021). Mardia's skewness and kurtosis for assessing normality assumption in multivariate regression. *Enthusiastic: International Journal of Applied Statistics and Data Science*, 1(1), 1–6. <https://doi.org/10.20885/enthusiastic.vol1.iss1.art1>

Appendix 1

Output of skewness and kurtosis calculation

```
Sample size: 319
Number of variables: 8

Univariate skewness and kurtosis
```

	Skewness	SE_skew	Z_skew	Kurtosis	SE_kurt	Z_kurt
DN	-0.083	0.137	-0.608	-0.264	0.272	-0.970
IN	-0.180	0.137	-1.318	0.164	0.272	0.602
LS	-0.656	0.137	-4.805	1.057	0.272	3.883
PN	-0.013	0.137	-0.094	-0.486	0.272	-1.784
SN	-0.174	0.137	-1.273	-0.245	0.272	-0.899
TF	-0.521	0.137	-3.819	0.743	0.272	2.730
TM	-0.343	0.137	-2.513	0.330	0.272	1.211
TTA	-0.667	0.137	-4.888	1.188	0.272	4.363

```

Mardia's multivariate skewness and kurtosis
```

	b	z	p-value
Skewness	9.460815	502.99998	0
Kurtosis	107.437410	19.37084	0