

# Mapping Anxiety, Developing Solutions: A Statistical Study of Student Anxiety Using The K-Modes Clustering Method

Fadhilah Fitri<sup>1\*</sup>, Fitri Mudia Sari<sup>1</sup>, Fauziah Taslim<sup>2</sup>, dan Sri Wahyuni<sup>1</sup>

<sup>1</sup>Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

<sup>2</sup>Departemen Psikologi, Universitas Negeri Padang, Padang, Indonesia

\*Corresponding author: fadhilahfitri@fmipa.unp.ac.id

Submitted : 19 April 2026

Revised : 18 Mei 2026

Accepted : 30 Mei 2026

## ABSTRACT

*Statistics anxiety is a common issue among university students that can negatively affect their learning process and academic performance. This study aims to identify patterns of statistics anxiety among undergraduate students at Universitas Negeri Padang using the Statistics Anxiety Rating Scale (STARS), which consists of six dimensions. A total of 479 valid responses were analyzed using the k-modes clustering method, which is appropriate for categorical data. The optimal number of clusters was determined using the elbow and silhouette methods, resulting in three clusters. The clustering results reveal three distinct groups of students characterized by high, moderate, and low levels of statistics anxiety. The average silhouette value of 0.52 indicates a moderately well-defined cluster structure. Further analysis shows that each cluster exhibits different patterns across the six anxiety dimensions, highlighting the heterogeneity of students' responses to statistics. These findings suggest that clustering provides a more informative approach than conventional descriptive analysis in understanding statistics anxiety. The results of this study can serve as a basis for developing targeted strategies to reduce student anxiety in statistics learning.*

**Keywords:** Higher Education, K-modes Clustering, STARS, Statistics Anxiety



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

## I. PENDAHULUAN

Statistika merupakan salah satu mata kuliah yang wajib ditempuh oleh mahasiswa di berbagai program studi, terutama di bidang sains, teknik, sosial, dan pendidikan. Namun, tidak sedikit mahasiswa yang menganggap statistika sebagai mata kuliah yang sulit dan menakutkan (Maricic *et al.*, 2023). Stigma negatif terhadap statistika ini sering kali muncul sebelum mahasiswa benar-benar memahami materi yang diajarkan, yang pada akhirnya dapat memengaruhi motivasi belajar dan pencapaian akademik mereka (Paechter *et al.*, 2017; Macher *et al.*, 2015). Fenomena ini dikenal sebagai *statistics anxiety* atau kecemasan terhadap statistika, yaitu ketakutan atau ketidaknyamanan saat belajar atau menggunakan statistika dalam berbagai konteks akademik. Mahasiswa yang mengalami *statistics anxiety* selalu mengalami kecemasan ketika melakukan hal yang berhubungan dengan statistika (Onwuegbuzie, 2004).

Untuk memahami fenomena tersebut secara lebih terukur, diperlukan instrumen yang mampu mengidentifikasi tingkat kecemasan mahasiswa secara sistematis. Salah satu instrumen yang banyak digunakan untuk mengukur kecemasan terhadap statistika adalah *Statistics Anxiety Rating Scale* (STARS) yang dikembangkan oleh Cruise, Cash, dan Bolton (1985). STARS mengukur kecemasan berdasarkan enam dimensi utama: *Worth of Statistics*, *Interpretation Anxiety*, *Test and Class Anxiety*, *Computational Self-concept*, *Fear of Asking for Help*, dan *Fear of Statistics Teachers* (Hanna *et al.*, 2008; Nielsen & Kreiner, 2018). Dengan adanya pengukuran berbasis dimensi ini, kecemasan mahasiswa dapat dipahami tidak hanya sebagai satu konstruk tunggal, tetapi sebagai fenomena multidimensional.

Universitas Negeri Padang (UNP) sebagai salah satu perguruan tinggi yang menawarkan mata kuliah statistika bagi berbagai program studi juga menghadapi tantangan serupa. Berdasarkan *pilot survey* yang dilakukan penulis, banyak mahasiswa yang mengalami kesulitan dalam memahami konsep statistika serta menunjukkan sikap yang cenderung menghindari mata kuliah ini. Di beberapa tempat lain dilakukan asesmen serupa memberikan hasil yang beragam, diantaranya mahasiswi memiliki tingkat kecemasan yang lebih tinggi dibanding mahasiswa (Levpušček & Cukon, 2020). Kemudian pada penelitian lain diketahui hasil kecemasan statistik, umur, jenis kelamin, model

pembelajaran dan program khusus menjelaskan 27% variasi pencapaian statistik mahasiswa (Maat et al., 2022). Kemudian pada penelitian yang dilakukan (Frias-Navarro *et al.*, 2018) disimpulkan bahwa mahasiswa memiliki tingkat kecemasan yang tinggi namun menurun secara bertahap seiring berlangsungnya perkuliahan. Untuk penelitian yang secara sistematis mengukur tingkat kecemasan mahasiswa dalam belajar statistika di UNP sudah dilakukan (Afdal *et al.*, 2019), namun, belum ada yang melakukan pengelompokan mereka berdasarkan pola kecemasan yang dialami. Kondisi ini menunjukkan bahwa pemahaman mengenai kecemasan mahasiswa masih terbatas pada tingkat umum dan belum menggambarkan variasi karakteristik kecemasan secara lebih spesifik.

Selain itu, pendekatan analisis yang bersifat deskriptif atau inferensial yang umum digunakan dalam penelitian sebelumnya cenderung hanya memberikan gambaran umum mengenai tingkat kecemasan mahasiswa, tanpa mampu mengidentifikasi pola heterogenitas kecemasan secara lebih mendalam. Padahal, mahasiswa dapat memiliki kombinasi tingkat kecemasan yang berbeda pada setiap dimensi, sehingga diperlukan pendekatan analisis yang mampu mengelompokkan karakteristik tersebut secara simultan. Dengan demikian, diperlukan metode analisis yang tidak hanya menjelaskan tingkat kecemasan, tetapi juga mampu mengungkap struktur pola kecemasan yang tersembunyi dalam data.

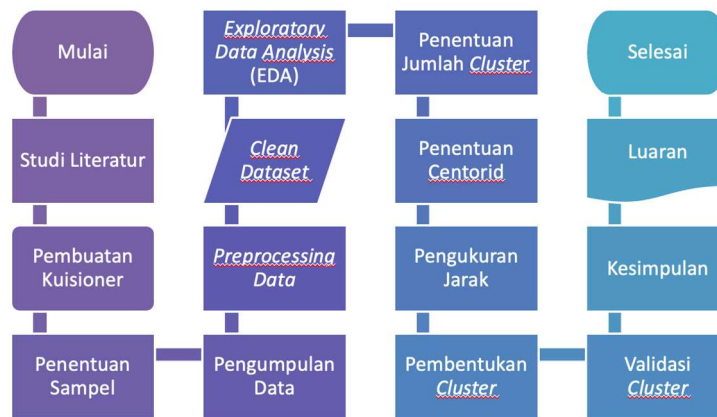
Pengelompokan dapat dilakukan menggunakan metode *clustering*. Salah satu metode yang unggul dan cocok digunakan adalah *k-modes clustering*, yang merupakan pengembangan dari metode *k-means* dengan menggunakan modus sebagai pusat *cluster* (Huang, 1998). Penggunaan metode *k-modes* dalam penelitian ini menjadi relevan karena data yang diperoleh dari instrumen STARS berskala ordinal atau data kategorik. *K-means* tidak sesuai untuk kondisi data karena merupakan metode berbasis rata-rata yang menggunakan data numerik. Dengan demikian, *k-modes* memberikan pendekatan yang lebih sesuai dalam mengidentifikasi kemiripan karakteristik antar objek berdasarkan nilai kategorik yang dimiliki. Selain itu, metode ini memungkinkan pengelompokan data secara efisien serta menghasilkan interpretasi *cluster* yang lebih mudah dipahami berdasarkan karakteristik dominan setiap kelompok.

*K-modes* dipilih dalam penelitian ini karena kemampuannya dalam menangani data kategorik secara efisien serta kemudahan interpretasi hasil *cluster* berdasarkan modus atribut. Dengan mengidentifikasi pola pengelompokan kecemasan mahasiswa, penelitian ini tidak hanya memberikan gambaran tingkat kecemasan, tetapi juga mengungkap struktur variasi kecemasan yang tidak dapat ditangkap melalui analisis konvensional. Pendekatan ini diharapkan dapat memberikan kontribusi dalam penerapan metode *clustering* untuk mengidentifikasi heterogenitas kecemasan mahasiswa berbasis data kategorik.

Oleh karena itu, metode *k-modes* digunakan untuk mengelompokkan mahasiswa berdasarkan enam dimensi kecemasan yang diukur menggunakan instrumen *Statistical Anxiety Rating Scale* (STARS). Keenam dimensi tersebut merepresentasikan berbagai aspek kecemasan mahasiswa terhadap statistik, sehingga menghasilkan data dalam bentuk kategorik. Dengan demikian, pendekatan ini memungkinkan identifikasi pola pengelompokan mahasiswa berdasarkan tingkat kecemasan yang dimiliki secara lebih komprehensif dibandingkan analisis konvensional.

## II. METODE PENELITIAN

Penelitian ini dilakukan di lingkungan Universitas Negeri Padang. Populasi target adalah mahasiswa yang telah mengambil mata kuliah statistika pada tahun ajaran 2024/2025. Responden dalam penelitian ini merupakan mahasiswa aktif pada tahun ajaran 2025/2026 yang telah menyelesaikan mata kuliah tersebut. Hal ini dilakukan agar responden memiliki pengalaman langsung dalam mengikuti pembelajaran statistika sehingga dapat memberikan penilaian yang lebih akurat terhadap tingkat kecemasan yang dialami. Variabel yang digunakan adalah *Statistics Anxiety* yang akan dinilai menggunakan *Statistics Anxiety Rating Scale* (STARS). Enam dimensi utama yang dinilai adalah *Worth of Statistics*( $X_1$ ), *Interpretation Anxiety*( $X_2$ ), *Test and Class Anxiety*( $X_3$ ), *Computational Self-concept*( $X_4$ ), *Fear of Asking for Help*( $X_5$ ), dan *Fear of Statistics Teachers*( $X_6$ ). Kuisisioner ini akan diubah ke dalam Bahasa Indonesia terlebih dahulu sebelum dibagikan. Berikut diagram alir dari penelitian yang dapat dilihat pada Gambar 1.



Gambar 1. Diagram alir penelitian

### A. Populasi dan Sampel

Populasi dalam penelitian ini adalah mahasiswa S1 Universitas Negeri Padang dengan tahun masuk 2018 hingga 2024 yang masih aktif pada semester ganjil Tahun Ajaran 2025/2026. Pemilihan rentang tahun masuk tersebut didasarkan pada pertimbangan bahwa mahasiswa dalam kelompok ini telah memiliki peluang untuk menempuh mata kuliah statistika dalam kurikulum studinya. Jumlah populasi berdasarkan kriteria tersebut sebanyak 27.888 mahasiswa. Sampel dalam penelitian ini diambil dari mahasiswa yang memenuhi kriteria populasi dan telah menyelesaikan mata kuliah statistika pada tahun ajaran 2024/2025. Pemilihan kriteria ini bertujuan agar responden memiliki pengalaman langsung dalam mengikuti pembelajaran statistika, sehingga dapat memberikan penilaian yang lebih akurat terhadap tingkat kecemasan yang dialami.

Teknik sampling yang digunakan adalah *purposive random sampling*. Teknik ini dipilih untuk memastikan keterwakilan mahasiswa dari berbagai program studi S1 di Universitas Negeri Padang sesuai dengan proporsi jumlah mahasiswanya, serta mempertimbangkan kemudahan dalam pengumpulan data pada populasi yang besar (Campbell *et al.*, 2020; Nyimbili & Nyimbili, 2024; Ames *et al.*, 2019). Dengan pendekatan ini, distribusi responden antar program studi dapat diatur secara lebih merata sehingga karakteristik populasi tetap terwakili dalam sampel penelitian. Selain itu, penggunaan teknik ini juga memungkinkan peneliti untuk mengontrol jumlah responden pada setiap kelompok yang telah ditentukan sebelumnya.

Jumlah sampel minimum ditentukan menggunakan rumus Slovin sebagai pendekatan untuk memperoleh ukuran sampel yang memadai, yang dirumuskan sebagai berikut:

$$n = \frac{N}{1+N(e)^2} \quad (1)$$

dengan  $N = 27.888$  dan tingkat kesalahan  $e = 5\%$ . Berdasarkan perhitungan tersebut diperoleh jumlah sampel minimum sebanyak 395 responden. Penggunaan rumus ini tidak dimaksudkan untuk generalisasi statistik, melainkan sebagai acuan awal dalam menentukan jumlah responden pada penelitian yang bersifat eksploratif (Sugiyono, 2017). Selanjutnya ditentukan proporsi jumlah mahasiswa per program studi terhadap jumlah keseluruhan sebagai dasar dalam penentuan jumlah sampel tiap program studi. Karena adanya pembulatan pada setiap program studi, jumlah sampel yang digunakan dalam penelitian ini menjadi 428 responden yang tersebar secara proporsional pada 73 program studi sarjana di Universitas Negeri Padang.

### B. Instrumen Penelitian

Instrumen yang digunakan dalam penelitian ini adalah *Statistics Anxiety Rating Scale* (STARS) yang dikembangkan oleh Cruise, Cash, dan Bolton (1985). Instrumen ini digunakan untuk mengukur tingkat kecemasan mahasiswa terhadap statistika dalam berbagai konteks pembelajaran. STARS merupakan salah satu instrumen yang banyak digunakan dalam penelitian terkait kecemasan statistik karena memiliki cakupan dimensi yang komprehensif serta telah digunakan secara luas dalam berbagai studi empiris.

Instrumen STARS mengukur kecemasan statistik berdasarkan enam dimensi utama, yaitu *Worth of Statistics*, *Interpretation Anxiety*, *Test and Class Anxiety*, *Computational Self-concept*, *Fear of Asking for Help*, dan *Fear of Statistics Teachers* (Hanna *et al.*, 2008; Nielsen & Kreiner, 2018). Setiap dimensi merepresentasikan aspek yang berbeda dari kecemasan mahasiswa terhadap statistika, sehingga memungkinkan analisis yang lebih mendalam terhadap pola kecemasan yang dialami.

Kuesioner dalam penelitian ini diadaptasi ke dalam Bahasa Indonesia untuk menyesuaikan dengan konteks responden. Proses adaptasi dilakukan dengan mempertimbangkan kesesuaian bahasa dan makna agar setiap pernyataan tetap merepresentasikan konstruk yang diukur. Instrumen disusun dalam bentuk skala Likert dengan beberapa kategori respon yang menunjukkan tingkat persetujuan responden terhadap setiap pernyataan, sehingga data yang diperoleh mencerminkan tingkat kecemasan mahasiswa pada masing-masing dimensi yang diukur.

### C. Tahapan Analisis Data

Sebelum dilakukan proses analisis, data hasil pengisian kuesioner terlebih dahulu diperiksa untuk memastikan kelengkapan dan konsistensi data. Proses ini bertujuan untuk memastikan bahwa data yang digunakan dalam analisis memenuhi kriteria yang telah ditetapkan. Tahapan analisis data dalam penelitian ini dilakukan melalui beberapa langkah sebagai berikut:

1. Pemeriksaan dan pembersihan data  
Mengidentifikasi data yang tidak lengkap, tidak konsisten, atau tidak memenuhi kriteria responden, kemudian mengeluarkannya dari proses analisis.
2. Penentuan variabel penelitian  
Variabel yang digunakan merupakan enam dimensi kecemasan statistik yang diukur menggunakan instrumen STARS, yang kemudian direpresentasikan sebagai atribut dalam proses *clustering*.
3. Representasi data  
Data yang diperoleh dari skala Likert diperlakukan sebagai data kategorik, di mana setiap kategori respon dianggap sebagai kategori yang berbeda tanpa mempertimbangkan jarak antar nilai.
4. Persiapan data untuk *clustering*  
Data yang telah diproses disusun dalam bentuk matriks data yang siap digunakan sebagai input dalam metode *k-modes clustering*.

### D. *K-modes clustering*

Metode *k-modes* merupakan pengembangan dari algoritma *k-means* yang digunakan untuk mengelompokkan data kategorik. Algoritma ini menggantikan rata-rata (*mean*) dengan modus (*mode*) serta proses pengelompokan didasarkan pada ukuran ketidaksamaan yang sederhana atau *simple matching dissimilarity* (Huang, 1998). Dalam penelitian ini, setiap objek merepresentasikan responden, sedangkan atribut yang digunakan adalah enam dimensi kecemasan statistik yang diukur menggunakan instrumen STARS.

Ukuran ketidaksamaan antara  $X$  dan  $Y$  dapat didefinisikan sebagai total ketidakcocokan dari variabel kategorik yang sesuai dari kedua objek tersebut. Semakin kecil nilai ketidaksamaan, semakin mirip kedua objek tersebut. Secara matematis dapat di tuliskan dalam bentuk persamaan berikut :

$$d(X, Y) = \sum_{j=1}^m \delta(x_j, y_j) \quad (2)$$

dimana

$$d(x_j, y_j) = \begin{cases} 0, & x_j = y_j \\ 1, & x_j \neq y_j \end{cases} \quad (3)$$

dengan  $x_j$  dan  $y_j$  merupakan nilai atribut ke- $j$  dari data  $X$  dan  $Y$ , dan  $m$  menyatakan jumlah atribut yang digunakan.

Proses *clustering* menggunakan metode *k-modes* dilakukan melalui tahapan sebagai berikut:

1. Menentukan jumlah *cluster* ( $k$ )  
Penentuan jumlah *cluster* optimal dilakukan menggunakan metode *elbow* dan *silhouette*. Pada penelitian ini, metode *elbow* digunakan sebagai pendekatan eksploratif untuk mengamati perubahan nilai cost atau ketidaksamaan total terhadap jumlah *cluster*, sedangkan metode *silhouette* digunakan untuk mengevaluasi kualitas *cluster* berdasarkan tingkat kemiripan objek di dalam *cluster* dibandingkan dengan *cluster* lain. Nilai *silhouette* berada pada rentang -1 hingga 1, di mana nilai yang lebih tinggi menunjukkan kualitas *cluster* yang lebih baik.
2. Inisialisasi pusat *cluster*  
Memilih secara acak  $k$  objek dari data sebagai pusat *cluster* awal.
3. Perhitungan jarak (*dissimilarity*)  
Menghitung jarak antara setiap objek dengan masing-masing pusat *cluster* menggunakan ukuran ketidaksamaan sederhana (*simple matching dissimilarity*), yaitu dengan menjumlahkan jumlah atribut yang tidak sama.
4. Pengelompokan objek

Mengelompokkan setiap objek ke dalam *cluster* dengan nilai ketidaksamaan terkecil atau dengan pusat *cluster* yang paling mirip.

5. Pembaruan pusat *cluster*

Menentukan pusat *cluster* baru untuk setiap *cluster* dengan memilih nilai kategori yang paling sering muncul (modus) pada masing-masing atribut.

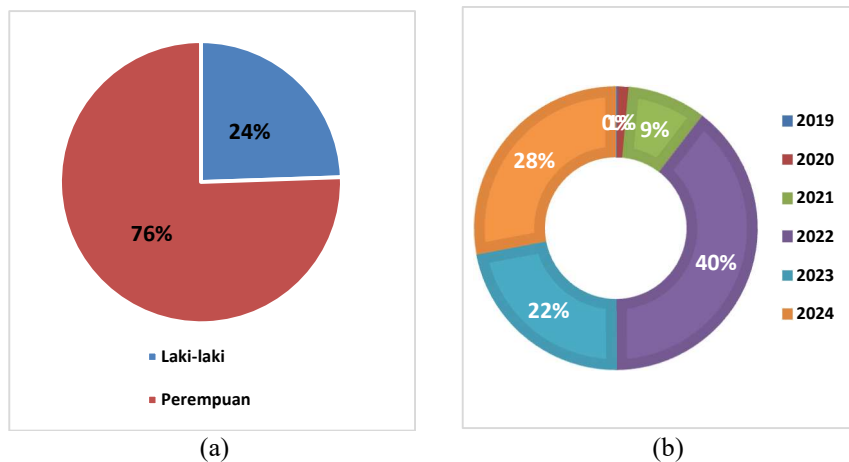
Hasil *clustering* selanjutnya dievaluasi menggunakan beberapa indeks validasi, yaitu *Silhouette Score*, *Dunn Index*, *Davies-Bouldin Index*, dan *Calinski-Harabasz Index*. Indeks-indeks tersebut digunakan untuk menilai kekompakan objek dalam *cluster* dan keterpisahan antar*cluster* yang terbentuk.

### III. HASIL DAN PEMBAHASAN

Berikut hasil dan pembahasan dari penelitian ini.

#### A. Analisis Deskriptif

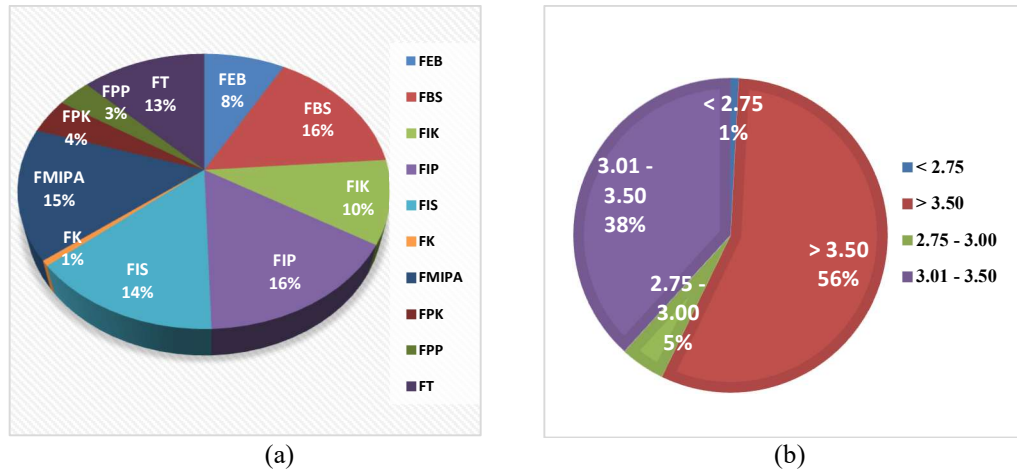
Responden yang berpartisipasi dalam survei ini berjumlah 480 orang. Setelah dilakukan proses *data cleaning*, terdapat 1 data yang tidak valid sehingga jumlah responden yang digunakan dalam analisis sebanyak 479 orang. Jumlah ini telah melampaui ukuran sampel minimum yang ditentukan sebelumnya, sehingga diharapkan mampu memberikan gambaran yang lebih stabil terhadap karakteristik responden. Karakteristik responden dalam penelitian ini disajikan berdasarkan jenis kelamin, tahun masuk, fakultas asal, IPK, dan jalur masuk perguruan tinggi.



Gambar 2. (a) Responden berdasarkan Jenis Kelamin, (b) Responden berdasarkan Tahun Masuk

Berdasarkan Gambar 2(a), responden dalam penelitian ini didominasi oleh perempuan dengan persentase sebesar 76%, sedangkan laki-laki sebesar 24%. Komposisi ini menunjukkan adanya ketimpangan distribusi responden, yang berpotensi memengaruhi hasil analisis, terutama apabila terdapat perbedaan kecemasan statistik berdasarkan gender. Sejumlah penelitian menunjukkan bahwa faktor psikologis, termasuk kecemasan, dapat berbeda antar gender, sehingga distribusi ini perlu diperhatikan dalam interpretasi hasil penelitian.

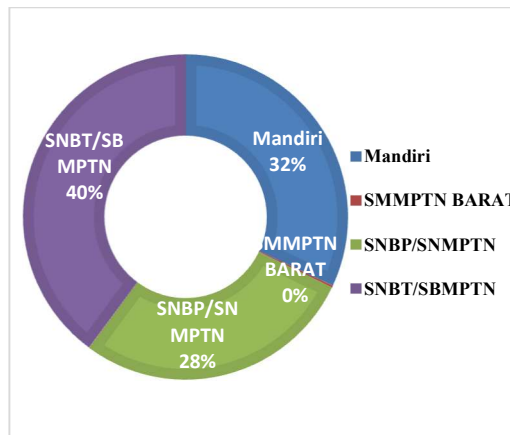
Pada Gambar 2(b), distribusi responden berdasarkan tahun masuk menunjukkan bahwa sebagian besar responden berasal dari angkatan 2022, yaitu sebesar 40%. Sementara itu, jumlah responden paling sedikit berasal dari angkatan 2019. Pola ini mengindikasikan bahwa mayoritas responden berada pada tahap studi menengah, di mana mahasiswa umumnya telah menyelesaikan mata kuliah statistika dan mulai menghadapi tuntutan akademik yang lebih kompleks, seperti penelitian atau penyusunan tugas akhir. Pada tahap ini, penggunaan statistika menjadi lebih intensif sehingga potensi munculnya kecemasan statistik juga dapat meningkat.



Gambar 3. (a) Responden berdasarkan Fakultas Asal, (b) Responden berdasarkan IPK

Distribusi responden berdasarkan fakultas asal pada Gambar 3(a) menunjukkan bahwa sampel telah terdistribusi secara proporsional mengikuti komposisi mahasiswa di Universitas Negeri Padang. Hal ini penting untuk memastikan bahwa variasi latar belakang keilmuan mahasiswa, baik dari rumpun sains dan teknologi maupun sosial dan humaniora, dapat terwakili dengan baik dalam penelitian. Perbedaan latar belakang ini berpotensi memengaruhi cara mahasiswa memahami dan merespons pembelajaran statistika.

Selanjutnya, pada Gambar 3(b), distribusi responden berdasarkan IPK menunjukkan bahwa sebagian besar responden memiliki IPK di atas 3,50, yaitu sebesar 56%, sedangkan hanya sebagian kecil yang memiliki IPK di bawah 2,75. Temuan ini menunjukkan bahwa mayoritas responden memiliki capaian akademik yang tinggi. Namun demikian, capaian akademik yang baik tidak selalu berkaitan dengan rendahnya kecemasan statistik, karena kecemasan juga dipengaruhi oleh faktor lain seperti kepercayaan diri, pengalaman belajar, dan persepsi terhadap statistika.

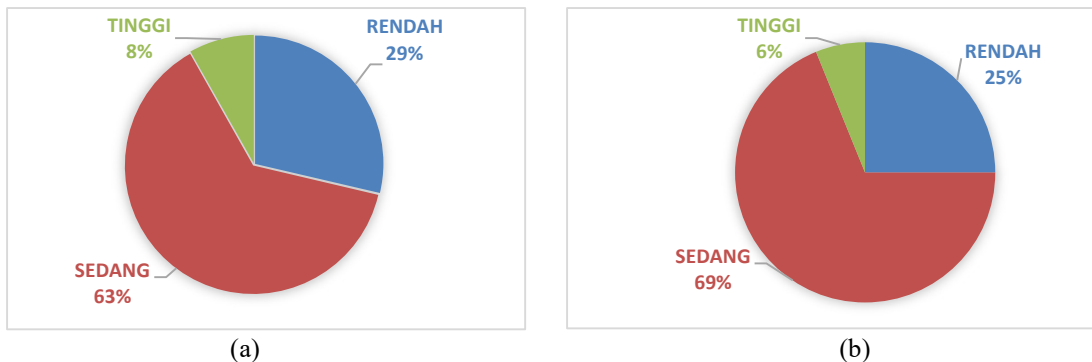


Gambar 4. Responden berdasarkan Jalur Masuk

Pada Gambar 4, distribusi responden berdasarkan jalur masuk menunjukkan bahwa sebagian besar responden berasal dari jalur SNBT/SBMPTN sebesar 40%, diikuti oleh jalur mandiri dan SNBP. Dominasi jalur seleksi berbasis tes ini mengindikasikan bahwa sebagian besar responden telah melalui proses seleksi akademik yang kompetitif. Meskipun demikian, jalur masuk tidak secara langsung mencerminkan kesiapan mahasiswa dalam memahami statistika, sehingga hubungannya dengan kecemasan perlu dianalisis lebih lanjut.

Selain itu, perbedaan latar belakang keilmuan mahasiswa dianalisis melalui pengelompokan rumpun ilmu Saintek dan Soshum. Berdasarkan Gambar 5(a), mahasiswa rumpun Saintek didominasi oleh kategori kecemasan sedang sebesar 63%, diikuti kategori rendah sebesar 29%, dan kategori tinggi sebesar 8%. Sementara itu, pada Gambar 5(b), mahasiswa

rumpun Soshum juga menunjukkan pola yang serupa dengan dominasi kategori sedang sebesar 69%, diikuti kategori rendah sebesar 25%, dan kategori tinggi sebesar 6%.



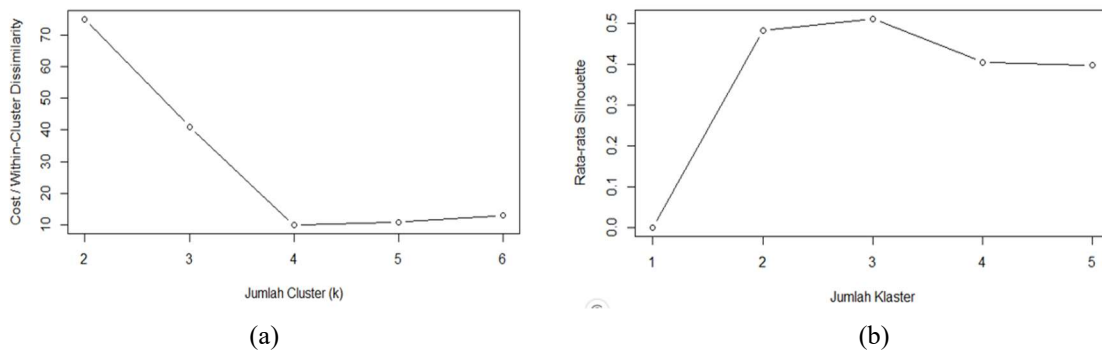
**Gambar 5.** Kategorisasi Tingkat Kecemasan Statistik Berdasarkan Rumpun Ilmu; (a) Rumpun Ilmu Saintek; (b) Rumpun Ilmu Soshum

Perbandingan kedua kelompok menunjukkan bahwa distribusi kecemasan statistik relatif serupa, dengan dominasi pada kategori sedang. Namun, terdapat perbedaan pada kategori kecemasan tinggi, di mana mahasiswa Saintek memiliki proporsi yang sedikit lebih besar dibandingkan Soshum. Hal ini mengindikasikan bahwa intensitas penggunaan materi kuantitatif pada kelompok Saintek dapat berkontribusi terhadap meningkatnya kecemasan pada sebagian mahasiswa.

Secara keseluruhan, hasil analisis deskriptif menunjukkan bahwa responden memiliki karakteristik yang beragam dari sisi demografis dan akademik, serta didominasi oleh tingkat kecemasan sedang. Variasi ini menjadi dasar penting dalam analisis lanjutan menggunakan metode *clustering*, karena memungkinkan identifikasi pola kecemasan yang lebih spesifik pada kelompok mahasiswa dengan karakteristik yang berbeda.

### B. K-Modes

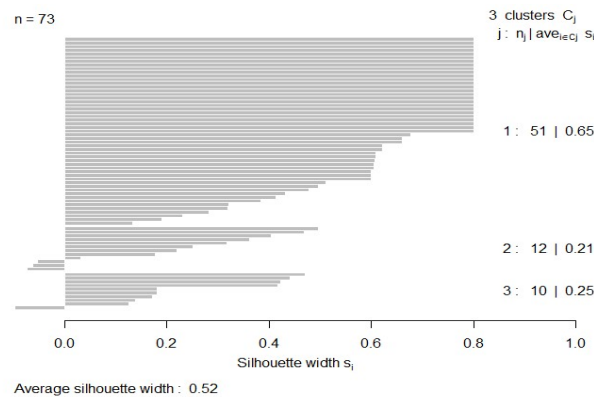
Penentuan jumlah *cluster* optimal merupakan tahap penting dalam proses *clustering* menggunakan metode *k-modes*. Pada penelitian ini, digunakan dua pendekatan, yaitu metode *elbow* dan metode *silhouette*, untuk memperoleh jumlah *cluster* yang paling sesuai dengan karakteristik data



**Gambar 6.** Penentuan jumlah *cluster* optimal; (a) metode Elbow; (b) metode Silhouette

Berdasarkan Gambar 6(a), penurunan nilai *cost* yang cukup tajam terjadi ketika jumlah *cluster* meningkat dari 2 menjadi 3, kemudian cenderung melandai setelahnya. Pola ini menunjukkan bahwa penambahan *cluster* di atas  $k = 3$  tidak memberikan peningkatan yang berarti dalam membedakan karakteristik data. Dengan demikian, jumlah *cluster* sebanyak 3 dianggap telah mampu merepresentasikan variasi pola kecemasan mahasiswa secara memadai. Hasil ini diperkuat oleh Gambar 6(b), di mana nilai rata-rata *silhouette* tertinggi juga diperoleh pada  $k = 3$ . Nilai ini menunjukkan bahwa objek dalam *cluster* memiliki tingkat kemiripan yang baik dengan anggota dalam *cluster* yang sama serta cukup terpisah dari *cluster* lainnya. Oleh karena itu, jumlah *cluster* yang digunakan dalam penelitian ini ditetapkan sebanyak 3 *cluster*.

Untuk mengevaluasi kualitas hasil *clustering* yang telah dilakukan menggunakan metode *k-modes*, digunakan metode *Silhouette Coefficient*. Metode ini bertujuan untuk mengukur tingkat kesamaan data dalam satu *cluster* serta perbedaannya dengan *cluster* lain. Hasil evaluasi tersebut divisualisasikan dalam bentuk *Silhouette Plot* yang dapat dilihat pada Gambar 7.



**Gambar 7.** *Silhouette Plot* dari *k-modes* dengan  $k = 3$

Berdasarkan Gambar 7, diperoleh nilai rata-rata *silhouette* sebesar 0,52 yang menunjukkan bahwa kualitas pengelompokan berada pada kategori cukup baik. Nilai ini mengindikasikan bahwa sebagian besar objek telah dikelompokkan ke dalam *cluster* yang sesuai, meskipun masih terdapat beberapa objek yang berada di area perbatasan antar *cluster*. Kondisi ini menunjukkan bahwa struktur data memiliki pemisahan yang cukup jelas, namun masih terdapat kemiripan karakteristik pada sebagian kecil data.

Berdasarkan hasil pengelompokan, diperoleh tiga *cluster* yang menunjukkan pola kecemasan statistik yang berbeda pada mahasiswa. Setiap *cluster* tidak hanya berbeda dalam tingkat kecemasan secara umum, tetapi juga menunjukkan kombinasi karakteristik kecemasan yang khas pada masing-masing dimensi. Adapun anggota untuk tiap *cluster* dapat dilihat pada Tabel 1 sebagai berikut.

**Tabel 1.** Hasil *K-Modes Clustering*  $k = 3$

Cluster	Anggota	Jumlah Anggota
1	Administrasi Pendidikan (S1), Agribisnis (S1), Agroteknologi (S1), Bahasa dan Sastra Inggris (S1), Biologi (S1), Bisnis Digital (S1), Desain Komunikasi Visual (S1), Fisika (S1), Geografi (S1), Ilmu Administrasi Negara (S1), Ilmu Aktuaria (S1), Ilmu Hukum (S1), Ilmu Keolahragaan (S1), Ilmu Komunikasi (S1), Kedokteran (S1), Keperawatan (S1), Kimia (S1), Manajemen Bisnis Syariah (S1), Matematika (S1), Pariwisata (S1), Pendidikan Agama Islam (S1), Pendidikan Bahasa dan Sastra Indonesia (S1), Pendidikan Bahasa Inggris (S1), Pendidikan Bahasa Jepang (S1), Pendidikan Biologi (S1), Pendidikan Ekonomi (S1), Pendidikan Geografi (S1), Pendidikan Guru Pendidikan Anak Usia Dini (S1), Pendidikan Ilmu Pengetahuan Sosial (S1), Pendidikan Kepelatihan Olahraga (S1), Pendidikan Kesejahteraan Keluarga (S1), Pendidikan Kimia (S1), Pendidikan Luar Biasa (S1), Pendidikan Matematika (S1), Pendidikan Musik (S1), Pendidikan Seni Drama, Tari dan Musik (S1), Pendidikan Seni Rupa (S1), Pendidikan Sosiologi (S1), Pendidikan Tari (S1), Pendidikan Teknik Elektro (S1), Pendidikan Teknik Mesin (S1), Pendidikan Vokasional Desain Fashion (S1), Perpustakaan dan Ilmu Informasi (S1), Peternakan (S1), Sastra Indonesia (S1), Teknik Elektro (S1), Teknik Geologi (S1), Teknik Mesin (S1), Teknik Pertambangan (S1), Teknik Sipil (S1), Teknologi Pendidikan (S1)	51

2	Ekonomi Pembangunan (S1), Ilmu Lingkungan (S1), Manajemen (S1), Pendidikan Ilmu Pengetahuan Alam (S1), Pendidikan Jasmani, Kesehatan dan Rekreasi (S1), Pendidikan Non Formal (S1), Pendidikan Pancasila dan Kewarganegaraan (S1), Pendidikan Sejarah (S1), Pendidikan Teknik Bangunan (S1), Pendidikan Teknik Otomotif (S1), Pendidikan Vokasional Seni Kuliner (S1), Statistika (S1)	12
3	Akuntansi (S1), Bimbingan dan Konseling (S1), Ekonomi dan Keuangan Islam (S1), Informatika (S1), Kedokteran Hewan (S1), Pendidikan Fisika (S1), Pendidikan Guru Sekolah Dasar (S1), Pendidikan Teknik Elektronika (S1), Psikologi (S1), Pendidikan Teknik Informatika (S1)	10

Untuk memastikan bahwa hasil pengelompokan yang diperoleh memiliki kualitas yang memadai, dilakukan evaluasi terhadap *cluster* menggunakan beberapa indeks validasi, yaitu *Silhouette Score*, *Dunn Index*, *Davies-Bouldin Index*, dan *Calinski-Harabasz Index*. Hasil perhitungan tersebut disajikan pada Tabel 2.

**Tabel 2.** Validasi *Cluster* k = 3

Ukuran	Nilai
Rata-rata <i>Silhouette</i>	0,5215144
<i>Dunn Index</i>	0,1666667
<i>Davies-Bouldin Index</i>	1,481991
<i>Calinski-Harabasz Index</i>	39,53342

Berdasarkan Tabel 2. nilai rata-rata *Silhouette* sebesar 0,5215 menunjukkan bahwa struktur *cluster* yang terbentuk sudah cukup baik, dengan pemisahan antar *cluster* yang relatif jelas meskipun masih terdapat beberapa data yang berada di area perbatasan. Nilai *Dunn Index* sebesar 0,1667 mengindikasikan bahwa jarak antar *cluster* masih relatif terbatas dibandingkan dengan variasi di dalam *cluster*, sehingga pemisahan antar kelompok belum sepenuhnya kuat. Sementara itu, nilai *Davies-Bouldin Index* sebesar 1,4819 menunjukkan bahwa tingkat kemiripan antar klaster masih cukup tinggi, yang menandakan adanya kedekatan karakteristik antar beberapa kelompok. Di sisi lain, nilai *Calinski-Harabasz Index* sebesar 39,5334 menunjukkan bahwa struktur *cluster* yang terbentuk sudah cukup mampu membedakan antar kelompok. Secara keseluruhan, hasil ini menggambarkan bahwa *clustering* telah menghasilkan pengelompokan yang cukup baik, namun masih perlu peningkatan kualitas pemisahan *cluster*.

Berdasarkan hasil pengelompokan yang diperoleh, terbentuk tiga *cluster* dengan tingkat kecemasan statistik yang berbeda, yaitu kecemasan tinggi, sedang, dan rendah. Setiap *cluster* tidak hanya merepresentasikan perbedaan tingkat kecemasan, tetapi juga mencerminkan variasi respons mahasiswa dalam menghadapi pembelajaran statistika. Perbedaan ini menunjukkan bahwa pengalaman mahasiswa dalam mempelajari statistika tidak bersifat homogen, melainkan bervariasi sesuai dengan karakteristik kecemasan yang dimiliki. Oleh karena itu, pengelompokan ini menjadi dasar penting untuk memahami pola kecemasan mahasiswa secara lebih terstruktur sebelum dilakukan analisis lebih lanjut terhadap masing-masing *cluster*.

*Cluster 1* merepresentasikan kelompok mahasiswa dengan tingkat kecemasan yang relatif tinggi. Mahasiswa dalam kelompok ini cenderung mengalami tekanan yang lebih besar dalam memahami konsep statistika serta dalam situasi evaluasi seperti ujian dan tugas berbasis analisis data. Selain itu, kelompok ini juga berpotensi memiliki kepercayaan diri yang lebih rendah dalam kemampuan komputasi statistika, sehingga dapat memengaruhi kemampuan mereka dalam menyelesaikan permasalahan statistik. Kondisi ini menunjukkan bahwa kecemasan pada *cluster* ini tidak hanya berkaitan dengan aspek kognitif, tetapi juga melibatkan aspek afektif seperti rasa takut dan kekhawatiran dalam pembelajaran statistika.

*Cluster 2* menunjukkan kelompok mahasiswa dengan tingkat kecemasan yang relatif rendah. Mahasiswa dalam kelompok ini cenderung memiliki persepsi yang lebih positif terhadap statistika serta tingkat kepercayaan diri yang lebih baik dalam memahami dan menggunakan konsep statistik. Mereka relatif lebih mampu beradaptasi dengan tuntutan pembelajaran statistika, baik dalam memahami materi maupun dalam menghadapi evaluasi akademik. Hal ini menunjukkan bahwa kelompok ini memiliki kesiapan yang lebih baik dalam mengikuti pembelajaran statistika.

*Cluster 3* menggambarkan kelompok mahasiswa dengan tingkat kecemasan sedang, yang berada di antara dua *cluster* lainnya. Mahasiswa dalam kelompok ini masih mengalami kecemasan pada beberapa aspek tertentu, terutama dalam situasi yang menuntut pemahaman dan penerapan konsep statistik, namun tidak secara dominan. Kelompok ini

menunjukkan adanya kondisi yang relatif stabil, tetapi tetap memerlukan dukungan dalam proses pembelajaran agar tidak berkembang menjadi kecemasan yang lebih tinggi.

Untuk memperoleh pemahaman yang lebih mendalam mengenai hasil pengelompokan, dilakukan analisis terhadap distribusi program studi pada masing-masing *cluster*. Hasilnya menunjukkan bahwa setiap *cluster* terdiri dari berbagai program studi dengan latar belakang keilmuan yang beragam, baik dari rumpun sains dan teknologi maupun sosial dan humaniora. Sebaran ini menunjukkan bahwa tidak terdapat pemisahan yang tegas antar bidang ilmu dalam pembentukan *cluster* kecemasan statistik mahasiswa.

Kondisi tersebut mengindikasikan bahwa kecemasan terhadap statistika merupakan fenomena yang dialami oleh mahasiswa dari berbagai disiplin ilmu, tidak terbatas pada kelompok keilmuan tertentu. Hal ini menunjukkan bahwa kecemasan statistik tidak secara langsung berkaitan dengan bidang studi tertentu, melainkan dapat muncul pada berbagai kelompok mahasiswa.

Dari sudut pandang pembelajaran, temuan ini memberikan implikasi bahwa pendekatan pembelajaran statistika yang bersifat umum dan seragam untuk berbagai program studi dengan rumpun keilmuan yang berbeda-beda mungkin kurang efektif untuk mengakomodasi kebutuhan seluruh mahasiswa. Mahasiswa dari berbagai program studi dapat mengalami kecemasan dengan pola yang berbeda, sehingga diperlukan strategi pembelajaran yang lebih fleksibel dan adaptif. Selain itu, upaya untuk mengurangi kecemasan statistik tidak cukup hanya difokuskan pada kelompok mahasiswa tertentu, melainkan perlu diterapkan secara lebih luas pada berbagai program studi. Pendekatan yang dapat dipertimbangkan antara lain peningkatan dukungan pembelajaran, penyederhanaan penyampaian konsep, serta penciptaan lingkungan belajar yang lebih suportif bagi mahasiswa dalam mempelajari statistika.

#### IV. KESIMPULAN

Berdasarkan hasil analisis *clustering* menggunakan metode *k-modes* terhadap enam dimensi kecemasan statistika dalam instrumen *Statistical Anxiety Rating Scale* (STARS), diperoleh tiga kelompok mahasiswa dengan karakteristik tingkat kecemasan yang berbeda, yaitu kecemasan tinggi, sedang, dan rendah. Temuan ini menunjukkan bahwa kecemasan statistik pada mahasiswa tidak bersifat homogen, melainkan memiliki pola yang beragam pada setiap kelompok. Perbedaan ini mengindikasikan bahwa mahasiswa memiliki respons yang berbeda dalam menghadapi pembelajaran statistika, baik dari aspek pemahaman, evaluasi, maupun kepercayaan diri dalam menggunakan konsep statistik.

Hasil evaluasi menggunakan beberapa indeks validasi, yaitu *Silhouette*, *Dunn Index*, *Davies-Bouldin Index*, dan *Calinski-Harabasz Index*, menunjukkan bahwa struktur *cluster* yang terbentuk tergolong cukup baik, meskipun pemisahan antar *cluster* belum sepenuhnya optimal. Hal ini mengindikasikan bahwa masih terdapat kedekatan karakteristik antar sebagian kelompok, sehingga batas antar *cluster* belum sepenuhnya tegas. Oleh karena itu, pengembangan metode *clustering* yang lebih lanjut masih diperlukan untuk memperoleh hasil pengelompokan yang lebih optimal dan representatif.

Secara umum, hasil penelitian ini memberikan gambaran mengenai variasi tingkat kecemasan statistik pada mahasiswa serta pentingnya mempertimbangkan perbedaan karakteristik tersebut dalam proses pembelajaran. Pendekatan pembelajaran yang tidak seragam berpotensi lebih efektif dalam membantu mahasiswa mengatasi kecemasan statistik sesuai dengan kondisi masing-masing kelompok. Dengan demikian, hasil penelitian ini dapat menjadi dasar dalam merancang strategi pembelajaran statistika yang lebih adaptif dan tepat sasaran.

#### UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Lembaga Penelitian dan Pengabdian Masyarakat Universitas Negeri Padang atas pendanaan penelitian ini dengan nomor kontrak:1956/UN34.15/LT/2025.

#### DAFTAR PUSTAKA

- Afdal, A., Alizamar, A., Ilyas, A., Zikra, Z., Taufik, T., Erkamsyah, E., & Dkk. (2019). Contribution of statistical anxiety to student learning outcomes : Study in Universitas Negeri Padang. *Journal of Physics: Conference Series*. <https://doi.org/10.1088/1742-6596/1157/4/042126>
- Ames, H., Glenton, C., & Lewin, S. (2019). Purposive sampling in a qualitative evidence synthesis: A worked example from a synthesis on parental perceptions of vaccination communication. *BMC Medical Research Methodology*, 19(1), 26. <https://doi.org/10.1186/s12874-019-0665-4>

- Campbell, S., Greenwood, M., Prior, S., Shearer, T., Walkem, K., Young, S., Bywaters, D., & Walker, K. (2020). Purposive sampling: Complex or simple? Research case examples. *Journal of Research in Nursing*, 25(8), 652–661
- Cruise, J. ., Cash, R. ., & Bolton, L. . (1985). Development and validation of an instrument to measure statistical anxiety. *Proceedings of the Section on Statistical Education*.
- Frias-Navarro, D., Monterde-i-Bort, H., Navarro-Gonzalez, N., Molina-Palomero, O., Pascual- Soler, M., PerezGonzalez, J., & Longobardi, C. (2018). Statistics anxiety in university students in assessment situations. *Proceedings of the 4th International Conference on Higher Education Advances (HEAd'18)*, 1–8. <https://doi.org/10.4995/HEAD18.2018.7990>
- Han, J., Kamber, M., & PeI, J. (2012). *Data Mining: Concepts and Techniques*. *Data Mining: Concepts and Techniques*. <https://doi.org/10.1016/C2009-0-61819-5>
- Hanna, D., Shevlin, M., & Dempster, M. (2008). The structure of the statistics anxiety rating scale: A confirmatory factor analysis using UK psychology students. *Personality and Individual Differences*, 45(1), 68–74. <https://doi.org/10.1016/j.paid.2008.02.021>
- Huang, Z. (1998). Extensions to the k -Means Algorithm for Clustering Large Data Sets with Categorical Values. *Data Mining and Knowledge Discovery*, 2(3), 283–304.
- Levpušček, M.P, & Cukon, M. (2020). That Old Devil Called ‘Statistics’: Statistics Anxiety in University Students and Related Factors. *Center for Educational Policy Studies Journal*. <https://doi.org/10.26529/cepsj.826>
- Macher, D., Papousek, I., Ruggeri, K., & Paechter, M. (2015). Statistics anxiety and performance: Blessings in disguise. *Frontiers in Psychology*, 6(AUG), 4–7. <https://doi.org/10.3389/fpsyg.2015.01116>
- Maricic, M., Rajkovic, T., Lecic-cvetkovic, D., & Ignjatovic, M. (2023). Statistics Anxiety: What are the antecedents ?, (10), 1–8. <https://doi.org/10.11159/icsta23.104>
- Maat, S.M., Mohd, N., Che Ahmad, C. N., & Puteh, M. (2022). Statistical Anxiety and Achievement among University Students. *TEM Journal*, 420–426. <https://doi.org/10.18421 /TEM111-53>
- Nielsen, T., & Kreiner, S. (2018). Measuring statistical anxiety and attitudes toward statistics: The development of a comprehensive Danish instrument (HFS-R). *Cogent Education*, 5(1), 1–19. <https://doi.org/10.1080/2331186X.2018.1521574>
- Nyimbili, F., & Nyimbili, L. (2024). Types of Purposive Sampling Techniques with Their Examples and Application in Qualitative Research Studies. *British Journal of Multidisciplinary and Advanced Studies*, 5(1), 90–99. <https://doi.org/10.37745 /bjmas.2022.0419>
- Onwuegbuzie, A. J. (2004). Academic procrastination and Statistics Anxiety. *Assessment and Evaluation in Higher Education*, 29, 3–19.
- Paechter, M., Macher, D., Martskvishvili, K., Wimmer, S., & Papousek, I. (2017). Mathematics Anxiety and statistics anxiety. Shared but also unshared components and antagonistic contributions to performance in statistics. *Frontiers in Physiology*, 8, 1–13. <https://doi.org/10.3389/fpsyg.2017.01196>
- Sugiyono. (2017). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.