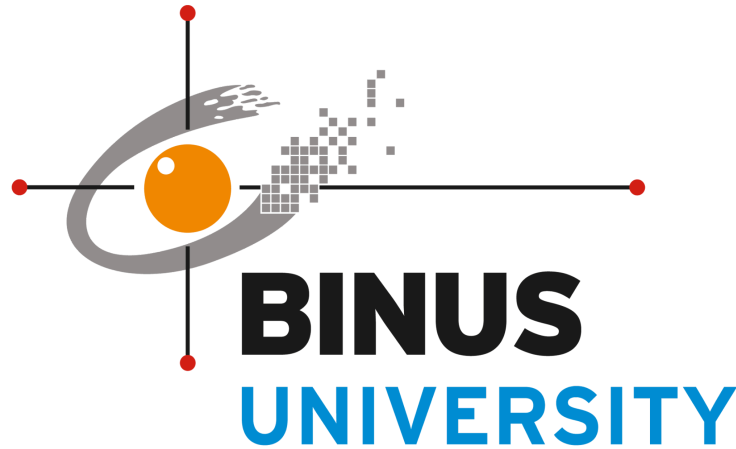


LAPORAN AKHIR
ANALISIS PREFERENSI SESEORANG DALAM PEMBELIAN KOPI
MATAKULIAH SURVEY AND SAMPLING METHODS
Erna Fransisca Angela Sihotang, S.Stat, M.Kom.



Disusun oleh :
Kelompok 10 LA09

Nama anggota :

Cherylene Callista Reksohartono	2602087024
Crysantha Monica Lim	2602090076
Sharon Selina Yoe	2602059595

DATA SCIENCE PROGRAM
SCHOOL OF COMPUTER SCIENCE

UNIVERSITAS BINA NUSANTARA

JAKARTA

2023/2024

Pembagian Jobdesk :

1. **Crysantha Monica Lim : Mengerjakan BAB 1, 2, dan 6**
2. **Cherylene Calista Reksohartono : Mengerjakan BAB 4, 5**
3. **Sharon Selina Yoe : Mengerjakan BAB 3, 7, dan membantu merapikan BAB 6**

I. INTRODUCTION

A. Background

Kopi memiliki peran yang signifikan dalam budaya global modern, dengan berbagai jenis dan merek yang memenuhi preferensi konsumen terkait rasa, harga, dan merek. Pemahaman mendalam terhadap preferensi ini sangat penting bagi produsen dan pemasar untuk meningkatkan kepuasan pelanggan dan loyalitas merek. Penelitian ini dalam konteks *Survey and Sampling* memberikan data akurat untuk mendukung keputusan strategis dalam industri kopi, dengan menggunakan metode yang tepat untuk memahami faktor-faktor yang mempengaruhi keputusan pembelian, menerapkan strategi pemasaran efektif, dan mengembangkan keterampilan analitik yang relevan.

B. Problem Statement

Dengan meningkatnya konsumsi kopi global, pasar semakin kompetitif, mendorong produsen dan pemasar untuk memahami preferensi konsumen secara mendalam. Kurangnya data representatif dan valid mengenai faktor-faktor yang mempengaruhi keputusan pembelian menjadi hambatan utama. Penelitian ini bertujuan mengidentifikasi faktor seperti rasa, harga, dan merek yang mempengaruhi preferensi pembelian kopi, dengan menggunakan metode *survey dan sampling* untuk data yang akurat dan representatif. Hasilnya memberikan wawasan penting bagi industri kopi dan berkontribusi pada pemahaman akademis tentang preferensi konsumen.

C. Research Objectives and Measurements

Dalam survei ini, pengukuran akan melibatkan penggunaan pertanyaan dan instrumen untuk mengumpulkan data tentang preferensi pembelian kopi dari responden. Variabel yang akan diukur meliputi usia, gender, tempat tinggal, frekuensi konsumsi kopi, preferensi harga kopi olah sendiri, preferensi harga kopi beli di luar, metode pembuatan kopi, jenis biji kopi, faktor pembelian kopi, cara pembuatan kopi, tempat pembelian kopi, frekuensi eksplorasi kedai/kopi baru, dan pengalaman pembelian kopi. Penelitian ini bertujuan untuk menilai hubungan

antara variabel-variabel ini dan preferensi pembelian kopi, serta memahami pola konsumsi kopi di kalangan responden.

II. SAMPLING DESIGN

A. Sampling Design

Penelitian ini menggunakan *Simple Random Sampling* untuk mempelajari preferensi pembelian kopi di kalangan siswa/i, mahasiswa, dan pekerja. Metode ini dipilih karena memberikan peluang yang sama bagi setiap anggota populasi, memastikan data representatif. Dalam *Life Cycle of A Survey*, *Simple Random Sampling* adalah langkah awal penting dalam fase *Measurement* untuk memastikan validitas hasil. Setelah sampel terpilih, data preferensi pembelian kopi dikumpulkan. Tahap berikutnya, seperti *Postsurvey Adjustments* dan *Survey Statistics*, menggunakan data ini untuk analisis lebih lanjut, menjadikan *Simple Random Sampling* sebagai fondasi keakuratan dan validitas hasil survei. Gambar dapat dilihat pada lampiran (Gambar 1.1 *Survey lifecycle from a design perspective (Measurement)*).

Penelitian ini mengukur preferensi pembelian kopi melalui variabel seperti usia, tempat tinggal, frekuensi konsumsi kopi, preferensi harga kopi, metode pembuatan kopi, jenis biji kopi, faktor pembelian, tempat pembelian, frekuensi eksplorasi kedai/kopi baru, dan pengalaman pembelian. Pengukuran dilakukan menggunakan pertanyaan tertutup dan beberapa pertanyaan terbuka. Setelah respons diperoleh, dilakukan proses edited response yang mencakup review, perbaikan, dan validasi data sebelum analisis lebih lanjut.

B. Sampling Frame

Sampling frame adalah daftar anggota populasi target yang berpeluang dipilih sebagai sampel. Dalam penelitian ini, *sampling frame* mencakup kontak lengkap dari siswa, mahasiswa, dan pekerja. Tujuannya adalah memastikan setiap anggota populasi memiliki peluang yang sama untuk dipilih, sehingga sampel representatif dan hasil survei dapat digeneralisasi.

C. Sample Size Calculation

Sample size atau ukuran sampel dalam penelitian ini dihitung berdasarkan beberapa faktor kunci seperti *Confidence Level* (tingkat kepercayaan), *Margin of Error* (marginal error), dan *Population Proportion* (proporsi populasi). Dalam penelitian ini, ukuran sampel yang diperlukan adalah 80 responden. Perhitungan ini mempertimbangkan tingkat kepercayaan 95% ($\alpha = 5\%$), *margin of error* sebesar 10% ($e = 10\%$), dan proporsi populasi yang diestimasi dari data sebelumnya sebesar 70% ($\hat{p} = 70\%$).

Perhitungan *sample size* didapatkan dari rumus berikut :
$$n = \frac{z_{\alpha/2}^2 \hat{p}(1 - \hat{p})}{e^2}$$

Dengan perhitungan, $\frac{(1,96)^2 \times 0,70(1 - 0,70)}{(0,1^2)}$

Setelah melakukan perhitungan, ukuran sampel yang diperoleh adalah 80,6736 yang dibulatkan menjadi 80 responden. Angka ini dianggap memadai untuk memberikan representasi yang *valid* terhadap populasi target, meminimalkan kesalahan sampling, dan memungkinkan analisis statistik yang dapat diandalkan terhadap preferensi konsumsi kopi di antara siswa/i, mahasiswa, dan pekerja yang menjadi fokus penelitian ini.

D. Random Sample Selection

Proses pemilihan sampel secara acak dengan *Simple Random Sampling* penting dalam survei untuk mewakili populasi siswa, mahasiswa, dan pekerja dengan tepat. Metode ini memberikan setiap individu peluang yang sama untuk dipilih, mengurangi bias dan memungkinkan hasil survei yang dapat digeneralisasi.

Berikut merupakan penjelasan *Life Cycle of A Survey from a Design Perspective* dari proses desain sampling yang dilakukan pada penelitian ini (bagian *representation*). Gambar dapat dilihat pada lampiran (Gambar 1.2 *Survey lifecycle from a design perspective (Representation)*).

Target population survei ini terdiri dari siswa/i, mahasiswa, dan pekerja yang akan dipelajari mengenai preferensi konsumsi kopi mereka. *Sampling frame* mencakup daftar kontak lengkap dari ketiga kelompok ini yang berpotensi menjadi responden survei. Peneliti akan menggunakan metode *simple random sampling* untuk memilih 80 individu dari populasi ini, berdasarkan *Confidence*

Level, Margin of Error, Ukuran Populasi, dan Proporsi Populasi, untuk berpartisipasi dalam survei. Responden adalah individu yang memberikan jawaban lengkap dan sesuai dengan pertanyaan survei, sementara nonresponden adalah mereka yang tidak merespons atau memberikan jawaban tidak lengkap atau ambigu. Data dari nonresponden akan dicatat untuk analisis lebih lanjut atau pengecualian dari analisis jika diperlukan. Setelah pengumpulan data, tahap *Postsurvey Adjustments* diperlukan untuk menangani *non-response bias*, *undercoverage*, *misclassification*, dan *outlier*, serta untuk membersihkan data guna memastikan keakuratan hasil survei mengenai preferensi konsumsi kopi.

III. QUESTIONNAIRE DESIGN

A. Mapping Questions to Measurements

Dalam penelitian ini, dengan 13 variabel *measurement* terdapat pertanyaan-pertanyaan yang merupakan turunan dari *measurements*. Penentuan variabel terbagi menjadi dua, yaitu Variabel Independen, dan Variabel Dependen. Berikut merupakan pemetaan dari setiap pertanyaan kuesioner:

A. Usia (Variabel Independen)

Usia dapat mempengaruhi preferensi kopi karena berbeda usia seringkali memiliki kebiasaan, selera, dan kebutuhan yang berbeda

1. Berapa Usia Anda? (Di bawah 18 tahun, 18 - 24 tahun, 25 - 34 tahun, 35 - 44 tahun, 45 - 54 tahun, 55 tahun ke atas)

B. Gender (Variabel Independen)

1. Apa jenis kelamin Anda? (Laki - laki, Perempuan)

C. Tempat Tinggal (Variabel Independen)

Lokasi geografis atau tempat tinggal seseorang dapat mempengaruhi aksesibilitas terhadap berbagai jenis kopi dan kedai kopi

1. Dimanakah tempat tinggal Anda saat ini? (Kota Besar (Contoh: Pusat Kota, Kawasan Perkotaan, Daerah Penyangga), Kota Kecil

(Contoh: Ibukota Kabupaten, Kecamatan Pusat, Kawasan Permukiman Padat Penduduk, Kawasan Non-Kota (Contoh: Kawasan Perumahan, Kawasan Industri))

D. Frekuensi Konsumsi Kopi (Variabel Dependen)

Mengukur seberapa sering seseorang mengonsumsi kopi memberikan wawasan tentang kebiasaan mereka dan seberapa besar minat mereka terhadap kopi. Ini juga bisa membantu dalam segmentasi pasar.

1. Seberapa sering Anda mengonsumsi kopi? (Setiap hari, Beberapa kali dalam seminggu, Seminggu sekali, Dalam situasi tertentu / tidak pasti)
2. Dalam seminggu, berapa banyak kopi yang Anda konsumsi? (rating 1 - 10)

E. Preferensi Harga Kopi Olah Sendiri (Variabel Dependen)

Mengetahui preferensi harga untuk kopi yang diolah sendiri membantu memahami nilai yang diberikan konsumen pada kopi rumahan dan mempengaruhi strategi harga penjualan biji kopi ataupun alat mesinnya.

1. Berapa nominal yang Anda dapat keluarkan jika Anda mengolah kopi sendiri? (< Rp. 8.000,00, Rp. 8.000,00 - Rp. 39.999,00, Rp. 40.000,00 - Rp. 79.999,00, Rp. 80.000,00 - Rp. 119.999,00, Rp. 120.000,00 - Rp. 299.999,00, Rp. 300.000,00 - Rp. 500.000,00, > Rp. 500.000,00)

F. Preferensi Harga Kopi Beli di Luar (Variabel Dependen)

Mengetahui informasi tentang seberapa banyak konsumen bersedia membayar untuk kopi di luar rumah adalah untuk melihat sensitivitas harga atau perilaku belanja konsumen.

1. Berapa nominal yang Anda dapat keluarkan jika membeli kopi di luar? (< Rp. 8.000,00, Rp. 8.000,00 - Rp. 14.999,00, Rp. 15.000,00 - Rp. 29.999,00, Rp. 30.000,00 - Rp. 49.999,00, Rp. 50.000,00 - Rp. 99.999,00, > Rp. 100.000,00)

15.000,00 - Rp. 19.999,00, Rp. 20.000,00 - Rp. 29.999,00, Rp.
30.000,00 - Rp. 40.000,00, > Rp. 40.000,00)

G. Metode Pembuatan Kopi (Variabel Dependen)

Metode Pembuatan Kopi menunjukkan preferensi seseorang dalam mengolah kopi dan membantu produsen alat kopi untuk memahami permintaan pasar, serta mengarahkan inovasi produk.

1. Metode pembuatan kopi apa yang biasa Anda gunakan dalam mengkonsumsi kopi tersebut? (Diseduh, Manual brew, Machine brew)

H. Jenis Biji Kopi (Variabel Dependen)

Jenis biji kopi yang disukai konsumen memberikan informasi tentang preferensi rasa dan kualitas. Hal ini dapat membantu petani dan pemasok kopi dalam memenuhi kebutuhan pasar yang spesifik. Selain itu juga dapat melihat perusahaan apa yang paling disukai oleh konsumen

1. Jenis biji kopi apa yang paling sesuai dengan preferensi Anda? (Arabica, Robusta, Liberica, Excelsa, Campuran (Blend), Tidak tahu / tidak ada preferensi khusus)
2. Dari perusahaan kopi di bawah ini, manakah yang paling cocok dengan preferensi kopi yang Anda konsumsi? (Starbucks, Kopi kenangan, Janji Jiwa, Fore, Tidak ada yang cocok)

I. Faktor Pembelian Kopi (Variabel Dependen)

Mengetahui faktor-faktor yang mempengaruhi pembelian kopi membantu dalam memahami motivasi konsumen dan dapat digunakan untuk mengembangkan strategi pemasaran menjadi lebih efektif.

1. Apa faktor terpenting bagi Anda saat membeli kopi? (Rasa, Harga, Kualitas bahan baku, Kenyamanan lokasi, Aroma, Others...)

J. Cara Pembuatan Kopi (Variabel Dependen)

Preferensi terhadap cara pembuatan kopi memberikan wawasan tentang kebiasaan dan kenyamanan konsumen, penting juga untuk bisnis kopi ataupun kedai kopi.

1. Jenis kopi apa yang biasa Anda minum? (Espresso, Latte / Cappuccino, Kopi hitam, Kopi instan, Frappuccino, Others...)

K. Tempat Pembelian Kopi (Variabel Dependen)

Tempat pembelian kopi mengungkap dimana konsumen lebih sering membeli kopi untuk membantu pemasok dan pengecer dalam menentukan saluran distribusi yang efektif.

1. Dimana Anda biasa membeli kopi? (Kedai kopi, Toko kelontong / supermarket, Kafe / Restoran, Mesin penjual otomatis, Others...)

L. Frekuensi Eksplorasi Kedai / Kopi baru (Variabel Dependen)

Mengukur seberapa sering konsumen mencoba kedai kopi atau varian kopi baru menunjukkan tingkat minat mereka terhadap inovasi dan pengalaman baru. Ini dapat membantu kedai kopi dalam menarik dan mempertahankan pelanggan dengan menawarkan pengalaman yang menarik.

1. Seberapa sering Anda mencoba jenis kopi baru atau kedai kopi baru? (Sangat sering, Cukup sering, Jarang, Tidak pernah)
2. Apakah Anda mengikuti tren kopi terbaru atau tetap dengan pilihan tradisional? (Mengikuti tren, Tetap dengan tradisi, Kombinasi keduanya)

M. Pengalaman Pembelian Kopi (Variabel Dependen)

Pengalaman pembelian kopi, baik positif maupun negatif, dapat mempengaruhi keputusan pembelian di masa depan. Memahami hal ini dapat membantu dalam meningkatkan kualitas layanan dan produk untuk meningkatkan kepuasan pelanggan.

1. Bagaimana pengalaman Anda saat mengunjungi kedai kopi atau membeli kopi secara online? (Sangat baik, Cukup baik, Kurang baik, Sangat buruk)
2. Apakah Anda akan merekomendasikan kedai kopi favorit Anda kepada teman atau keluarga? (Ya, Tidak)
3. Apa yang dapat dilakukan oleh kedai kopi untuk meningkatkan pengalaman Anda sebagai pelanggan? (Long answer text)

B. Additional Questions

Adapun beberapa pertanyaan yang dirancang untuk mengumpulkan data demografis dan preferensi konsumsi kopi dari responden untuk proses analisis yang relevan dan segmentasi pasar yang efektif. Dengan meminta data pribadi singkat, kuesioner ini memastikan validitas data, mencegah duplikasi, dan mengarahkan pertanyaan hanya kepada responden yang relevan. Berikut pertanyaan tambahan dalam kuesioner:

1. Cantumkan inisial nama Anda! (short answer)

Meminta inisial nama adalah cara untuk melihat tanggapan responden untuk pengolahan data dan bisa membantu dalam memastikan bahwa tidak ada duplikasi data tanpa melanggar privasi.

2. Apakah Anda mengonsumsi kopi? (Ya, Tidak)

Mengetahui apakah seseorang mengonsumsi kopi menentukan apakah responden relevan dengan penelitian kami. Jika mereka tidak mengonsumsi kopi, banyak pertanyaan lain yang mungkin tidak relevan, dan mereka diarahkan ke bagian lain atau diakhiri lebih awal.

3. Apakah Anda alergi terhadap kopi atau tidak minum kopi karena alasan kesehatan? (Ya, Tidak)

Memahami apakah seseorang menghindari kopi karena alergi atau alasan kesehatan penting untuk memahami batasan yang mempengaruhi preferensi atau

konsumsi mereka. Hal ini dapat membantu dalam menginterpretasikan data dengan lebih akurat, karena alasan kesehatan bisa menjadi faktor penting dalam mengapa seseorang tidak mengonsumsi kopi.

IV. DATA COLLECTION

A. Data Collection Method

Peneliti menggunakan *web surveys* dengan *Google Form* yang dibagikan online kepada siswa, mahasiswa, dan pekerja. Metode ini dipilih karena efisien untuk mengumpulkan data dari target responden yang tinggi dan memungkinkan jangkauan nasional, termasuk daerah terpencil, selama ada akses internet. Dengan demikian, data yang dikumpulkan lebih representatif dan tidak *bias* wilayah.

B. Data Collection Process

Selama pengumpulan data, jumlah responden melebihi target dengan variabilitas umur yang tinggi, namun mayoritas berasal dari kota besar, kurang mewakili wilayah terpencil. Terdapat *measurement error* karena beberapa pertanyaan dan pilihan jawaban ambigu, serta *scaling* yang digunakan tidak sesuai dengan teknik analisis, menyebabkan masalah dalam proses analisis yang akan dibahas di bab berikutnya.

C. Efforts to Increase Response Rate

Untuk pengisian survey, kuesioner dikirimkan secara *personal* ke kontak yang terdapat dalam daftar kontak, dan kemudian disebarkan ke teman - teman terdekat mereka untuk memperbanyak peluang meningkatnya responden. Untuk meningkatkan peluang tersebut, diberikan *incentive* atau hadiah berupa kopi gratis kepada satu responden beruntung. Dengan metode ini, jumlah responden berhasil meningkat hingga mencapai sekitar dua kali lebih banyak dari target responden.

D. Response and Non-Response Data

Responden yang tidak minum kopi karena alasan kesehatan akan diarahkan ke akhir kuesioner dan diminta untuk tidak melanjutkan pengisian karena jawaban mereka tidak akan valid. Secara total, terkumpul sebanyak 178 responden. Dari jumlah tersebut, 171 respons merupakan data yang berhasil diambil, sementara sisanya berasal dari responden yang tidak minum kopi karena

alasan kesehatan. Selain itu, tidak terdapat *non-response data*, karena setiap pertanyaan dalam kuesioner bersifat wajib dan tidak dapat dilewati tanpa diisi.

V. MEASUREMENT QUALITY

A. Validity Calculation

Validity adalah proses evaluasi untuk menentukan sejauh mana suatu instrumen mengukur konstruk yang dimaksud secara akurat. Validity dapat dihitung dengan korelasi antar jawaban. Korelasi tinggi menunjukkan bahwa pertanyaan mengukur konstruk yang sama, sedangkan korelasi rendah menunjukkan sebaliknya. Korelasi positif (menuju 1) menunjukkan hubungan searah, korelasi negatif (menuju -1) menunjukkan hubungan berlawanan, dan 0 menunjukkan tidak ada korelasi. Hasil perhitungan validity pada kuesioner adalah sebagai berikut. (Gambar 1.3. Hasil perhitungan *validity* dengan korelasi)

Berdasarkan hasil korelasi tersebut, terdapat variabel yang memiliki korelasi cukup tinggi terhadap variabel lain. Namun untuk beberapa variabel, memiliki nilai korelasi yang cukup rendah dengan variabel lainnya. Ini menunjukkan bahwa data yang dihasilkan dari kuesioner kurang berhubungan satu dengan yang lain dan belum menggambarkan konstruk yang direncanakan di awal. (Gambar 1.4. dan 1.5 Bagian pertama dan kedua hasil pemeriksaan *validity* dengan Pearson's Correlation Coefficient)

Perhitungan *validity* juga dilakukan dengan Pearson's Correlation Coefficient. Validitas dilihat dari apabila r dari data lebih besar daripada r yang sebenarnya dari tabel r korelasi Pearson. Berdasarkan hasil di atas, ditunjukkan bahwa kolom gender, tempat_tinggal, jenis_kopi, jenis_biji, dan pengalaman bersifat tidak *valid*. Kesimpulannya, kelima variabel tersebut kurang mampu dalam mengukur konstruk yang direncanakan secara akurat.

B. Reliability Calculation

Reliability mengukur konsistensi jawaban responden. Jika survei yang sama diulang dalam kondisi yang sama dan menghasilkan hasil serupa, survei

$$\alpha = \frac{k\bar{r}}{1 + (k-1)\bar{r}}$$

tersebut memiliki reliabilitas tinggi. Reliabilitas diukur menggunakan Cronbach's alpha, dengan nilai mendekati 1 menunjukkan reliabilitas tinggi.

Perhitungan reliabilitas survei ini dilakukan dua kali. Pertama, menggunakan Excel dengan 12 dari 20 variabel, menghasilkan reliabilitas 0.1391 yang termasuk *unacceptable*. Kedua, dengan 15 variabel, menghasilkan reliabilitas 0.3534, juga *unacceptable*. Hasil ini menunjukkan data dari kuesioner belum dapat diandalkan. *Measurement error* terjadi karena ketidaksesuaian skala jawaban, menyulitkan penentuan hubungan antar variabel dan nilai reliabilitas yang akurat. Pertanyaan yang ambigu menyebabkan kebingungan dan inkonsistensi jawaban responden. Meskipun semua pertanyaan wajib diisi, kemungkinan responden memberikan jawaban yang tidak sesuai dengan preferensi sebenarnya tetap ada.

C. Response Bias Calculation

Response bias terjadi ketika respon yang diberikan tidak sesuai dengan nilai sebenarnya (*true value*) dari apa yang diukur. Variabel utama yang dipilih untuk diperiksa *bias* nya yaitu jumlah_kopi (berapa jumlah gelas kopi yang diminum dalam seminggu). Untuk pendekatan true value dilakukan dengan dua cara :

$$Y_{it} = \mu_i + \varepsilon_{it}$$

1. Perhitungan dengan rumus berikut :

Hasil perhitungan dengan rumus tersebut menghasilkan *true value* bernilai 5.0411, dan untuk rata - rata dari data untuk jumlah_kopi yaitu 5.1411. Dari perhitungan tersebut dapat dilihat bahwa hanya terdapat perbedaan yang kecil dari *true value* dan data dari kuesioner.

2. Penelitian yang dilakukan pihak eksternal

Response bias terjadi ketika respon tidak mencerminkan nilai sebenarnya yang diukur. Variabel utama yang diperiksa adalah jumlah_kopi (jumlah gelas kopi yang diminum per minggu). Dengan true value 5.0411 dan rata-rata data 5.1411, terdapat perbedaan kecil. Berdasarkan penelitian eksternal, konsumsi kopi responden serupa dengan data kuesioner.

Perhitungan bias dilakukan dengan rumus berikut :

$$Bias(\bar{Y}) = \sum_i \left(\frac{\sum_t Y_{it}}{N} \right) - \frac{\sum_i \mu_i}{N}$$

Ada tidaknya *bias* dalam data ditunjukkan dari tinggi atau rendahnya hasil perhitungan skor *bias*.

Perhitungan *bias* menggunakan dua pendekatan. Skor *bias* dari pendekatan pertama adalah 0.0999, menunjukkan *bias* rendah dan akurasi tinggi. Namun, pendekatan kedua dengan asumsi true value dua kopi per hari menghasilkan skor *bias* -8.8588, menunjukkan data yang terlalu rendah. Kesimpulannya, terdapat *bias* pada variabel jumlah_kopi karena pilihan jawaban terbatas hingga 10, sehingga responden tidak dapat menjawab sesuai preferensi sebenarnya.

VI. PRELIMINARY ANALYSIS

A. Descriptive Analysis

1. Missing Value Calculation

Dalam analisis data survei, menangani nilai yang hilang adalah penting untuk keakuratan dan keandalan hasil. Nilai yang hilang dapat terjadi karena responden tidak menjawab atau kesalahan dalam pengumpulan data. Langkah pertama adalah mengidentifikasi lokasi dan jumlah nilai yang hilang dengan menghitung jumlah atau persentasenya untuk setiap variabel. Dalam analisis ini, digunakan kode ``columnData.isna().sum()`` untuk menghitung jumlah nilai yang hilang di setiap kolom dataset. Terdapat beberapa variabel dalam survei ini yang memiliki nilai yang hilang. Secara khusus, variabel Alergi_kopi memiliki nilai yang hilang sebanyak 152, sementara variabel lainnya memiliki nilai yang hilang sebanyak 7. Variabel-variabel dengan nilai yang hilang termasuk Frekuensi_kopi, Jumlah_kopi, Nominal_olah, Nominal_beli, Metode, Jenis_biji, Faktor_beli, Jenis_kopi, Perusahaan_kopi, Preferensi_tempat, Tingkat_percobaan, Preferensi_tren, Pengalaman, Preferensi_merekomendasikan, dan Masukan_peningkatan. Variabel Umur, Gender, dan Tempat_tinggal tidak memiliki nilai yang hilang.

Dalam penelitian ini, nilai yang hilang muncul karena konteks pertanyaan “Apakah Anda mengonsumsi kopi?”. Responden yang menjawab "tidak" diarahkan ke pertanyaan konfirmasi tentang alergi atau alasan kesehatan, yang

jika dijawab "ya" langsung menyelesaikan survei. Ini menyebabkan 7 nilai hilang karena 7 responden alergi atau tidak minum kopi karena alasan kesehatan.

2. Summary Statistics

Untuk mendapatkan pemahaman yang lebih mendalam mengenai karakteristik populasi dan preferensi responden dalam survei ini, dilakukan analisis deskriptif terhadap data yang dikumpulkan. Analisis deskriptif membantu dalam mengidentifikasi pola, distribusi, dan tren dalam data, serta menyediakan gambaran umum tentang variabel-variabel utama yang diteliti.

Dataset ini mencakup informasi dari 178 responden yang terdiri dari siswa, mahasiswa, dan pekerja, yang mengungkapkan preferensi mereka terhadap konsumsi kopi. Variabel umur menunjukkan mayoritas responden berusia antara 45 hingga 54 tahun, dengan sebagian besar (90 orang) adalah laki-laki. Sebagian besar responden (133 orang) tinggal di kota besar. Frekuensi konsumsi kopi bervariasi, dengan 85 responden mengonsumsi kopi setiap hari. Jumlah kopi yang dikonsumsi per minggu memiliki rata-rata sebesar 5.12 dengan variasi yang signifikan dari 1 hingga 10 cangkir. Preferensi biaya untuk mengolah kopi sendiri dan membeli kopi dari luar juga bervariasi, dengan sebagian besar memilih kategori biaya tertentu. Metode yang paling umum digunakan untuk membuat kopi adalah dengan cara diseduh, sementara jenis biji kopi yang paling disukai adalah yang tidak memiliki preferensi khusus. Faktor yang paling dipertimbangkan dalam pembelian kopi adalah rasa, dengan sebagian besar responden lebih menyukai jenis kopi seperti Latte/Cappuccino dari perusahaan kopi terkemuka seperti Starbucks. Tingkat eksplorasi kedai atau jenis kopi baru cenderung jarang, sementara pengalaman pembelian kopi umumnya dinilai cukup baik. Sebagian besar responden juga bersedia merekomendasikan kopi yang mereka sukai kepada orang lain. Dataset ini memberikan wawasan mendalam tentang preferensi konsumsi kopi di kalangan populasi yang diteliti. (Hasil statistik deskriptif dapat dilihat pada Tabel 1.1 *Statistic Descriptive*.)

B. Preprocessing Methods

Sebelum melakukan analisis mendalam terhadap data yang telah dikumpulkan, langkah penting yang perlu dilakukan adalah preprocessing. Preprocessing merupakan

tahap persiapan data yang sangat krusial untuk memastikan bahwa data yang digunakan bersih, konsisten, dan siap untuk dianalisis lebih lanjut. Proses ini mencakup berbagai teknik untuk mengatasi inkonsistensi, kesalahan, dan nilai yang hilang dalam dataset, sehingga dapat memberikan hasil yang lebih akurat dan reliabel. Berikut ini adalah langkah-langkah preprocessing yang telah dilakukan pada dataset ini.

Pertama, nama kolom yang panjang dan kompleks diubah menjadi lebih sederhana dan mudah dimengerti (contoh code dapat dilihat pada lampiran Code 1.1 Mengganti nama kolom). Kolom "Berapa usia Anda?" diubah menjadi "umur" untuk memudahkan referensi dan manipulasi data. Selanjutnya, kolom "masukan_peningkatan" dihapus karena hanya berfungsi sebagai informasi tambahan dan tidak relevan untuk analisis utama. Kemudian, beberapa kolom dengan jawaban terbuka, seperti "faktor_beli", "jenis_kopi", dan "preferensi_tempat", dibersihkan dan nilai-nilainya distandarisasi untuk menjaga konsistensi data (contoh code dapat dilihat pada lampiran Code 1.2 Menghapus kolom tidak relevan). Langkah terakhir mencakup penanganan nilai hilang menggunakan berbagai metode imputasi seperti mean value imputation, sehingga dataset menjadi lebih bersih dan siap untuk analisis statistik mendalam, memastikan keakuratan, konsistensi, dan reliabilitas data.

VII. CONCLUSION

A. Research Conclucions

Proses riset telah memberikan gambaran komprehensif tentang preferensi konsumsi kopi di kalangan siswa, mahasiswa, dan pekerja. Pengumpulan data melalui kuesioner telah mengidentifikasi variabel penting seperti frekuensi konsumsi, preferensi harga, metode pembuatan kopi, dan faktor pembelian. Penanganan missing values dan normalisasi data telah meningkatkan kualitas dataset, memastikan analisis statistik dapat diandalkan. Pemilihan sampel acak mengurangi bias dan memungkinkan generalisasi hasil survei. Namun, validasi lebih lanjut diperlukan untuk menjaga representasi akurat, serta analisis mendalam terhadap subkelompok yang dapat memberikan wawasan lebih spesifik. Hasil korelasi menunjukkan `frekuensi_kopi` dan `jumlah_kopi` memiliki hubungan erat, sementara `gender`, `tempat_tinggal`, `jenis_kopi`, `jenis_biji`, dan `pengalaman` memiliki korelasi rendah. Reliabilitas survei yang diukur dengan Cronbach's alpha

menunjukkan data kurang konsisten, dengan skor 0.1391 dan 0.3534, disebabkan ketidaksesuaian skala jawaban dan pertanyaan ambigu. Analisis response bias pada variabel jumlah_kopi menunjukkan bias dengan skor pertama 0.1 dan kedua -8.858823529411747, yang mungkin disebabkan pilihan jawaban terbatas.

B. Future Analysis Plans

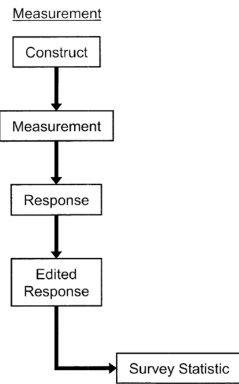
Untuk memahami preferensi konsumen kopi, langkah pertama adalah analisis segmentasi pasar. Data yang dikumpulkan akan digunakan untuk mengelompokkan konsumen berdasarkan demografi (usia, jenis kelamin, tempat tinggal) dan perilaku konsumsi (frekuensi konsumsi kopi, metode pembuatan, dan preferensi jenis biji kopi) menggunakan teknik clustering seperti K-Means atau Hierarchical Clustering. Segmentasi ini akan mengidentifikasi segmen pasar yang berbeda, sehingga produsen dapat menargetkan strategi pemasaran yang lebih spesifik dan efektif. Langkah berikutnya adalah analisis regresi untuk menentukan faktor-faktor utama yang mempengaruhi keputusan pembelian kopi, dengan menggunakan regresi berganda untuk menganalisis variabel dependen seperti frekuensi konsumsi, preferensi harga, dan tempat pembelian terhadap variabel independen seperti usia dan tempat tinggal. Selain itu, analisis konjoin akan digunakan untuk mengukur nilai relatif dari berbagai atribut kopi dalam keputusan pembelian. Hasil analisis ini akan memberikan wawasan mendalam bagi produsen kopi dalam mengembangkan produk dan strategi pemasaran yang sesuai dengan preferensi konsumen.

LAMPIRAN

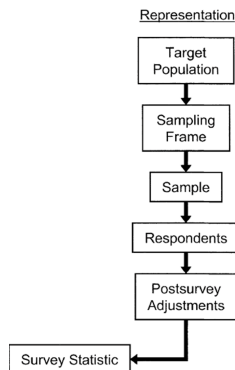
- Google Form :

<https://forms.gle/fzwapgVM1SKZHbtp7>

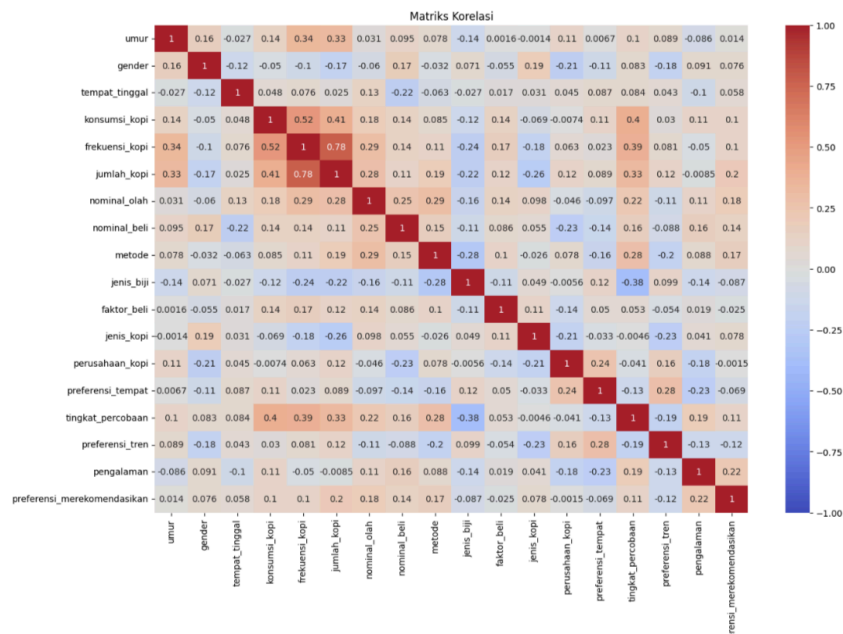
- Gambar 1.1 *Survey lifecycle from a design perspective (Measurement)*



- Gambar 1.2 *Survey lifecycle from a design perspective (Representation)*



- Gambar 1.3. Hasil perhitungan *validity* dengan korelasi



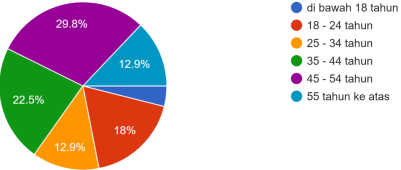
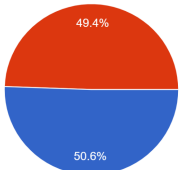
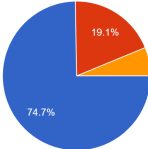
- Gambar 1.4. Bagian pertama hasil pemeriksaan *validity* dengan Pearson's Correlation Coefficient

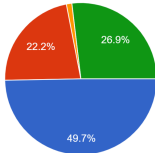
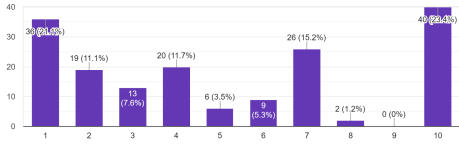
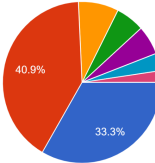
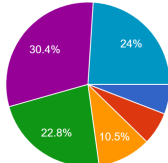
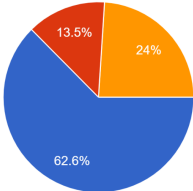
Variabel	umur	gender	tempat_tinggal	konsumsi_kopi	frekuensi_kopi	jumlah_kopi	nominal_olah	nominal_beli	metode
R hitung	0.45134824	-0.013302196	0.094615264	0.475510934	0.711744786	0.77070679	0.472550917	0.30075994	0.23994
R tabel	0.151	0.151	0.151	0.151	0.151	0.151	0.151	0.151	0.151
Validity	Valid	Tidak Valid	Tidak Valid	Valid	Valid	Valid	Valid	Valid	Valid

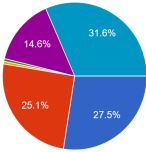
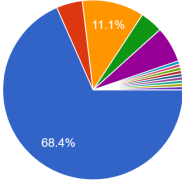
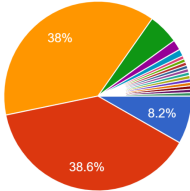
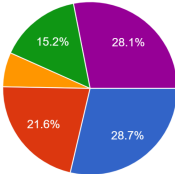
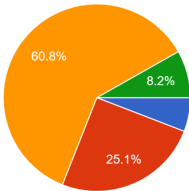
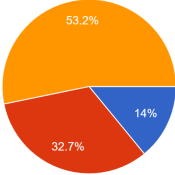
- Gambar 1.5. Bagian kedua hasil pemeriksaan *validity* dengan Pearson's Correlation Coefficient

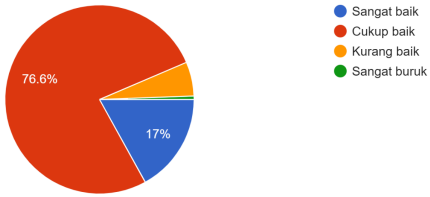
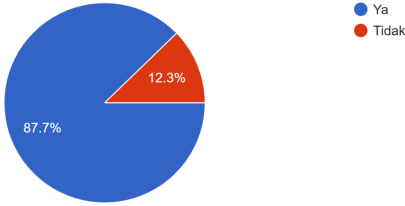
jenis_biji	faktor_beli	jenis_kopi	perusahaan_kopi	preferensi_tempat	tingkat_percobaan	preferensi_tren	pengalaman	preferensi_merekomendasikan	
0.00696696	0.321515321	0.03540849	0.199793361	0.277512062	0.323452108	0.170993981	0.006675783	0.22811	
0.151	0.151	0.151	0.151	0.151	0.151	0.151	0.151	0.151	
Tidak Valid	Valid	Tidak Valid	Valid	Valid	Valid	Valid	Tidak Valid	Valid	

- Tabel 1.1 *Summary Statistic*

Variabel	Visual Statistik	Statistik Deskriptif
umur		<p>N/A values: 0</p> <p>Description:</p> <ol style="list-style-type: none"> Count: 178 Unique: 6 Top: 45 - 54 tahun Frequency: 53
gender		<p>N/A values: 0</p> <p>Description:</p> <ol style="list-style-type: none"> Count: 178 Unique: 2 Top: Laki-laki Frequency: 90
tempat_tinggal		<p>N/A values: 0</p> <p>Description:</p> <ol style="list-style-type: none"> Count: 178 Unique: 3 Top: Kota Besar (Contoh : Pusat Kota, Kawasan Perkotaan, Daerah Penyangga) Frequency: 133

frekuensi_kopi	 <p> • Setiap hari • Beberapa kali dalam seminggu • Seminggu sekali • Dalam situasi tertentu / tidak pasti </p>	N/A values: 7 Description: <ol style="list-style-type: none"> Count: 171 Unique: 4 Top: Setiap hari Frequency: 85
jumlah_kopi		N/A values: 7 Description: <ol style="list-style-type: none"> Count: 171.000000 Mean: 5.116959 Std: 3.384794 Min: 1.000000 25%: 2.000000 50%: 4.000000 75%: 7.000000 Max: 10.000000
nominal_olah	 <p> • < Rp8.000,00 • Rp8.000,00 - Rp39.999,00 • Rp40.000,00 - Rp79.999,00 • Rp80.000,00 - Rp119.999,00 • Rp120.000,00 - Rp299.999,00 • Rp300.000,00 - Rp500.000,00 • > Rp500.000,00 </p>	N/A values: 7 Description: <ol style="list-style-type: none"> Count: 171 Unique: 7 Top: Rp. 8.000,00 - Rp. 39.999,00 Frequency: 70
nominal_beli	 <p> • < Rp8.000,00 • Rp8.000,00 - Rp14.999,00 • Rp15.000,00 - Rp19.999,00 • Rp20.000,00 - Rp29.999,00 • Rp30.000,00 - Rp40.000,00 • > Rp40.000,00 </p>	N/A values: 7 Description: <ol style="list-style-type: none"> Count: 171 Unique: 6 Top: Rp. 30.000,00 - Rp. 40.000,00 Frequency: 52
metode	 <p> • Diseduh • Manual brew • Machine brew </p>	N/A values: 7 Description: <ol style="list-style-type: none"> Count: 171 Unique: 3 Top: Diseduh Frequency: 107

jenis_biji	 <ul style="list-style-type: none"> Arabica Robusta Liberica Excelsa Campuran (Blend) Tidak tahu / Tidak ada preferensi khusus 	<p>N/A values: 7</p> <p>Description:</p> <ol style="list-style-type: none"> Count: 171 Unique: 6 Top: Tidak tahu / tidak ada preferensi khusus Frequency: 54
faktor_beli	 <ul style="list-style-type: none"> Rasa Harga Kualitas bahan baku Kenyamanan lokasi Aroma Penat dalam berfikir Fungsinya Rasa, Harga <p>▲ 1/2 ▼</p>	<p>N/A values: 7</p> <p>Description:</p> <ol style="list-style-type: none"> Count: 171 Unique: 14 Top: Rasa Frequency: 117
jenis_kopi	 <ul style="list-style-type: none"> Espresso Latte/Cappuccino Kopi hitam Kopi instan Frappuccino Kopi gula aren Coffee latte Americano <p>▲ 1/3 ▼</p>	<p>N/A values: 7</p> <p>Description:</p> <ol style="list-style-type: none"> Count: 171 Unique: 18 Top: Latte/Cappuccino Frequency: 66
perusahaan_kopi	 <ul style="list-style-type: none"> Starbucks Kopi Kenangan Janji Jiwa Fore Tidak ada yang cocok 	<p>N/A values: 7</p> <p>Description:</p> <ol style="list-style-type: none"> Count: 171 Unique: 5 Top: Starbucks Frequency: 49
tingkat_percobaan	 <ul style="list-style-type: none"> Sangat sering Cukup sering Jarang Tidak pernah 	<p>N/A values: 7</p> <p>Description:</p> <ol style="list-style-type: none"> Count: 171 Unique: 4 Top: Jarang Frequency: 104
preferensi_tren	 <ul style="list-style-type: none"> Mengikuti tren Tetap dengan tradisi Kombinasi keduanya 	<p>N/A values: 7</p> <p>Description:</p> <ol style="list-style-type: none"> Count: 171 Unique: 3 Top: Kombinasi keduanya Frequency: 91

pengalaman	 <p> ● Sangat baik ● Cukup baik ● Kurang baik ● Sangat buruk </p>	N/A values:7 Description: 1. Count: 171 2. Unique: 4 3. Top: Cukup baik 4. Frequency: 131
preferensi_mereko mendasikan	 <p> ● Ya ● Tidak </p>	N/A values: 7 Description: 1. Count: 171 2. Unique: 2 3. Top: Ya 4. Frequency: 150

- Code 1.1 Mengganti nama kolom

```
df = df.rename(columns={
    'Berapa usia Anda?': 'umur',
    'Apa jenis kelamin Anda?': 'gender',
    'Dimanakah tempat tinggal Anda saat ini?': 'tempat_tinggal',
    'Apakah Anda mengonsumsi kopi?': 'konsumsi_kopi',
    'Apakah Anda alergi terhadap kopi atau tidak minum kopi karena alasan kesehatan?': 'alergi_kopi',
    'Seberapa sering Anda mengonsumsi kopi?': 'frekuensi_kopi',
    'Dalam seminggu, berapa banyak kopi yang Anda konsumsi?': 'jumlah_kopi',
    'Berapa nominal yang Anda dapat keluarkan jika Anda mengolah kopi sendiri?': 'nominal_olah',
    'Berapa nominal yang Anda dapat keluarkan jika membeli minuman kopi di luar?': 'nominal_beli',
    'Metode pembuatan kopi apa yang biasa Anda gunakan dalam mengonsumsi kopi tersebut?': 'metode',
    'Jenis biji kopi apa yang paling sesuai dengan preferensi Anda?': 'jenis_biji',
    'Apa faktor terpenting bagi Anda saat membeli kopi?': 'faktor_beli',
    'Jenis kopi apa yang biasa Anda minum?': 'jenis_kopi',
```

```

'Dari perusahaan kopi di bawah ini, manakah yang paling cocok
dengan preferensi kopi yang Anda konsumsi?': 'perusahaan_kopi',
'Dimana Anda biasa membeli kopi?': 'preferensi_tempat',
'Seberapa sering Anda mencoba jenis kopi baru atau kedai kopi
baru?': 'tingkat_percobaan',
'Apakah Anda mengikuti tren kopi terbaru atau tetap dengan
pilihan tradisional?': 'preferensi_tren',
'bagaimana pengalaman Anda saat mengunjungi kedai kopi atau
membeli kopi secara online?': 'pengalaman',
'Apakah Anda akan merekomendasikan kedai kopi favorit Anda
kepada teman atau keluarga?': 'preferensi_merekomendasikan',
'Apa yang dapat dilakukan oleh kedai kopi untuk meningkatkan
pengalaman Anda sebagai pelanggan?': 'masukan_peningkatan'
})

```

- Code 1.2 Menghapus kolom tidak relevan

- Kolom faktor_beli:

```

valid_choices = ["rasa", "harga", "kualitas bahan baku",
"kenyamanan lokasi", "aroma"] def
standardize_factor_beli(value): value = value.lower() if
value in valid_choices: return value return 'other'
df['faktor_beli'] =
df['faktor_beli'].apply(standardize_factor_beli)

```

- Kolom jenis_kopi:

```

replacement_dict = {
    "Latte/Cappuccino": "latte/cappuccino",
    "Kopi hitam": "kopi hitam",
    "Kopi gula aren": "latte/cappuccino",
    "Coffee latte": "latte/cappuccino",
    "Espresso": "espresso",
    "Kopi instan": "kopi instan",
    "Americano": "kopi hitam",
}

```

```

"Frappuccino": "frappuccino",
"Kopsus": "latte/cappuccino",
"Indocafe": "kopi instan",
"Jarang minum": "other",
"Vietnam drep ": "other",
"kopi robusta + susu": "latte/cappuccino",
"Kopi susu": "latte/cappuccino",
"Torabika ": "other",
"macchiato ": "other",
"ga minum kopi ;-;": None,
"v60": "kopi hitam"
}

```

```

df['jenis_kopi'] =
df['jenis_kopi'].replace(replacement_dict)
df = df.dropna(subset=['jenis_kopi'])

```

- Kolom preferensi_tempat:

```

place_replacement_dict = {
    'Kafe/restoran': 'kafe/restoran',
    'Kedai kopi': 'kedai kopi',
    'Toko kelontong/supermarket': 'toko
kelontong/supermarket',
    'Selewatnya': 'other',
    'Kopi jo': 'kedai kopi',
    'Mesin penjual otomatis': 'mesin penjual otomatis',
    'Mall': 'other',
    'On line': 'other',
    'Peracik kopi': 'other',
    'Warkop pangku': 'kedai kopi',
    'Pesan khusus': 'other',
    'starbucks': 'kedai kopi',

```



```
'Home Industri Kopi HAKA': 'other',  
'buat sendiri': 'other'  
}  
  
df['preferensi_tempat'] =  
df['preferensi_tempat'].replace(place_replacement_dict)
```

Referensi

(N.d.). <https://real-statistics.com/statistics-tables/pearsons-correlation-table/>

Khasanah, T. A., & Safilah, F. (2023). Hubungan Kebiasaan Konsumsi Kopi Dan kualitas Tidur Dengan status Gizi Pada remaja di kelurahan Jagakarsa Jakarta Selatan. Jurnal Andaliman: Jurnal Gizi Pangan, Klinik Dan Masyarakat, 3(1). <https://doi.org/10.24114/jgpkkm.v3i1.46401>