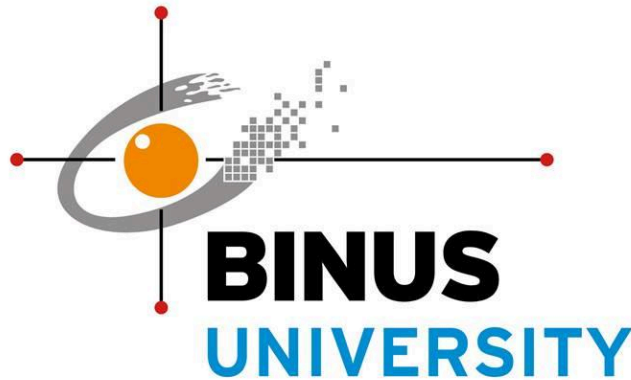


Pengaruh Gaya Hidup terhadap Indeks Prestasi Kumulatif Mahasiswa Data Science B27



Kelompok 4

No	Name	Student ID	Contributions
1	Benjamin Eleazar Manafe	2702340704	<ul style="list-style-type: none">- Bab 2: Life Cycle Design Perspective & Perhitungan- Bab 3: Desain Kuesioner- Bab 4: Pengumpulan Data
2	Rayyan	2702257495	<ul style="list-style-type: none">- Bab 6: Hasil dan Analisa- Mengumpulkan Data kusioner- Code
3	Patrick Surya Pranata	2702367716	<ul style="list-style-type: none">- Bab 1: Pendahuluan- Bab 2: Life Cycle Quality Perspective- Membuat Form Kuesioner- Mengumpulkan Data kusioner- Bab 6: Proses dan Analisa Data
4	Richard Dean Tanjaya	2702262652	<ul style="list-style-type: none">- Membuat Form Kuesioner- Code- Bab 5: Kualitas Pengukuran- Bab 6: Proses dan Analisa Data- Bab 7: Kesimpulan

I. Pendahuluan

Gaya hidup merupakan salah satu faktor penting yang dapat mempengaruhi berbagai aspek kehidupan seseorang, termasuk pencapaian akademik. Di kalangan mahasiswa, gaya hidup seperti pola tidur, kebiasaan belajar, hingga aktivitas sosial diluar/didalam kampus dapat memberikan dampak langsung maupun tidak langsung terhadap Indeks Prestasi Kumulatif (IPK). Survei ini bertujuan untuk mengetahui sejauh mana gaya hidup mahasiswa berpengaruh terhadap IPK. Melalui data yang terkumpul, diharapkan hasil survei ini akan memberikan gambaran yang lebih jelas mengenai hubungan antara gaya hidup dan prestasi akademik mahasiswa.

II. Desain Pengambilan Sampel

Dalam pengambilan sampel untuk survey ini, kami menerapkan metode Simple Random Sampling (SRS), yaitu metode pengambilan sampel secara acak. Metode ini dipilih karena anggota populasi memiliki peluang yang sama dan independen untuk terpilih menjadi bagian dari sampel. Kita memilih metode ini karena hanya mencakup mahasiswa B27 Data Science dan pemilihan satu individu tidak akan mempengaruhi peluang individu lain untuk terpilih sehingga mendapatkan data yang tidak bias.

Selanjutnya, untuk menentukan berapa sampel yang kami perlukan, kami melakukan perhitungan sample size terlebih dahulu. Rumus sampel ada sebagai berikut:

$$n = \frac{(Z_{(0.95)})^2 \hat{P}(1-\hat{P})}{\epsilon^2}$$

Dari rumus sampel tersebut, diketahui bahwa kita memerlukan untuk menentukan signifikan alpha, proporsi dan errornya. Untuk itu, value untuk masing-masing variable yang kami tentukan adalah:

$$\alpha = 10\% \mid \hat{p} = 0.5 \mid error = 10\%$$

Alasan kami menentukan value-value tersebut adalah untuk signifikan alpha 10% agar untuk tidak menentukan jumlah sampel jika alpha meningkat dan jumlah sampel tidak sangat minimal jika alpha menurun. Dengan 10%, jumlah sampel yang kami dapatkan mencukupi kebutuhan kita. Selanjutnya, karena kita tidak tahu tentang data kami, kami menentukan proporsi 0.5 sebagai defaultnya dan dengan error 10%.

Setelah itu, kami menghitung sample size dan reduced sample size. Kami menggunakan reduced sample size sebagai patokan sample size kami untuk menghemat energi dan mendapatkan hasil yang baik dengan sample size minimal.

$$n = \frac{(Z_{(0.95)})^2 \hat{P}(1-\hat{P})}{\epsilon^2} = \frac{(1.645)^2 \times 0.5(1-0.5)}{(0.1)^2} = 67,6 \approx 68$$
$$n' = \frac{n}{1 + \frac{(Z_{(0.95)})^2 \hat{P}(1-\hat{P})}{\epsilon^2 N}} = \frac{68}{1 + \frac{(1.645)^2 \times 0.5(1-0.5)}{(0.1)^2 \times 192}} = 50,2 \approx 51 \text{ size}$$

Dari perhitungannya, kami menemukan bahwa sample size dengan population 192 adalah 51 Mahasiswa Aktif Data Science B27.

Life Cycle Design Perspective

Measurement:

1. **Construct:** Hubungan antara berbagai aspek gaya hidup mahasiswa yang diukur melalui IPK.

2. **Measurement:**

- Pelajaran = pertanyaan seputar jam belajar, mengatur waktu belajar, pemahaman mata kuliah, seberapa rutin revisi pelajaran, dll
 - Kebiasaan = pertanyaan seputar Jam tidur, aktif sosial media, organisasi yang diikuti, kerja sampingan, dan aktivitas diluar
 - Lingkungan = pertanyaan seputar lingkungan perkuliahan, pertemanan, pengaruh dosen terhadap tingkat pemahaman dan IPK
3. **Response:** pada tahap ini, kami memilih jenis respon yang berorientasi pada “close-ended response”.
4. **Edited Response:** proses pengeditan jawaban responden sebelum dilakukan analisis lebih lanjut bertujuan untuk memastikan semua data yang sudah dikumpulkan konsisten, akurat, dan menghindari respon yang invalid.

Representation:

1. **Target Population:** Mahasiswa Aktif Data Science B27.
2. **Sampling Frame:** Daftar nama semua mahasiswa B27 yang aktif dan bisa dihubungi.
3. **Sample:** 51 mahasiswa.
4. **Respondents:** Beberapa bagian dari sample yang dipilih .
5. **Postsurvey Adjustments:** Memperbaiki data dengan imputasi dan pembobotan jika diperlukan.

Life Cycle Quality Perspective

Measurement:

1. **Validity:** Survei mengukur aktivitas sehari-hari untuk mendapatkan korelasi dengan IPK setiap mahasiswa berdasarkan data resmi universitas.

Untuk mendapatkan data dengan validitas tinggi variabel seperti jam belajar, jam menggunakan medsos, keaktifan dalam kelas akan kami korelasikan dengan IPK mahasiswa sesuai dengan data resmi universitas. Dengan itu validitas survei akan tinggi, sesuai dengan hasil yang didapatkan pada dunia nyata. Hal ini akan mengurangi measurement error yang terjadi. Validitas data akan bergantung pada kejujuran responden dalam mengisi data IPK-nya masing masing.

2. **Measurement Error:** Mengidentifikasi observational gap yang terjadi untuk mengetahui deviasi dari nilai sebenarnya.

Observation Gap terjadi karena beberapa hal seperti kekurangan dalam pilihan yang tersedia atau tidak dapat mencerminkan pendapat responden, interpretasi pertanyaan yang berbeda karena ambigu/kurang jelas, dan kecerobohan responden yaitu jika responden tidak sengaja memilih pilihan yang tidak sesuai karena terburu-buru.

3. **Processing Error:** Kesalahan yang terjadi saat pengolahan data sebelum analisis.

Processing error terjadi saat mengolah input dari responden seperti entri data/transkripsi menggunakan ekstensi pada google form (autocrat) yang mengalami bug sehingga entri data seperti nilai IPK 3,75 menjadi 3,57 atau responden melakukan typo pada jawabannya/format yang salah.

Selanjutnya data cleaning yaitu proses menghapus data outlier yang seharusnya dibiarkan saja (valid) dan tidak menghapus data duplikat sehingga menyebabkan bias.

Transformasi data adalah saat salah mengkonversi data kategorikal/object menjadi numerical, menggunakan skala yang tidak sesuai dengan distribusi data, dan tidak mentransformasi distribusi saat datanya terlalu skewed sehingga model menjadi bias.

Imputation error terjadi jika menginputasi data tanpa mempertimbangkan distribusi data sehingga menyebabkan bias pada model/hasil prediksi

Representation:

1. **Coverage Error:** Coverage error terjadi ketika ada perbedaan antara populasi target dan populasi yang benar-benar tercakup dalam survei, sehingga hasil yang diperoleh bisa menjadi bias.

Beberapa faktor penyebabnya adalah sampling frame tidak mencakup seluruh populasi yaitu Jika daftar mahasiswa aktif yang digunakan sebagai dasar pemilihan sampel tidak lengkap, maka ada mahasiswa yang seharusnya bisa menjadi bagian dari survei tetapi tidak terdata. Selanjutnya adalah bias akibat keterlibatan responden yaitu Bisa saja mahasiswa yang lebih aktif atau lebih peduli dengan akademik lebih cenderung mengisi survei dibandingkan mereka yang kurang aktif, sehingga hasil survei lebih condong ke kelompok tertentu. Selanjutnya adalah terbatasnya keterlibatan mahasiswa dalam group chat yaitu Meskipun semua mahasiswa bisa dihubungi, mayoritas hanya menggunakan group chat sekedar untuk melihat event tertentu. Akibatnya, banyak yang mungkin mengabaikan survei, sementara yang mengisi bisa jadi hanya mereka yang kebetulan melihat, iseng, atau sekedar membantu teman, sehingga hasilnya bisa kurang merepresentasikan seluruh populasi.

2. **Sampling Error:** Sampling error terjadi ketika hasil survei tidak sepenuhnya mencerminkan populasi karena variasi acak dalam pemilihan sampel. Meskipun tidak bisa dihilangkan sepenuhnya dapat diminimalisir.

Beberapa faktor penyebabnya adalah ukuran sampel terbatas yaitu, dengan populasi 192 dan sampel hanya 51 mahasiswa, ada kemungkinan bahwa karakteristik yang muncul dalam sampel tidak mencerminkan seluruh populasi. Selanjutnya, variasi acak dalam pemilihan sampel yaitu meskipun pemilihan dilakukan berdasarkan metode tertentu, ada kemungkinan distribusi gaya hidup dan kebiasaan belajar dalam sampel berbeda dari populasi sebenarnya. Selanjutnya adalah ketidakseimbangan distribusi karakteristik yaitu jika kelompok mahasiswa dengan kebiasaan belajar tertentu lebih banyak dalam sampel dibandingkan kelompok lain, hasilnya bisa menjadi bias.

3. **Nonresponse Error:** Kejadian saat sebagian sampel responden yang ditentukan tidak mengisi survei.

Kualitas data dapat bias jika hasil dari non responden berbeda secara signifikan, seperti jika 25 Mahasiswa mengisi survei menjawab bahwa mereka aktif saat di kelas dan ternyata 26 Mahasiswa tidak menjawab sebenarnya tidak aktif di kelas, hasil survei akan bias terhadap 25 Mahasiswa tersebut karena 26 Mahasiswa tersebut tidak menjawab survei. Selanjutnya, kurangnya generalisasi dari survei dapat terjadi jika hanya 25 Mahasiswa menjawab dan 26 sisanya tidak, maka akan terjadi kurangnya generalisasi karena jumlah sampel yang diinginkan kurang. Selanjutnya, meningkatnya variabilitas akan terjadi jika hasil dari non respondent berbeda secara

signifikan dari responden, maka akan meningkatkan variabilitas yang akan berdampak dengan kurangnya representatif sampel data terhadap populasi.

4. **Adjustment Error:** Kesalahan saat teknik-teknik yang digunakan untuk memperbaiki nonresponse error menimbulkan bias dan kesalahan hasil akurasi.

Adjustment error dapat terjadi dari beberapa hal seperti, asumsi-asumsi yang kurang akurat atau salah tentang hasil non responden dan imputasi salah terhadap data yang tidak terisi.

III. Desain Kuesioner

a. Variable Utama dari Measurements

Kuesioner ini dirancang untuk mengukur variabel utama "Gaya Hidup Mahasiswa", yang dipecah menjadi tiga sub-variabel (measurement): Pelajaran, Kebiasaan, dan Lingkungan. Setiap pertanyaan kuesioner (selain IPK) merupakan turunan dari salah satu dari tiga sub-variabel ini. Berikut adalah pemetaannya:

1. Pelajaran Variabel ini bertujuan untuk mengukur strategi, manajemen waktu, dan efektivitas kegiatan belajar mahasiswa. Pertanyaan yang diturunkan adalah:

- **Jam Belajar UTS/UAS (1-4):** Mengukur alokasi waktu belajar pada periode krusial
- **Ngatur Waktu Kuliah, Tugas, Belajar (1-5):** Mengukur kemampuan manajemen waktu akademik secara umum
- **Mudah memahami materi (1-5):** Mengukur persepsi efektivitas belajar dan pemahaman konsep
- **Notes/ringkasan Kuliah (1-5):** Mengukur salah satu metode belajar aktif
- **Jadwal belajar teratur (1-5):** Mengukur konsistensi dan disiplin dalam belajar
- **Memanfaatkan AI (1-5):** Mengukur adaptasi terhadap teknologi sebagai alat bantu belajar modern
- **Revisi Pelajaran (1-5):** Mengukur frekuensi pengulangan materi sebagai indikator ketekunan atau kesulitan pemahaman awal

2. Kebiasaan Variabel ini bertujuan untuk mengukur rutinitas dan aktivitas non-akademik mahasiswa yang dapat mempengaruhi energi, fokus, dan alokasi waktu mereka. Pertanyaan yang diturunkan adalah:

- **Jumlah Organisasi (0-3):** Mengukur tingkat keterlibatan dalam kegiatan ekstrakurikuler
- **Terbebani Organisasi (1-5):** Mengukur persepsi dampak dari keterlibatan organisasi terhadap beban pribadi
- **Kerja Samping (0/1):** Mengidentifikasi apakah mahasiswa memiliki komitmen kerja di luar kuliah
- **Kerja Sampingan mengganggu (1-5):** Mengukur persepsi dampak dari kerja sampingan terhadap fokus akademik
- **Sosial Media (1-4):** Mengukur kuantitas penggunaan media sosial
- **Medsos Mengganggu (1-5):** Mengukur persepsi dampak dari media sosial terhadap produktivitas
- **Waktu Tidur (1-4):** Mengukur kuantitas istirahat yang esensial untuk fungsi kognitif
- **Waktu Tidur Cukup (1-5):** Mengukur persepsi kualitas dan kecukupan istirahat
- **Aktivitas Diluar Rumah (1-5):** Mengukur tingkat aktivitas sosial atau hobi di luar kampus.

3. Variabel: Lingkungan Variabel ini bertujuan untuk mengukur pengaruh faktor eksternal, baik dari lingkungan sosial maupun akademik, terhadap motivasi dan prestasi mahasiswa. Pertanyaan yang diturunkan adalah:

- **Pengaruh Lingkungan mendukung Pembelajaran (1-5):** Mengukur persepsi umum terhadap lingkungan sekitar (kos, rumah, dll)
- **Teman2 mendukung Akademik (1-5):** Mengukur pengaruh lingkungan pertemanan terhadap motivasi belajar
- **Lingkungan pertemanan mempengaruhi IPK (1-5):** Mengukur persepsi langsung hubungan antara pertemanan dan hasil akademik
- **Dosen mempengaruhi akademik (1-5):** Mengukur pengaruh dosen pengajar terhadap pemahaman dan motivasi akademik

b. Variable bukan dari Measurements

Terdapat satu pertanyaan kunci dalam kuesioner yang tidak bisa kategorikan ke dalam variabel gaya hidup, yaitu pertanyaan mengenai **IPK (Indeks Prestasi Kumulatif)**. Seluruh tujuan dari penelitian ini adalah untuk mengetahui pengaruh gaya hidup terhadap prestasi akademik. Sementara itu, IPK adalah Variabel Dependen, yaitu variabel hasil yang ingin kita ukur dan jelaskan. Tanpa mengukur variabel dependen, penelitian ini tidak dapat menjawab rumusan masalahnya.

IV. Pengumpulan Data

Metode data collection yang diterapkan pada survey ini adalah menggunakan Online Survey Forms dengan Google Forms berisi dengan data quantitative. Pemilihan platform survey digital ini dikarenakan efisiensi, aksesibilitas yang mudah untuk para responden, merekam data secara *real-time* dan mempermudah proses tabulasi dan analisis data awal. Untuk memastikan partisipasi yang tinggi, kami menghubungi langsung mahasiswa B27 Data Science agar response rate kami meningkat serta memastikan kepada mereka bahwa seluruh data yang mereka masukan tidak akan disebar dan hanya digunakan untuk tujuan pendidikan. Cara penyebaran tersebut kami lakukan secara berulang sampai kami mendapatkan jumlah responden yang diperlukan, yaitu 51 responden.

Hasil dari proses tersebut adalah pencapaian dalam mendapatkan jumlah sampel minimal yang dibutuhkan. Jumlah non-response dari cara penyebaran tersebut juga adalah nol. Dari hasil ini, proses penyebaran tersebut terbukti efektif dalam menaikkan response rate dan meminimalisir non-response pada populasi yang kecil.

V. Kualitas Pengukuran

Survei ini akan diukur kualitasnya dari pengukuran validitas dan reliabilitasnya. Validitas adalah pengukuran untuk melihat seberapa valid atau akurat hasil survey datanya dengan apa yang ingin dibandingkan. Reliabilitas adalah pengukuran untuk melihat kualitas data di mana yang akan dilihat adalah konsistensi dan stabilitas datanya.

Validitas akan diukur dengan membandingkan p-value dari sebuah variabel dengan signifikansi alpha. Jika p-value kurang atau sama dengan alpha, maka variabel tersebut dapat disimpulkan sebagai valid. Signifikansi alpha yang kami gunakan adalah 0.1 sehingga pada rumusnya, untuk mendapatkan nilai valid, p-value dari variabel-variabel kami seharusnya dibawah 0.1.

$$Pvalue \leq \alpha$$

Untuk menghitung reliabilitas data dan konsistensi data untuk mengukur IPK kami menggunakan cronbach's alpha.

$$n = \frac{k}{k-1} \left(1 - \frac{\Sigma var(item)}{var(totalscore)}\right) = \frac{20}{20-1} \left(1 - \frac{20.639}{49.443}\right) \approx 0.613$$

Dari hasil perhitungan diatas dapat disimpulkan bahwa data kami memiliki **reliabilitas tinggi** karena berada pada kisaran 0.6 - 0.8.

VI. Proses Analisa data

1. Missing Value

Pada data yang kelompok kami kumpulkan tidak didapati adanya missing value dari setiap variabel yang akan kita gunakan. Oleh karena itu tidak ada keperluan untuk melakukan *imputation* terhadap data.

2. Descriptive Statistics

Dari hasil analisa, kita dapat menyimpulkan bahwa:

- Mayoritas IPK Mahasiswa berada diantara 3.13 - 3.87
- Jadwal Belajar Mahasiswa tidak teratur
- Mayoritas Mahasiswa tidak Bekerja
- Banyak Mahasiswa memanfaatkan AI
- Dosen mempengaruhi akademik

3. Preprocessing

Ada beberapa kesalahan input dari responden pada variable “Terbebani Organisasi” dan “Kerja Sampingan mengganggu” jika mereka tidak ikut organisasi dan tidak ikut kerja ia seharusnya mengisi 3 (netral) pada kedua variable tersebut.

4. Multikolinearitas

Multikolinearitas juga digunakan untuk mengetahui apakah adanya korelasi antar variabel independen pada model regresi. Untuk mengetahui adanya multikolinearitas kita harus menghitung Variance Inflation Factor (VIF) dimana nilai tolerance ini akan mengukur variabilitas dari variabel independen yang terpilih yang tidak dapat dijelaskan oleh variabel bebas lainnya. Dengan nilai tolerance yang rendah maka nilai VIF tinggi, dikarenakan $VIF = 1/\text{tolerance}$, dan menunjukkan terdapat kolinearitas yang tinggi.

Dari hasil pada gambar (lihat dalam lampiran) dapat dilihat bahwa korelasi antar variabel masih bernilai moderat sehingga dapat digunakan dalam model regresi

Kondisi VIF:

- **VIF \approx 1:** hampir tidak ada korelasi antara variabel ini dengan variabel lain.
- **VIF 1–5:** ada korelasi ringan sampai moderat, biasanya masih aman.
- **VIF $>$ 5 :** menandakan multikolinearitas yang tinggi, maka perlu dikaji ulang

5. Modelling

a. Regression Model

Dalam melakukan Regression Modelling, untuk memprediksi Indeks Prestasi Kumulatif (IPK) berdasarkan variabel-variabel gaya hidup mahasiswa, langkah

pertama adalah menghitung statistik regresi dasar. Dari hasil analisis (lihat dalam lampiran) diperoleh nilai Multiple R sebesar 0,7568, R-Square sebesar 0,5727, Adjusted R-Square sebesar 0,2878, Standard Error sebesar 0,3125, dan jumlah observasi sebanyak 51 mahasiswa. Nilai R-Square (koefisien determinasi) sebesar 0,5727 menunjukkan bahwa sekitar 57,27 % variasi IPK dapat dijelaskan oleh variabel-variabel gaya hidup yang dimasukkan ke dalam model. Karena rentang interpretasi R-Square antara 0,4 hingga 0,6 dikategorikan sebagai kecocokan model “moderate” maka kita dapat menyimpulkan bahwa model ini memiliki kemampuan moderat dalam menjelaskan perubahan IPK mahasiswa Data Science B27.

b. ANOVA Table (Type I)

ANOVA table (Type I) menunjukkan efek penambahan satu variabel gaya hidup secara berurutan ke dalam model regresi (lihat dalam lampiran). Misalnya, “jam_belajar_uts_uas_1_4” dimasukkan terlebih dahulu dan menghasilkan nilai F sebesar 3,0185 dengan p-value 0,0926, sedangkan variabel “ngatur_waktu_kuliah_tugas_belajar_1_5” dimasukkan berikutnya dan memberikan kontribusi tambahan ($F = 7,6729$; $p\text{-value} = 0,0095$). Karena ANOVA Type I menghitung jumlah kuadrat yang dijelaskan secara kumulatif, setiap p-value menunjukkan besar penurunan residual sum of squares saat penambahan variabel baru ke model. Dengan kata lain, p-value pada ANOVA Type I tidak menilai kontribusi variabel secara simultan dengan variabel lain yang selevelnya, melainkan kontribusi marginalnya pada titik penambahan tertentu dalam urutan.

Dalam konteks penelitian “Pengaruh Gaya Hidup terhadap Indeks Prestasi Kumulatif Mahasiswa Data Science B27,” hal ini berarti p-value ANOVA Type I akan tergantung urutan variabel dalam model. Sebagai contoh, variabel “kerja_samping_0_1” menunjukkan p-value 0,0205 karena pada urutan penambahan tersebut sisa variasi relatif besar dan berkurang cukup banyak ketika variabel itu dimasukkan. Jika urutan variabel diubah, p-value ini bisa berbeda dengan data yang sama. Maka itu, p-value ANOVA Type I kurang relevan untuk menilai kekuatan pengaruh variabel secara keseluruhan dan hanya berguna ketika ada alasan logis atau temporal untuk memasukkan variabel dalam urutan tertentu.

c. Model Significance

Secara keseluruhan, nilai F-statistic sebesar 2,0105 dengan Significance F (p-value) 0,0406 menunjukkan bahwa model regresi bersifat signifikan pada signifikansi $\alpha = 0,1$. Artinya kita menerima risiko kesalahan sebesar 10%. Dengan p-value 0,0406 yang lebih kecil dari 0,1, kita menolak hipotesis nol yang menyatakan bahwa tidak ada hubungan linier antara variabel-variabel gaya hidup dan IPK.

d. Coefficient Table

Secara keseluruhan, dari tabel koefisien terdapat lima variabel gaya hidup yang memiliki p-value $< 0,1$, yaitu “jadwal_belajar_teratur_1_5” (0,0174), “revisi_pelajaran_1_5” (0,0016), “terbebani_organisasi_1_5” (0,0840), “kerja_samping_0_1” (0,0751), dan “dosen_mempengaruhi_akademik_1_5” (0,0433). Koefisien regresi menunjukkan besaran perubahan rata-rata IPK ketika satu unit skor pada variabel gaya hidup diubah, dengan asumsi variabel lain konstan. Misalnya, koefisien positif 0,1753 pada “jadwal_belajar_teratur” berarti setiap peningkatan satu tingkat keteraturan jadwal belajar diharapkan menaikkan IPK sebesar 0,1753 poin, sedangkan koefisien negatif $-0,2420$ pada “revisi_pelajaran”

menunjukkan bahwa semakin sering revisi justru berhubungan dengan penurunan IPK sebesar 0,2420 poin, mungkin karena beban revisi mengganggu efektivitas belajar.

P-value yang ada di tabel koefisien berbeda dengan p-value pada ANOVA Type I karena di sini setiap p-value menghitung signifikansi variabel secara bersamaan (setelah semua variabel lain dimasukkan), bukan secara berurutan. Maka itu, p-value pada tabel koefisien menjadi acuan utama untuk menilai pengaruh tiap variabel gaya hidup secara realistik dalam model.

e. Result Visualization

1) Linear Regression Visual

Hasil Regression Model yang kami buat (lihat pada lampiran) telah menunjukkan terbentuknya hubungan positif linear moderat yang mengindikasikan garis linear terhadap variabel-variabel Pengaruh Gaya Hidup Mahasiswa. Secara overall, dapat dilihat bahwa mayoritas actual IPK mendekati predicted IPK dengan R Squared 0.572 dapat menghasilkan output predicted IPK yang cukup reliable.

2) Residual Plot

Dari Histogram dan Q-Q Plot terhadap Residual, dapat dilihat bahwa secara Histogram Plot hasil menunjukkan bahwa simpangan rata-rata terdapat di range $-0.2 < x < +0.2$. Ini menunjukkan bahwa tidak terdapat simpangan yang signifikan antara residual yang menandakan reliabilitas variabel. Selain itu dari Q-Q Plot dapat dilihat bahwa Residual (yang berwarna biru) memiliki hubungan yang cukup linear terhadap garis merah. Hal ini berarti secara overall, setiap residual memiliki signifikansi, meskipun terdapat beberapa residual yang menyimpang atau outlier, ini dapat disimpulkan bahwa hasil cukup signifikan.

VII. Kesimpulan

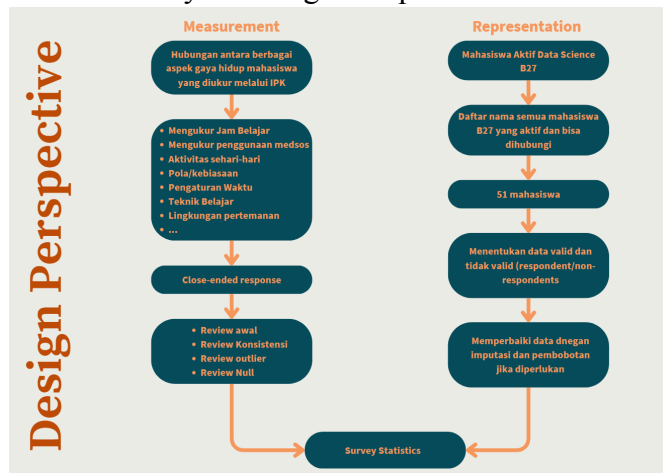
Berdasarkan analisis yang telah dilakukan, dapat disimpulkan bahwa gaya hidup memang memiliki pengaruh yang signifikan secara statistik terhadap Indeks Prestasi Kumulatif (IPK) mahasiswa. Proses survei dan analisis data ini cukup handal, dibuktikan dengan nilai Cronbach's Alpha (0.613). Model regresi yang dibangun mampu menjelaskan 57.2% variansi data ($R\text{-squared} = 0.572$), yang mengindikasikan bahwa meskipun gaya hidup adalah faktor penting, masih ada 42.8% faktor lain di luar lingkup penelitian ini yang mempengaruhi IPK. Dari proses ini, peningkatan dapat dilakukan dengan memperluas jumlah sampel dan variabel penelitian untuk menangkap faktor-faktor lain tersebut, seperti kesehatan mental atau latar belakang sosial-ekonomi, untuk meningkatkan kekuatan prediksi model.

Sebagai rencana analisis lanjutan, fokus akan diarahkan pada dua hal utama. Pertama, melakukan analisis kualitatif melalui wawancara mendalam pada variabel dengan penemuan yang menarik, seperti "Kerja Sampingan" dan "Terbebani Organisasi". Meskipun sering dianggap penghambat, keduanya justru berkorelasi positif dengan IPK, sehingga perlu digali lebih dalam mengapa hal ini terjadi, kemungkinan terkait manajemen waktu atau motivasi yang lebih tinggi. Kedua, mengembangkan model clustering untuk mengelompokkan mahasiswa berdasarkan gaya hidup mereka. Ini bertujuan untuk mengidentifikasi pola-pola spesifik pada tiap kelompok dan bagaimana pengaruh gaya hidup terhadap IPK bisa berbeda antar segmen.

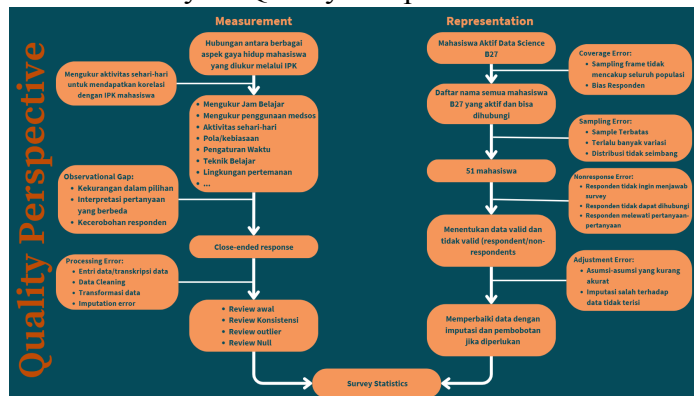
Lampiran

BabII. Desain Pengambilan Sampel

- Life Cycle Design Perspective



- Life Cycle Quality Perspective



BabVI. Hasil dan Analisa

- Missing value :

Jenis kelamin	0
Tempat Lahir	0
IPK	0
Jam Belajar UTS/UAS (1-4)	0
Ngatur Waktu Kuliah, Tugas, Belajar (1-5)	0
Mudah memahami materi (1-5)	0
Notes/ringkasan Kuliah (1-5)	0
Jadwal belajar teratur (1-5)	0
Memanfaatkan AI (1-5)	0
Revisi Pelajaran (1-5)	0
Jumlah Organisasi (0-3)	0
Terbebani Organisasi (1-5)	0
Kerja Samping (0/1)	0
Kerja Sampingan mengganggu (1-5)	7
Sosial Media (1-4)	0
Medsos Mengganggu (1-5)	0
Waktu Tidur (1-4)	0
Waktu Tidur Cukup (1-5)	0
Aktivitas Diluar Rumah (1-5)	0
Pengaruh Lingkungan mendukung Pembelajaran (1-5)	0
Teman2 mendukung Akademik (1-5)	0
Lingkungan pertemanan mempengaruhi IPK (1-5)	0
Dosen mempengaruhi akademik (1-5)	0

dtype: int64

- Statistic Summary :

	mean	std	min	max	range
ipk	3.503333	0.370317	2.49	3.99	1.5
jam_belajar_uts_uas_1_4	2.313725	1.009756	1.00	4.00	3.0
ngatur_waktu_kuliah_tugas_belajar_1_5	3.490196	1.046376	1.00	5.00	4.0
mudah_memahami_materi_1_5	3.372549	1.038476	1.00	5.00	4.0
notes_ringkasan_kuliah_1_5	3.529412	1.270479	1.00	5.00	4.0
jadwal_belajar_teratur_1_5	2.254902	1.293725	1.00	5.00	4.0
memanfaatkan_ai_1_5	4.411765	0.828890	2.00	5.00	3.0
revisi_pelajaran_1_5	2.529412	1.189167	1.00	5.00	4.0
jumlah_organisasi_0_3	1.156863	0.966903	0.00	3.00	3.0
terbebani_organisasi_1_5	2.843137	1.065318	1.00	5.00	4.0
kerja_samping_0_1	0.215686	0.415390	0.00	1.00	1.0
kerja_sampingan_mengganggu_1_5	2.941176	0.675626	1.00	5.00	4.0
sosial_media_1_4	2.588235	0.983391	1.00	4.00	3.0
medsos_mengganggu_1_5	3.333333	1.336663	1.00	5.00	4.0
waktu_tidur_1_4	2.529412	0.643520	1.00	4.00	3.0
waktu_tidur_cukup_1_5	3.352941	1.308794	1.00	5.00	4.0
aktivitas_diluar_rumah_1_5	3.411765	1.042621	1.00	5.00	4.0
pengaruh_lingkungan_mendukung_pembelajaran_1_5	4.019608	0.969334	1.00	5.00	4.0
teman2_mendukung_akademik_1_5	4.098039	0.806347	2.00	5.00	3.0
lingkungan_pertemanan_mempengaruhi_ipk_1_5	3.921569	1.074116	1.00	5.00	4.0
dosen_mempengaruhi_akademik_1_5	4.411765	0.804400	2.00	5.00	3.0

- Data preprocessing :

```
df[df['Jumlah Organisasi (0-3)'] == 0]['Terbebani Organisasi (1-5)'].value_counts()
```

```
Terbebani Organisasi (1-5)
```

```
3    13
```

```
1     1
```

```
4     1
```

```
5     1
```

```
Name: count, dtype: int64
```

```
df.loc[df['Jumlah Organisasi (0-3)'] == 0, 'Terbebani Organisasi (1-5)'] = 3
```

```
df[df['Jumlah Organisasi (0-3)'] == 0]['Terbebani Organisasi (1-5)'].value_counts()
```

```
Terbebani Organisasi (1-5)
```

```
3    16
```

```
Name: count, dtype: int64
```

```
df[df['Kerja Samping (0/1)'] == "Tidak"]['Kerja Sampingan mengganggu (1-5)'].value_counts()
```

```
Kerja Sampingan mengganggu (1-5)
```

```
3.0    31
```

```
1.0     2
```

```
Name: count, dtype: int64
```

```
df.loc[df['Kerja Samping (0/1)'] == "Tidak", 'Kerja Sampingan mengganggu (1-5)'] = 3
```

```
df[df['Kerja Samping (0/1)'] == "Tidak"]['Kerja Sampingan mengganggu (1-5)'].value_counts()
```

```
Kerja Sampingan mengganggu (1-5)
```

```
3.0    40
```

```
Name: count, dtype: int64
```

Multikolinearitas :

	Variable	VIF
0	const	232.979149
1	jadwal_belajar_teratur_1_5	4.152720
2	ngatur_waktu_kuliah_tugas_belajar_1_5	3.852043
3	pengaruh_lingkungan_mendukung_pembelajaran_1_5	3.593186
4	revisi_pelajaran_1_5	3.539671
5	kerja_sampingan_mengganggu_1_5	3.242123
6	waktu_tidur_cukup_1_5	3.024700
7	notes_ringkasan_kuliah_1_5	2.847112
8	waktu_tidur_1_4	2.763154
9	mudah_memahami_materi_1_5	2.533482
10	jumlah_organisasi_0_3	2.472184
11	teman2_mendukung_akademik_1_5	1.982384
12	sosial_media_1_4	1.928211
13	memanfaatkan_ai_1_5	1.863914
14	medsos_mengganggu_1_5	1.851847
15	lingkungan_pertemanan_mempengaruhi_ipk_1_5	1.780311
16	dosen_mempengaruhi_akademik_1_5	1.774543
17	kerja_samping_0_1	1.770862
18	terbebani_organisasi_1_5	1.699300
19	jam_belajar_uts_uas_1_4	1.552535
20	aktivitas_diluar_rumah_1_5	1.463990

- Regression Model :

```

--- Regression Statistics ---
Multiple R:          0.756774
R Square:            0.572707
Adjusted R Square:   0.287844
Standard Error:      0.312508
Observations:        51

```

- ANOVA Table (Type I)

	df	SS	MS	F	Significance F
jam_belajar_uts_uas_1_4	1.0	0.294791	0.294791	3.018503	0.092577
ngatur_waktu_kuliah_tugas_belajar_1_5	1.0	0.749353	0.749353	7.672984	0.009525
mudah_memahami_materi_1_5	1.0	0.024361	0.024361	0.249444	0.621110
notes_ringkasan_kuliah_1_5	1.0	0.202713	0.202713	2.075675	0.160020
jadwal_belajar_teratur_1_5	1.0	0.023759	0.023759	0.243276	0.625442
memanfaatkan_ai_1_5	1.0	0.000324	0.000324	0.003319	0.954439
revisi_pelajaran_1_5	1.0	0.086305	0.086305	0.883721	0.354693
jumlah_organisasi_0_3	1.0	0.000696	0.000696	0.007123	0.933299
terbebani_organisasi_1_5	1.0	0.479349	0.479349	4.908288	0.034462
kerja_samping_0_1	1.0	0.584128	0.584128	5.981166	0.020543
kerja_sampingan_mengganggu_1_5	1.0	0.024019	0.024019	0.245941	0.623562
sosial_media_1_4	1.0	0.004822	0.004822	0.049372	0.825665
medsos_mengganggu_1_5	1.0	0.261457	0.261457	2.677187	0.112245
waktu_tidur_1_4	1.0	0.000105	0.000105	0.001077	0.974036
waktu_tidur_cukup_1_5	1.0	0.002928	0.002928	0.029985	0.863686
aktivitas_diluar_rumah_1_5	1.0	0.064805	0.064805	0.663570	0.421722
pengaruh_lingkungan_mendukung_pembelajaran_1_5	1.0	0.412123	0.412123	4.219926	0.048754
teman2_mendukung_akademik_1_5	1.0	0.000119	0.000119	0.001217	0.972403
lingkungan_pertemanan_mempengaruhi_ipk_1_5	1.0	0.275992	0.275992	2.826013	0.103132
dosen_mempengaruhi_akademik_1_5	1.0	0.434747	0.434747	4.451585	0.043318
Residual	30.0	2.929837	0.097661	NaN	NaN

- Model Significant :

```

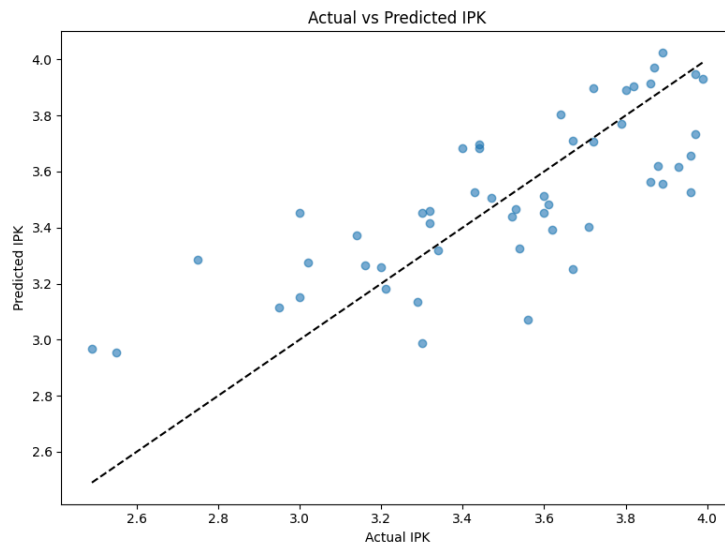
F-statistic: 2.010469
Significance F (p-value): 0.040624
The regression model is SIGNIFICANT at  $\alpha = 0.1$ 

```

- Coefficient Table :

	Coefficients	P-value
Intercept	1.050105	0.126401
jam_belajar_uts_uas_1_4	0.077481	0.165705
ngatur_waktu_kuliah_tugas_belajar_1_5	-0.013654	0.870273
mudah_memahami_materi_1_5	0.023054	0.735977
notes_ringkasan_kuliah_1_5	0.051279	0.389258
jadwal_belajar_teratur_1_5	0.175299	0.017367
memanfaatkan_ai_1_5	-0.084530	0.254706
revisi_pelajaran_1_5	-0.242040	0.001635
jumlah_organisasi_0_3	-0.083814	0.252711
terbebani_organisasi_1_5	0.096651	0.084013
kerja_samping_0_1	0.261090	0.075070
kerja_sampingan_mengganggu_1_5	0.180137	0.136644
sosial_media_1_4	-0.020391	0.746131
medsos_mengganggu_1_5	0.033289	0.465146
waktu_tidur_1_4	0.030841	0.788891
waktu_tidur_cukup_1_5	-0.022807	0.700502
aktivitas_diluar_rumah_1_5	0.010195	0.843773
pengaruh_lingkungan_mendukung_pembelajaran_1_5	0.143298	0.107730
teman2_mendukung_akademik_1_5	0.042337	0.587328
lingkungan_pertemanan_mempengaruhi_ipk_1_5	0.092694	0.101702
dosen_mempengaruhi_akademik_1_5	0.154420	0.043318

- Result Visualization (Linear Regression Visual) :



- Result Visualization (residual plot) :

