

FINAL PROJECT

ANALISA PENGARUH HUBUNGAN ANTARA RIWAYAT MEDIS DENGAN PENYAKIT DIABETES

Mata Kuliah IS240 – Probability And Statistic

Dosen Pengampu: Rudi Sutomo, S.Kom., M.Si., M.Kom.







Disusun Oleh :

Catherine Olivia	00000076229
Haura Putry Yasha	00000075900
Melissa Triana	00000077664
Zevanaya Beverly Drew	00000077121

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2023

HALAMAN IDENTITAS ANGGOTA KELOMPOK

NIM	Nama	Peran	Tanda Tangan
00000076229	Catherine Olivia	Mengkoordinasi tugas dan berdiskusi bersama anggota kelompok, bekerja sama dengan anggota kelompok, membuat bab 4, membuat kesimpulan, merapikan laporan dan daftar pustaka, melakukan pengecekan pada laporan dan PPT, membuat jalan kerja dan menguji hipotesa dalam R Studio, membuat hipotesa, menentukan data dan tipe data.	
00000075900	Haura Putry Yasha	Mengkoordinasi tugas dan berdiskusi bersama anggota kelompok, bekerja sama dengan anggota kelompok, membuat bab I (latar belakang, rumusan masalah, tujuan), membuat landasan teori dibagian pengertian, membuat rumus rumus bagian chi-square dan wilcoxon, membuat saran pada bagian penutup serta merapikan daftar pustaka.	
00000076664	Melissa Triana	Mengkoordinasi tugas dan berdiskusi bersama anggota kelompok, bekerja sama dengan anggota kelompok, membuat halaman kata pengantar, membuat deskripsi topik penelitian, membuat dan merapikan daftar isi, membuat bab 2 yang berisikan rumus bagian Tukey Kramer dan Korelasi Spearman serta hipotesis yang akan diuji, membuat bab 3 yang berisikan metode penelitian, metode pengumpulan data, data beserta tipe datanya, merapikan bab 5 bagian saran, membantu membuat slide PPT, melakukan pengecekan laporan dan merapikan daftar pustaka.	
00000077121	Zevanaya Beverly Drew	Mengkoordinasi tugas dan berdiskusi bersama anggota kelompok, bekerja sama dengan anggota kelompok, berdiskusi menentukan latar belakang, rumusan masalah dan tujuan,	

		menentukan hipotesis yang akan diuji, membuat diagram hipotesis, mencari definisi rumus (Kruskal-Wallis, ANOVA, TurkeyHSD), membuat bab 3 (Metode penelitian, instrumen penelitian), membuat PPT untuk video presentasi.	
--	--	--	--



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

KATA PENGANTAR

Dengan rasa syukur yang dalam, kami menyampaikan kata pengantar ini sebagai bagian dari tugas akhir kami dalam mata kuliah Probability and Statistic. Dalam kesempatan ini, kami ingin mengucapkan terima kasih kepada semua pihak yang telah membantu dan mendukung kami dalam perjalanan ini.

1. Ibu Ririn Ikana Desanti, S. Kom., M. Kom. selaku Ketua Program Studi, Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara.
2. Bapak Rudi Sutomo, S.Kom., M.Si., M. Kom. selaku Dosen Pengampu Mata Kuliah Probability and Statistic sebagai pembimbing yang telah memberikan bimbingan, materi, arahan, dan motivasi atas terselesainya laporan ini.
3. Ibu Ir. Endah Tri Esthi H.,MMSI selaku Dosen Pengampu Mata Kuliah Probability and and Statistic Universitas Nasional sebagai pembimbing yang telah memberikan bimbingan dan materi atas terselesainya laporan ini.
4. Kepada kelompok 3 yang sudah berpartisipasi dalam memberikan kontribusi pada proses penyelesaian laporan ini.
5. Teman-teman jurusan Sistem Informasi yang telah membantu dan memberikan semangat dalam penyusunan laporan ini,
6. Keluarga yang telah memberikan bantuan dukungan material dan moral, sehingga laporan ini dapat terselesaikan.
7. Serta kepada semua pihak yang telah memberikan motivasi serta semangat yang tidak dapat penulis sebutkan satu per satu.

Semoga karya ilmiah ini dapat menjadi acuan bagi penelitian berikutnya.

Tangerang, 12 Juni 2023

Kelompok 3

TOPIK PENELITIAN

ANALISA PENGARUH HUBUNGAN ANTARA RIWAYAT MEDIS DENGAN PENYAKIT DIABETES

DESKRIPSI TOPIK PENELITIAN

Penelitian ini bertujuan untuk menganalisis secara mendalam pengaruh hubungan antara riwayat medis dengan penyakit diabetes. Dapat diketahui bahwa riwayat medis merupakan kumpulan data yang berperan penting dalam dunia kesehatan. Data tersebut mencatat beberapa kondisi kesehatan yang pernah dialami oleh pasien. Kondisi kesehatan yang tercatat dalam riwayat medis pasien itu meliputi riwayat penyakit yang diderita seperti adanya penyakit jantung, penggunaan obat-obatan tertentu serta riwayat pola hidup yang buruk. Kondisi yang tercermin dalam riwayat medis termasuk pola hidup yang buruk bisa meningkatkan resiko pasien untuk terkena berbagai macam penyakit, salah satunya diabetes. Banyak orang yang berusaha menghindari diabetes dikarenakan diabetes adalah penyakit kronis yang memiliki dampak serius pada kesehatan dan kualitas hidup seseorang. Fokus utama dalam penelitian ini adalah untuk mengidentifikasi hubungan dari beberapa aspek yang terdapat di riwayat medis meliputi jenis kelamin, umur, kebiasaan merokok, dan beberapa penyakit yang meliputi penyakit jantung dan kadar gula darah dengan diabetes. Penelitian ini akan meninjau pengaruh dari jenis kelamin serta umur dalam riwayat kebiasaan merokok, selanjutnya penelitian ini akan meninjau pengaruh kebiasaan merokok tersebut dengan penyakit jantung dan gula darah dan yang terakhir penelitian ini akan meninjau pengaruh dari kedua penyakit tersebut dengan diabetes. Melalui analisis data yang dikumpulkan dari hubungan antara riwayat medis dengan penyakit diabetes, penelitian ini diharapkan dapat memberi wawasan yang lebih baik dan dapat membantu dalam mengidentifikasi faktor resiko yang dapat diintervensi sehingga memungkinkan upaya pencegahan yang lebih efektif terhadap diabetes.

DAFTAR ISI

HALAMAN IDENTITAS ANGGOTA KELOMPOK.....	1
KATA PENGANTAR.....	3
TOPIK PENELITIAN.....	4
DESKRIPSI TOPIK PENELITIAN.....	4
DAFTAR ISI.....	5
BAB I.....	6
PENDAHULUAN.....	7
1.1 Latar Belakang.....	7
1.2 Rumusan Masalah.....	8
1.3 Tujuan.....	8
BAB II.....	9
LANDASAN TEORI.....	9
2.1 Pengertian Diabetes.....	9
2.2 Teori Autoimunitas pada Diabetes Tipe 1.....	10
2.3 Teori Resistensi Insulin pada Diabetes Tipe 2.....	10
2.4 Teori Inflamasi dan Stress Oksidatif.....	10
2.5 Teori Stres Endoplasmic Reticulum (ER).....	10
2.6 Hipotesis.....	10
2.7 Rumus.....	12
2.7.1 Rumus Kruskal-Wallis.....	12
2.7.2 Rumus Anova.....	12
2.7.3 Rumus Tukey HSD.....	13
2.7.4 Rumus Tukey - Kramer.....	14
2.7.6 Rumus Wilcoxon.....	15
2.7.7 Rumus Korelasi Spearman.....	16
BAB III.....	17
METODE PENELITIAN.....	17
3.1 Metodologi Penelitian.....	17
3.2 Metode Pengumpulan Data.....	17
3.3 Instrumen Penelitian.....	17
3.4 Data dan Tipe Data.....	17
BAB IV.....	19
ANALISA DAN HASIL.....	19
4.1 Visualisasi Data.....	19
4.1.1 Jenis Kelamin.....	19
4.1.2 Umur.....	19
4.1.3 Riwayat Merokok.....	20
4.1.4 Penyakit Jantung.....	20

4.1.5 Gula Darah.....	21
4.1.6 Diabetes.....	21
4.2 Statistika Deskriptif.....	22
4.2.1 Mean/Rerata.....	22
4.2.2 Simpangan Baku/Standar Deviasi.....	22
4.2.3 Coefficient of Variation/Koefisien Variasi.....	22
4.3 Uji Normalitas.....	23
4.3.1 Skewness.....	23
4.3.2 Kurtosis.....	23
4.3.3 Anderson-Darling Test.....	23
4.3.4 Grafik Normalitas.....	24
4.4 Regresi.....	25
4.5 Hasil Uji Hipotesa.....	29
1) Hipotesa 1.....	29
2) Hipotesa 2.....	30
3) Hipotesa 3.....	31
4) Hipotesa 4.....	32
5) Hipotesa 5.....	34
6) Hipotesa 6.....	35
BAB V.....	36
KESIMPULAN DAN SARAN.....	36
5.1 Kesimpulan.....	36
5.2 Saran.....	37
DAFTAR PUSTAKA.....	38
LAMPIRAN.....	40

BAB I

PENDAHULUAN

1.1 Latar Belakang

Diabetes mellitus merupakan salah satu penyakit kronis yang memiliki prevalensi yang tinggi dan menjadi masalah kesehatan global. Penyakit ini ditandai dengan tingginya kadar gula darah dalam tubuh akibat ketidakmampuan tubuh dalam memproduksi atau menggunakan insulin dengan efektif. Diabetes mellitus dapat menyebabkan berbagai komplikasi serius, seperti penyakit jantung, kerusakan saraf, gangguan mata, gagal ginjal, dan gangguan sirkulasi.

Banyak faktor yang dapat berkontribusi terhadap risiko terjadinya diabetes mellitus, termasuk faktor genetik, pola makan yang tidak sehat, gaya hidup yang tidak aktif, kelebihan berat badan, dan riwayat medis seseorang. Riwayat medis mencakup berbagai informasi penting tentang kondisi kesehatan individu, termasuk riwayat penyakit yang pernah diderita, penggunaan obat-obatan tertentu, dan kebiasaan hidup yang buruk. Informasi ini dapat memberikan wawasan yang berharga dalam memahami faktor risiko yang terkait dengan diabetes mellitus.

Beberapa faktor yang sering dikaitkan dengan peningkatan risiko diabetes mellitus adalah jenis kelamin, umur, kebiasaan merokok, penyakit jantung, dan kadar gula darah. Namun, masih diperlukan penelitian yang lebih mendalam untuk menganalisis secara mendalam hubungan antara faktor-faktor ini dan diabetes mellitus.

Maka dari itu, kelompok kami memutuskan untuk menganalisis probabilitas untuk mengidentifikasi hubungan yang signifikan antara riwayat medis dan diabetes mellitus. Dengan menggunakan data yang disediakan oleh Kaggle, kami akan melakukan analisis mendalam tentang pengaruh hubungan antara riwayat medis dengan penyakit diabetes. serta kami juga akan menggunakan metode analisis statistik yang tepat untuk mengidentifikasi hubungan yang signifikan antara faktor-faktor riwayat medis dan risiko diabetes mellitus.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, penulis merumuskan masalah antara lain:

1. Apakah ada hubungan antara gender dengan riwayat merokok?
2. Apakah ada hubungan antara umur dan riwayat merokok?
3. Apakah ada hubungan antara riwayat merokok dan penyakit jantung?
4. Apakah ada hubungan antara riwayat merokok dan kadar gula darah?
5. Apakah ada hubungan antara penyakit jantung dan diabetes?
6. Apakah ada hubungan antara kadar gula darah dan diabetes?

1.3 Tujuan

Tujuan dari penelitian adalah sebagai berikut:

1. Mengetahui hubungan antara gender dengan riwayat merokok.
2. Mengetahui hubungan antara umur dengan riwayat merokok.
3. Mengetahui hubungan antara riwayat merokok dengan penyakit jantung.
4. Mengetahui hubungan antara riwayat merokok dengan kadar gula darah.
5. Mengetahui hubungan antara penyakit jantung dengan diabetes.
6. Mengetahui hubungan antara gula darah dengan penyakit diabetes.



BAB II

LANDASAN TEORI

2.1 Pengertian Diabetes

Dikutip dari Warsi Maryati, Aris Ocktavian Wannay, Devi Permani Suci (2018) Jurnal Rekam Medis dan Informasi Kesehatan. Diabetes merupakan penyakit metabolik yang ditandai dengan kadar gula darah yang tinggi (hiperglikemia) yang diakibatkan oleh gangguan sekresi insulin, dan resistensi insulin atau keduanya. Prevalensi DM menurut WHO, bahwa lebih dari 382 juta jiwa orang di dunia telah mengidap penyakit diabetes mellitus. Prevalensi DM di dunia dan Indonesia akan mengalami peningkatan, secara epidemiologi diperkirakan bahwa pada tahun 2030 prevalensi diabetes melitus (DM) di Indonesia mencapai 21,3 juta orang. Selain itu diabetes melitus menduduki peringkat keenam penyebab kematian terbesar di Indonesia.

Sedangkan menurut, American Diabetes Association. Tahun 202 (Standards of medical care in diabetes), Diabetes adalah kondisi kronis yang ditandai oleh tingginya kadar glukosa (gula) dalam darah akibat gangguan produksi atau penggunaan hormon insulin dalam tubuh. Insulin, hormon yang dihasilkan oleh pankreas, berperan penting dalam mengatur kadar glukosa dalam darah. Terdapat dua jenis utama diabetes, yaitu diabetes tipe 1 dan diabetes tipe 2. Pada diabetes tipe 1, sistem kekebalan tubuh menyerang dan merusak sel-sel pankreas yang memproduksi insulin. Sebagai hasilnya, tubuh tidak dapat menghasilkan insulin yang cukup, sehingga penderita diabetes tipe 1 memerlukan suntikan insulin seumur hidup untuk menjaga kadar gula darah normal. Sementara itu, diabetes tipe 2 terjadi ketika tubuh tidak efektif dalam menggunakan insulin atau tidak menghasilkan insulin dengan cukup. Faktor risiko yang berkontribusi pada diabetes tipe 2 termasuk obesitas, gaya hidup tidak sehat, faktor keturunan, dan faktor lingkungan. Diabetes tipe 2 dapat dikendalikan melalui perubahan gaya hidup, pola makan sehat, rutin berolahraga, dan penggunaan obat-obatan yang diresepkan. Gejala diabetes meliputi rasa haus yang berlebihan, sering buang air kecil, kelelahan yang berlebihan, penurunan berat badan yang tidak dapat dijelaskan, luka yang sulit sembuh, dan penglihatan kabur. Jika diabetes tidak terkelola dengan baik, dapat menyebabkan komplikasi serius seperti kerusakan saraf, gangguan ginjal, masalah kardiovaskular, bahkan mengancam nyawa. Oleh karena itu, penting untuk mendiagnosis dan mengelola diabetes dengan baik melalui pemantauan rutin, perubahan gaya hidup, pola makan yang

sehat, rutin berolahraga, dan penggunaan obat-obatan yang direkomendasikan oleh dokter.

2.2 Teori Autoimunitas pada Diabetes Tipe 1

Teori ini menyatakan bahwa diabetes tipe 1 disebabkan oleh kerusakan pada sistem kekebalan tubuh yang menyebabkan serangan pada sel-sel pankreas yang memproduksi insulin. Proses ini disebut sebagai reaksi autoimun. Autoantibodi dan sel-sel kekebalan tubuh lainnya menyerang dan merusak sel-sel beta di pankreas, yang mengakibatkan penurunan produksi insulin.

2.3 Teori Resistensi Insulin pada Diabetes Tipe 2

Teori ini mengemukakan bahwa diabetes tipe 2 terjadi ketika sel-sel tubuh menjadi kurang responsif terhadap insulin yang diproduksi, sehingga glukosa tidak dapat diserap dengan efektif oleh sel-sel tubuh. Ini dapat terjadi karena adanya perubahan dalam mekanisme sinyal insulin atau akumulasi lemak di dalam sel yang mengganggu fungsi insulin.

2.4 Teori Inflamasi dan Stress Oksidatif

Teori ini menyatakan bahwa peradangan kronis dan stres oksidatif dalam tubuh dapat mempengaruhi pengaturan glukosa dan resistensi insulin, serta merusak sel-sel beta pankreas yang memproduksi insulin. Faktor-faktor seperti obesitas, pola makan tidak sehat, dan gaya hidup tidak aktif dapat memicu peradangan dan stres oksidatif yang berkontribusi pada perkembangan diabetes.

2.5 Teori Stres Endoplasmic Reticulum (ER)

Teori ini berfokus pada peran stres endoplasmik retikulum (ER) dalam perkembangan diabetes. Stres ER dapat terjadi akibat kondisi seperti obesitas, resistensi insulin, dan lingkungan metabolik yang tidak normal. Stres ER dapat menyebabkan kerusakan pada sel-sel beta pankreas dan mempengaruhi produksi insulin.

2.6 Hipotesis

Terdapat 6 hipotesis yang akan diuji dalam penelitian, antara lain :

1) Hipotesa 1

H₀ : Tidak terdapat hubungan antara gender dengan riwayat merokok.

H_a : Terdapat hubungan antara gender dengan riwayat merokok.

Asumsi : Terdapat hubungan antara gender dengan riwayat merokok.

Rumus : Kruskal Wallis

2) Hipotesa 2

H₀ : Tidak terdapat hubungan antara umur dengan riwayat merokok.

H_a : Terdapat hubungan antara umur dengan riwayat merokok.

Asumsi : Terdapat hubungan antara umur dengan riwayat merokok.

Rumus : Anova, Turkey HSD

3) Hipotesa 3

H₀ : Tidak terdapat hubungan antara riwayat merokok dengan penyakit jantung.

H_a : Terdapat hubungan antara riwayat merokok dengan penyakit jantung.

Asumsi : Terdapat hubungan antara riwayat merokok dengan penyakit jantung.

Rumus : Chi-Square

4) Hipotesa 4

H₀ : Tidak terdapat hubungan antara riwayat merokok dengan gula darah.

H_a : Terdapat hubungan antara riwayat merokok dengan gula darah.

Asumsi : Terdapat hubungan antara riwayat merokok dengan gula darah.

Rumus : Anova, Tukey Kramer (HSD Kramer)

5) Hipotesa 5

H₀ : Tidak terdapat hubungan antara penyakit jantung dengan diabetes.

H_a : Terdapat hubungan antara penyakit jantung dengan diabetes.

Asumsi : Terdapat hubungan antara penyakit jantung dengan diabetes.

Rumus : Wilcoxon Mann Whitney

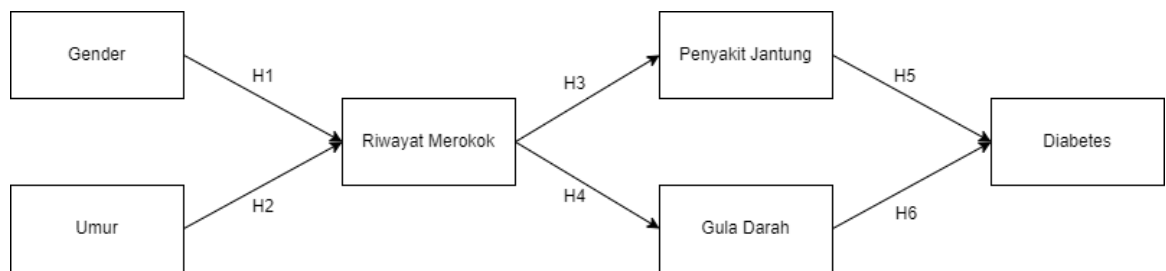
6) Hipotesa 6

H₀ : Tidak terdapat hubungan antara gula darah dengan diabetes.

H_a : Terdapat hubungan antara gula darah dengan diabetes.

Asumsi : Terdapat hubungan antara gula darah dengan diabetes.

Rumus : Kruskal Wallis



Gambar 1. Kerangka Pemikiran

2.7 Rumus

2.7.1 Rumus Kruskal-Wallis

Uji Kruskal-Wallis merupakan salah satu metode statistik non-parametrik yang digunakan untuk menguji apakah terdapat perbedaan signifikan antara kelompok-kelompok yang merupakan variabel independen dengan variabel dependen lainnya. Metode ini cocok digunakan ketika kita ingin membandingkan lebih dari dua kelompok populasi dengan data berbentuk ranking atau ordinal. Berikut ini merupakan rumus Kruskal-Wallis :

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1)$$

Dengan keterangan sebagai berikut :

H : nilai Kruskal-Wallis dari hasil penghitungan

R_i : jumlah rank dari kelompok/kategori ke-i

n_i : banyaknya kasus dalam sampel pada kelompok /kategori ke-i

k : banyaknya kelompok/kategori

N : jumlah seluruh observasi (N = n₁ + n₂ + n₃ + ... + n_k)

2.7.2 Rumus Anova

ANOVA adalah teknik uji statistik yang digunakan untuk membandingkan dua atau lebih variabel penelitian. Menurut G.Keller (2014), ANOVA umumnya digunakan untuk menguji signifikansi dari perbedaan

pengamatan eksperimental atau variabel yang dibandingkan. Statistik uji ANOVA dapat disajikan dalam bentuk tabel berikut.

Sumber Variasi	Sum of Squares (SS)	Degree of Freedom (df)	Mean Squares (MS)	F_{hit}
Perlakuan (Tr)	$SSTr$	$k - 1$	$MSTr$	$\frac{MSTr}{MSE}$
Error (E)	SSE	$n - k$	MSE	
Total (T)	SST	$n - 1$		

Dengan keterangan sebagai berikut :

- k = banyaknya perlakuan
 n = banyaknya sampel gabungan
 $SSTr$ = jumlah kuadrat perlakuan atau treatment
 SSE = jumlah kuadrat error
 $MSTr$ = rata-rata kuadrat perlakuan
 F_{hit} = statistik uji ANOVA

2.7.3 Rumus Tukey HSD

Pengujian Tukey HSD (Honestly Significant Difference) adalah metode pengujian perbandingan jamak yang digunakan untuk menentukan apakah terdapat perbedaan signifikan antara rata-rata dari setiap kelompok dalam analisis varian. Tujuan dari penelitian ini adalah untuk mengetahui kelompok mana yang memiliki rata-rata yang berbeda secara signifikan. Metode ini membantu mengidentifikasi perbedaan yang spesifik antara kelompok-kelompok tersebut, sehingga dapat memberikan pemahaman yang lebih mendalam tentang hubungan antara variabel dalam dataset yang diteliti.

$$\omega = q_{\alpha}(p, v) \sqrt{\frac{KTG}{r}}$$

Dengan keterangan sebagai berikut :

- p = Jumlah perlakuan
 v = Derajat batas galat
 r = Banyaknya ulangan
 α = Taraf nyata
 $q_{\alpha}(p, v)$ = Nilai kritis

2.7.4 Rumus Tukey - Kramer

Menurut dari Purwanto, 2011: 205, Uji Tukey - Kramer adalah uji lanjut setelah uji anova yang membandingkan beberapa kelompok dengan jumlah sampel yang tidak sama besar. Uji Tukey - Kramer sering digunakan dalam analisis perbandingan pasangan ganda antara kelompok-kelompok yang ada di dalam satu variabel untuk mengidentifikasi perbedaan signifikan antara semua pasangan rata-rata kelompok secara keseluruhan.

Uji Tukey - Kramer dapat dirumuskan sebagai berikut :

$$Q = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{\frac{MS_{error}}{n}}}$$

Dimana :

Q = Nilai uji

\bar{X}_i = Rata - rata kelompok ke - i

\bar{X}_j = Rata - rata kelompok ke - j

MS_{error} = Mean square error yaitu varians yang dihitung dari selisih antara nilai observasi dengan rata-rata kelompok.

n = jumlah observasi dalam setiap kelompok

2.7.5 Rumus Chi Square

Dikutip dari Igo Cahya.N. dan Agung Prabowo, Chi-square disebut juga dengan Kai Kuadrat. Uji Chi-square adalah salah satu jenis uji komparatif non parametris yang dilakukan pada dua variabel, di mana skala data kedua variabel adalah nominal (Sutrisno, 2000). Apabila dari 2 variabel, ada 1 variabel dengan skala nominal maka dilakukan uji Chi-square dengan merujuk bahwa harus digunakan uji pada derajat yang terendah. Uji Chi Square merupakan uji non parametrik yang paling banyak digunakan.

Uji Chi-square dapat dirumuskan sebagai berikut:

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}$$

Dimana :

χ^2 = Distribusi Chi-square

O_i = Nilai observasi (pengamatan) ke-i

E_i = Nilai ekspektasi ke-i

2.7.6 Rumus Wilcoxon

Dikutip dari Jurnal Wahyuni Saputri, Sigit Nugroho, dan Fachri Faisal Wilcoxon Signed Rank Test atau Uji Wilcoxon-Mann-Whitney digunakan untuk menguji hipotesis nol tentang kesamaan parameter lokasi. Selain itu, uji ini juga dapat diterapkan untuk menguji hipotesis tentang sembarang perbedaan dalam populasi antara kedua sampel. Asumsi-asumsi yang digunakan untuk metode ini (Conover, 1980) adalah:

- Masing-masing sampel adalah sampel acak
- Skala pengukuran yang digunakan sekurang-kurangnya ordinal.
- Variabel yang diamati kontinu.
- Apabila terdapat perbedaan antara fungsi distribusi populasi, maka perbedaan yang dimaksud adalah perbedaan distribusi lokasi.

Rumus uji Wilcoxon Signed Rank Test sebagai berikut :

$$Z = \frac{T - \sigma_T}{\sigma_T} = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

T = jumlah ranking bertanda kecil

n = banyaknya pasangan yang tidak sama nilainya

Signifikansi, nilai Z dibandingkan dengan tabel kurva normal, H_0 ditolak jika $Z_{hitung} \geq Z_{tabel}$, H_0 diterima jika $Z_{hitung} < Z_{tabel}$.

2.7.7 Rumus Korelasi Spearman

Korelasi Spearman merupakan pengukuran non-parametrik. Pengukuran dengan koefisien korelasi spearman ini digunakan untuk mengukur hubungan atau korelasi antara dua variabel ordinal atau nominal berdasarkan data yang berbentuk *ranking*. Terdapat dua metode dalam perhitungan pengujian korelasi spearman.

1. Apabila tidak terdapat ranking yang sama / kembar (*tied rank*)

$$r_s = 1 - \frac{6 \sum d_i^2}{(n^3 - n)}$$

Dimana :

d_i = perbedaan peringkat berpasangan

dan n = jumlah kasus

2. Apabila terdapat ranking yang sama / kembar (*tied rank*)

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

Dimana :

i = skor berpasangan

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

BAB III

METODE PENELITIAN

3.1 Metodologi Penelitian

Dalam penelitian ini, kelompok peneliti menggunakan penelitian berbasis data sekunder. Penelitian berbasis data sekunder adalah penelitian dengan menggunakan data yang sudah ada dan data tersebut diperoleh dari sumber terkait yang telah dikumpulkan oleh pihak lain pada penelitian sebelumnya untuk tujuan penelitian baru. Kelompok peneliti memilih menggunakan penelitian berbasis data sekunder untuk menjawab pertanyaan penelitian yang baru dan juga untuk menguji hipotesis baru.

3.2 Metode Pengumpulan Data

Metode pengumpulan data yang kelompok peneliti gunakan adalah metode pendekatan studi literatur yaitu mengumpulkan dan mengolah data - data yang berkaitan dan sudah ada sebagai referensi dalam menyelesaikan penelitian. Kelompok peneliti menggunakan dataset yang bersumber dari *website* Kaggle dengan data yang berjudul *Diabetes Prediction Dataset*. *Diabetes Prediction Dataset* merupakan dataset yang berisi kumpulan data medis dan demografis dari pasien, beserta status diabetes mereka (positif atau negatif). Kelompok peneliti memilih dataset ini untuk mengeksplorasi hubungan antara berbagai faktor medis dan demografis dengan kemungkinan mengembangkan diabetes.

3.3 Instrumen Penelitian

Dalam penelitian ini, kelompok peneliti menggunakan dua alat analisis data, yaitu Microsoft Excel dan R Studio. Microsoft Excel digunakan untuk mengelola dan memvisualisasikan output dataset yang diambil dari *website* Kaggle, sementara R Studio digunakan untuk melakukan pemrosesan dan analisis data yang lebih lanjut. Kombinasi kedua alat tersebut memungkinkan tim peneliti untuk mengoptimalkan penelitian dari data yang sudah ada dengan lebih efektif.

3.4 Data dan Tipe Data

Tipe data yang digunakan dalam penelitian ini adalah numerik dan kategorikal. Berikut rincian tipe data yang digunakan :

- Numerik :
 - 1) Blood Glucose Level atau Riwayat Gula Darah
 - 2) Age atau Umur
- Kategorikal :
 - 3) Gender atau Jenis Kelamin
 - 4) Smoking History atau Riwayat Merokok
 - 5) Heart Disease atau Penyakit Jantung
 - 6) Diabetes



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

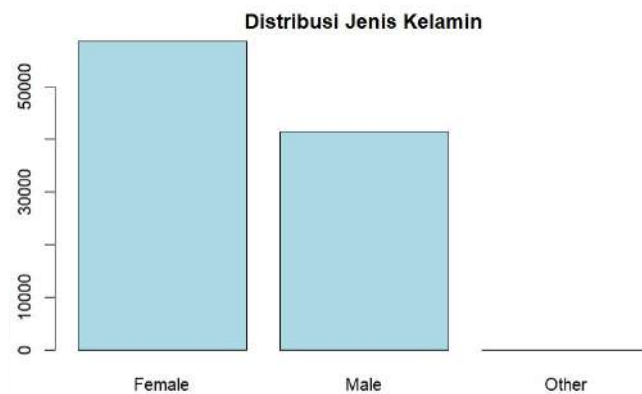
BAB IV

ANALISA DAN HASIL

4.1 Visualisasi Data

4.1.1 Jenis Kelamin

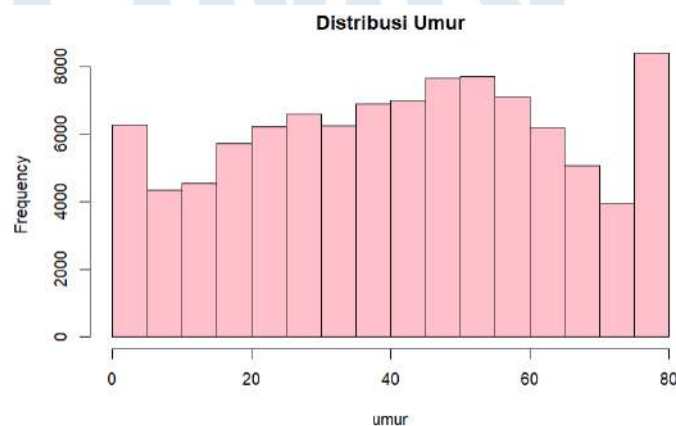
```
# Gender~  
barplot(table(diabetes1$gender), main = "Distribusi Jenis Kelamin", col = "lightblue")~
```



Berdasarkan gambar barplot diatas, jenis kelamin yang ada di dalam dataset Diabetes ini dibagi menjadi 3 yaitu Female, Male, dan Other. Jenis kelamin yang paling banyak dalam dataset ini adalah Female atau Perempuan.

4.1.2 Umur

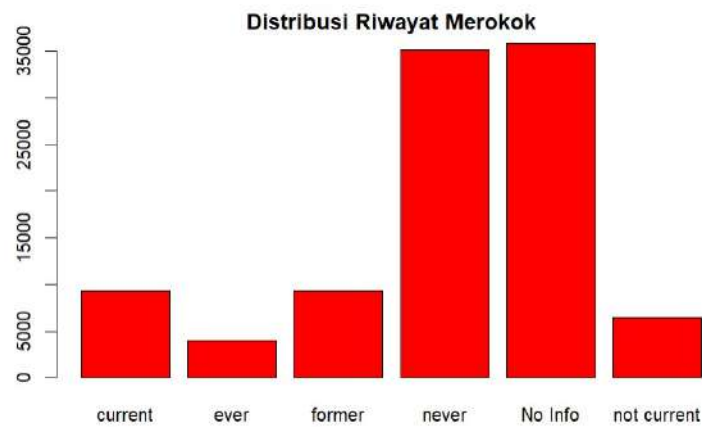
```
# Umur~  
hist(umur, main = "Distribusi Umur", col = "pink")~
```



Tabulasi di atas menunjukkan distribusi yang ada di dalam dataset Diabetes. Distribusi umur ini berada pada *range* mulai dari 0.08 tahun hingga 80.0 tahun. Data didominasi oleh umur yang berada pada range 75.0 hingga 80.0 tahun.

4.1.3 Riwayat Merokok

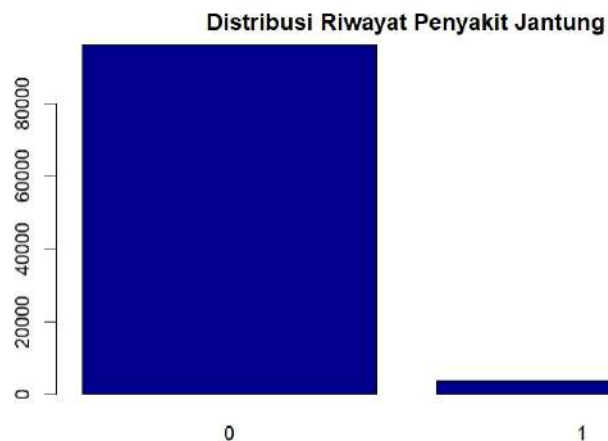
```
# Riwayat Merokok-  
barplot(table(diabetes1$smoking_history), main = "Distribusi Riwayat Merokok", col = "red")
```



Gambar di atas menunjukkan distribusi Riwayat Merokok dari populasi yang dibagi menjadi 6 range yaitu “Current”, “Ever”, “Former”, “Never”, “No Info”, dan “Not Current”. Data didominasi oleh range “No Info” yang berbeda sedikit dengan “Never”.

4.1.4 Penyakit Jantung

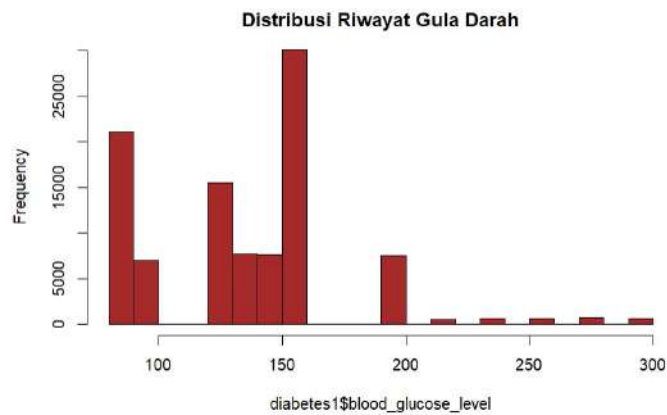
```
# Penyakit Jantung-  
barplot(table(diabetes1$heart_disease), main = "Distribusi Riwayat Penyakit Jantung", col = "darkblue")
```



Distribusi Riwayat Penyakit Jantung digambarkan oleh tabulasi di atas. Mayoritas populasi memilih “0” yang berarti tidak memiliki riwayat penyakit jantung. Namun, tetap ada sebagian kecil dari populasi yang memilih “1” yang berarti memiliki riwayat penyakit jantung.

4.1.5 Gula Darah

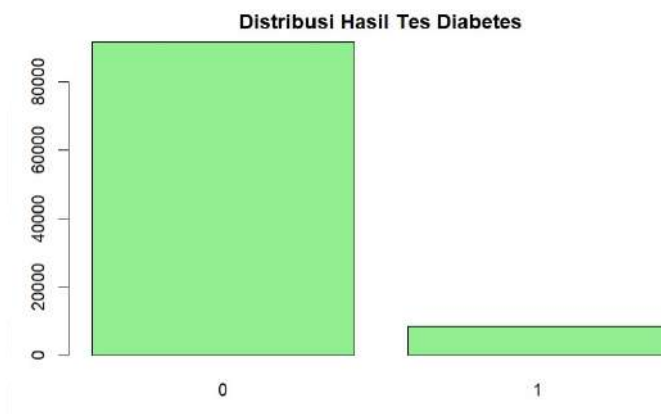
```
# Gula Darah  
hist(diabetes1$blood_glucose_level, main = "Distribusi Riwayat Gula Darah", col = "brown")
```



Tabulasi di atas merepresentasikan Riwayat Gula Darah dari populasi. Riwayat Gula Darah mempunyai *range* dari angka 80 hingga 300. Sebagian besar Riwayat Gula Darah populasi berada di *range* angka 150 hingga 160.

4.1.6 Diabetes

```
# Diabetes  
barplot(table(diabetes1$diabetes), main = "Distribusi Hasil Tes Diabetes", col = "lightgreen")
```



Tabulasi di atas menggambarkan distribusi Hasil Tes Diabetes yang dilakukan oleh populasi. Mayoritas populasi tidak memiliki penyakit diabetes yang direpresentasikan dengan angka “0”. Namun, tetap ada sebagian kecil dari populasi yang memiliki hasil tes diabetes positif, yang dinyatakan dengan angka “1”.

4.2 Statistika Deskriptif

4.2.1 Mean/Rerata

```
{r}
# Rerata Variabel Numerik
(meanUmur <- mean(umur))
(meanGulaDarah <- mean(diabetes1$blood_glucose_level))

[1] 41,88586
[1] 138,0581
```

Rata-rata individu pada sampel berumur 41,8 tahun. Sedangkan, rata-rata riwayat gula darah individu berada di angka 138,05.

4.2.2 Simpangan Baku/Standar Deviasi

```
{r}
# Simpangan Baku atau Standar Deviasi Variabel Numerik
(sdUmur <- sd(umur))
(sdGulaDarah <- sd(diabetes1$blood_glucose_level))

[1] 22,51684
[1] 40,70814
```

Simpangan baku digunakan untuk mengukur kesamaan sebaran suatu data dengan *mean* data itu sendiri. Simpangan baku umur individu pada sampel adalah 22,51. Sedangkan, simpangan baku riwayat gula darah individu pada sampel adalah 40,70.

4.2.3 Coefficient of Variation/Koefisien Variasi

```
{r}
# Koefisien Variasi Variabel Numerik
(cvUmur <- sdUmur / meanUmur) * 100
(cvGulaDarah <- sdGulaDarah / meanGulaDarah) * 100

[1] 53,75762
[1] 29,48624
```

Data yang mempunyai koefisien variasi yang tinggi memiliki arti bahwa data tersebut tersebar dengan luas. Koefisien variasi umur individu pada sampel adalah 53,75. Sedangkan, koefisien variasi riwayat gula darah individu pada sampel adalah 29,48. Berdasarkan hasilnya, dapat disimpulkan bahwa data umur mempunyai koefisien variasi paling tinggi yang menandakan bahwa data umur mempunyai persebaran yang paling luas. Berbeda dengan data riwayat gula darah yang kurang tersebar dengan luas.

4.3 Uji Normalitas

4.3.1 Skewness

```
## {r}-  
# Skewness Variabel Numerik-  
(skewnessUmur <- skewness(umur))-  
(skewnessGulaDarah <- skewness(diabetes1$blood_glucose_level))-  
##-  
[1] -0,05197822  
[1] 0,8216426
```

Skewness merupakan ukuran ketidaksimetrisan dalam distribusi nilai. Skewness dapat bernilai positif, negatif, dan nol. Data yang normal adalah data yang memiliki hasil skewness mendekati 0. Data umur memiliki hasil -0,05 yang berarti data umur berdistribusi secara simetris. Sedangkan, data riwayat gula darah memiliki hasil 0,8 yang berarti kemiringan data ini moderat atau sedang.

4.3.2 Kurtosis

```
## {r}-  
# Kurtosis Variabel Numerik-  
(kurtosisUmur <- kurtosis(umur))-  
(kurtosisGulaDarah <- kurtosis(diabetes1$blood_glucose_level))-  
##-  
[1] 1,996155  
[1] 4,737477
```

Kurtosis adalah indikator untuk menunjukkan derajat keruncingan dari suatu data. Semakin besar nilai kurtosis maka kurva akan semakin runcing. Nilai referensi kurtosis adalah 3. Data umur memiliki kurtosis 1,99 yang lebih rendah dari 3. Ini berarti data umur memiliki kurva distribusi yang rendah. Sedangkan, data riwayat gula darah memiliki kurtosis 4,73 yang lebih tinggi dari 3. Berbeda dengan data umur, data riwayat gula darah ini memiliki kurva distribusi yang sangat runcing.

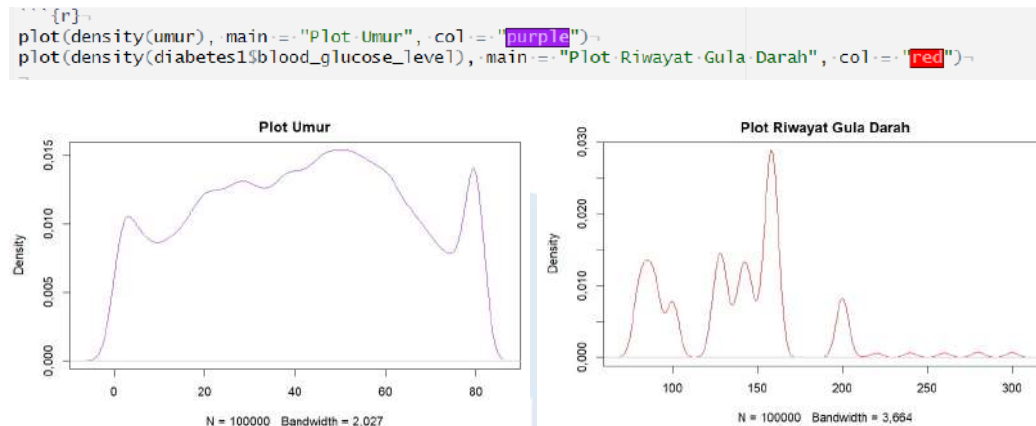
4.3.3 Anderson-Darling Test

```
# Anderson-Darling Test-  
ad.test(umur)-  
ad.test(diabetes1$blood_glucose_level)-  
##-  
  
Anderson-Darling normality test  
data: umur  
A = 610,11, p-value < 2,2e-16  
  
Anderson-Darling normality test  
data: diabetes1$blood_glucose_level  
A = 2556,6, p-value < 2,2e-16
```

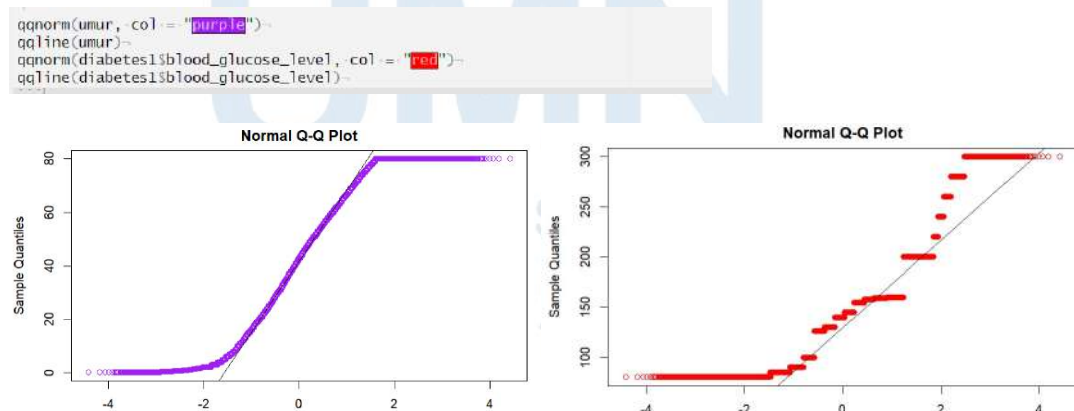
Selain menggunakan *skewness* dan *kurtosis*, uji Anderson-Darling dapat digunakan untuk mengukur normalitas distribusi suatu data ketika sampel data lebih dari 3000. Suatu data dapat dikatakan normal apabila *p-value* >

alpha (0,05). Data umur dan riwayat gula darah memiliki $p\text{-value} < 2,2\text{e-}16$ atau 0.000000000000000022, yang berarti kedua data tersebut tidak berdistribusi dengan normal karena $p\text{-value}$ yang kurang dari 0,05.

4.3.4 Grafik Normalitas



Berdasarkan gambar kedua density plot di atas yang merepresentasikan data umur dan data riwayat gula darah, terlihat pola yang dapat menandakan normalitas suatu sebaran data. Plot umur lebih dari satu puncak yang menandakan bahwa distribusi data tersebut tidak normal. Begitu pula dengan plot riwayat gula darah. Kedua gambar plot ini mendukung hasil dari uji normalitas Anderson-Darling yang menyatakan bahwa $p\text{-value} < 0.000000000000000022$ kurang dari alpha (0,05) yang berarti data tidak berdistribusi dengan normal.



Selain menggunakan *density plot*, normalitas suatu data juga dapat divisualisasikan dengan *Q-Q Plot*. Hasil dari kedua *Q-Q Plot* serupa, dimana terlihat bahwa data umur dan data riwayat gula darah tidak berada pada garis dengan tepat, yang berarti bahwa kedua data tersebut tidak mengikuti distribusi normal.

4.4 Regresi

Analisis regresi merupakan analisis yang digunakan untuk mempelajari dan mengukur hubungan statistik yang terjadi antara dua variabel atau lebih variabel. Variabel tersebut adalah variabel independent (variabel yang mempengaruhi) dan variabel dependent (variabel yang dipengaruhi). Analisis ini bertujuan untuk mengukur “seberapa kuat” atau “derajat kedekatan”, suatu relasi yang terjadi antar variabel.

Pengukuran dapat menggunakan koefisien korelasi (r) yaitu bilangan yang menyatakan kekuatan hubungan antar variabel ataupun arah hubungan dari variabel-variabel. Nilai r bergerak dari -1 sampai 1. Kekuatan hubungan diketahui dari nilai angka, sedangkan arah dinyatakan dalam bentuk positif (+) atau negatif (-). Berikut tabel tingkatan korelasi dan kekuatan hubungannya.

Nilai Korelasi (r)	Tingkat Hubungan
0	Tidak berkorelasi
0,01 - 0,20	Sangat rendah
0,21 - 0,40	Rendah
0,41 - 0,60	Agak rendah
0,61 - 0,80	Cukup
0,81 - 0,99	Tinggi
1	Sangat tinggi

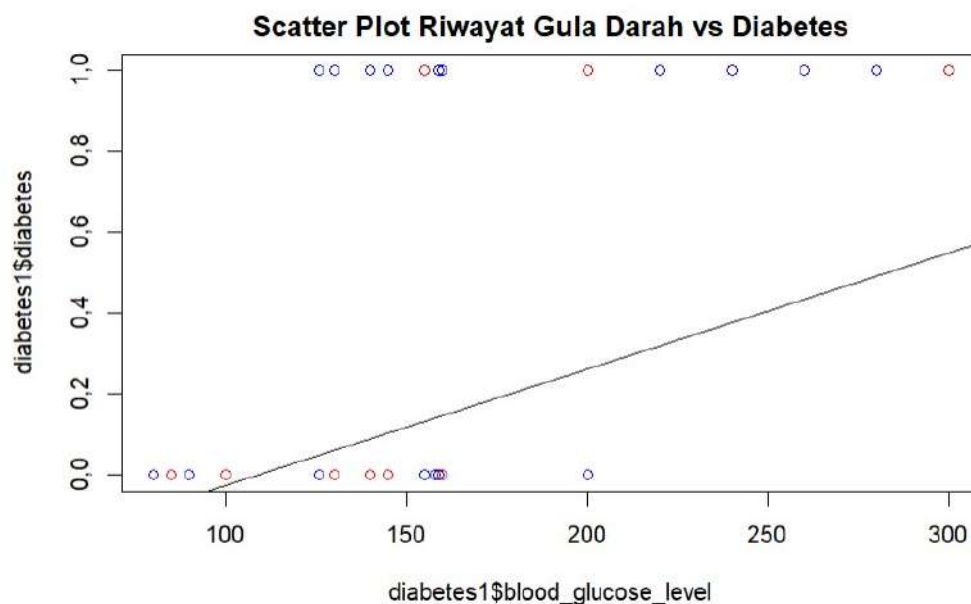
```
cor.test(diabetes1$blood_glucose_level, diabetes1$diabetes, method = "spearman")
```

```
Warning: Cannot compute exact p-value with ties  
Spearman's rank correlation rho
```

```
data: diabetes1$blood_glucose_level and diabetes1$diabetes  
S = 1,2096e+14, p-value < 2,2e-16  
alternative hypothesis: true rho is not equal to 0  
sample estimates:  
rho  
0,2742168
```

Korelasi antara variabel `blood_glucose_level` dengan variabel `diabetes` diuji menggunakan metode Spearman karena kedua variabel tersebut tidak berdistribusi dengan normal. Berdasarkan hasil uji diatas, nilai korelasi berada di angka 0,27. Dengan demikian, korelasi antara kedua variabel tersebut terbukti berada di kategori 0,21 - 0,40 yang berarti rendah.

```
judul = "Scatter Plot Riwayat Gula Darah vs Diabetes"
plot(diabetes1$blood_glucose_level, diabetes1$diabetes, col = c("red", "blue"), main = judul)
abline(lm(diabetes1$diabetes~diabetes1$blood_glucose_level))
```



Scatter plot di atas menggambarkan pola linear yang sangat lemah dikarenakan plot tidak menyebar dengan rata mengikuti garis. Hal ini dapat dibuktikan melalui uji korelasi yang telah dilakukan sebelumnya sekaligus memperkuat bukti bahwa korelasi antara dua variabel ini rendah.

```
#modeling
model <- lm(diabetes1$blood_glucose_level~diabetes1$diabetes)
summary(model)
```

```
Call:
lm(formula = diabetes1$blood_glucose_level ~ diabetes1$diabetes)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-68,095 -34,095   7,148  25,148 105,905
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  132,8525     0,1222  1087,5  <2e-16 ***
diabetes1$diabetes  61,2422     0,4190   146,2  <2e-16 ***
---
Signif. codes:  0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1
```

```
Residual standard error: 36,95 on 99998 degrees of freedom
Multiple R-squared:  0,176,    Adjusted R-squared:  0,176
F-statistic: 2,136e+04 on 1 and 99998 DF,  p-value: < 2,2e-16
```

Terdapat berbagai hal yang dapat disimpulkan berdasarkan ringkasan dari hasil modeling yang telah dilakukan, antara lain:

1. *Intercept* = $b_0 = 132,8525$

Diabetes = $b_1 = 61,2422$

2. Tiga bintang di samping kanan bawah $Pr(>|t|)$ yang menandakan bahwa tingkat signifikansi yang cukup tinggi.

3. *Residual Standard Error* (RSE) = 36,95

4. *Degrees of Freedom* (DoF) = 99.998

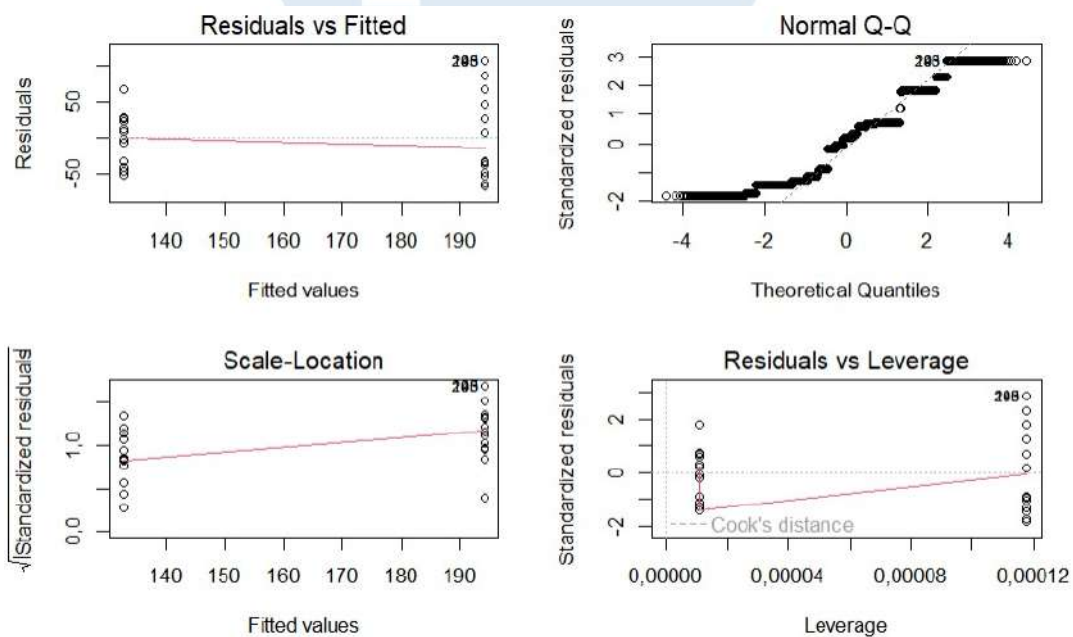
5. *R-squared* < 0,75 mempresentasikan variabilitas yang lemah, terdapat 17% variabilitas.

- *Multiple R-squared* = 0,176

- *Adjusted R-squared* = 0,176

6. *P-value* ($<2,2e-16$) < 0,05 menandakan signifikansi dan data dapat dipercaya kebenarannya..

```
#Pengecekan Asumsi Regresi Linear~  
par(mfrow = c(2,2))~  
plot(model)~  
par(mfrow = c(1,1))~
```



- Plot 1: Terdapat pelanggaran linieritas (TIDAK TERPENUHI)
- Plot 2: Data tidak berdistribusi dengan normal (TIDAK TERPENUHI)
- Plot 3: Varians residual homoskedastis (TIDAK TERPENUHI)
- Plot 4 : Ada pengaruh *influential point* atau *outliers* yang serius dalam hasil analisis (TIDAK TERPENUHI)

```
#uji otokorelasi  
lmtest::dwtest(model)
```

Durbin-Watson test

```
data: model  
DW = 2,0036, p-value = 0,7137  
alternative hypothesis: true autocorrelation is greater than 0
```

Berdasarkan data di atas, residual tidak berotokorelasi karena $p\text{-value} (0,7137) > 0,05$.

```
#uji homoskedastisitas  
lmtest::bptest(model)
```

studentized Breusch-Pagan test

```
data: model  
BP = 14896, df = 1, p-value < 2,2e-16
```

Menurut uji Breusch-Pagan, varians berubah-ubah dan residual tidak homoskedastis karena $p\text{-value} (<2,2e-16) < 0,05$.

```
#uji normalitas  
ad.test(model$residuals)
```

Anderson-Darling normality test

```
data: model$residuals  
A = 2582,9, p-value < 2,2e-16
```

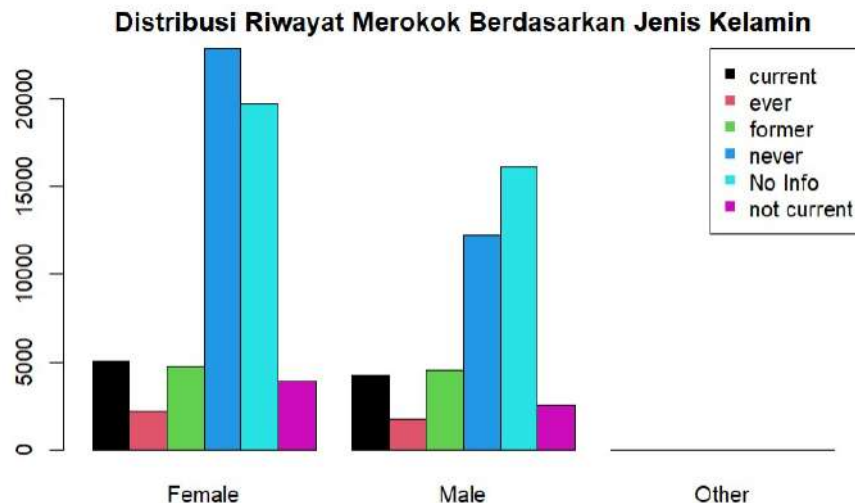
Dari uji Anderson-Darling, didapat bahwa data tidak berdistribusi dengan normal karena $p\text{-value} (<2,2e-16) < 0,05$.

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

4.5 Hasil Uji Hipotesa

1) Hipotesa 1

```
# Uji Hipotesa 1
judul <- "Distribusi Riwayat Merokok Berdasarkan Jenis Kelamin"
tabel <- table(diabetes1$smoking_history, diabetes1$gender)
barplot(tabel, main = judul, col = 1:nrow(tabel), beside = TRUE)
legend("topright", legend = rownames(tabel), pch = 15, col = 1:nrow(tabel))
```



Barplot di atas menggambarkan distribusi riwayat merokok berdasarkan jenis kelamin yang ada di dalam data sampel. Pada jenis kelamin *Female*, riwayat merokok “*Never*” paling banyak. Sedangkan, pada *gender Male*, riwayat merokok yang mendominasi adalah “*No Info*”.

```
kruskal.test(diabetes1$gender~diabetes1$smoking_history)
```

kruskal-wallis rank sum test

data: diabetes1\$gender by diabetes1\$smoking_history
Kruskal-wallis chi-squared = 1121,5, df = 5, p-value < 2,2e-16

Decision Rules :

- H_0 ditolak saat $p\text{-value} < 0,05$ dan terdapat hubungan antara gender dengan riwayat merokok.
- H_0 diterima atau tidak ditolak saat $p\text{-value} > 0,05$ dan tidak terdapat hubungan antara gender dengan riwayat merokok.

Hasil :

- $p\text{-value} < \alpha$ ($2,2e-16 < 0,05$)
- H_0 ditolak pada $\alpha = 0,05$.
- Terdapat hubungan antara gender dengan riwayat merokok.

2) Hipotesa 2

```
# Uji Hipotesa 2
result1 <- aov(umur ~ diabetes1$smoking_history)
result1
summary(result1)

th <- TukeyHSD(result1, conf.level = 0.95)
th
plot(th)
```

```
Call:
aov(formula = umur ~ diabetes1$smoking_history)

Terms:
             diabetes1$smoking_history Residuals
Sum of Squares              5262593  45437708
Deg. of Freedom                  5    99994

Residual standard error: 21,31676
Estimated effects may be unbalanced
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
diabetes1\$smoking_history	5	5262593	1052519	2316	<2e-16 ***
Residuals	99994	45437708	454		

```
---
Signif. codes:  0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1
Tukey multiple comparisons of means
 95% family-wise confidence level

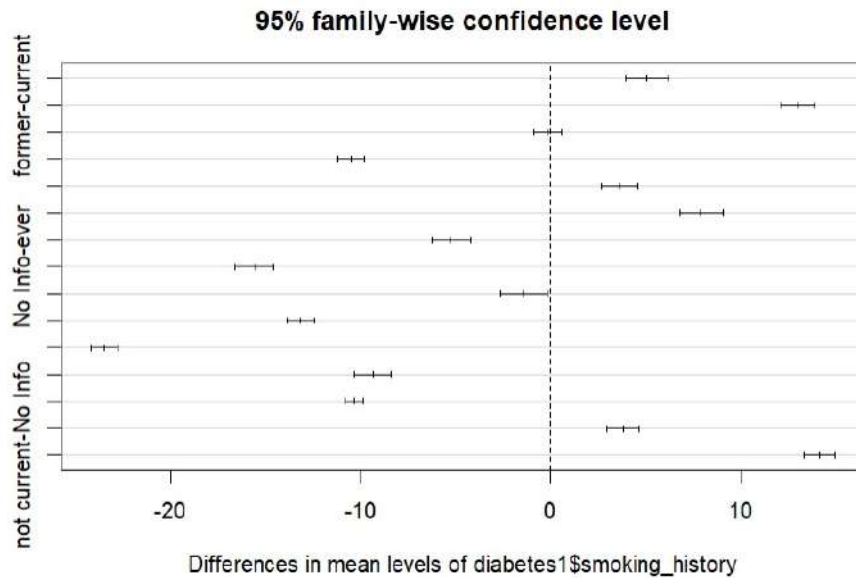
Fit: aov(formula = umur ~ diabetes1$smoking_history)

$`diabetes1$smoking_history`
              diff              lwr              upr              p adj
ever-current      5,0731673      3,9246902      6,2216443 0,0000000
former-current    12,9980021     12,1080754     13,8879289 0,0000000
never-current     -0,1722048     -0,8811012      0,5366916 0,9828677
No Info-current  -10,5256593    -11,2330612     -9,8182573 0,0000000

not current-current  3,6259148      2,6411471      4,6106825 0,0000000
former-ever         7,9248349      6,7775794      9,0720904 0,0000000
never-ever        -5,2453720     -6,2586635     -4,2320806 0,0000000
No Info-ever      -15,5988265    -16,6110730    -14,5865801 0,0000000
not current-ever   -1,4472525     -2,6695425     -0,2249624 0,0096467
never-former      -13,1702069    -13,8771225    -12,4632913 0,0000000
No Info-former    -23,5236614    -24,2290783    -22,8182445 0,0000000
not current-former  -9,3720874     -10,3554301     -8,3887446 0,0000000
No Info-never    -10,3534545    -10,8097191     -9,8971898 0,0000000
not current-never   3,7981196      2,9749990      4,6212401 0,0000000
not current-No Info 14,1515740     13,3297402     14,9734078 0,0000000
```

Berikut adalah ringkasan dari hasil uji anova :

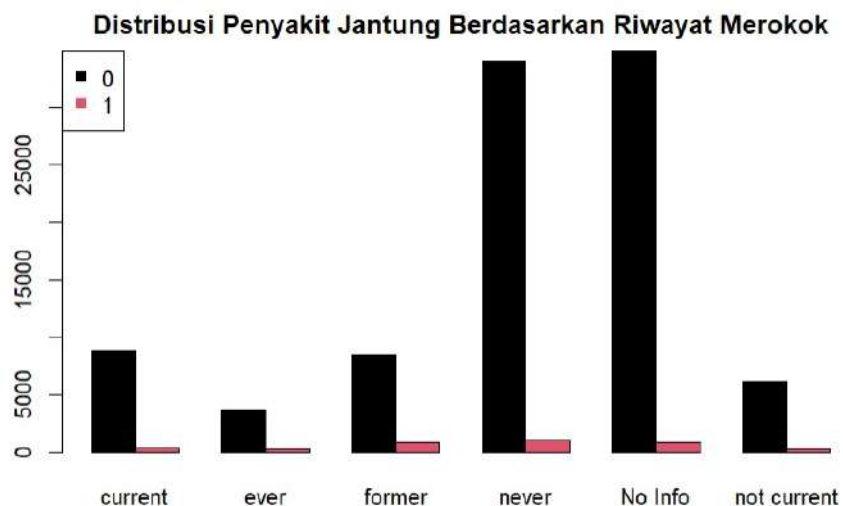
- Variabel smoking_history memiliki 5 tingkat atau kategori.
- Jumlah kuadrat untuk smoking_history adalah 5.262.593 dan residual sisa adalah 45.437.708.
- Derajat kebebasan untuk smoking_history adalah 5 dan untuk residualnya adalah 99.994.
- Standard error residual adalah 21,31676.
- *F-value* adalah 2.316 dan *p-value* < 2e-16.
- Terdapat hubungan antara umur dengan riwayat merokok (*p-value* < *alpha*) dengan *p-value* < 2e-16



Berdasarkan plot Tukey HSD di atas, selisih dari masing-masing smoking_history tidak mengandung angka 0, sehingga dapat dikatakan semuanya tidak memiliki mean atau rata-rata yang sama. Kesimpulannya adalah terdapat hubungan antara umur dengan riwayat merokok.

3) Hipotesa 3

```
# Uji Hipotesa 3
judul <- "Distribusi Penyakit Jantung Berdasarkan Riwayat Merokok"
(tabel <- table(diabetes1$heart_disease, diabetes1$smoking_history))
barplot(tabel, main = judul, col = 1:nrow(tabel), beside = TRUE,
legend("topleft", legend = rownames(tabel), pch = 15, col = 1:nrow(tabel))
```



Berdasarkan barplot di atas, pada kelima kategori riwayat merokok data yang dominan adalah data populasi yang tidak memiliki riwayat penyakit jantung. Namun, tetap masih ada sebagian kecil yang memiliki riwayat penyakit jantung.


```
chisq.test(diabetes1$smoking_history, diabetes1$heart_disease)
```

	current	ever	former	never	No Info	not current
0	8877	3691	8444	33998	34893	6155
1	409	313	908	1097	923	292

Pearson's Chi-squared test

data: diabetes1\$smoking_history and diabetes1\$heart_disease
X-squared = 1229,3, df = 5, p-value < 2,2e-16

Decision Rules :

- H_0 ditolak saat $p\text{-value} < 0,05$ dan terdapat hubungan antara riwayat merokok dengan penyakit jantung.
- H_0 diterima atau tidak ditolak saat $p\text{-value} > 0,05$ dan tidak terdapat hubungan antara riwayat merokok dengan penyakit jantung.

Hasil :

- $p\text{-value} < \alpha$ ($2,2e-16 < 0,05$)
- H_0 ditolak pada $\alpha = 0,05$.
- Terdapat hubungan antara riwayat merokok dengan penyakit jantung.

4) Hipotesa 4

```
# Uji Hipotesa 4
result2 <- aov(diabetes1$blood_glucose_level ~ diabetes1$smoking_history)
result2
summary(result2)
```

```
Call:
aov(formula = diabetes1$blood_glucose_level ~ diabetes1$smoking_history)

Terms:
              diabetes1$smoking_history Residuals
Sum of Squares              590029 165123548
Deg. of Freedom                5      99994

Residual standard error: 40,63662
Estimated effects may be unbalanced

              Df Sum Sq Mean Sq F value Pr(>F)
diabetes1$smoking_history  5    590029   118006   71,46 <2e-16 ***
Residuals              99994 165123548    1651

---
Signif. codes:  0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1 ' ' 1
Tukey multiple comparisons of means
 95% family-wise confidence level
factor levels have been ordered

Fit: aov(formula = diabetes1$blood_glucose_level ~ diabetes1$smoking_history)

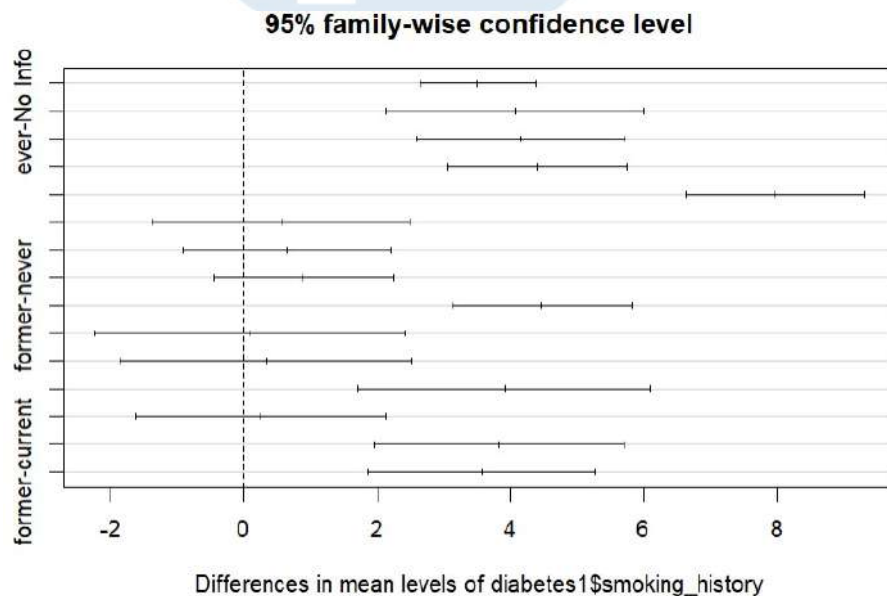
$`diabetes1$smoking_history`
              diff              lwr              upr              p adj
never-No Info    3,50440527    2,6346178  4,374193  0,0000000
ever-No Info     4,06294443    2,1332766  5,992612  0,0000000
not current-No Info 4,14942309    2,5827431  5,716103  0,0000000
current-No Info  4,39747034    3,0489344  5,746006  0,0000000
```

former-No Info	7,97046505	6,6257132	9,315217	0,0000000
ever-never	0,55853915	-1,3731208	2,490199	0,9631992
not current-never	0,64501782	-0,9241152	2,214151	0,8505790
current-never	0,89306506	-0,4583199	2,244450	0,4124754
former-never	4,46605978	3,1184509	5,813669	0,0000000
not current-ever	0,08647867	-2,2435999	2,416557	0,9999982
current-ever	0,33452591	-1,8548414	2,523893	0,9980262
former-ever	3,90752062	1,7204821	6,094559	0,0000053
current-not current	0,24804724	-1,6292373	2,125332	0,9990201
former-not current	3,82104196	1,9464739	5,695610	0,0000001
former-current	3,57299471	1,8765076	5,269482	0,0000000

Berikut adalah ringkasan dari hasil uji anova :

- Variabel `smoking_history` memiliki 5 tingkat atau kategori.
- Jumlah kuadrat untuk `smoking_history` adalah 590.029 dan residual sisa adalah 165.123.548.
- Derajat kebebasan untuk `smoking_history` adalah 5 dan untuk residualnya adalah 99.994.
- Standard error residual adalah 40,63662.
- *F-value* adalah 71,46 dan *p-value* < 2e-16.
- Terdapat hubungan antara riwayat merokok dengan gula darah (*p-value* < *alpha*) dengan *p-value* < 2e-16

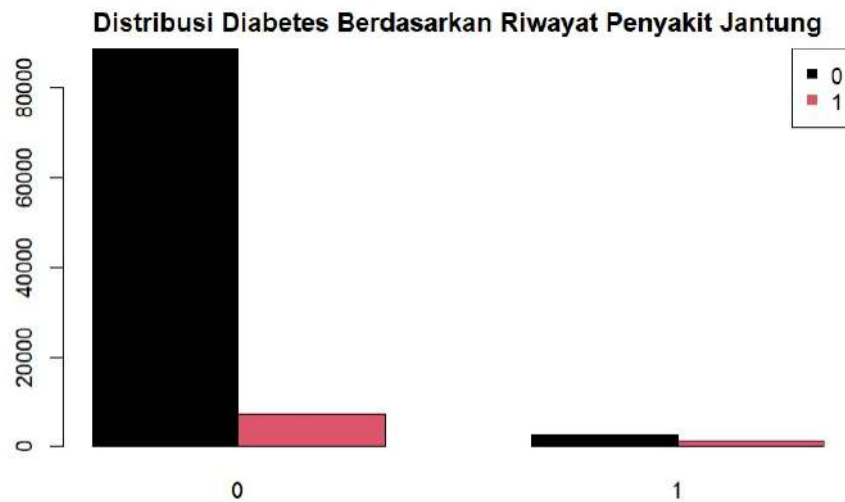
```
tk <- TukeyHSD(result2, ordered = TRUE)
tk
plot(tk)
```



Berdasarkan plot Tukey Kramer di atas, selisih dari masing-masing `smoking_history` tidak mengandung angka 0, sehingga dapat dikatakan secara keseluruhan kelompok tidak memiliki *mean* atau rata-rata yang sama. Kesimpulannya adalah terdapat hubungan antara riwayat merokok dengan gula darah.

5) Hipotesa 5

```
# Uji Hipotesa 5
judul <- "Distribusi Diabetes Berdasarkan Riwayat Penyakit Jantung"
(tabel <- table(diabetes1$diabetes, diabetes1$heart_disease))
barplot(tabel, main = judul, col = 1:nrow(tabel), beside = TRUE)
legend("topright", legend = rownames(tabel), pch = 15, col = 1:nrow(tabel))
```



Berdasarkan barplot di atas, baik populasi yang tidak memiliki maupun yang memiliki riwayat penyakit jantung didominasi dengan data hasil diabetes yang negatif atau tidak diabetes.

```
wilcox.test(diabetes1$heart_disease, diabetes1$diabetes)
```

	0	1
0	88825	2675
1	7233	1267

wilcoxon rank sum test with continuity correction

data: diabetes1\$heart_disease and diabetes1\$diabetes

W = 4772100000, p-value < 2,2e-16

alternative hypothesis: true location shift is not equal to 0

Decision Rules :

- H_0 ditolak saat $p\text{-value} < 0,05$ dan terdapat hubungan antara penyakit jantung dengan diabetes.
- H_0 diterima atau tidak ditolak saat $p\text{-value} > 0,05$ dan tidak terdapat hubungan antara riwayat penyakit jantung dengan diabetes.

Hasil :

- $p\text{-value} < \alpha$ ($2,2e-16 < 0,05$)
- H_0 ditolak pada $\alpha = 0,05$.
- Terdapat hubungan antara penyakit jantung dengan diabetes.

6) Hipotesa 6

```
kruskal.test(diabetes1$blood_glucose_level~diabetes1$diabetes)~  
# data: diabetes1  
Kruskal-Wallis rank sum test  
data: diabetes1$blood_glucose_level by diabetes1$diabetes  
Kruskal-Wallis chi-squared = 7519,4, df = 1, p-value < 2,2e-16
```

Decision Rules :

- H_0 ditolak saat $p\text{-value} < 0,05$ dan terdapat hubungan antara gula darah dengan diabetes.
- H_0 diterima atau tidak ditolak saat $p\text{-value} > 0,05$ dan tidak terdapat hubungan antara gula darah dengan diabetes.

Hasil :

- $p\text{-value} < \alpha$ ($< 2,2e-16 < 0,05$)
- H_0 ditolak pada $\alpha = 0,05$.
- Terdapat hubungan antara gula darah dengan diabetes.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Terdapat berbagai hal yang dapat disimpulkan melalui penelitian di atas, antara lain :

1. Terdapat hubungan yang rendah antara gula darah dan diabetes. Hal tersebut dibuktikan melalui uji Spearman dengan koefisien korelasi yang rendah sebesar 0,27.
2. Uji multilinear regression antara gula darah dan diabetes menghasilkan hal berikut.
 - Syarat 1 : Terdapat pelanggaran linieritas (TIDAK TERPENUHI)
 - Syarat 2 : Data tidak berdistribusi dengan normal dalam uji Anderson-Darling dengan $p\text{-value}$ ($<2,2e-16$) $< 0,05$. (TIDAK TERPENUHI)
 - Syarat 3 : Varians residual homoskedastis dalam uji Breusch-Pagan dengan $p\text{-value}$ ($<2,2e-16$) $< 0,05$. (TIDAK TERPENUHI)
 - Syarat 4 : Ada pengaruh *influential point* atau *outliers* yang serius dalam hasil analisis (TIDAK TERPENUHI)
 - Syarat 5 : Residual tidak berotokorelasi karena $p\text{-value}$ (0,7137) $> 0,05$ (TERPENUHI)
3. Terdapat hubungan antara gender dengan riwayat merokok. Hal tersebut dibuktikan melalui uji Kruskal-Wallis dengan $p\text{-value} < \alpha$ ($2,2e=16 < 0,05$) dan H_0 ditolak.
4. Terdapat hubungan antara umur dengan riwayat merokok. Hal tersebut dibuktikan melalui uji Anova dan Tukey HSD dengan $p\text{-value} < \alpha$ ($2,2e=16 < 0,05$) dan H_0 ditolak.
5. Terdapat hubungan antara riwayat merokok dengan penyakit jantung. Hal tersebut dibuktikan melalui uji Chi-Square dengan $p\text{-value} < \alpha$ ($2,2e=16 < 0,05$) dan H_0 ditolak.

6. Terdapat hubungan antara riwayat merokok dengan gula darah. Hal tersebut dibuktikan melalui uji Anova dan Tukey Kramer dengan $p\text{-value} < \alpha$ ($2,2e=16 < 0,05$) dan H_0 ditolak.
7. Terdapat hubungan antara penyakit jantung dengan diabetes. Hal tersebut dibuktikan melalui uji Mann Whitney Wilcoxon dengan $p\text{-value} < \alpha$ ($2,2e=16 < 0,05$) dan H_0 ditolak.
8. Terdapat hubungan antara gula darah dengan diabetes. Hal tersebut dibuktikan melalui uji Kruskal-Wallis dengan $p\text{-value} < \alpha$ ($2,2e=16 < 0,05$) dan H_0 ditolak.

5.2 Saran

Berdasarkan hasil analisis data, saran dalam pencegahan dan pengelolaan penyakit diabetes dapat diberikan kepada pihak terkait. Pertama, penting untuk mengendalikan faktor risiko seperti merokok, penyakit jantung, dan kadar gula darah dengan program pencegahan yang mengedukasi masyarakat dan melakukan deteksi dini penyakit jantung serta pemantauan kadar gula darah secara rutin. Kedua, diperlukan upaya meningkatkan kesadaran masyarakat tentang pentingnya pola hidup sehat melalui pola makan yang seimbang, olahraga teratur, dan menghindari merokok. Terakhir, penelitian lanjutan dengan sampel yang lebih besar dan melibatkan faktor risiko lain perlu dilakukan untuk mendapatkan wawasan yang lebih mendalam.



DAFTAR PUSTAKA

- Anderson, M., Thompson, R., Roberts, K. (2018). "Analysis of the Relationship between Medical History and Diabetes: A Statistical Approach."
- Atkinson, M. A., et al. (2014). "Type 1 diabetes." *The Lancet*, 383(9911), 69-82. doi: 10.1016/S0140-6736(13)60591-7
- Chen, L., Lee, S., Wang, Y. (2019). "Exploring the Impact of Medical History on Diabetes: A Probabilistic Study."
- Dapubeang, A., Talan, M. R., & Adan, L. N. (2022). "Sikap Bahasa Generasi Muda Etnis Sulawesi di Desa Balauring terhadap Bahasa Kedang." *Jurnal Onoma: Pendidikan, Bahasa, dan Sastra*, 8(2), 901-916.
- Donath, M. Y., Shoelson, S. E. (2011). "Type 2 diabetes as an inflammatory disease." *Nature Reviews Immunology*, 11(2), 98-107. doi: 10.1038/nri2925
- Eizirik, D. L., Cardozo, A. K., Cnop, M. (2008). "The role for endoplasmic reticulum stress in diabetes mellitus." *Endocrine Reviews*, 29(1), 42-61. doi: 10.1210/er.2007-0015
- Fanani, A. (2020). "Hubungan Faktor Risiko dengan Kejadian Diabetes Mellitus." *Jurnal Keperawatan*, 12(3), 371-378.
- Igo Cahya Negara, Agung Prabowo (2018). "PENGUNAAN UJI CHI-SQUARE UNTUK MENGETAHUI PENGARUH TINGKAT PENDIDIKAN DAN UMUR TERHADAP PENGETAHUAN PENASUN MENGENAI HIV-AIDS DI PROVINSI DKI JAKARTA." *Prosiding Seminar Nasional Matematika dan Terapannya 2018* p-ISSN: 2550-0384; e-ISSN: 2550-0392
- Jamco, J., & Balami, A. M. (2022). "Analisis Kruskal-Wallis Untuk Mengetahui Konsentrasi Belajar Mahasiswa Berdasarkan Bidang Minat Program Studi Statistika Fmipa Unpatti." *PARAMETER: Jurnal Matematika, Statistika dan Terapannya*, 1(1), 39-44.
- Marlina, M., & Masnur, M. (2021). "Aplikasi E-Learning Siswa SMK Berbasis Web." *Jurnal Sintaks Logika*, 1(1), 8-17.

- Maulana, M. R., & Kurniawan, I. (2019). "Pengaruh Banyaknya Chunks dan Priority Render Animation 3D Terhadap Waktu Render Menggunakan Anova One Way Dengan Menerapkan Parallel Computing Render Engine." IC-Tech,
- Quraissy, A., & Hasni, N. (2021). "Analisis Kruskal-Wallis Terhadap Kemampuan Numerik Siswa." VARIANSI J. Stat. Its Appl. Teach. Res, 3(3), 156-161.
- Samuel, V. T., Shulman, G. I. (2016). "Mechanisms for insulin resistance: Common threads and missing links." Cell, 164(1-2), 10-20. doi: 10.1016/j.cell.2015.12.05
- Smith, J., Johnson, A., Davis, L. (2020). "Association of Medical History with Diabetes: A Probabilistic and Statistical Analysis."
- Wahyuni Saputri, S., Nugroho, S., & Faisal, F. (n.d.). "KAJIAN DUA SAMPEL INDEPENDEN DENGAN UJI MEDIAN, MANNWHITNEY-WILCOXON, DAN KOLMOGOROV-SMIRNOV." e-jurnal statistika.
- Warsi Maryati, Aris Ocktavian Wannay, Devi Permani Suci (2018). "Relationship to Completeness of Medical Information and Accuracy of the Diagnosis Code of Diabetes Mellitus." Jurnal Rekam Medis dan Informasi Kesehatan, Volume 1, Nomor 2.
- Yasril, A. I., Fatma, F., & Febrianti, D. (2021). "PENERAPAN UJI KORELASI SPEARMAN UNTUK MENGAJI FAKTOR YANG BERHUBUNGAN DENGAN KEJADIAN DIABETES MELITUS DI PUSKESMAS SICINCIN KABUPATEN PADANG PARIAMAN." Jurnal Human Care, Volume 6, Nomor 3.

LAMPIRAN

Link Google Drive yang berisi dataset, video presentasi, lembar kerja R Studio, dan PPT.

https://drive.google.com/drive/folders/1Prq_ozNwszcel2p7tvi_EiyXhAdUP0QP?usp=sharing

Diabetes Prediction Dataset

	A	B	C	D	E	F	G	H	I
1	gender	age	hypertension	heart_disease	smoking_	bmi	HbA1c_level	blood_glucose_level	diabetes
2	Female	80	0	1	never	25.19	6.6	140	0
3	Female	54	0	0	No Info	27.32	6.6	80	0
4	Male	28	0	0	never	27.32	5.7	158	0
5	Female	36	0	0	current	23.45	5	155	0
6	Male	76	1	1	current	20.14	4.8	155	0
7	Female	20	0	0	never	27.32	6.6	85	0
8	Female	44	0	0	never	19.31	6.5	200	1
9	Female	79	0	0	No Info	23.86	5.7	85	0
10	Male	42	0	0	never	33.64	4.8	145	0
11	Female	32	0	0	never	27.32	5	100	0
12	Female	53	0	0	never	27.32	6.1	85	0
13	Female	54	0	0	former	54.7	6	100	0
14	Female	78	0	0	former	36.05	5	130	0
15	Female	67	0	0	never	25.69	5.8	200	0
16	Female	76	0	0	No Info	27.32	5	160	0
17	Male	78	0	0	No Info	27.32	6.6	126	0
18	Male	15	0	0	never	30.36	6.1	200	0
19	Female	42	0	0	never	24.48	5.7	158	0
20	Female	42	0	0	No Info	27.32	5.7	80	0
21	Male	37	0	0	ever	25.72	3.5	159	0
22	Male	40	0	0	current	36.38	6	90	0

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA