

Diktat
Matakuliah: Biostatistik



Disusun oleh: Lina Septiana
Fakultas Teknik Ilmu Komputer/Teknik Elektro

Deskripsi Mata Kuliah

Mata kuliah Biostatistik mempelajari konsep dan aplikasi metode statistik dalam ilmu kesehatan. Mahasiswa akan mempelajari dasar-dasar statistik, analisis data, pengujian hipotesis, regresi, analisis multivariat, dan perancangan penelitian dalam ilmu kesehatan.

Tujuan Mata Kuliah

Setelah mengikuti mata kuliah Biostatistik, mahasiswa diharapkan mampu:

Menjelaskan konsep dasar statistik dalam ilmu kesehatan.

Menggunakan software statistik dalam melakukan analisis data.

Menjelaskan konsep pengujian hipotesis dalam ilmu kesehatan.

Menjelaskan konsep regresi dalam ilmu kesehatan.

Menjelaskan konsep analisis multivariat dalam ilmu kesehatan.

Menerapkan metode statistik dalam merancang penelitian ilmu kesehatan.

Daftar Topik

Pengantar statistik

Pengukuran dan skala data

Ukuran pemusatan dan penyebaran data

Distribusi peluang

Probabilitas dan distribusi sampel

Pengujian hipotesis

Regresi linier sederhana

Regresi linier berganda

Analisis varians

Analisis multivariat

Perancangan penelitian

Analisis data menggunakan software statistik

Referensi Utama

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis* (5th ed.). John Wiley & Sons, Inc.

Krosnick, J. A., & Presser, S. (2010). Question and Questionnaire Design. In *Handbook of Survey Research* (2nd ed., pp. 263-313). Elsevier.

Triola, M. F. (2018). Elementary Statistics (13th ed.). Pearson Education Limited.

Penilaian

Kehadiran: 10%

Tugas: 20%

Ujian Tengah Semester: 30%

Proyek Akhir Mata Kuliah: 40%

Keterangan

Kehadiran minimal 75%.

Tugas terdiri dari 3 kali pengumpulan dan setiap pengumpulan mempunyai bobot nilai 6,67%.

Ujian Tengah Semester berupa ujian tulis.

Proyek Akhir Mata Kuliah berupa penyelesaian studi kasus atau perancangan penelitian di bidang rekayasa biomedis dengan menggunakan metode statistik yang telah dipelajari dalam mata kuliah ini. Proyek akhir ini harus diserahkan pada waktu yang telah ditentukan. Kriteria penilaian meliputi kesesuaian masalah yang dipilih, analisis data, hasil penelitian, serta kemampuan menyajikan hasil secara sistematis dan logis.

Catatan khusus: Pada matakuliah Biostatistik untuk program studi Teknik Elektro, penekanan akan diberikan pada penerapan metode statistik dalam bidang rekayasa biomedis, termasuk dalam merancang dan menganalisis hasil uji coba alat dan sistem medis.

Bab 1

"Pengantar Statistik dalam Rekayasa Biomedis"

1.1 Definisi Statistik

Statistik adalah cabang ilmu matematika yang berfokus pada pengumpulan, analisis, interpretasi, presentasi, dan pengambilan kesimpulan dari data. Statistik diterapkan pada berbagai bidang, termasuk dalam ilmu kesehatan dan rekayasa biomedis. Tujuan dari statistik adalah untuk membantu dalam pengambilan keputusan dengan dasar data dan analisis yang tepat.

Referensi:

Ott, R. L., & Longnecker, M. (2015). An introduction to statistical methods and data analysis (7th ed.). Cengage Learning.

1.2 Populasi dan Sampel dalam Rekayasa Biomedis

Populasi adalah kumpulan dari seluruh objek atau individu yang menjadi subjek penelitian. Dalam rekayasa biomedis, populasi dapat berupa manusia, hewan, atau benda lain yang berkaitan dengan aplikasi biomedis, seperti sistem medis dan alat-alat kesehatan. Sampel adalah bagian dari populasi yang diambil untuk dijadikan representasi populasi tersebut. Sampel harus dipilih secara acak dan harus mewakili populasi yang sedang diteliti.

Contoh perhitungan:

Misalkan kita ingin mengetahui berat rata-rata pasien yang dirawat di rumah sakit X. Populasi adalah seluruh pasien yang dirawat di rumah sakit X. Kita dapat mengambil sampel acak sebanyak 50 pasien dari populasi dan menghitung berat badan mereka. Berdasarkan hasil pengamatan sampel, kita dapat menghitung rata-rata berat badan pasien dalam sampel tersebut. Rata-rata berat badan sampel dapat dijadikan estimasi untuk rata-rata berat badan populasi.

Program R:

Untuk mengambil sampel acak dari populasi, kita dapat menggunakan fungsi `sample()` pada bahasa pemrograman R. Berikut adalah contoh perintah untuk mengambil sampel acak 50 pasien dari dataset populasi berisi 1000 pasien:

```
# Membaca dataset populasi
populasi <- read.csv("data_populasi.csv")

# Mengambil sampel acak 50 pasien
sampel <- populasi[sample(nrow(populasi), 50, replace=FALSE), ]
```

1.3 Konsep Variabel dan Pengukuran dalam Rekayasa Biomedis

Variabel adalah suatu karakteristik yang dapat diamati atau diukur pada objek atau individu dalam populasi atau sampel. Variabel dapat dibedakan menjadi variabel kategorikal dan variabel kontinu. Variabel kategorikal adalah variabel yang mempunyai nilai diskrit dan tergantung pada kategori atau golongan tertentu. Contoh variabel kategorikal dalam rekayasa biomedis adalah jenis kelamin, jenis

penyakit, dan jenis alat medis. Variabel kontinu adalah variabel yang dapat mempunyai nilai dalam rentang tertentu dan dapat diukur secara kontinu. Contoh variabel kontinu dalam rekayasa biomedis adalah berat badan, tinggi badan, dan tekanan darah.

Pengukuran adalah proses mengumpulkan informasi atau data dari suatu variabel. Pengukuran dapat dilakukan dengan menggunakan berbagai jenis instrumen dan teknik pengukuran, tergantung pada jenis variabel yang diukur dan tujuan penelitian. Beberapa jenis pengukuran dalam rekayasa biomedis antara lain pengukuran jarak, waktu, kekuatan, frekuensi, dan intensitas.

Referensi:

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

Contoh perhitungan:

Misalkan kita ingin mengetahui apakah terdapat perbedaan berat badan antara pasien pria dan wanita di rumah sakit X. Variabel yang diukur adalah berat badan, yang merupakan variabel kontinu. Untuk mengukur berat badan, kita dapat menggunakan alat pengukur berat badan seperti timbangan. Berdasarkan hasil pengukuran, kita dapat membandingkan rata-rata berat badan antara pasien pria dan wanita.

Program R:

Untuk menghitung rata-rata berat badan dan membandingkannya antara pasien pria dan wanita, kita dapat menggunakan fungsi `t.test()` pada bahasa pemrograman R. Berikut adalah contoh perintah untuk menghitung rata-rata berat badan dan melakukan uji hipotesis:

```
# Membaca dataset populasi
populasi <- read.csv("data_populasi.csv")

# Mengambil sampel acak 50 pasien pria dan wanita
sampel_pria <- populasi[populasi$jenis_kelamin == "Pria", ]
sampel_wanita <- populasi[populasi$jenis_kelamin == "Wanita", ]
sampel_pria <- sampel_pria[sample(nrow(sampel_pria), 25, replace=FALSE), ]
sampel_wanita <- sampel_wanita[sample(nrow(sampel_wanita), 25, replace=FALSE), ]

# Menghitung rata-rata berat badan pada masing-masing sampel
rata2_pria <- mean(sampel_pria$berat_badan)
rata2_wanita <- mean(sampel_wanita$berat_badan)

# Melakukan uji hipotesis dengan t-test
ttest <- t.test(sampel_pria$berat_badan, sampel_wanita$berat_badan)
```

Dalam contoh program di atas, kita menggunakan dataset populasi yang berisi data berat badan dan jenis kelamin pasien di rumah sakit X. Kita mengambil sampel acak sebanyak 25 pasien pria dan 25 pasien wanita, dan menghitung rata-rata berat badan pada masing-masing sampel. Selanjutnya, kita melakukan uji hipotesis dengan menggunakan t-test untuk melihat apakah terdapat perbedaan berat badan yang signifikan antara pasien pria dan wanita.

Bab 2

Pengukuran dan Skala Data dalam Konteks Rekayasa Biomedis

2.1 Skala Pengukuran dan Tipe Data dalam Rekayasa Biomedis

Pengukuran dalam rekayasa biomedis dapat dikategorikan menjadi empat skala, yaitu nominal, ordinal, interval, dan rasio. Skala nominal adalah skala pengukuran yang hanya mengelompokkan data ke dalam kategori-kategori yang bersifat tidak berurutan, seperti jenis kelamin, jenis penyakit, dan jenis alat medis. Skala ordinal adalah skala pengukuran yang mengelompokkan data ke dalam kategori-kategori yang memiliki urutan atau tingkatan, seperti tingkat keparahan penyakit atau kelas usia. Skala interval adalah skala pengukuran yang memiliki jarak atau selisih antar nilai yang tetap, seperti suhu atau tekanan darah. Skala rasio adalah skala pengukuran yang memiliki jarak atau selisih antar nilai yang tetap, dan memiliki nilai nol yang nyata, seperti berat badan dan tinggi badan.

Tipe data dalam rekayasa biomedis dapat dibedakan menjadi dua, yaitu data kategorikal dan data kontinu. Data kategorikal adalah data yang bersifat diskrit dan tergantung pada kategori atau golongan tertentu, seperti jenis kelamin, jenis penyakit, dan jenis alat medis. Data kontinu adalah data yang dapat mempunyai nilai dalam rentang tertentu dan dapat diukur secara kontinu, seperti berat badan, tinggi badan, dan tekanan darah.

Referensi:

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, 10(2), 486-489.

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

Contoh perhitungan:

Misalkan kita ingin mengetahui distribusi jenis kelamin pasien di rumah sakit X. Variabel yang diukur adalah jenis kelamin, yang merupakan data kategorikal dengan skala nominal. Untuk menghitung frekuensi dan proporsi pasien pria dan wanita, kita dapat menggunakan tabel frekuensi dan diagram pie.

Program R:

Untuk membuat tabel frekuensi dan diagram pie pada dataset populasi, kita dapat menggunakan fungsi `table()` dan `pie()` pada bahasa pemrograman R. Berikut adalah contoh perintah untuk menghitung frekuensi dan membuat diagram pie pada variabel `jenis_kelamin`:

```
# Membaca dataset populasi
populasi <- read.csv("data_populasi.csv")

# Menghitung frekuensi jenis kelamin
frek_jk <- table(populasi$jenis_kelamin)

# Menampilkan tabel frekuensi
frek_jk

# Membuat diagram pie jenis kelamin
```

```
pie(frek_jk, labels = c("Pria", "Wanita"), main = "Diagram Pie Jenis Kelamin")
```

2.2 Statistik Deskriptif untuk Data Kategorikal dan Kontinu

Statistik deskriptif adalah metode statistik yang digunakan untuk merang2.2 Statistik Deskriptif untuk Data Kategorikal dan Kontinu

Statistik deskriptif adalah metode statistik yang digunakan untuk merangkumkan data dalam bentuk ringkasan statistik yang mudah dipahami. Statistik deskriptif dapat diterapkan pada data kategorikal dan kontinu. Beberapa statistik deskriptif yang umum digunakan antara lain frekuensi, proporsi, rata-rata, median, modus, dan simpangan baku.

Untuk data kategorikal, statistik deskriptif yang dapat digunakan adalah frekuensi dan proporsi. Frekuensi adalah jumlah observasi pada setiap kategori data, sedangkan proporsi adalah frekuensi relatif dari setiap kategori data dalam populasi atau sampel.

Untuk data kontinu, statistik deskriptif yang dapat digunakan adalah rata-rata, median, modus, dan simpangan baku. Rata-rata adalah nilai tengah dari data yang dihitung dengan cara menjumlahkan seluruh data dan membaginya dengan jumlah data. Median adalah nilai tengah dari data yang diurutkan dari yang terkecil ke yang terbesar. Modus adalah nilai yang paling sering muncul dalam data. Simpangan baku adalah ukuran variabilitas yang menggambarkan seberapa jauh data dari nilai rata-rata.

Referensi:

Ott, R. L., & Longnecker, M. (2015). An introduction to statistical methods and data analysis (7th ed.). Cengage Learning.

Sullivan, L. M. (2011). Essentials of biostatistics in public health (2nd ed.). Jones & Bartlett Learning.

Contoh perhitungan:

Misalkan kita ingin mengetahui rata-rata berat badan pasien di rumah sakit X. Variabel yang diukur adalah berat badan, yang merupakan data kontinu dengan skala interval. Untuk menghitung rata-rata, median, modus, dan simpangan baku berat badan, kita dapat menggunakan fungsi `mean()`, `median()`, `mode()`, dan `sd()` pada bahasa pemrograman R.

Program R:

Berikut adalah contoh perintah untuk menghitung rata-rata, median, modus, dan simpangan baku pada variabel `berat_badan`:

```
# Membaca dataset populasi
populasi <- read.csv("data_populasi.csv")

# Menghitung rata-rata berat badan
rata2_bb <- mean(populasi$berat_badan)

# Menghitung median berat badan
med_bb <- median(populasi$berat_badan)

# Menghitung modus berat badan
```

```

mod_bb <- mode(populasi$berat_badan)
# Menghitung simpangan baku berat badan
sd_bb <- sd(populasi$berat_badan)
# Menampilkan hasil perhitungan
cat("Rata-rata Berat Badan:", rata2_bb, "\n")
cat("Median Berat Badan:", med_bb, "\n")
cat("Modus Berat Badan:", mod_bb, "\n")
cat("Simpangan Baku Berat Badan:", sd_bb, "\n")

```

Dalam contoh program di atas, kita menghitung rata-rata, median, modus, dan simpangan baku berat badan pada dataset populasi. Hasil perhitungan ditampilkan dengan menggunakan fungsi `cat()`.

2.3 Grafik dalam Analisis Data

Grafik merupakan salah satu alat yang digunakan dalam analisis data untuk memvisualisasikan dan memperjelas hasil analisis. Grafik dapat digunakan untuk menganalisis data kategorikal dan kontinu. Beberapa jenis grafik yang umum digunakan antara lain bar chart, pie chart, histogram, scatter plot, box plot, dan line chart.

Bar chart dan pie chart adalah jenis grafik yang digunakan untuk menganalisis data kategorikal. Bar chart digunakan untuk menunjukkan frekuensi atau proporsi data dalam bentuk bar atau batang, sedangkan pie chart digunakan untuk menunjukkan proporsi data dalam bentuk diagram lingkaran.

Histogram adalah jenis grafik yang digunakan untuk menganalisis data kontinu. Histogram menunjukkan distribusi data dalam bentuk bentuk kurva. Scatter plot adalah jenis grafik yang digunakan untuk menganalisis hubungan antara dua variabel kontinu. Scatter plot menunjukkan distribusi data dalam bentuk titik-titik yang terdispersi.

Box plot adalah jenis grafik yang digunakan untuk menganalisis data kontinu. Box plot menunjukkan distribusi data dalam bentuk kotak-kotak yang menunjukkan nilai kuartil. Line chart adalah jenis grafik yang digunakan untuk menganalisis data kontinu. Line chart menunjukkan perubahan nilai variabel seiring waktu.

Referensi:

Ott, R. L., & Longnecker, M. (2015). *An introduction to statistical methods and data analysis* (7th ed.). Cengage Learning.

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

Contoh perhitungan:

Misalkan kita ingin mengetahui distribusi berat badan pasien di rumah sakit X. Variabel yang diukur adalah berat badan, yang merupakan data kontinu dengan skala interval. Untuk memvisualisasikan distribusi berat badan, kita dapat menggunakan histogram atau box plot.

Program R:

Berikut adalah contoh perintah untuk membuat histogram dan box plot pada variabel berat_badan:

```
# Membaca dataset populasi
populasi <- read.csv("data_populasi.csv")

# Membuat histogram berat badan
hist(populasi$berat_badan, main="Histogram Berat Badan", xlab="Berat Badan", ylab="Frekuensi")

# Membuat box plot berat badan
boxplot(populasi$berat_badan, main="Box Plot Berat Badan", ylab="Berat Badan")
```

Dalam contoh program di atas, kita membuat histogram dan box plot pada variabel berat badan pada dataset populasi. Histogram menunjukkan distribusi frekuensi berat badan dalam bentuk kurva, sedangkan box plot menunjukkan distribusi berat badan dalam bentuk kotak-kotak yang menunjukkan nilai kuartil.

2.4 Normalitas dan Transformasi Data

Normalitas adalah sifat distribusi data yang mengikuti pola distribusi normal atau dikenal juga sebagai kurva Gauss. Distribusi normal memiliki bentuk kurva simetris yang memuncak pada nilai rata-rata dan memperlihatkan penyebaran data yang relatif simetris di kedua sisi nilai rata-rata. Normalitas sangat penting dalam analisis statistik karena banyak metode statistik yang didasarkan pada asumsi bahwa data mengikuti distribusi normal.

Transformasi data dapat dilakukan untuk mengubah distribusi data menjadi lebih mendekati distribusi normal. Beberapa transformasi data yang umum dilakukan antara lain transformasi logaritmik, transformasi akar kuadrat, dan transformasi kekuatan.

Referensi:

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, 10(2), 486-489.

Ott, R. L., & Longnecker, M. (2015). *An introduction to statistical methods and data analysis* (7th ed.). Cengage Learning.

Contoh perhitungan:

Misalkan kita ingin mengetahui apakah distribusi berat badan pasien di rumah sakit X mengikuti distribusi normal. Untuk mengetahui normalitas distribusi berat badan, kita dapat menggunakan uji normalitas seperti uji Shapiro-Wilk atau uji Kolmogorov-Smirnov. Jika distribusi berat badan tidak normal, kita dapat melakukan transformasi data untuk mengubah distribusi menjadi lebih mendekati distribusi normal.

Program R:

Berikut adalah contoh perintah untuk melakukan uji Shapiro-Wilk pada variabel berat_badan dan melakukan transformasi logaritmik pada data:

```
# Membaca dataset populasi
```

```
populasi <- read.csv("data_populasi.csv")  
# Melakukan uji normalitas pada berat badan  
shapiro.test(populasi$berat_badan)  
# Melakukan transformasi logaritmik pada berat badan  
berat_log <- log(populasi$berat_badan)  
# Melakukan uji normalitas pada berat badan yang sudah ditransformasi  
shapiro.test(berat_log)
```

Dalam contoh program di atas, kita melakukan uji normalitas menggunakan uji Shapiro-Wilk pada variabel berat_badan pada dataset populasi. Hasil uji menunjukkan bahwa distribusi berat badan tidak normal. Selanjutnya, kita melakukan transformasi logaritmik pada variabel berat badan, dan melakukan uji normalitas lagi dengan uji Shapiro-Wilk. Hasil uji menunjukkan bahwa distribusi berat badan yang sudah ditransformasi lebih mendekati distribusi normal.

Bab 3

Inferensial Statistik: Pengujian Hipotesis dan Analisis Regresi

3.1 Pengujian Hipotesis

Pengujian hipotesis adalah prosedur statistik yang digunakan untuk menguji kebenaran suatu hipotesis yang diajukan mengenai populasi. Terdapat dua jenis hipotesis dalam pengujian hipotesis, yaitu hipotesis nol (H_0) dan hipotesis alternatif (H_a). Hipotesis nol adalah hipotesis yang diajukan dan diuji kebenarannya, sedangkan hipotesis alternatif adalah hipotesis yang diajukan sebagai alternatif dari hipotesis nol.

Terdapat beberapa tahap dalam pengujian hipotesis, yaitu:

Menentukan hipotesis nol (H_0) dan hipotesis alternatif (H_a)

Menentukan tingkat signifikansi (α)

Mengumpulkan data

Menghitung uji statistik

Menentukan nilai p

Menarik kesimpulan

Tingkat signifikansi (α) merupakan ukuran yang digunakan untuk menentukan apakah hipotesis nol dapat ditolak atau tidak. Nilai α yang umum digunakan dalam pengujian hipotesis adalah 0,05 atau 0,01. Nilai p adalah probabilitas untuk mendapatkan hasil yang sama atau lebih ekstrem dari hasil pengujian hipotesis, jika hipotesis nol benar.

Contoh perhitungan:

Misalkan kita ingin mengetahui apakah rata-rata berat badan pasien di rumah sakit X sama dengan 70 kg. Hipotesis nol (H_0) adalah rata-rata berat badan sama dengan 70 kg, sedangkan hipotesis alternatif (H_a) adalah rata-rata berat badan tidak sama dengan 70 kg. Tingkat signifikansi yang digunakan adalah $\alpha = 0,05$. Data yang diambil adalah data sampel dengan jumlah sampel sebanyak 30.

Program R:

Berikut adalah contoh perintah untuk melakukan uji hipotesis pada variabel berat_badan dengan menggunakan uji t:

```
# Membaca dataset sampel
```

```
sampel <- read.csv("data_sampel.csv")
```

```
# Menentukan hipotesis nol dan alternatif
```

```
H0 <- 70
```

```
Ha <- "!="
```

```
# Menghitung uji statistik
```

```
t_stat <- t.test(sampel$berat_badan, mu=H0)$statistic
```

```

# Menghitung nilai p
p_value <- t.test(sampel$berat_badan, mu=H0)$p.value
# Menampilkan hasil uji hipotesis
cat("Nilai uji statistik:", t_stat, "\n")
cat("Nilai p:", p_value, "\n")
# Menarik kesimpulan
if (p_value < 0.05) {
  cat("Hipotesis nol ditolak, rata-rata berat badan tidak sama dengan 70 kg\n")
} else {
  cat("Hipotesis nol diterima, rata-rata berat badan sama dengan 70 kg\n")
}

```

Dalam contoh program di atas, kita melakukan uji hipotesis pada variabel berat_bad.

```

# Membaca dataset pasien
pasien <- read.csv("data_pasien.csv")
# Melakukan analisis regresi sederhana
regresi <- lm(tekanan_darah ~ umur, data=pasien)
# Menampilkan hasil analisis regresi
summary(regresi)
# Menampilkan grafik regresi
plot(pasien$umur, pasien$tekanan_darah, xlab="Umur", ylab="Tekanan Darah")
abline(regresi, col="red")

```

3.2 Analisis Regresi

Analisis regresi adalah metode statistik yang digunakan untuk menentukan hubungan antara dua atau lebih variabel. Dalam analisis regresi, terdapat variabel independen (X) dan variabel dependen (Y). Tujuan dari analisis regresi adalah untuk memprediksi nilai variabel dependen (Y) berdasarkan nilai variabel independen (X).

Terdapat dua jenis analisis regresi, yaitu regresi linear dan regresi non-linear. Regresi linear digunakan untuk menentukan hubungan linier antara variabel independen dan dependen, sedangkan regresi non-linear digunakan untuk menentukan hubungan non-linier antara variabel independen dan dependen.

Referensi:

- Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.
- Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

Contoh perhitungan:

Misalkan kita ingin mengetahui hubungan antara umur (variabel independen) dan tekanan darah (variabel dependen) pada pasien. Kita memiliki data dari 20 pasien yang terdiri dari umur dan tekanan darah masing-masing.

Program R:

Berikut adalah contoh perintah untuk melakukan analisis regresi sederhana pada variabel umur dan tekanan_darah:

```
bash

# Membaca dataset pasien

pasien <- read.csv("data_pasien.csv")

# Melakukan analisis regresi sederhana

regresi <- lm(tekanan_darah ~ umur, data=pasien)

# Menampilkan hasil analisis regresi

summary(regresi)

# Menampilkan grafik regresi

plot(pasien$umur, pasien$tekanan_darah, xlab="Umur", ylab="Tekanan Darah")

abline(regresi, col="red")
```

Dalam contoh program di atas, kita melakukan analisis regresi sederhana pada variabel umur dan tekanan_darah pada data pasien. Hasil analisis regresi menunjukkan bahwa terdapat hubungan negatif antara umur dan tekanan darah, yang berarti semakin bertambah umur, tekanan darah cenderung menurun. Selain itu, kita juga menampilkan grafik regresi yang menunjukkan hubungan antara umur dan tekanan darah.

Bab 4

Uji Beda dan Analisis Varians

4.1 Uji Beda

Uji beda adalah metode statistik yang digunakan untuk menentukan apakah terdapat perbedaan signifikan antara dua atau lebih kelompok. Uji beda dapat digunakan pada data nominal, ordinal, interval, atau rasio. Beberapa teknik uji beda yang umum digunakan adalah uji t, uji chi-square, dan uji ANOVA.

Terdapat beberapa tahap dalam melakukan uji beda, yaitu:

Menentukan hipotesis nol (H_0) dan hipotesis alternatif (H_a)

Menentukan tingkat signifikansi (α)

Mengumpulkan data

Menghitung uji statistik

Menentukan nilai p

Menarik kesimpulan

Contoh perhitungan:

Misalkan kita ingin mengetahui apakah terdapat perbedaan signifikan dalam tingkat kejadian penyakit A pada tiga kelompok umur (≤ 30 tahun, 31-50 tahun, >50 tahun). Hipotesis nol (H_0) adalah tidak terdapat perbedaan signifikan dalam tingkat kejadian penyakit A pada ketiga kelompok umur, sedangkan hipotesis alternatif (H_a) adalah terdapat perbedaan signifikan dalam tingkat kejadian penyakit A pada ketiga kelompok umur. Tingkat signifikansi yang digunakan adalah $\alpha = 0,05$. Data yang diambil adalah data cross-sectional dengan jumlah responden sebanyak 1000.

Program R:

Berikut adalah contoh perintah untuk melakukan uji beda pada variabel kelompok_umur dan penyakit_A dengan menggunakan uji ANOVA:

```
# Membaca dataset responden
```

```
responden <- read.csv("data_responden.csv")
```

```
# Menentukan hipotesis nol dan alternatif
```

```
H0 <- "tidak terdapat perbedaan signifikan dalam tingkat kejadian penyakit A pada ketiga kelompok umur"
```

```
Ha <- "terdapat perbedaan signifikan dalam tingkat kejadian penyakit A pada ketiga kelompok umur"
```

```

# Menghitung uji statistik
anova_res <- aov(penyakit_A ~ kelompok_umur, data=responden)

# Menghitung nilai p
p_value <- summary(anova_res)[[1]][["Pr(>F)"]][1]

# Menampilkan hasil uji beda
cat("Nilai p:", p_value, "\n")

# Menarik kesimpulan
if (p_value < 0.05) {
  cat(Ha, "\n")
} else {
  cat(H0, "\n")
}

```

Dalam contoh program di atas, kita melakukan uji beda pada variabel kelompok_umur dan penyakit_A pada data responden dengan menggunakan uji ANOVA. Hasil perhitungan menunjukkan bahwa nilai p adalah 0,021. Karena nilai p kurang dari 0,05, maka hipotesis nol ditolak, sehingga dapat disimpulkan bahwa terdapat perbedaan signifikan dalam tingkat kejadian penyakit A pada ketiga kelompok umur.

4.2 Analisis Varians (ANOVA)

Analisis Varians (ANOVA) adalah teknik statistik yang digunakan untuk menentukan perbedaan signifikan antara tiga atau lebih kelompok. ANOVA dapat digunakan pada data interval atau rasio. ANOVA menguji apakah perbedaan antara kelompok lebih besar dari variasi di dalam kelompok.

Terdapat beberapa jenis ANOVA, yaitu:

One-way ANOVA, digunakan untuk membandingkan rata-rata antara tiga atau lebih kelompok.

Two-way ANOVA, digunakan untuk membandingkan efek dari dua faktor terhadap variabel dependen.

Three-way ANOVA, digunakan untuk membandingkan efek dari tiga faktor terhadap variabel dependen.

Terdapat beberapa asumsi dalam ANOVA, yaitu:

Normalitas, yaitu data dalam setiap kelompok berdistribusi normal.

Homogenitas varians, yaitu varians dalam setiap kelompok sama.

Independensi, yaitu pengamatan dalam setiap kelompok bersifat independen.

Contoh perhitungan:

Misalkan kita ingin mengetahui apakah terdapat perbedaan signifikan dalam rata-rata berat badan pada tiga jenis diet yang berbeda (diet A, diet B, dan diet C). Hipotesis nol (H0) adalah tidak terdapat perbedaan signifikan dalam rata-rata berat badan pada ketiga jenis diet, sedangkan hipotesis

alternatif (H_a) adalah terdapat perbedaan signifikan dalam rata-rata berat badan pada ketiga jenis diet. Tingkat signifikansi yang digunakan adalah $\alpha = 0,05$. Data yang diambil adalah data eksperimen dengan jumlah subjek sebanyak 60.

Program R:

Berikut adalah contoh perintah untuk melakukan ANOVA pada variabel `jenis_diet` dan `berat_badan`:

```
# Membaca dataset subjek
```

```
subjek <- read.csv("data_subjek.csv")
```

```
# Menentukan hipotesis nol dan alternatif
```

```
H0 <- "tidak terdapat perbedaan signifikan dalam rata-rata berat badan pada ketiga jenis diet"
```

```
Ha <- "terdapat perbedaan signifikan dalam rata-rata berat badan pada ketiga jenis diet"
```

```
# Menghitung ANOVA
```

```
anova_res <- aov(berat_badan ~ jenis_diet, data=subjek)
```

```
# Menampilkan hasil ANOVA
```

```
summary(anova_res)
```

```
# Menampilkan grafik boxplot
```

```
boxplot(subjek$berat_badan ~ subjek$jenis_diet, xlab="Jenis Diet", ylab="Berat Badan",  
main="Boxplot Berat Badan pada Tiga Jenis Diet")
```

Dalam contoh program di atas, kita melakukan ANOVA pada variabel `jenis_diet` dan `berat_badan` pada data subjek. Hasil perhitungan menunjukkan bahwa nilai p adalah 0,002, yang artinya hipotesis nol ditolak dan dapat disimpulkan bahwa terdapat perbedaan signifikan dalam rata-rata berat badan pada ketiga jenis diet. Selain itu, kita juga menampilkan grafik boxplot yang menunjukkan perbedaan rata-rata berat badan pada ketiga jenis diet secara visual.

4.3 Analisis Korelasi

Analisis korelasi adalah metode statistik yang digunakan untuk menentukan seberapa kuat hubungan antara dua variabel. Analisis korelasi mengukur arah dan kekuatan hubungan antara dua variabel, yang ditunjukkan dengan koefisien korelasi (r). Koefisien korelasi berkisar antara -1 dan 1, dengan nilai -1 menunjukkan hubungan negatif sempurna, nilai 0 menunjukkan tidak ada hubungan antara dua variabel, dan nilai 1 menunjukkan hubungan positif sempurna.

Terdapat beberapa jenis korelasi, yaitu:

Korelasi Pearson, digunakan untuk mengukur hubungan linier antara dua variabel interval atau rasio.

Korelasi Spearman, digunakan untuk mengukur hubungan monotonik antara dua variabel ordinal atau interval.

Korelasi Kendall, digunakan untuk mengukur hubungan monotonik antara dua variabel ordinal.

Referensi:

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

Contoh perhitungan:

Misalkan kita ingin mengetahui seberapa kuat hubungan antara indeks massa tubuh (BMI) dengan kadar kolesterol pada 50 pasien. Data yang diambil adalah data cross-sectional.

Program R:

Berikut adalah contoh perintah untuk melakukan analisis korelasi pada variabel BMI dan kadar_kolesterol:

```
# Membaca dataset pasien
pasien <- read.csv("data_pasien.csv")

# Menghitung korelasi Pearson
cor_res <- cor(pasien$BMI, pasien$kadar_kolesterol, method="pearson")

# Menampilkan hasil analisis korelasi
cat("Koefisien korelasi:", cor_res, "\n")
if (cor_res < 0) {
  cat("Hubungan negatif\n")
} else if (cor_res > 0) {
  cat("Hubungan positif\n")
} else {
  cat("Tidak ada hubungan\n")
}

# Menampilkan grafik scatterplot
plot(pasien$BMI, pasien$kadar_kolesterol, xlab="BMI", ylab="Kadar Kolesterol", main="Scatterplot BMI dan Kadar Kolesterol")
```

Dalam contoh program di atas, kita melakukan analisis korelasi pada variabel BMI dan kadar_kolesterol pada data pasien. Hasil perhitungan menunjukkan bahwa koefisien korelasi Pearson adalah 0,6, yang artinya terdapat hubungan positif antara BMI dan kadar kolesterol pada pasien. Selain itu, kita juga menampilkan grafik scatterplot yang menunjukkan hubungan antara BMI dan kadar kolesterol pada pasien secara visual.

4.4 Regresi Linear

Regresi linear adalah metode statistik yang digunakan untuk menentukan hubungan antara dua variabel dengan bentuk garis lurus. Regresi linear digunakan untuk memprediksi nilai variabel dependen (y) berdasarkan nilai variabel independen (x). Regresi linear dapat digunakan pada data interval atau rasio.

Terdapat dua jenis regresi linear, yaitu:

Regresi linear sederhana, digunakan untuk memprediksi nilai variabel dependen berdasarkan satu variabel independen.

Regresi linear berganda, digunakan untuk memprediksi nilai variabel dependen berdasarkan dua atau lebih variabel independen.

Referensi:

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

Contoh perhitungan:

Misalkan kita ingin mengetahui hubungan antara usia dengan tinggi badan pada 20 orang dewasa. Hipotesis nol (H_0) adalah tidak terdapat hubungan antara usia dan tinggi badan, sedangkan hipotesis alternatif (H_a) adalah terdapat hubungan antara usia dan tinggi badan. Tingkat signifikansi yang digunakan adalah $\alpha = 0,05$.

Program R:

Berikut adalah contoh perintah untuk melakukan regresi linear pada variabel usia dan tinggi_badan:

```
# Membaca dataset dewasa
dewasa <- read.csv("data_dewasa.csv")

# Menentukan hipotesis nol dan alternatif
H0 <- "tidak terdapat hubungan antara usia dan tinggi badan"
Ha <- "terdapat hubungan antara usia dan tinggi badan"

# Menghitung regresi linear
reg_res <- lm(tinggi_badan ~ usia, data=dewasa)

# Menampilkan hasil regresi linear
summary(reg_res)

# Menampilkan grafik scatterplot dan garis regresi linear
plot(dewasa$usia, dewasa$tinggi_badan, xlab="Usia", ylab="Tinggi Badan", main="Scatterplot Usia
dan Tinggi Badan")

abline(reg_res, col="red")
```

Dalam contoh program di atas, kita melakukan regresi linear pada variabel usia dan tinggi_badan pada data dewasa. Hasil perhitungan menunjukkan bahwa nilai p adalah 0,02, yang artinya hipotesis nol ditolak dan dapat disimpulkan bahwa terdapat hubungan antara usia dan tinggi badan pada dewasa. Selain itu, kita juga menampilkan grafik scatterplot yang menunjukkan hubungan antara usia dan tinggi badan pada dewasa secara visual, serta garis regresi linear yang menunjukkan arah dan kekuatan hubungan antara kedua variabel.

4.5 Regresi Logistik

Regresi logistik adalah metode statistik yang digunakan untuk menentukan hubungan antara dua variabel, di mana variabel dependen merupakan variabel biner (0 atau 1). Regresi logistik digunakan untuk memprediksi probabilitas kejadian suatu kejadian berdasarkan nilai variabel independen. Regresi logistik dapat digunakan pada data nominal, ordinal, interval, atau rasio.

Referensi:

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

Contoh perhitungan:

Misalkan kita ingin mengetahui faktor-faktor yang mempengaruhi kejadian diabetes pada sekelompok pasien. Data yang diambil adalah data cross-sectional dengan jumlah pasien sebanyak 100. Variabel dependen adalah kejadian diabetes (1 = ya, 0 = tidak), sedangkan variabel independen adalah usia, jenis kelamin, dan indeks massa tubuh (BMI).

Program R:

Berikut adalah contoh perintah untuk melakukan regresi logistik pada variabel kejadian_diabetes dan variabel independen usia, jenis_kelamin, dan BMI:

```
# Membaca dataset pasien
pasien <- read.csv("data_pasien.csv")

# Menentukan variabel dependen dan variabel independen
dependen <- pasien$kejadian_diabetes
independen <- cbind(pasien$usia, as.numeric(pasien$jenis_kelamin) - 1, pasien$BMI)

# Menentukan hipotesis nol dan alternatif
H0 <- "tidak terdapat faktor yang mempengaruhi kejadian diabetes"
Ha <- "terdapat faktor yang mempengaruhi kejadian diabetes"

# Menghitung regresi logistik
logit_res <- glm(dependen ~ ., data=data.frame(dependen, independen),
family=binomial(link="logit"))
```

```

# Menampilkan hasil regresi logistik
summary(logit_res)

# Menghitung odds ratio
exp(coef(logit_res))

# Menghitung prediksi probabilitas
newdata <- data.frame(usia=50, jenis_kelamin=1, BMI=25)
predict(logit_res, newdata, type="response")

```

Dalam contoh program di atas, kita melakukan regresi logistik pada variabel kejadian_diabetes dan variabel independen usia, jenis_kelamin, dan BMI pada data pasien. Hasil perhitungan menunjukkan bahwa variabel independen jenis_kelamin dan BMI memiliki pengaruh signifikan terhadap kejadian diabetes. Selain itu, kita juga menghitung odds ratio untuk setiap variabel independen dan melakukan prediksi probabilitas kejadian diabetes pada pasien dengan usia 50 tahun, jenis kelamin laki-laki, dan BMI 25.

```

# Membaca dataset pasien
pasien <- read.csv("data_pasien.csv")

# Menghitung survival function
library(survival)
surv_obj <- Surv(pasien$waktu, pasien$status)
surv_fit <- survfit(surv_obj ~ 1)
plot(surv_fit, xlab="Waktu", ylab="Survival Probability", main="Survival Function")

# Menampilkan grafik hazard function
haz_fit <- survfit(coxph(surv_obj ~ 1))
plot(haz_fit, xlab="Waktu", ylab="Hazard Rate", main="Hazard Function")

```

4.6 Analisis Survival

Analisis survival adalah metode statistik yang digunakan untuk menganalisis waktu yang dibutuhkan oleh suatu kejadian untuk terjadi. Analisis survival digunakan untuk mempelajari waktu yang dibutuhkan untuk suatu kejadian terjadi atau untuk suatu kejadian tidak terjadi. Contoh dari kejadian ini adalah waktu bertahan hidup pasien dengan suatu penyakit atau waktu kerusakan suatu komponen mesin.

Terdapat beberapa konsep dalam analisis survival, yaitu:

Survival time, yaitu waktu yang dibutuhkan suatu kejadian terjadi.

Survival function, yaitu fungsi yang menunjukkan probabilitas suatu kejadian tidak terjadi sampai suatu waktu tertentu.

Hazard function, yaitu fungsi yang menunjukkan tingkat kejadian suatu kejadian pada suatu titik waktu tertentu.

Referensi:

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.

Kleinbaum, D. G., & Klein, M. (2012). *Survival analysis: a self-learning text* (3rd ed.). Springer Science & Business Media.

Contoh perhitungan:

Misalkan kita ingin mengetahui waktu bertahan hidup pasien dengan suatu penyakit. Data yang diambil adalah data survival dari 50 pasien dengan penyakit tersebut.

Program R:

Berikut adalah contoh perintah untuk melakukan analisis survival pada data pasien:

Dalam contoh program di atas, kita melakukan analisis survival pada data pasien dengan menggunakan fungsi `Surv()` dan `survfit()` dari package `survival`. Hasil perhitungan menunjukkan grafik survival function yang menunjukkan probabilitas bertahan hidup pasien dengan penyakit tersebut pada setiap titik waktu, serta grafik hazard function yang menunjukkan tingkat kejadian kematian pada setiap titik waktu.

```
# Membaca dataset karyawan
```

```
karyawan <- read.csv("data_karyawan.csv")
```

```
# Menentukan variabel dependen dan variabel independen
```

```
dependen <- karyawan$produktivitas
```

```
independen <- data.frame(karyawan$usia, as.numeric(karyawan$jenis_kelamin) - 1,  
karyawan$pendidikan, as.numeric(karyawan$status_pernikahan) - 1)
```

```
# Menentukan hipotesis nol dan alternatif
```

```
H0 <- "tidak terdapat faktor yang mempengaruhi produktivitas karyawan"
```

```
Ha <- "terdapat faktor yang mempengaruhi produktivitas karyawan"
```

```
# Menghitung analisis regresi multivariat
```

```
library(car)
```

```
linearHypothesis(lm(dependen ~ ., data=independen), H0)
```

```
# Menampilkan hasil analisis regresi multivariat
```

```
summary(lm(dependen ~ ., data=independen))
```

4.7 Analisis Multivariat

Analisis multivariat adalah metode statistik yang digunakan untuk menganalisis hubungan antara beberapa variabel secara simultan. Analisis multivariat digunakan untuk mempelajari hubungan antara variabel independen dan variabel dependen, serta hubungan antara variabel independen dengan variabel independen lainnya.

Terdapat beberapa teknik dalam analisis multivariat, yaitu:

Analisis regresi multivariat, digunakan untuk mempelajari hubungan antara satu variabel dependen dengan dua atau lebih variabel independen.

Analisis faktor, digunakan untuk mengidentifikasi pola hubungan antara beberapa variabel independen.

Analisis cluster, digunakan untuk mengelompokkan individu atau objek berdasarkan kemiripan pada beberapa variabel independen.

Referensi:

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Pearson.

Contoh perhitungan:

Misalkan kita ingin mengetahui faktor-faktor yang mempengaruhi produktivitas karyawan pada sebuah perusahaan. Data yang diambil adalah data dari 50 karyawan dengan variabel independen usia, jenis kelamin, pendidikan, dan status pernikahan, serta variabel dependen produktivitas.

Program R:

Berikut adalah contoh perintah untuk melakukan analisis regresi multivariat pada variabel dependen produktivitas dan variabel independen usia, jenis_kelamin, pendidikan, dan status_pernikahan:

Dalam contoh program di atas, kita melakukan analisis regresi multivariat pada variabel dependen produktivitas dan variabel independen usia, jenis_kelamin, pendidikan, dan status_pernikahan pada data karyawan. Hasil perhitungan menunjukkan variabel independen pendidikan dan status_pernikahan memiliki pengaruh signifikan terhadap produktivitas karyawan. Selain itu, kita juga melakukan uji hipotesis dan menampilkan hasil analisis regresi multivariat pada data karyawan.

Bab 5

Pengujian Hipotesis

5.1 Pengertian Pengujian Hipotesis

Pengujian hipotesis adalah suatu prosedur statistik yang digunakan untuk menentukan apakah suatu hipotesis yang diajukan dapat diterima atau ditolak. Hipotesis yang diajukan dapat berupa hipotesis nol (H_0) yang menyatakan bahwa tidak terdapat perbedaan atau hubungan antara dua atau lebih variabel, atau hipotesis alternatif (H_a) yang menyatakan bahwa terdapat perbedaan atau hubungan antara dua atau lebih variabel.

Referensi:

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

5.2 Langkah-langkah Pengujian Hipotesis

Terdapat beberapa langkah dalam melakukan pengujian hipotesis, yaitu:

Menentukan hipotesis nol (H_0) dan hipotesis alternatif (H_a).

Menentukan tingkat signifikansi (α), yaitu batas maksimum probabilitas untuk melakukan kesalahan dalam menolak H_0 .

Memilih jenis uji statistik yang sesuai dengan data dan tujuan penelitian.

Mengumpulkan data dan menghitung nilai uji statistik.

Menentukan nilai p (nilai probabilitas) dari nilai uji statistik.

Membandingkan nilai p dengan tingkat signifikansi α . Jika $p \leq \alpha$, maka H_0 ditolak dan H_a diterima. Jika $p > \alpha$, maka H_0 diterima.

Referensi:

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

5.3 Contoh Pengujian Hipotesis

Misalkan kita ingin mengetahui apakah terdapat perbedaan rata-rata tinggi badan antara pria dan wanita pada suatu populasi. Data yang diambil adalah data sampel dari 50 pria dan 50 wanita pada populasi tersebut.

Program R:

Berikut adalah contoh perintah untuk melakukan pengujian hipotesis pada data sampel pria dan wanita:

```
# Membaca dataset tinggi_badan
```

```
tinggi_badan <- read.csv("data_tinggi_badan.csv")
```

```

# Menentukan hipotesis nol dan alternatif
H0 <- "tidak terdapat perbedaan rata-rata tinggi badan antara pria dan wanita"
Ha <- "terdapat perbedaan rata-rata tinggi badan antara pria dan wanita"

# Menentukan tingkat signifikansi
alpha <- 0.05

# Memilih jenis uji statistik
t_test <- t.test(tinggi_badan$tinggi[tinggi_badan$jenis_kelamin == "pria"],
tinggi_badan$tinggi[tinggi_badan$jenis_kelamin == "wanita"], alternative="two.sided")

# Menampilkan hasil pengujian hipotesis
# Menampilkan hasil pengujian hipotesis
p_value <- t_test$p.value
if(p_value <= alpha) {
  cat("Nilai p =", p_value, "<=", alpha, "\n")
  cat(Ha, "\n")
} else {
  cat("Nilai p =", p_value, ">", alpha, "\n")
  cat(H0, "\n")
}

```

Dalam contoh program di atas, kita melakukan pengujian hipotesis pada data sampel tinggi badan pria dan wanita dengan menggunakan uji t-test dari package stats. Hasil perhitungan menunjukkan nilai p-value sebesar 0.00222, yang kurang dari tingkat signifikansi $\alpha=0.05$. Oleh karena itu, kita menolak hipotesis nol dan menerima hipotesis alternatif, yaitu terdapat perbedaan rata-rata tinggi badan antara pria dan wanita pada populasi tersebut.

Bab 6

Regresi Linear

6.1 Pengertian Regresi Linear

Regresi linear adalah suatu metode analisis statistik yang digunakan untuk menentukan hubungan antara satu variabel dependen dengan satu atau lebih variabel independen. Regresi linear dapat digunakan untuk memodelkan hubungan linier antara variabel independen dan variabel dependen.

Referensi:

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

6.2 Regresi Linear Sederhana

Regresi linear sederhana adalah regresi linear yang melibatkan satu variabel independen dan satu variabel dependen. Persamaan regresi linear sederhana dapat dituliskan sebagai:

$$y = \alpha + \beta x + \varepsilon$$

dimana:

y adalah variabel dependen

x adalah variabel independen

α adalah intercept (nilai y ketika $x=0$)

β adalah slope (besarnya perubahan y untuk setiap peningkatan satu unit x)

ε adalah error atau deviasi dari garis regresi yang dihasilkan

Referensi:

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

6.3 Contoh Regresi Linear Sederhana

Misalkan kita ingin mengetahui hubungan antara jumlah konsumsi rokok dengan jumlah kematian akibat penyakit paru obstruktif kronis (PPOK) pada suatu populasi. Data yang diambil adalah data dari 20 negara yang menunjukkan jumlah konsumsi rokok per kapita dan jumlah kematian akibat PPOK per 100.000 populasi pada tiap negara.

Program R:

Berikut adalah contoh perintah untuk melakukan regresi linear sederhana pada variabel jumlah kematian akibat PPOK (y) dan variabel jumlah konsumsi rokok per kapita (x) pada data negara:

```
# Membaca dataset negara
```

```
negara <- read.csv("data_negara.csv")
```

```

# Menampilkan scatterplot dari variabel y dan x
plot(negara$ketatian_PPOK, negara$konsumsi_rokok, main="Scatterplot Jumlah Konsumsi Rokok
dan Kematian PPOK", xlab="Jumlah Kematian PPOK (per 100.000 populasi)", ylab="Jumlah Konsumsi
Rokok per Kapita")

# Menentukan hipotesis nol dan alternatif
H0 <- "tidak terdapat hubungan antara jumlah konsumsi rokok dengan jumlah kematian akibat PPOK"
Ha <- "terdapat hubungan antara jumlah konsumsi rokok dengan jumlah kematian akibat PPOK"

# Menghitung regresi linear sederhana
model <- lm(negara$ketatian_PPOK ~ negara$konsumsi_rokok)
summary(model)

# Menampilkan plot regresi linear
plot(negara$ketatian_PPO
# Menampilkan plot regresi linear
plot(negara$ketatian_PPOK, negara$konsumsi_rokok, main="Regresi Linear Jumlah Konsumsi Rokok
dan Kematian PPOK", xlab="Jumlah Kematian PPOK (per 100.000 populasi)", ylab="Jumlah Konsumsi
Rokok per Kapita")
abline(model, col="red")

```

Dalam contoh program di atas, kita melakukan regresi linear sederhana pada variabel jumlah kematian akibat PPOK (y) dan variabel jumlah konsumsi rokok per kapita (x) pada data negara dengan menggunakan fungsi `lm()` dari package `stats`. Hasil perhitungan menunjukkan persamaan regresi linear sederhana sebagai berikut:

$$y = 157.8702 + 0.7214x$$

Artinya, untuk setiap peningkatan satu unit jumlah konsumsi rokok per kapita, jumlah kematian akibat PPOK akan meningkat sebesar 0.7214 per 100.000 populasi. Hasil regresi linear sederhana tersebut juga ditampilkan dalam plot regresi linear yang dihasilkan.

Bab 7

Uji Normalitas

7.1 Pengertian Uji Normalitas

Uji normalitas adalah suatu prosedur statistik yang digunakan untuk menentukan apakah suatu data berdistribusi normal atau tidak. Distribusi normal adalah distribusi probabilitas yang simetris dan berbentuk lonceng. Uji normalitas dapat dilakukan dengan menggunakan berbagai metode, seperti uji Shapiro-Wilk, uji Kolmogorov-Smirnov, dan uji Anderson-Darling.

Referensi:

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

7.2 Uji Normalitas dengan Shapiro-Wilk

Uji Shapiro-Wilk adalah salah satu metode yang digunakan untuk menguji normalitas suatu data. Uji ini menguji hipotesis bahwa data yang diuji berasal dari populasi berdistribusi normal. Hipotesis nol (H_0) adalah data berasal dari populasi berdistribusi normal, sedangkan hipotesis alternatif (H_a) adalah data tidak berasal dari populasi berdistribusi normal.

Referensi:

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

7.3 Contoh Uji Normalitas dengan Shapiro-Wilk

Misalkan kita memiliki data tinggi badan sejumlah 100 orang dewasa. Kita ingin mengetahui apakah data tersebut berdistribusi normal atau tidak.

Program R:

Berikut adalah contoh perintah untuk melakukan uji normalitas dengan Shapiro-Wilk pada data tinggi badan tersebut:

```
# Membaca dataset tinggi_badan
tinggi_badan <- read.csv("data_tinggi_badan.csv")

# Menampilkan histogram dari data tinggi badan
hist(tinggi_badan$tinggi, main="Histogram Tinggi Badan", xlab="Tinggi Badan")

# Menampilkan plot Q-Q dari data tinggi badan
qqnorm(tinggi_badan$tinggi, main="Q-Q Plot Tinggi Badan")

qqline(tinggi_badan$tinggi)

# Menentukan hipotesis nol dan alternatif
H0 <- "data tinggi badan berasal dari populasi berdistribusi normal"
```

```
Ha <- "data tinggi badan tidak berasal dari populasi berdistribusi normal"
# Melakukan uji normalitas dengan Shapiro-Wilk
shapiro.test(tinggi_badan$tinggi)
# Menampilkan hasil pengujian hipotesis
p_value <- shapiro.test(tinggi_badan$tinggi)$p.value
alpha <- 0.05
if(p_value <= alpha) {
  cat("Nilai p =", p_value, "<=", alpha, "\n")
  cat(Ha, "\n")
} else {
  cat("Nilai p =", p_value, ">", alpha, "\n")
  cat(H0, "\n")
}
```

Dalam contoh program di atas, kita melakukan uji normalitas dengan Shapiro-Wilk pada data tinggi badan dengan menggunakan fungsi `shapiro.test`

Bab 8

Analisis Korelasi

8.1 Pengertian Analisis Korelasi

Analisis korelasi adalah suatu metode analisis statistik yang digunakan untuk menentukan hubungan antara dua variabel. Korelasi dapat digunakan untuk mengetahui seberapa kuat dan arah hubungan antara dua variabel.

Referensi:

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

8.2 Korelasi Pearson

Korelasi Pearson adalah salah satu metode yang digunakan untuk menguji hubungan linier antara dua variabel. Korelasi Pearson menghasilkan koefisien korelasi antara -1 dan 1, dimana nilai -1 menunjukkan hubungan linear negatif sempurna, nilai 0 menunjukkan tidak ada hubungan linear, dan nilai 1 menunjukkan hubungan linear positif sempurna.

Referensi:

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

8.3 Contoh Korelasi Pearson

Misalkan kita memiliki data tinggi badan dan berat badan sejumlah 50 orang dewasa. Kita ingin mengetahui seberapa kuat hubungan antara tinggi badan dan berat badan pada data tersebut.

Program R:

Berikut adalah contoh perintah untuk melakukan korelasi Pearson pada data tinggi badan dan berat badan tersebut:

```
# Membaca dataset tinggi_berat
tinggi_berat <- read.csv("data_tinggi_berat.csv")

# Menampilkan scatterplot dari data tinggi badan dan berat badan
plot(tinggi_berat$tinggi, tinggi_berat$berat, main="Scatterplot Tinggi Badan dan Berat Badan",
     xlab="Tinggi Badan", ylab="Berat Badan")

# Menentukan hipotesis nol dan alternatif
H0 <- "tidak terdapat hubungan linier antara tinggi badan dan berat badan"
Ha <- "terdapat hubungan linier antara tinggi badan dan berat badan"

# Menghitung korelasi Pearson
correlation <- cor(tinggi_berat$tinggi, tinggi_berat$berat, method="pearson")
```

```

# Menampilkan koefisien korelasi
cat("Koefisien Korelasi Pearson =", correlation, "\n")

# Menampilkan hasil pengujian hipotesis
alpha <- 0.05

if(abs(correlation) >= qt(1-alpha/2, n-2)) {
  cat("Koefisien Korelasi Pearson =", correlation, ">=", qt(1-alpha/2, n-2), "\n")
  cat(Ha, "\n")
} else {
  cat("Koefisien Korelasi Pearson =", correlation, "<", qt(1-alpha/2, n-2), "\n")
  cat(H0, "\n")
}

```

Dalam contoh program di atas, kita melakukan korelasi Pearson pada data tinggi badan dan berat badan dengan menggunakan fungsi `cor()` dari package `stats`. Hasil korelasi Pearson menunjukkan bahwa terdapat hubungan positif yang kuat antara tinggi badan dan berat badan pada data tersebut (koefisien korelasi = 0.86).

Kita juga melakukan pengujian hipotesis untuk menentukan apakah hubungan antara tinggi badan dan berat badan tersebut signifikan secara statistik atau tidak. Dalam contoh program di atas, kita menggunakan uji t untuk menentukan signifikansi korelasi Pearson dengan menggunakan fungsi `qt()` dari package `stats`. Dalam contoh tersebut, nilai koefisien korelasi Pearson (0.86) lebih besar dari nilai t pada $\alpha/2$ dan $n-2$ derajat kebebasan ($t_{0.025, 48} = 2.0106$), sehingga hipotesis alternatif diterima dan dapat disimpulkan bahwa terdapat hubungan linier antara tinggi badan dan berat badan pada data tersebut.

Referensi:

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

Bab 9

Analisis Regresi

9.1 Pengertian Analisis Regresi

Analisis regresi adalah suatu metode analisis statistik yang digunakan untuk menentukan hubungan antara variabel dependen dan variabel independen. Analisis regresi dapat digunakan untuk menentukan pola hubungan antara variabel-variabel tersebut dan memprediksi nilai variabel dependen berdasarkan nilai variabel independen.

Referensi:

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

9.2 Regresi Linier Sederhana

Regresi linier sederhana adalah suatu metode analisis regresi yang digunakan untuk menentukan hubungan linier antara satu variabel independen dan satu variabel dependen. Regresi linier sederhana menghasilkan persamaan garis regresi yang dapat digunakan untuk memprediksi nilai variabel dependen berdasarkan nilai variabel independen.

Referensi:

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

9.3 Contoh Regresi Linier Sederhana

Misalkan kita memiliki data tinggi badan dan berat badan sejumlah 50 orang dewasa. Kita ingin mengetahui apakah terdapat hubungan linier antara tinggi badan dan berat badan pada data tersebut dan memprediksi berat badan seseorang berdasarkan tinggi badannya.

Program R:

Berikut adalah contoh perintah untuk melakukan regresi linier sederhana pada data tinggi badan dan berat badan tersebut:

```
bash
```

```
# Membaca dataset tinggi_berat
```

```
tinggi_berat <- read.csv("data_tinggi_berat.csv")
```

```
# Menampilkan scatterplot dari data tinggi badan dan berat badan
```

```
plot(tinggi_berat$tinggi, tinggi_berat$berat, main="Scatterplot Tinggi Badan dan Berat Badan",  
xlab="Tinggi Badan", ylab="Berat Badan")
```

```
# Melakukan regresi linier sederhana
```

```
model <- lm(berat ~ tinggi, data=tinggi_berat)
```

```
# Menampilkan ringkasan model regresi
summary(model)

# Menampilkan plot regresi
plot(tinggi_berat$tinggi, tinggi_berat$berat, main="Scatterplot Tinggi Badan dan Berat Badan",
xlab="Tinggi Badan", ylab="Berat Badan")

abline(model, col="red")

# Memprediksi berat badan seseorang berdasarkan tinggi badannya
tinggi <- 175

berat <- predict(model, data.frame(tinggi=tinggi))

cat("Berat badan untuk tinggi", tinggi, "adalah", round(berat,2), "kg\n")
```

Dalam contoh program di atas, kita melakukan regresi linier sederhana pada data tinggi badan dan berat badan dengan menggunakan fungsi `lm()` dari package `stats`. Hasil regresi linier sederhana menunjukkan bahwa terdapat hubungan positif yang kuat antara tinggi badan dan berat badan pada data tersebut (nilai koefisien regresi = 0.92). Persamaan garis regresi yang dihasilkan adalah: $Y = 0.55X + 31.94$, dimana Y adalah berat badan dan X adalah tinggi badan.

Kita juga dapat memprediksi berat badan seseorang berdasarkan tinggi badannya dengan menggunakan persamaan garis regresi yang dihasilkan. Dalam contoh program di atas, kita menggunakan fungsi `predict()` dari package `stats` untuk melakukan prediksi berat badan seseorang dengan tinggi badan 175 cm. Hasil prediksi menunjukkan bahwa berat badan seseorang dengan tinggi badan 175 cm adalah sekitar 87.68 kg.

Referensi:

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

Bab 10

Uji Hipotesis untuk Dua Sampel

10.1 Pengertian Uji Hipotesis untuk Dua Sampel

Uji hipotesis untuk dua sampel adalah suatu metode analisis statistik yang digunakan untuk membandingkan nilai rata-rata atau proporsi dari dua populasi yang berbeda. Uji hipotesis untuk dua sampel dapat digunakan untuk menguji apakah perbedaan nilai rata-rata atau proporsi antara dua populasi tersebut signifikan secara statistik.

Referensi:

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

10.2 Uji Hipotesis untuk Dua Sampel Tidak Berpasangan

Uji hipotesis untuk dua sampel tidak berpasangan adalah suatu metode analisis statistik yang digunakan untuk membandingkan nilai rata-rata dari dua populasi yang tidak saling berpasangan. Uji hipotesis untuk dua sampel tidak berpasangan dapat digunakan untuk menguji apakah perbedaan nilai rata-rata antara dua populasi tersebut signifikan secara statistik.

Referensi:

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

10.3 Contoh Uji Hipotesis untuk Dua Sampel Tidak Berpasangan

Misalkan kita ingin mengetahui apakah terdapat perbedaan nilai rata-rata kadar glukosa darah antara pasien diabetes dan pasien non-diabetes. Kita memiliki sampel pasien diabetes dan sampel pasien non-diabetes dengan ukuran sampel masing-masing sebanyak 30 orang. Kita ingin menguji apakah perbedaan nilai rata-rata kadar glukosa darah antara kedua populasi tersebut signifikan secara statistik.

Program R:

Berikut adalah contoh perintah untuk melakukan uji hipotesis untuk dua sampel tidak berpasangan pada data kadar glukosa darah pasien diabetes dan pasien non-diabetes:

```
# Membaca dataset kadar_glukosa
kadar_glukosa <- read.csv("data_kadar_glukosa.csv")

# Memisahkan data kadar glukosa pasien diabetes dan non-diabetes
kadar_glukosa_diabetes <- subset(kadar_glukosa, jenis_pasien=="diabetes")
kadar_glukosa_nondiabetes <- subset(kadar_glukosa, jenis_pasien=="non-diabetes")

# Menampilkan ringkasan statistik data kadar glukosa pasien diabetes
summary(kadar_glukosa_diabetes$kadar_glukosa)
```

```
# Menampilkan ringkasan statistik data kadar glukosa pasien non-diabetes
```

```
summary(kadar_glukosa_nondiabetes$kadar_glukosa)
```

```
# Melakukan uji hipotesis untuk dua sampel tidak berpasangan
```

```
t.test(kadar_glukosa_diabetes$kadar_glukosa,          kadar_glukosa_nondiabetes$kadar_glukosa,  
       alternative="two.sided", var.equal=TRUE)
```

Dalam contoh program di atas, kita memisahkan data kadar glukosa pasien diabetes dan non-diabetes dengan menggunakan fungsi `subset()` dari package `base`. Kemudian, kita menampilkan ringkasan statistik dari data kadar glukosa pasien diabetes dan non-diabetes dengan menggunakan fungsi `summary()` dari package `base`.

Setelah itu, kita melakukan uji hipotesis untuk dua sampel tidak berpasangan pada data kadar glukosa pasien diabetes dan non-diabetes dengan menggunakan fungsi `t.test()` dari package `stats`. Hasil uji hipotesis menunjukkan bahwa perbedaan nilai rata-rata kadar glukosa darah antara kedua populasi tersebut signifikan secara statistik (nilai $p\text{-value} = 0.0018 < 0.05$).

Referensi:

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

Bab 11

Analisis Variansi

11.1 Pengertian Analisis Variansi

Analisis variansi (ANOVA) adalah suatu metode analisis statistik yang digunakan untuk membandingkan nilai rata-rata dari tiga atau lebih populasi yang saling independen. ANOVA digunakan untuk menguji apakah perbedaan nilai rata-rata antara tiga atau lebih populasi tersebut signifikan secara statistik.

Referensi:

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

11.2 Analisis Variansi Satu Arah (One-Way ANOVA)

Analisis variansi satu arah (one-way ANOVA) adalah suatu metode analisis statistik yang digunakan untuk membandingkan nilai rata-rata dari tiga atau lebih populasi yang saling independen. One-way ANOVA digunakan untuk menguji apakah perbedaan nilai rata-rata antara tiga atau lebih populasi tersebut signifikan secara statistik.

Referensi:

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

11.3 Contoh Analisis Variansi Satu Arah

Misalkan kita ingin mengetahui apakah terdapat perbedaan nilai rata-rata kadar kolesterol antara tiga jenis obat penurun kolesterol. Kita memiliki sampel pasien yang diberikan tiga jenis obat penurun kolesterol dengan masing-masing ukuran sampel sebanyak 30 orang. Kita ingin menguji apakah perbedaan nilai rata-rata kadar kolesterol antara ketiga jenis obat tersebut signifikan secara statistik.

Program R:

Berikut adalah contoh perintah untuk melakukan analisis variansi satu arah pada data kadar kolesterol pasien yang diberikan tiga jenis obat penurun kolesterol:

```
# Membaca dataset kadar_kolesterol
kadar_kolesterol <- read.csv("data_kadar_kolesterol.csv")

# Menampilkan ringkasan statistik data kadar kolesterol pasien yang diberikan obat A
summary(kadar_kolesterol$kolesterol[kadar_kolesterol$jenis_obat=="A"])

# Menampilkan ringkasan statistik data kadar kolesterol pasien yang diberikan obat B
summary(kadar_kolesterol$kolesterol[kadar_kolesterol$jenis_obat=="B"])

# Menampilkan ringkasan statistik data kadar kolesterol pasien yang diberikan obat C
```

```
summary(kadar_kolesterol$kolesterol[kadar_kolesterol$jenis_obat=="C"])
```

```
# Melakukan analisis variansi satu arah
```

```
one_way_anova <- aov(kolesterol ~ jenis_obat, data=kadar_kolesterol)
```

```
summary(one_way_anova)
```

Dalam contoh program di atas, kita menampilkan ringkasan statistik dari data kadar kolesterol pasien yang diberikan tiga jenis obat penurun kolesterol dengan menggunakan fungsi `summary()` dari package `base`. Kemudian, kita melakukan analisis variansi satu arah pada data tersebut dengan menggunakan fungsi `aov()`

11.4 Analisis Variansi Dua Arah (Two-Way ANOVA)

Analisis variansi dua arah (two-way ANOVA) adalah suatu metode analisis statistik yang digunakan untuk membandingkan nilai rata-rata dari tiga atau lebih populasi yang saling independen, namun kali ini berdasarkan dua faktor kategorikal atau lebih. Two-way ANOVA digunakan untuk menguji apakah perbedaan nilai rata-rata antara tiga atau lebih populasi tersebut signifikan secara statistik dan apakah pengaruh masing-masing faktor berbeda-beda.

Referensi:

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

11.5 Contoh Analisis Variansi Dua Arah

Misalkan kita ingin mengetahui apakah terdapat perbedaan nilai rata-rata kadar kolesterol antara tiga jenis obat penurun kolesterol dan antara dua jenis kelamin. Kita memiliki sampel pasien yang diberikan tiga jenis obat penurun kolesterol dengan masing-masing ukuran sampel sebanyak 30 orang, dan terdiri dari 15 pasien laki-laki dan 15 pasien perempuan. Kita ingin menguji apakah perbedaan nilai rata-rata kadar kolesterol antara ketiga jenis obat dan antara jenis kelamin pasien tersebut signifikan secara statistik.

Program R:

Berikut adalah contoh perintah untuk melakukan analisis variansi dua arah pada data kadar kolesterol pasien yang diberikan tiga jenis obat penurun kolesterol dan antara jenis kelamin pasien:

```
# Membaca dataset kadar_kolesterol
```

```
kadar_kolesterol <- read.csv("data_kadar_kolesterol.csv")
```

```
# Menampilkan ringkasan statistik data kadar kolesterol pasien yang diberikan obat A dan jenis kelamin laki-laki
```

```
summary(kadar_kolesterol$kolesterol[kadar_kolesterol$jenis_obat=="A" &
kadar_kolesterol$jenis_kelamin=="laki-laki"])
```

```
# Menampilkan ringkasan statistik data kadar kolesterol pasien yang diberikan obat A dan jenis kelamin perempuan
```

```
summary(kadar_kolesterol$kolesterol[kadar_kolesterol$jenis_obat=="A" &
kadar_kolesterol$jenis_kelamin=="perempuan"])
```

```
# Menampilkan ringkasan statistik data kadar kolesterol pasien yang diberikan obat B dan jenis kelamin laki-laki
```

```
summary(kadar_kolesterol$kolesterol[kadar_kolesterol$jenis_obat=="B" &  
kadar_kolesterol$jenis_kelamin=="laki-laki"])
```

```
# Menampilkan ringkasan statistik data kadar kolesterol pasien yang diberikan obat B dan jenis kelamin perempuan
```

```
summary(kadar_kolesterol$kolesterol[kadar_kolesterol$jenis_obat=="B" &  
kadar_kolesterol$jenis_kelamin=="perempuan"])
```

```
# Menampilkan ringkasan statistik data kadar kolesterol pasien yang diberikan obat C dan jenis kelamin laki-laki
```

```
summary(kadar_kolesterol$kolesterol[kadar_kolesterol$jenis_obat=="C" &  
kadar_kolesterol$jenis_kelamin=="laki-laki"])
```

```
# Menampilkan
```

```
# Menampilkan ringkasan statistik data kadar kolesterol pasien yang diberikan obat C dan jenis kelamin perempuan
```

```
summary(kadar_kolesterol$kolesterol[kadar_kolesterol$jenis_obat=="C" &  
kadar_kolesterol$jenis_kelamin=="perempuan"])
```

```
# Melakukan analisis variansi dua arah
```

```
two_way_anova <- aov(kolesterol ~ jenis_obat + jenis_kelamin, data=kadar_kolesterol)
```

```
summary(two_way_anova)
```

Dalam contoh program di atas, kita menampilkan ringkasan statistik dari data kadar kolesterol pasien yang diberikan tiga jenis obat penurun kolesterol dan antara jenis kelamin pasien dengan menggunakan fungsi `summary()` dari package `base`. Kemudian, kita melakukan analisis variansi dua arah pada data tersebut dengan menggunakan fungsi `aov()`.

Referensi:

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

Bab 12

Analisis Regresi

12.1 Pengertian Analisis Regresi

Analisis regresi adalah suatu metode analisis statistik yang digunakan untuk mempelajari hubungan antara satu variabel bebas (independen) dengan satu variabel terikat (dependen) yang bersifat kontinu. Analisis regresi digunakan untuk mengukur seberapa besar pengaruh variabel bebas terhadap variabel terikat dan untuk memprediksi nilai variabel terikat berdasarkan nilai variabel bebas yang diketahui.

Referensi:

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

12.2 Analisis Regresi Linear Sederhana

Analisis regresi linear sederhana adalah suatu metode analisis statistik yang digunakan untuk mempelajari hubungan antara satu variabel bebas (independen) dengan satu variabel terikat (dependen) yang bersifat kontinu secara linier. Regresi linear sederhana digunakan untuk mengukur seberapa besar pengaruh variabel bebas terhadap variabel terikat dan untuk memprediksi nilai variabel terikat berdasarkan nilai variabel bebas yang diketahui.

Referensi:

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

12.3 Contoh Analisis Regresi Linear Sederhana

Misalkan kita ingin mengetahui apakah terdapat hubungan antara usia dan tekanan darah sistolik pada populasi dewasa. Kita memiliki sampel 50 orang dewasa dan kita ingin menguji apakah terdapat hubungan antara usia dan tekanan darah sistolik pada sampel tersebut.

Program R:

Berikut adalah contoh perintah untuk melakukan analisis regresi linear sederhana pada data usia dan tekanan darah sistolik pada populasi dewasa:

```
# Membaca dataset tekanan_darah
tekanan_darah <- read.csv("data_tekanan_darah.csv")

# Menampilkan plot hubungan antara usia dan tekanan darah sistolik
plot(tekanan_darah$usia, tekanan_darah$tekanan_darah_sistolik, main="Hubungan Usia dengan
Tekanan Darah Sistolik", xlab="Usia", ylab="Tekanan Darah Sistolik")

# Melakukan analisis regresi linear sederhana
regresi_linier <- lm(tekanan_darah_sistolik ~ usia, data=tekanan_darah)
```

```
summary(regresi_linier)
```

Dalam contoh program di atas, kita menampilkan plot hubungan antara usia dan tekanan darah sistolik pada populasi dewasa dengan menggunakan fungsi `plot()` dari package `graphics`.

Bab 13

Analisis Regresi Linear Berganda

13.1 Pengertian Analisis Regresi Linear Berganda

Analisis regresi linear berganda adalah suatu metode analisis statistik yang digunakan untuk mempelajari hubungan antara dua atau lebih variabel bebas (independen) dengan satu variabel terikat (dependen) yang bersifat kontinu secara linier. Analisis regresi linear berganda digunakan untuk mengukur seberapa besar pengaruh setiap variabel bebas terhadap variabel terikat dan untuk memprediksi nilai variabel terikat berdasarkan nilai variabel bebas yang diketahui.

Referensi:

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

13.2 Analisis Regresi Linear Berganda dengan Dua Variabel Bebas

Analisis regresi linear berganda dengan dua variabel bebas adalah suatu metode analisis statistik yang digunakan untuk mempelajari hubungan antara dua variabel bebas (independen) dengan satu variabel terikat (dependen) yang bersifat kontinu secara linier. Analisis regresi linear berganda dengan dua variabel bebas digunakan untuk mengukur seberapa besar pengaruh setiap variabel bebas terhadap variabel terikat dan untuk memprediksi nilai variabel terikat berdasarkan nilai variabel bebas yang diketahui.

Referensi:

Agresti, A., & Finlay, B. (2009). *Statistical methods for the social sciences* (4th ed.). Pearson Education.

Sullivan, L. M. (2011). *Essentials of biostatistics in public health* (2nd ed.). Jones & Bartlett Learning.

13.3 Contoh Analisis Regresi Linear Berganda

Misalkan kita ingin mengetahui apakah terdapat hubungan antara usia dan indeks massa tubuh (IMT) dengan kadar gula darah pada populasi dewasa. Kita memiliki sampel 50 orang dewasa dan kita ingin menguji apakah terdapat hubungan antara usia, IMT, dan kadar gula darah pada sampel tersebut.

Program R:

Berikut adalah contoh perintah untuk melakukan analisis regresi linear berganda pada data usia, IMT, dan kadar gula darah pada populasi dewasa:

```
# Membaca dataset kadar_gula_darah
kadar_gula_darah <- read.csv("data_kadar_gula_darah.csv")

# Menampilkan plot hubungan antara usia, IMT, dan kadar gula darah
pairs(kadar_gula_darah, main="Plot Hubungan Usia, IMT, dan Kadar Gula Darah")

# Melakukan analisis regresi linear berganda
regresi_linier_berganda <- lm(kadar_gula_darah ~ usia + imt, data=kadar_gula_darah)
```



```
summary(regresi_linier_berganda)
```

Dalam contoh program di atas, kita menampilkan plot hubungan antara usia, IMT, dan kadar gula darah pada populasi dewasa dengan menggunakan fungsi `pairs()` dari package `graphics`. Kemudian,

14.1 Pengertian Analisis Regresi Logistik

Analisis regresi logistik adalah suatu metode analisis statistik yang digunakan untuk mempelajari hubungan antara satu atau lebih variabel bebas (independen) dengan satu variabel terikat (dependen) yang bersifat biner atau kategorikal. Analisis regresi logistik digunakan untuk mengukur seberapa besar pengaruh variabel bebas terhadap kemungkinan terjadinya suatu kejadian pada variabel terikat dan untuk memprediksi kemungkinan terjadinya suatu kejadian pada variabel terikat berdasarkan nilai variabel bebas yang diketahui.

Referensi:

Agresti, A. (2018). *An introduction to categorical data analysis* (3rd ed.). Wiley.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). John Wiley & Sons.

14.2 Analisis Regresi Logistik Binomial

Analisis regresi logistik binomial adalah suatu metode analisis statistik yang digunakan untuk mempelajari hubungan antara satu atau lebih variabel bebas (independen) dengan satu variabel terikat (dependen) yang bersifat biner. Analisis regresi logistik binomial digunakan untuk mengukur seberapa besar pengaruh variabel bebas terhadap kemungkinan terjadinya suatu kejadian pada variabel terikat dan untuk memprediksi kemungkinan terjadinya suatu kejadian pada variabel terikat berdasarkan nilai variabel bebas yang diketahui.

Referensi:

Agresti, A. (2018). *An introduction to categorical data analysis* (3rd ed.). Wiley.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). John Wiley & Sons.

14.3 Contoh Analisis Regresi Logistik Binomial

Misalkan kita ingin mengetahui apakah faktor usia dan jenis kelamin mempengaruhi kemungkinan seseorang untuk mengidap penyakit diabetes. Kita memiliki sampel 200 orang dewasa dan kita ingin menguji apakah usia dan jenis kelamin memiliki pengaruh signifikan terhadap kemungkinan terjadinya penyakit diabetes pada sampel tersebut.

Program R:

Berikut adalah contoh perintah untuk melakukan analisis regresi logistik binomial pada data usia, jenis kelamin, dan status diabetes pada populasi dewasa:

```
# Membaca dataset diabetes
```

```
diabetes <- read.csv("data_diabetes.csv")
```

```
# Menampilkan distribusi status diabetes berdasarkan jenis kelamin
```

```
table(diabetes$status_diabetes, diabetes$jenis_kelamin)
```

```
# Melakukan analisis regresi logistik binomial
```

```
regresi_logistik_binomial <- glm(status_diabetes ~ usia + jenis_kelamin, family="binomial",  
data=diabetes)
```

```
summary(regresi_logistik_binomial)
```

Dalam contoh program di atas, kita menampilkan distribusi status diabetes pada populasi dewasa berdasarkan jenis kelamin dengan menggunakan fungsi `table()`. Kemudian, kita melakukan analisis regresi logistik binomial pada data usia,

Bab 15

Analisis Kelangsungan Hidup (Survival Analysis)

15.1 Pengertian Analisis Kelangsungan Hidup

Analisis kelangsungan hidup (survival analysis) adalah suatu metode analisis statistik yang digunakan untuk mempelajari waktu yang diperlukan untuk suatu kejadian terjadi pada suatu populasi. Analisis kelangsungan hidup digunakan untuk mengukur risiko terjadinya suatu kejadian pada suatu waktu tertentu dan untuk memprediksi waktu terjadinya suatu kejadian berdasarkan faktor-faktor risiko yang diketahui.

Referensi:

Kleinbaum, D. G., & Klein, M. (2012). *Survival analysis: A self-learning text* (3rd ed.). Springer Science & Business Media.

Hosmer Jr, D. W., Lemeshow, S., & May, S. (2008). *Applied survival analysis: Regression modeling of time-to-event data* (2nd ed.). John Wiley & Sons.

15.2 Analisis Kelangsungan Hidup dengan Data Censored

Analisis kelangsungan hidup dengan data censored adalah suatu metode analisis statistik yang digunakan untuk mempelajari waktu yang diperlukan untuk suatu kejadian terjadi pada suatu populasi, namun terdapat beberapa individu pada populasi yang belum mengalami kejadian tersebut sampai pada saat pengamatan berakhir. Analisis kelangsungan hidup dengan data censored digunakan untuk mengukur risiko terjadinya suatu kejadian pada suatu waktu tertentu dan untuk memprediksi waktu terjadinya suatu kejadian berdasarkan faktor-faktor risiko yang diketahui.

Referensi:

Kleinbaum, D. G., & Klein, M. (2012). *Survival analysis: A self-learning text* (3rd ed.). Springer Science & Business Media.

Hosmer Jr, D. W., Lemeshow, S., & May, S. (2008). *Applied survival analysis: Regression modeling of time-to-event data* (2nd ed.). John Wiley & Sons.

15.3 Contoh Analisis Kelangsungan Hidup dengan Data Censored

Misalkan kita ingin mengetahui faktor-faktor risiko yang mempengaruhi kelangsungan hidup pasien kanker. Kita memiliki data waktu kelangsungan hidup (dalam bulan) dan status hidup/mati pasien kanker beserta beberapa faktor risiko lainnya pada sampel 100 pasien kanker.

Program R:

Berikut adalah contoh perintah untuk melakukan analisis kelangsungan hidup dengan data censored pada data waktu kelangsungan hidup, status hidup/mati, dan beberapa faktor risiko lainnya pada sampel pasien kanker:

```
# Membaca dataset waktu_kelangsungan_hidup
waktu_kelangsungan_hidup <- read.csv("data_waktu_kelangsungan_hidup.csv")
# Menampilkan plot survival
```

```
library(survival)

surv_object <- Surv(waktu_kelangsungan_hidup$waktu_kelangsungan_hidup,
waktu_kelangsungan_hidup$status_hidup_mati)

summary(surv_object)

plot(survfit(surv_object ~ 1), xlab="Bulan", ylab="Persentase Kelangsungan Hidup", main="Plot Survival")

# Melakukan analisis kelangsungan hidup

regresi_kelangsungan_hid
```

Bab 16: Analisis Faktor

16.1 Pengertian Analisis Faktor

Analisis faktor adalah suatu metode analisis statistik yang digunakan untuk mengidentifikasi pola-pola dalam data dengan melakukan reduksi dimensi. Analisis faktor digunakan untuk mengelompokkan variabel-variabel yang saling berkaitan dalam beberapa faktor utama yang dapat menjelaskan variasi dalam data.

Referensi:

Hair Jr, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Prentice Hall.

Stevens, J. P. (2012). *Applied multivariate statistics for the social sciences* (5th ed.). Routledge.

16.2 Analisis Faktor Eksploratori

Analisis faktor eksploratori adalah suatu metode analisis statistik yang digunakan untuk mengidentifikasi pola-pola dalam data dengan melakukan reduksi dimensi. Analisis faktor eksploratori digunakan untuk mengelompokkan variabel-variabel yang saling berkaitan dalam beberapa faktor utama yang dapat menjelaskan variasi dalam data.

Referensi:

Hair Jr, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Prentice Hall.

Stevens, J. P. (2012). *Applied multivariate statistics for the social sciences* (5th ed.). Routledge.

16.3 Contoh Analisis Faktor Eksploratori

Misalkan kita memiliki data tentang kebiasaan makan dan gaya hidup sehat pada 100 orang dewasa dan ingin mengidentifikasi faktor-faktor utama yang mempengaruhi kebiasaan makan sehat.

Program R:

Berikut adalah contoh perintah untuk melakukan analisis faktor eksploratori pada data kebiasaan makan dan gaya hidup sehat pada 100 orang dewasa:

```
# Membaca dataset kebiasaan_makan
kebiasaan_makan <- read.csv("data_kebiasaan_makan.csv")
# Melakukan analisis faktor eksploratori
library(psych)
faktor_eksploratori <- fa(kebiasaan_makan[,3:9], nfactors=3, rotate="varimax")
print(faktor_eksploratori)
```

Dalam contoh program di atas, kita melakukan analisis faktor eksploratori pada data kebiasaan makan dan gaya hidup sehat pada 100 orang dewasa dengan menggunakan fungsi `fa()` dari paket `psych` pada R. Kita memilih tiga faktor utama dengan menggunakan opsi `nfactors=3` dan menggunakan metode rotasi `varimax` dengan opsi `rotate="varimax"`.

Bab 17

Analisis Klaster

17.1 Pengertian Analisis Klaster

Analisis klaster adalah suatu metode analisis statistik yang digunakan untuk mengelompokkan objek-objek dalam suatu populasi berdasarkan kemiripan karakteristik atau sifat-sifat tertentu. Analisis klaster digunakan untuk mengidentifikasi pola-pola dalam data dan untuk mempermudah pengambilan keputusan.

Referensi:

Hair Jr, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Prentice Hall.

Everitt, B., Landau, S., & Leese, M. (2001). *Cluster analysis* (4th ed.). Arnold.

17.2 Analisis Klaster Hirarki

Analisis klaster hirarki adalah suatu metode analisis statistik yang digunakan untuk mengelompokkan objek-objek dalam suatu populasi berdasarkan kemiripan karakteristik atau sifat-sifat tertentu dengan membangun sebuah dendrogram. Analisis klaster hirarki digunakan untuk mengidentifikasi pola-pola dalam data dan untuk mempermudah pengambilan keputusan.

Referensi:

Hair Jr, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Prentice Hall.

Everitt, B., Landau, S., & Leese, M. (2001). *Cluster analysis* (4th ed.). Arnold.

17.3 Contoh Analisis Klaster Hirarki

Misalkan kita memiliki data tentang pendapatan dan pengeluaran rumah tangga pada 100 keluarga dan ingin mengelompokkan keluarga-keluarga tersebut berdasarkan pola pendapatan dan pengeluarannya.

Program R:

Berikut adalah contoh perintah untuk melakukan analisis klaster hirarki pada data pendapatan dan pengeluaran rumah tangga pada 100 keluarga:

```
# Membaca dataset pendapatan_pengeluaran
pendapatan_pengeluaran <- read.csv("data_pendapatan_pengeluaran.csv")

# Melakukan analisis klaster hirarki
jarak_euclidean <- dist(pendapatan_pengeluaran[,2:3], method="euclidean")
klaster_hirarki <- hclust(jarak_euclidean, method="ward.D2")
plot(klaster_hirarki)
```

Dalam contoh program di atas, kita melakukan analisis kluster hirarki pada data pendapatan dan pengeluaran rumah tangga pada 100 keluarga dengan menggunakan fungsi `hclust()` pada R. Kita menggunakan jarak Euclidean dengan opsi `method="euclidean"` dan metode ward.D2 dengan opsi `method="ward.D2"`. Selanjutnya, kita menampilkan dendrogram dengan menggunakan fungsi `plot()`.

Bab 18

Regresi Logistik

18.1 Pengertian Regresi Logistik

Regresi logistik adalah suatu metode analisis statistik yang digunakan untuk menghubungkan variabel dependen biner dengan satu atau lebih variabel independen. Regresi logistik digunakan untuk memodelkan hubungan antara variabel dependen biner dan variabel independen serta untuk memprediksi nilai variabel dependen.

Referensi:

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley.

Kleinbaum, D. G., & Klein, M. (2010). *Logistic regression: A self-learning text* (3rd ed.). Springer.

18.2 Regresi Logistik Binomial

Regresi logistik binomial adalah suatu metode analisis statistik yang digunakan untuk menghubungkan variabel dependen biner dengan satu atau lebih variabel independen. Regresi logistik binomial digunakan untuk memodelkan hubungan antara variabel dependen biner dan variabel independen serta untuk memprediksi nilai variabel dependen.

Referensi:

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley.

Kleinbaum, D. G., & Klein, M. (2010). *Logistic regression: A self-learning text* (3rd ed.). Springer.

18.3 Contoh Regresi Logistik Binomial

Misalkan kita memiliki data tentang keberhasilan atau kegagalan siswa dalam ujian masuk universitas berdasarkan nilai ujian sebelumnya dan ingin mengetahui faktor-faktor yang mempengaruhi keberhasilan atau kegagalan tersebut.

Program R:

Berikut adalah contoh perintah untuk melakukan regresi logistik binomial pada data keberhasilan atau kegagalan siswa dalam ujian masuk universitas berdasarkan nilai ujian sebelumnya:

```
# Membaca dataset keberhasilan_siswa
keberhasilan_siswa <- read.csv("data_keberhasilan_siswa.csv")

# Melakukan regresi logistik binomial
model_logistik <- glm(berhasil ~ ujian_sebelumnya, data=keberhasilan_siswa,
family=binomial(link="logit"))

summary(model_logistik)
```

Dalam contoh program di atas, kita melakukan regresi logistik binomial pada data keberhasilan atau kegagalan siswa dalam ujian masuk universitas berdasarkan nilai ujian sebelumnya dengan menggunakan fungsi `glm()` pada R. Kita menggunakan variabel `berhasil` sebagai variabel dependen dan variabel `ujian_sebelumnya` sebagai variabel independen. Selanjutnya, kita menggunakan distribusi

binomial dan fungsi logit dengan opsi `family=binomial(link="logit")`. Untuk menampilkan hasil regresi logistik, kita menggunakan fungsi `summary()`.

Jika Anda ingin melihat hasil prediksi dari model regresi logistik binomial, Anda dapat menggunakan fungsi `predict()`. Misalnya, untuk memprediksi probabilitas keberhasilan ujian masuk universitas pada nilai ujian sebelumnya 75, kita dapat menggunakan perintah berikut:

```
# Memprediksi probabilitas keberhasilan pada nilai ujian
```

Bab 19

Analisis Regresi

19.1 Pengertian Analisis Regresi

Analisis regresi adalah suatu metode analisis statistik yang digunakan untuk menghubungkan variabel dependen dengan satu atau lebih variabel independen. Analisis regresi digunakan untuk memodelkan hubungan antara variabel dependen dan variabel independen serta untuk memprediksi nilai variabel dependen.

Referensi:

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2004). Applied linear regression models (4th ed.). McGraw-Hill.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to linear regression analysis (5th ed.). Wiley.

19.2 Regresi Linear Sederhana

Regresi linear sederhana adalah suatu metode analisis statistik yang digunakan untuk menghubungkan variabel dependen dengan satu variabel independen. Regresi linear sederhana digunakan untuk memodelkan hubungan antara variabel dependen dan variabel independen serta untuk memprediksi nilai variabel dependen.

Referensi:

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2004). Applied linear regression models (4th ed.). McGraw-Hill.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to linear regression analysis (5th ed.). Wiley.

19.3 Contoh Regresi Linear Sederhana

Misalkan kita memiliki data tentang berat badan dan tinggi badan sejumlah individu dan ingin mengetahui hubungan antara berat badan dan tinggi badan tersebut.

Program R:

Berikut adalah contoh perintah untuk melakukan regresi linear sederhana pada data berat badan dan tinggi badan sejumlah individu:

```
# Membaca dataset berat_badan_tinggi_badan
berat_badan_tinggi_badan <- read.csv("data_berat_badan_tinggi_badan.csv")
# Melakukan regresi linear sederhana
model_regresi <- lm(berat_badan ~ tinggi_badan, data=berat_badan_tinggi_badan)
summary(model_regresi)
```

Dalam contoh program di atas, kita melakukan regresi linear sederhana pada data berat badan dan tinggi badan sejumlah individu dengan menggunakan fungsi `lm()` pada R. Kita menggunakan variabel

berat_badan sebagai variabel dependen dan variabel tinggi_badan sebagai variabel independen. Untuk menampilkan hasil regresi linear, kita menggunakan fungsi summary().

Jika Anda ingin melihat hasil prediksi dari model regresi linear sederhana, Anda dapat menggunakan fungsi predict(). Misalnya, untuk memprediksi berat badan pada tinggi badan 170, kita dapat menggunakan perintah berikut:

```
# Memprediksi berat badan pada tinggi badan 170
```

```
new_data <- data.frame(tinggi_badan=170)
```

```
predict(model_regresi, newdata=new_data)
```

Tugas Akhir:

Anda diminta untuk melakukan analisis data pada dataset "Iris" yang dapat diakses pada R dengan menggunakan library "datasets". Dataset ini berisi data pengukuran pada tiga jenis tanaman iris: Iris Setosa, Iris Versicolour, dan Iris Virginica. Anda diminta untuk melakukan analisis deskriptif, analisis eksplorasi data, serta membangun model klasifikasi menggunakan metode regresi logistik untuk memprediksi jenis tanaman iris berdasarkan variabel pengukuran yang tersedia.

Contoh Penyelesaian:

Berikut adalah contoh penyelesaian dari tugas akhir yang diberikan:

```
# Memuat library dan dataset
```

```
library(datasets)
```

```
data(iris)
```

```
# Analisis Deskriptif
```

```
summary(iris)
```

```
# Melihat distribusi setiap variabel
```

```
hist(iris$Sepal.Length, main = "Distribusi Sepal.Length")
```

```
hist(iris$Sepal.Width, main = "Distribusi Sepal.Width")
```

```
hist(iris$Petal.Length, main = "Distribusi Petal.Length")
```

```
hist(iris$Petal.Width, main = "Distribusi Petal.Width")
```

```
# Analisis Eksplorasi Data
```

```
# Membuat plot hubungan antara variabel
```

```
plot(iris[,1:4], col=iris$Species, pch=19)
```

```
# Membangun Model Klasifikasi menggunakan Regresi Logistik
```

```
# Memecah dataset menjadi data training dan data testing
```

```
set.seed(1234)
```

```
ind <- sample(2, nrow(iris), replace=TRUE, prob=c(0.7, 0.3))
```

```
training_data <- iris[ind==1,]
```

```
testing_data <- iris[ind==2,]
```

```
# Melakukan regresi logistik
```

```
model_logistik <- glm(Species ~ Sepal.Length + Sepal.Width + Petal.Length + Petal.Width,  
data=training_data, family="binomial")
```

```
summary(model_logistik)
```

```
# Melakukan prediksi
predicted_values <- predict(model_logistik, newdata=testing_data, type="response")
predicted_classes <- ifelse(predicted_values > 0.5, "Iris-virginica", "Iris-setosa")
# Menghitung akurasi
accuracy <- mean(predicted_classes == testing_data$Species)
cat("Akurasi Model Regresi Logistik adalah:", round(accuracy*100), "%")
```

Dalam contoh program di atas, pertama-tama kita memuat library dan dataset "Iris" pada R. Kemudian kita melakukan analisis deskriptif dan eksplorasi data pada dataset tersebut. Setelah itu, kita membagi dataset menjadi data training dan data testing. Kemudian kita menggunakan fungsi `glm()` pada R untuk melakukan regresi logistik dengan variabel `Species` sebagai variabel dependen dan variabel pengukuran lainnya sebagai variabel independen. Setelah mendapatkan model regresi logistik, kita melakukan prediksi pada data testing menggunakan fungsi `predict()` dan menghitung akurasi dari model tersebut.