

Laporan Akhir Praktikum

Statistika Inferensial Lanjut Praktik

Dosen Pengampu: Selfi Artika S.Si., M.Si.



Disusun oleh:

Lathif Ramadhan (5231811022)

**PROGRAM STUDI SAINS DATA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS TEKNOLOGI YOGYAKARTA
YOGYAKARTA
2024**

Pendahuluan

Statistika inferensial merupakan cabang ilmu statistik yang berperan penting dalam pengambilan keputusan berdasarkan data sampel. Melalui berbagai metode dan teknik pengujian hipotesis, statistika inferensial memungkinkan para peneliti untuk membuat generalisasi tentang populasi, mengidentifikasi hubungan antar variabel, serta merumuskan prediksi yang lebih akurat. Dalam konteks ini, bahasa pemrograman R telah menjadi alat utama untuk menganalisis data secara efektif, dengan menyediakan berbagai paket dan fungsi yang dirancang untuk mendukung proses analisis statistik.

Laporan praktikum ini disusun sebagai bagian dari pembelajaran mata kuliah **Statistika Inferensial Lanjut Praktik**, dengan fokus pada penerapan teknik pengujian hipotesis menggunakan R. Setiap bab dalam laporan ini mencakup pengenalan teori, langkah-langkah praktikum, hingga interpretasi hasil, sehingga memberikan pemahaman menyeluruh tentang penerapan statistika inferensial dalam berbagai kasus.

Dalam laporan ini, materi yang dibahas meliputi pengujian hipotesis deskriptif, komparatif, hingga analisis tren dan time series. Setiap bab disusun berdasarkan pertemuan kelas yang terstruktur, dimulai dari pengujian sederhana seperti uji binomial dan uji Chi-Square, hingga analisis kompleks seperti pemodelan ARIMA pada data time series. Selain itu, laporan ini juga dilengkapi dengan studi kasus nyata yang memperkaya wawasan mahasiswa terhadap aplikasi statistika dalam kehidupan sehari-hari.

Melalui laporan ini, diharapkan pembaca dapat memahami konsep-konsep kunci dalam statistika inferensial, mengevaluasi data secara kritis, serta memanfaatkan kemampuan analisis statistik untuk pengambilan keputusan berbasis data yang lebih baik. Seluruh langkah analisis disertai dengan penjelasan yang rinci untuk memastikan pemahaman yang mendalam serta aplikatif.

Daftar Isi

Pendahuluan	II
Daftar Isi	III
BAB 1: PERTEMUAN 1-2-3 Uji Hipotesis Deskriptif 1 Sampel	1
Laporan 1: Uji Binomial	1
PPT Hal. 11 Contoh 1	1
PPT Hal. 17 Contoh 2	3
PPT Hal. 11 Contoh 3	5
PPT Hal. 29 Latihan 1	7
PPT Hal. 30 Latihan 2	9
PPT Hal. 31 Latihan 3	11
PPT Hal. 32 Latihan 4	13
PPT Hal. 33 Latihan 5	15
Laporan 2: Uji Chi Square	17
PPT Hal. 37 Contoh 1	17
PPT Hal. 41 Contoh 2	19
PPT Hal. 47 Latihan 7	21
PPT Hal. 46 Latihan 8	23
Laporan 3: Uji Runs (Uji Kerandoman)	25
PPT Hal. 55 Contoh 1	25
PPT Hal. 59 Contoh 2	28
PPT Hal. 64 Latihan 1	30
PPT Hal. 65 Latihan 2	32
PPT Hal. 66 Latihan 3	34
Modul Praktikum Hal. 19 Latihan 4	36
BAB 2: PERTEMUAN 4 Uji Hipotesis Komparatif 2 Sampel Dependen...	38
Laporan 4: Uji Mc Nemar	38
Contoh 1 PPT Hal. 8	38
Contoh 2 PPT Hal. 12	40
Latihan 1 PPT Hal. 16	42
Latihan 2 PPT Hal. 21	44

Laporan 5: Uji Tanda	46
Contoh 1 PPT Hal. 27	46
Contoh 2 PPT Hal. 35	51
Latihan 1 PPT Hal. 40.....	55
Laporan 6: Uji Wilcoxon	58
Contoh $n < 25$ PPT Hal. 46.....	58
Contoh $n > 25$ PPT Hal. 51.....	60
Latihan 1 PPT Hal. 57.....	64
Latihan 2 PPT Hal. 59.....	66
Latihan 3 PPT Hal. 61.....	68
BAB 3: PERTEMUAN 5-6 Uji Hipotesis Komparatif 2 Sampel Independen	71
Laporan 7: Uji Fisher	71
Contoh 1 PPT Pertemuan 4,5 Hal 10	71
Contoh 2 PPT Pertemuan 4,5 Hal 17	73
Soal Latihan PPT Pertemuan 4,5 Hal 21	75
Modul Praktikum Latihan 1 Hal 38	77
Laporan 8: Uji Chi Square	79
Contoh 1 PPT Pertemuan 4,5 Hal 26	79
Contoh 2 PPT Pertemuan 4,5 Hal 31	81
Soal Latihan PPT Pertemuan 4,5 Hal 37	83
Laporan 9: Uji Median.....	84
Contoh 1 PPT Pertemuan 5,6 Hal 6	84
Contoh 2 PPT Pertemuan 5,6 Hal 12	87
Latihan 1 PPT Pertemuan 5,6 Hal 18.....	90
Latihan 2 PPT Pertemuan 5,6 Hal 19.....	93
Laporan 10: Uji Mann-Whitney.....	96
Contoh 1 PPT Pertemuan 5,6 Hal 26	96
Contoh 2 PPT Pertemuan 5,6 Hal 34	98
Latihan Modul No. 2 Hal. 45	100
BAB 4: PERTEMUAN 6-7 Uji Hipotesis Komparatif k Sampel Dependen dan Independen	102
1. Uji Cochran (Q-test) (<i>Dependen</i>).....	102

PPT Pertemuan 6,7 Hal 5	102
2. Uji Friedman (Anava Ranking Two Way) (<i>Dependen</i>)	104
PPT Pertemuan 6,7 Hal 14.....	104
3. Uji Chi-Square k Sampel (<i>Independen</i>)	106
PPT Pertemuan 6,7 Hal 25.....	106
4. Uji Median Extension (Perluasan Uji Median) (<i>Independen</i>)	108
PPT Pertemuan 6,7 Hal 37.....	108
5. Anova One Way Kruskal-Wallis (<i>Independen</i>).....	111
PPT Pertemuan 6,7 Hal 50.....	111
BAB 5: PERTEMUAN 10 (Laporan 11: Angka Indeks).....	113
Pendahuluan Materi Angka Indeks	113
Langka Awal: Menginstal Package IndexNumber.....	114
Bagian 1: Pengantar Angka Indeks	115
Bagian 2: Jenis-Jenis Angka Indeks.....	117
1. Berdasarkan Penggunaan	117
2. Berdasarkan Cara Penentuan	119
Bagian 3: Metode Penghitungan Angka Indeks.....	122
a. Indeks Harga Relatif Sederhana.....	122
b. Indeks Kuantitas Relatif Sederhana	123
c. Angka Indeks Agregat Sederhana	124
d. Angka Indeks Kuantitas Agregat Sederhana.....	125
e. Angka Indeks Nilai Kuantitas Agregat Sederhana.....	126
f. Angka Indeks Harga Tertimbang	127
Bagian 4: Indeks Berantai.....	130
Bagian 5: Pergeseran Tahun Dasar	130
Bagian 6: Merangkai Angka Indeks (Splicing).....	132
Bagian 7: Angka Indeks untuk Proses Deflasi	134
Bagian 8: Daya Beli Mata Uang	135
BAB 6: PERTEMUAN 11 (Laporan 12: Time Series)	136
DEFINISI UMUM	136
TUJUAN	137
1. Mengenali Pola	137
2. Peramalan (Forecasting)	137

3. Memudahkan Pengambilan Keputusan.....	137
4. Pemahaman Penyebab Tren	137
5. Analisis Intervensi	137
Persyaratan data	137
LANGKAH PENGUJIAN TIME SERIES.....	138
LANGKAH PENGUJIAN TIME SERIES SECARA MANUAL.....	138
LANGKAH PENGUJIAN TIME SERIES MENGGUNAKAN RSTUDIO	138
KOMPONEN TIME SERIES.....	140
RSTUDIO VARIANSI TREN	140
RSTUDIO VARIANSI SIKLIS	142
RSTUDIO VARIANSI MUSIM	143
RSTUDIO VARIANSI MUSIM	143
RSTUDIO VARIANSI TAK BERATURAN	145
MOVING AVERAGE.....	147
1. SIMPLE MOVING AVERAGE (SMA)	147
2. WEIGHTED MOVING AVERAGE (WMA).....	150
3. EXPONENTIAL MOVING AVERAGE (EMA).....	154
BAB 7: PERTEMUAN 12 (Laporan 13: Analisis Tren).....	156
Studi Kasus 1: Data Genap Kelompok Genap(Asosiasi Industri Sepeda Motor Indonesia/AISI 2023).....	156
1. Memasukkan dan Memuat Data	156
2. Visualisasi Data Awal (2 Kolom, 2 Grafik)	159
3. Tren Linear.....	164
4. Tren Non-Linear.....	172
5. Tabel Hasil Keseluruhan	174
6. Menghitung Measure of Accuracy	175
Studi Kasus 2: Data Genap Kelompok Ganjil(Penjualan Mobil di Indonesia Tahun 2003-2012)	178
1. Memasukkan dan memuat data.....	178
2. Visualisasi data penjualan awal.....	179
3. Metode Semi Average	182
4. Visualisasi Data Awal dan Prediksi.....	183
5. Menghitung Measure of Accuracy	186
Studi Kasus 3: Data Ganjil (Data Jumlah Mahasiswa Tahun 1980-1988).....	188

1. Memasukkan Data	188
2. Visualisasi Data Awal.....	189
3. Analisis Tren dengan Regresi Linier.....	192
4. Analisis Tren dengan Semiavarage	194
5. Metode Least Squares (Linear).....	195
6. Metode Quadratic.....	196
7. Metode Exponential	197
8. Visualisasi Perbandingan Metode	197
BAB 8: PERTEMUAN 13 (Laporan: 14 Konsep Dasar Pemodelan Time Series)	201
Install Library yang akan Digunakan.....	201
Mengecek stasionaritas dan cara mengatasi data tidak stasioner.....	203
1. Load Dataset	203
2. Menfilter hanya stock A.....	207
3. Pemeriksaaan Kestasioneran ADF dari Data Time Series	211
4. Pemeriksaaan Kestasioneran ADF dari data Bond A.....	214
5. ACF dan PACF pada Data Time Series	217
6. Analisis dan Pemodelan ARIMA	219
BAB 9: PERTEMUAN 14 (Model AR, MA, ARMA, dan ARIMA).....	223
1. Mengimpor Library yang diperlukan.....	223
2. ACF dan PACF untuk AR(1)	224
3. Model AR(1) pada Data Sunspots.....	228
a. Memasukkan Data.....	228
b. Uji Stasioneritas.....	228
c. AIC	229
d. Diagnostik Residual.....	231
e. Prediksi AR	233
4. Model MA(2).....	235
a. Model MA(1) dengan Data Sunspots: Uji Stasioneritas	236
b. AIC	236
c. Cek Residual untk MA(1)	238
d. Prediksi MA(1)	239
5. Model ARMA(1,1).....	241

a. Model ARMA (1,1) dengan Data Sunspots.....	243
b. AIC	244
c. Cek Residual	245
d. Prediksi ARMA (1,1)	247
6. Model ARIMA Stasioner	249
a. Membaca Data	249
b. Mengubah Data Menjadi Data Time series dan Menggambar Grafik Data Contoh..	249
c. Uji Stasioner.....	250
d. Spesifikasi Model.....	251
e. Menentukan Model terbaik Berdasarkan AR1 dan MA1	254
f. Peramalan Model.....	257
7. Model ARIMA Non-Stasioner	259
a. Memuat dan Memvisualisasikan Data	259
b. Uji Stasioner.....	260
c. Proses Differencing karena Data Tidak Stasioner.....	261
d. Uji stasioner untuk data diff1	263
e. Proses diferensing ke 2.....	264
f. Uji Stasioner untuk diff2	266
g. Penentuan Model	267
h. Uji Signifikansi Model (Koefisien)	268
i. Menentukan Model Terbaik	269
j. Diagnostik & Akurasi Model.....	271
k. Peramalan model ARIMA 021 menggunakan Forecasting.....	273
8. Latihan Soal	276
a. Cara 1	276
b. Cara 2	276

BAB 1: PERTEMUAN 1-2-3 Uji Hipotesis

Deskriptif 1 Sampel

Laporan 1: Uji Binomial

PPT Hal. 11 Contoh 1

Contoh 1 : untuk $n < 25$

Sebuah perusahaan otomotif memproduksi 2 jenis mobil mini bus yaitu yang berbahan bakar bensin dan solar. Perusahaan tersebut ingin mengetahui apakah masyarakat lebih senang mobil berbahan bakar solar atau bensin. Berdasarkan 24 anggota sampel yang dipilih secara random ternyata 14 orang memilih mobil berbahan bakar bensin dan 10 orang memilih mobil berbahan bakar solar. (gunakan $\alpha = 0,05$)

Penyelesaian

1. Hipotesis:

- **H0:** Frekuensi Masyarakat yang memilih menggunakan mobil berbahan bakar solar *tidak berbeda/sama*. ($P_1 = P_2 = 0,5$)
- **H1:** Frekuensi Masyarakat yang memilih menggunakan mobil berbahan bakar solar *tidak sama*. ($P_1 \neq P_2 \neq 0,5$)

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic : $n = 24$ dan $x = 10$ (ambil yg terkecil antara 10 dengan 14)

Menggunakan R diperoleh nilai p-value sbb :

```
> binom.test(10, 24, 0.5)

Exact binomial test

data: 10 and 24
number of successes = 10, number of trials = 24, p-value = 0.5413
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
0.2210969 0.6335694
sample estimates:
probability of success
0.4166667
```

4. Kriteria Penolakan

H_0 ditolak jika $P(x < \alpha)$

Karena p-value = 0,5412 > $\alpha = 0,01$ maka **H_0 tidak ditolak**. Artinya, frekuensi masyarakat dalam memilih dua jenis mobil tersebut **relative sama**.

5. Kesimpulan

Jadi, terdapat kecenderungan yang **sama** di masyarakat dalam memilih mobil berbahan bakar bensin dan solar.

PPT Hal. 17 Contoh 2

Contoh 2 : Untuk $n < 25$

Di sebuah kecamatan, telah dilakukan imunisasi campak tahap 1 pada balita. Dari pelaksanaan imunisasi tersebut terdapat 2 kemungkinan untuk terjadinya demam dan tidak. Dari 20 balita yang diimunisasi, terdapat 13 balita yang tidak mengalami demam dan 7 balita mengalami demam. Bagaimana keputusan hipotesisnya? Jika derajat kepercayaan sebesar 95 % dan derajat signifikansi 5 %?

Penyelesaian

1. Hipotesis:

- **H₀**: *Tidak ada perbedaan* antara proporsi balita yang menderita demam setelah imunisasi dengan balita yang tidak mengalami demam setelah imunisasi ($P_1 = P_2 = 0,5$)
- **H₁**: *Ada perbedaan* antara proporsi balita yang menderita demam setelah imunisasi dengan balita yang tidak mengalami demam setelah imunisasi ($P_1 \neq P_2 \neq 0,5$)

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic: $n = 20$ dan $x = 7$ (ambil yg terkecil antara 7 dengan 13) Menggunakan R diperoleh nilai p-value sbb :

```
> binom.test(7, 20, 0.5, "two.sided", 0.95)

Exact binomial test

data: 7 and 20
number of successes = 7, number of trials = 20, p-value = 0.2632
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.1539092 0.5921885
sample estimates:
probability of success
                0.35
```

4. Kriteria Penolakan

H_0 ditolak jika $P(x < \alpha)$ Karena p-value = 0,2632 > $\alpha = 0,05$ maka **H_0 tidak ditolak.**

5. Kesimpulan

Jadi, **Tidak ada perbedaan** antara proporsi balita yang menderita demam setelah imunisasi dengan balita yang tidak mengalami demam setelah imunisasi.

PPT Hal. 11 Contoh 3

Contoh 3 Untuk $n > 25$

Seorang pengusaha restoran ingin melakukan penelitian mengenai selera masakan tradisional yang disukai mahasiswa. Hasil penelitian terhadap 30 responden di restoran tradisional memberikan data sebagai berikut :

- 22 orang menyukai masakan Jawa
- 8 orang menyukai masakan Padang.

Ujilah dugaan bahwa jumlah mahasiswa yang menyukai masakan Jawa berbeda dengan masakan Padang. Gunakan taraf nyata sebesar 5%.

Penyelesaian

1. Hipotesis:

- **H0:** Jumlah (frekuensi) mahasiswa yang menyukai masakan Jawa dan masakan Padang adalah sama atau tidak berbeda ($P1 = P2 = 0,5$).
- **H1:** Jumlah (frekuensi) mahasiswa yang menyukai masakan Jawa dan masakan Padang adalah berbeda ($P1 \neq P2 \neq 0,5$).

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic : $n = 30$ dan $x = 8$ (ambil yg terkecil antara 8 dengan 22).

Menggunakan R diperoleh nilai p-value sbb :

```
> binom.test(8, 30, 0.5, "two.sided", 0.95)

Exact binomial test

data: 8 and 30
number of successes = 8, number of trials = 30, p-value = 0.01612
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
0.1227948 0.4588937
sample estimates:
probability of success
0.2666667
```

4. Kriteria Penolakan

H_0 ditolak jika $P(x < \alpha)$ Karena p-value = 0,0178 < $\alpha = 0,05$ maka **H_0 ditolak.**

Atau H_0 ditolak jika $Z_{hitung} < -Z\alpha/2$ atau $Z_{hitung} > Z\alpha/2$ Karena $Z_{hitung} = -2,373 < -Z0,025 = -1,960$ maka H_0 ditolak

5. Kesimpulan

Jadi, Jumlah (frekuensi) mahasiswa yang menyukai masakan Jawa dan masakan Padang adalah **berbeda**.

PPT Hal. 29 Latihan 1

Latihan 1

Pengaruh didirikannya PLTN terhadap kesehatan pekerja dan penduduk yang tinggal di sekitarnya, akhir-akhir ini menjadi bahan perdebatan. Salah satu bahayanya adalah kemungkinan paparan radiasi akan meningkatkan kematian karena kanker. Problem yang biasa dijumpai ketika kita melakukan studi kasus seperti ini adalah sedikitnya jumlah kematian karena kanker (dari semua jenis) maupun suatu jenis kanker tertentu, sehingga kebermaknaan statistik akan sulit dicapai, kecuali jika studi dilakukan dalam jangka waktu lama. Salah satu alternatif adalah melakukan studi mortalitas proporsional, yang didalamnya proporsi kematian karena suatu penyebab pada kelompok yang terpapar dibandingkan dengan proporsi pada populasi umum. Andaikan hasil penelitian menemukan bahwa 4 dari 13 kematian pada para pekerja berusia 55-64 tahun di PLTN disebabkan karena kanker. **Dapatkah kita simpulkan bahwa hasil penelitian tersebut berbeda secara makna terhadap statistik populasi umum ? Taraf nyata pengujian 5%.**

Penyelesaian

1. Hipotesis:

- **H0: Tidak ada perbedaan** antara proporsi kematian akibat kanker pada pekerja PLTN dengan statistik populasi umum ($P1 = P2 = 0,5$).
- **H1: Ada perbedaan** antara proporsi kematian akibat kanker pada pekerja PLTN dengan statistik populasi umum ($P1 \neq P2 \neq 0,5$).

2. Tingkat Signifikansi: $\alpha = 0,05$ (5%)

3. Uji statistic : Jumlah seluruh sampel $n = 13$ dan jumlah kematian karena kanker $x = 4$ (ambil yang terkecil dari 4 dan 9)

Menggunakan fungsi binom.test() di R diperoleh nilai p-value sbb :

```
> binom.test(4, 13, 0.5)

Exact binomial test

data: 4 and 13
number of successes = 4, number of trials = 13, p-value = 0.2668
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
0.09092039 0.61426166
sample estimates:
probability of success
0.3076923
```

4. Kriteria Penolakan

H_0 ditolak jika $p\text{-value} < \alpha$.

Karena $p\text{-value} = 0.2668 > \alpha = 0,05$, maka **H_0 tidak ditolak.**

5. Kesimpulan

Tidak ada bukti yang cukup untuk menyimpulkan bahwa proporsi kematian akibat kanker pada pekerja PLTN berbeda secara signifikan dengan statistik populasi umum. Yang dimana artinya adalah proporsi kematian akibat kanker pada pekerja PLTN **sama dengan** statistik populasi umum.

PPT Hal. 30 Latihan 2

Latihan 2

Seorang mahasiswa Fakultas Peternakan, melakukan penelitian yang berkaitan dengan "Cakupan Pemakaian Dua Merk Vaksin ND di Suatu Daerah". Di daerah tersebut hanya ada vaksin Merk A dan Merk B yang biasa dipakai peternak. Kepada setiap peternak yang dipilih secara random diberikan pertanyaan mengenai merk vaksin apa yang biasa mereka gunakan. Menurut penilaian peneliti, kedua merk vaksin tersebut memiliki kesamaan dalam berbagai hal, baik kualitas, efektivitas, kemudahan mendapatkannya, maupun harganya. Namun ada dugaan bahwa peternak yang memakai merk vaksin A proporsinya lebih banyak dari peternak yang memakai vaksin B. Hasil penelitian terhadap 20 orang responden peternak memberikan hasil Terdapat 15 orang peternak yang menggunakan vaksin Merk A dan 5 orang peternak yang memakai Merk B. **Ujilah dugaan tersebut dengan menggunakan tingkat signifikansi 5%!**

Penyelesaian

1. Hipotesis:

- **H0:** Jumlah peternak yang menggunakan vaksin Merk A *sama dengan* jumlah yang menggunakan vaksin Merk B ($P1 = P2 = 0,5$).
- **H1:** Jumlah peternak yang menggunakan vaksin Merk A *lebih besar* jumlah yang menggunakan vaksin Merk B ($P1 > P2$).

2. Tingkat Signifikansi: $\alpha = 0,05$ (5%)

3. Uji statistic : Jumlah sampel $n = 20$ dan jumlah peternak yang menggunakan vaksin Merk B $x = 5$ (ambil yang terkecil antara 5 dan 15), atau juga bisa $x = 15$ namun dengan syarat tertentu.

Dibawah ini terdapat dua jenis kode yaitu menggunakan alternative “less” dan “greater”. Namun, hal tersebut tergantung pada nilai x-nya, jika x nya menggunakan Merk A yaitu 15, maka kita menggunakan parameter alternative = “greater” seperti dibawah ini:

```
> binom.test(15, 20, 0.5, alternative = "greater")
```

```
Exact binomial test

data: 15 and 20
number of successes = 15, number of trials = 20, p-value = 0.02069
alternative hypothesis: true probability of success is greater than 0.5
95 percent confidence interval:
 0.5444176 1.0000000
sample estimates:
probability of success
                  0.75
```

Lalu, jika kita menggunakan nilai Merk B yaitu 5, maka kita menggunakan parameter alternative “less” seperti dibawah ini:

```
> binom.test(5, 20, 0.5, alternative = "less")
```

```
Exact binomial test
```

```
data: 5 and 20
number of successes = 5, number of trials = 20, p-value = 0.02069
alternative hypothesis: true probability of success is less than 0.5
95 percent confidence interval:
0.0000000 0.4555824
sample estimates:
probability of success
0.25
```

Walaupun keduanya menggunakan parameter alternative yang berbeda, tetapi hasil nilai p-value nya tetap sama yaitu **0.02069**.

4. Kriteria Penolakan

H_0 ditolak jika $p\text{-value} < \alpha$.

Karena $p\text{-value} = 0.02069 < \alpha = 0,05$, maka **H_0 ditolak**.

5. Kesimpulan

Jadi, proporsi peternak yang menggunakan vaksin Merk A **benar/nyata lebih besar** dibandingkan dengan yang menggunakan vaksin Merk B.

PPT Hal. 31 Latihan 3

Latihan 3

Seorang mahasiswa Fakultas Peternakan, melakukan penelitian yang berkaitan dengan "Cakupan Pemakaian Dua Merk Vaksin ND di Suatu Daerah". Di daerah tersebut hanya ada vaksin Merk A dan Merk B yang biasa dipakai peternak. Kepada setiap peternak yang dipilih secara random diberikan pertanyaan mengenai merk vaksin apa yang biasa mereka gunakan. Menurut penilaian peneliti, kedua merk vaksin tersebut memiliki kesamaan dalam berbagai hal, baik kualitas, efektivitas, kemudahan mendapatkannya, maupun harganya. Namun peneliti **belum dapat menduga** bahwa peternak yang memakai merk vaksin A proporsinya lebih banyak dari peternak yang memakai vaksin B. Hasil penelitian terhadap 30 orang responden peternak memberikan hasil Terdapat 24 orang peternak yang menggunakan vaksin Merk A dan 6 orang peternak yang memakai Merk B. **Bantulah peneliti tersebut untuk menentukan adakah perbedaan antara proporsi pengguna vaksin A dengan B , dengan menggunakan tingkat signifikansi 5%!**

Penyelesaian

1. Hipotesis:

- **H0:** Proporsi peternak yang menggunakan vaksin Merk A *sama dengan* yang menggunakan vaksin Merk B ($P1 = P2 = 0,5$).
- **H1:** Proporsi peternak yang menggunakan vaksin Merk A *lebih besar* yang menggunakan vaksin Merk B ($P1 > P2$).

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic : Total responden $n = 30$ dan jumlah pengguna vaksin B $x = 6$.

Menggunakan R diperoleh nilai p-value sbb :

```
> binom.test(6, 30, 0.5, alternative = "less")  
  
Exact binomial test  
  
data: 6 and 30  
number of successes = 6, number of trials = 30, p-value = 0.0007155  
alternative hypothesis: true probability of success is less than 0.5  
95 percent confidence interval:  
 0.0000000 0.3570091  
sample estimates:  
probability of success  
 0.2
```

4. Kriteria Penolakan

H_0 ditolak jika $p\text{-value} < \alpha$.

Karena p-value = 0.0007155 < α = 0,05, maka **H0 ditolak.**

5. Kesimpulan

Jadi, proporsi peternak yang menggunakan vaksin Merk A **benar/nyata lebih besar** daripada vaksin Merk B pada tingkat signifikansi 5%.

PPT Hal. 32 Latihan 4

Latihan 4

Sekelompok mahasiswa berjumlah 10 orang akan berkunjung ke tempat pariwisata bekas Tsunami di Aceh yaitu “Museum Tsunami” dan “PLTD Apung”, jika tiga orang memilih Museum Tsunami dan sisanya memilih PLTD Apung. Ujilah pada taraf nyata 5% apakah kedua tempat pariwisata tersebut sama-sama disukai?

Penyelesaian

1. Hipotesis:

- **H0:** Banyak mahasiswa yang memilih Museum Tsunami *sama dengan* yang memilih PLTD Apung ($P_1 = P_2 = 0,5$).
- **H1:** Banyak mahasiswa yang memilih Museum Tsunami *berbeda* dengan yang memilih PLTD Apung ($P_1 \neq P_2 \neq 0,5$).

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic : Jumlah total mahasiswa $n = 10$ dan pemilih Museum Tsunami $x = 3$.

Menggunakan R diperoleh nilai p-value sbb :

```
> binom.test(3, 10, 0.5, "two.sided")  
  
Exact binomial test  
  
data: 3 and 10  
number of successes = 3, number of trials = 10, p-value = 0.3438  
alternative hypothesis: true probability of success is not equal to 0.5  
95 percent confidence interval:  
 0.06673951 0.65245285  
sample estimates:  
probability of success  
 0.3
```

Pada kode diatas menggunakan parameter alternative “two.sided” karena yang dicari adalah apakah kedua tempat wisata di sukai atau tidak.

4. Kriteria Penolakan

H_0 ditolak jika $p\text{-value} < \alpha$.

Karena $p\text{-value} = 0.3438 > \alpha = 0,05$, maka H_0 tidak ditolak.

5. Kesimpulan

Tidak ada cukup bukti untuk menyatakan bahwa proporsi mahasiswa yang memilih Museum Tsunami berbeda dari proporsi yang memilih PLTD Apung pada tingkat signifikansi 5%. Dengan kata lain, kedua tempat pariwisata tersebut **sama-sama disukai**.

PPT Hal. 33 Latihan 5

Latihan 5

Seorang peneliti akan melakukan penelitian terhadap 10 orang pembeli di mini market (Indomaret dan Alfamart) Banda Aceh. Jika 4 orang pembeli membeli di Alfamart dan sisanya membeli di Indomaret. Ujilah pada taraf nyata 5% apakah kedua mini market tersebut samasama disukai?

Penyelesaian

1. Hipotesis:

- **H0:** Jumlah pembeli yang memilih Alfamart *sama dengan* yang memilih Indomaret ($P_1 = P_2 = 0,5$).
- **H1:** Jumlah pembeli yang memilih Alfamart *berbeda/tidak sama dengan* dari yang memilih Indomaret ($P_1 \neq P_2 \neq 0,5$).

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic : Jumlah total pembeli $n = 10$ dan terdapat 4 orang yang memilih Alfamart dan 6 orang memilih Indomaret, sehingga $x = 4$ karena 4 merupakan yang terkecil.

Menggunakan R diperoleh nilai p-value sbb :

```
> binom.test(4, 10, 0.5, "two.sided")
Exact binomial test

data: 4 and 10
number of successes = 4, number of trials = 10, p-value = 0.7539
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
0.1215523 0.7376219
sample estimates:
probability of success
0.4
```

4. Kriteria Penolakan

H_0 ditolak jika p-value $< \alpha$.

Karena p-value = $0.7539 > \alpha = 0.05$, maka **H_0 tidak ditolak.**

5. Kesimpulan

Ternyata, **tidak ada cukup bukti** untuk menyatakan bahwa proporsi pembeli yang memilih Alfamart berbeda dari proporsi yang memilih Indomaret pada tingkat signifikansi 5%. Jadi, kedua mini market tersebut **sama-sama** disukai.

Laporan 2: Uji Chi Square

PPT Hal. 37 Contoh 1

Contoh 1

Seorang peneliti ingin melakukan survey terhadap respon praktisi mengenai kunjungan ke laboratorium. Di kota A terdapat 3 laboratorium yakni Lab A, B dan C. Dilakukan pengambilan sampel secara random pada 200 orang yang pernah melakukan pemeriksaan laboratorium hasilnya sebagai berikut:

- Lab A : 64
- Lab B : 76
- Lab C : 60

Bagaimanakah kesimpulan dari penelitian tersebut? ($\alpha=0,05$)

Penyelesaian

1. Hipotesis:

- **H0: Tidak ada perbedaan** kunjungan praktisi ke laboratorium.
- **H1: Ada perbedaan** kunjungan praktisi ke laboratorium.

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic :

Untuk nilai ketiga variablenya adalah sebagai berikut:

- Lab A : 64
- Lab B : 76
- Lab C : 60

Menggunakan R diperoleh nilai χ^2 dan p-value sbb :

```

> pengunjung<- matrix(c(64, 76, 60), dimnames=list(c("Lab A", "Lab B", "Lab C"), c("Jumlah")))
> pengunjung
   Jumlah
Lab A     64
Lab B     76
Lab C     60
> chisq.test(pengunjung)

Chi-squared test for given probabilities

data: pengunjung
X-squared = 2.08, df = 2, p-value = 0.3535

```

4. Kriteria Penolakan

H_0 ditolak jika $p - value < \alpha$ atau jika $\chi_{hitung}^2 > \chi_{tabel}^2$. Karena nilai $\chi_{hitung}^2 = 2,079 < \chi_{tabel}^2 = 5,99146$ maka H_0 tidak ditolak, atau karena nilai $p - value = 0,3535 > 0,05$ maka **H_0 tidak ditolak**.

5. Kesimpulan

Jadi, **Ada perbedaan** kunjungan praktisi ke laboratorium

PPT Hal. 41 Contoh 2

Contoh 2

Sebuah Mall yang baru saja dibuka memberi hadiah kepada para pembeli dengan 3 pilihan, yaitu: t-shirt, anting-ting dan mug. Jika dari 500 total hadiah yang dipilih pembeli ternyata yang memilih T-shirt adalah sebanyak 183 orang, anting-ting sebanyak 142 orang dan mug sebanyak 175 orang. Apakah terdapat perbedaan kesukaan pembeli terhadap ketiga pilihan hadiah? Berikut datanya:

Hadiah	T-Shirt	Anting-Anting	Mug
Jumlah	183	142	175

Penyelesaian

1. Hipotesis:

- **H0:** *Tidak terdapat perbedaan* kesukaan pembeli terhadap ketiga pilihan hadiah.
- **H1:** *Terdapat perbedaan* kesukaan pembeli terhadap ketiga pilihan hadiah.

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic :

Tabel data jumlah setiap ketiga hadiah

Hadiah	T-Shirt	Anting-Anting	Mug
Jumlah	183	142	175

Menggunakan R diperoleh nilai χ^2 dan p-value sbb :

```
> hadiah<- matrix(c(183, 142, 175), dimnames=list(c("T-Shirt", "Anting-anting", "mug"), c("Jumlah")))
> hadiah
      Jumlah
T-Shirt      183
Anting-anting 142
mug          175
> chisq.test(hadiah)

Chi-squared test for given probabilities

data: hadiah
X-squared = 5.668, df = 2, p-value = 0.05878
```

4. Kriteria Penolakan

H_0 ditolak jika $p - value < \alpha$ atau jika $\chi_{hitung}^2 > \chi_{tabel}^2$. Karena nilai $\chi_{hitung}^2 = 5,6673 < \chi_{tabel}^2 = 5,99146$ maka H_0 tidak ditolak, atau karena nilai $p - value = 0,05878 > 0,05$ maka **H_0 tidak ditolak**

5. Kesimpulan

Jadi, **Tidak terdapat perbedaan** kesukaan pembeli terhadap ketiga pilihan hadiah

Latihan 7

- Suatu perusahaan cat mobil ingin mengetahui warna cat apa yang harus lebih banyak diproduksi. Untuk itu dilakukan penelitian. Berdasarkan pengamatan selama 1 minggu di jalan protocol terhadap mobil-mobil pribadi di temukan 1000 berwarna biru, 900 merah, 600 putih dan 500 warna lain. Ujilah hipotesis yang dapat digunakan untuk menunjukkan ada atau tidaknya perbedaan jumlah cat yang diproduksi!

Penyelesaian

1. Hipotesis:

- **H₀:** *Tidak ada perbedaan* jumlah cat yang harus diproduksi untuk setiap warna (proporsi cat biru, merah, putih, dan warna lain sama).
- **H₁:** *Ada perbedaan* jumlah cat yang harus diproduksi untuk setiap warna (proporsi cat biru, merah, putih, dan warna lain berbeda).

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic :

Data observasi yang tersedia:

- Mobil biru: 1000
- Mobil merah: 900
- Mobil putih: 600
- Mobil warna lain: 500

$$\text{Total mobil} = 1000 + 900 + 600 + 500 = 3000$$

Frekuensi harapan: Jika tidak ada perbedaan, kita bisa asumsikan bahwa setiap warna memiliki frekuensi yang sama, yakni $3000/4 = 750$ untuk tiap kategori.

Menggunakan R diperoleh nilai χ^2 dan p-value sbb :

```

> pilihan_cat <- matrix(c(1000, 900, 600, 500),
+                           dimnames = list(c("Biru", "Merah", "Putih", "Lainnya"),
+                                         c("Jumlah")))
> pilihan_cat
      Jumlah
Biru     1000
Merah    900
Putih    600
Lainnya   500
> chisq.test(pilihan_cat)

Chi-squared test for given probabilities

data: pilihan_cat
X-squared = 226.67, df = 3, p-value < 2.2e-16

```

4. Kriteria Penolakan

H_0 ditolak jika nilai p-value $< \alpha$.atau jika nilai $\chi^2_{hitung} > \chi^2_{tabel}$.

Dari hasil R, didapatkan $\chi^2_{hitung} = 226.67$ dan p-value $< 2.2e-16$.

Karena p-value < 0.05 , maka **H_0 ditolak**.

5. Kesimpulan

Berdasarkan uji statistik, **terdapat perbedaan yang signifikan** pada jumlah cat yang harus diproduksi untuk warna biru, merah, putih, dan warna lain.

PPT Hal. 46 Latihan 8

Latihan 8

Seorang mahasiswa Fakultas Peternakan, melakukan penelitian yang berkaitan dengan "Volume Penjualan Feed Aditive yang Dijual dalam Kemasan Berbeda dari Sebuah Poultry Shop". Toko yang diteliti adalah yang menjual feed aditive dalam 4 macam kemasan yaitu Kemasan Jenis 1, Jenis 2, Jenis 3, dan Jenis 4. Feed aditive yang dijual berasal dari produsen yang sama serta memiliki kualitas yang sama pula, tetapi berdasarkan pengamatan sekilas di lapangan peneliti memperkirakan, bahwa feed aditive yang dijual dalam jenis kemasan tertentu lebih banyak terjual dibandingkan dengan jenis kemasan lainnya. Pengukuran dilakukan dengan cara mencatat jenis kemasan yang paling banyak terjual setiap hari. Penelitian dilakukan selama 40 hari.

Kemasan	Kemasan 1	Kemasan 2	Kemasan 3	Kemasan 4
Frekuensi (hari)	19	9	7	5

Penyelesaian

1. Hipotesis:

- **H0: Tidak ada perbedaan** volume penjualan feed additive berdasarkan jenis kemasan.
- **H1: Ada perbedaan** volume penjualan feed additive berdasarkan jenis kemasan.

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic :

Data penjualan feed additive di 40 hari berdasarkan jenis kemasan adalah sebagai berikut:

- Kemasan 1: 19 hari
- Kemasan 2: 9 hari
- Kemasan 3: 7 hari
- Kemasan 4: 5 hari

$$\text{Total hari} = 19 + 9 + 7 + 5 = 40$$

Karena diasumsikan bahwa setiap kemasan memiliki peluang yang sama untuk terjual, frekuensi harapan untuk masing-masing kemasan adalah:

$$\text{Total / Jumlah kategori} = 40 / 4 = 10$$

Menggunakan R diperoleh nilai χ^2 dan p-value sbb :

```

> kemasan <- matrix(c(19, 9, 7, 5),
+                         dimnames = list(c("Kemasan 1", "Kemasan 2", "Kemasan 3", "Kemasan 4"),
+                                         c("Frekuensi (hari)")))
> kemasan
      Frekuensi (hari)
Kemasan 1          19
Kemasan 2           9
Kemasan 3           7
Kemasan 4           5
> chisq.test(kemasan)

Chi-squared test for given probabilities

data: kemasan
X-squared = 11.6, df = 3, p-value = 0.008887

```

Hasil Uji:

- **X-squared:** 11.6
- **Degrees of Freedom (df):** 3
- **p-value:** 0.008887

4. Kriteria Penolakan

H_0 ditolak jika nilai p-value $< \alpha$.

Dari hasil R, didapatkan p-value = **0.008887**.

Karena p-value < 0.05 , maka **H_0 ditolak**.

5. Kesimpulan

Berdasarkan uji statistik, **terdapat perbedaan yang signifikan** dalam volume penjualan feed additive berdasarkan jenis kemasan. Artinya, ada jenis kemasan yang **lebih banyak** terjual dibandingkan dengan yang lainnya.

Laporan 3: Uji Runs (Uji Kerandoman)

PPT Hal. 55 Contoh 1

Contoh 1 Untuk $n < 20$

Suatu penelitian dilakukan terhadap 24 responden yang ditentukan secara random. Pengambilan data dilakukan untuk mengetahui keberhasilan kegiatan usahatani. Tingkat keberhasilan usahatani oleh setiap petani ditunjukkan oleh nilai total skor pada Tabel. Lakukan pengujian hipotesis apakah skor usahatani adalah random atau tidak.

Nomor Responden	Skor	Nomor Responden	Skor
1	31	13	15
2	23	14	18
3	36	15	78
4	43	16	24
5	51	17	13
6	44	18	27
7	12	19	86
8	26	20	61
9	43	21	13
10	75	22	7
11	2	23	6
12	3	24	8

Penyelesaian

1. Hipotesis:

- **H0:** Skor Usahatani *random*.
- **H1:** Skor Usahatani *tidak random*.

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic:

Nomor Responden	Skor	Skor thd Median	run	Nomor Responden	Skor	Skor thd Median	run
1	31	+	1	13	15	-	
2	23	-	2	14	18	-	
3	36	+		15	78	+	7
4	43	+		16	24	-	
5	51	+		17	13	-	
6	44	+		18	27	+	
7	12	-	4	19	86	+	
8	26	+		20	61	+	
9	43	+		21	13	-	
10	75	+		22	7	-	
11	2	-		23	6	-	
12	3	-	6	24	8	-	

- Jumlah Sampel (N) = 24 ,
- Median = $\frac{data\ ke-12+data\ ke-13}{2} = \frac{24+26}{2} = 25$
- n_1 = banyaknya tanda + = 12
- n_2 = banyaknya tanda - = 12
- r = banyaknya run observasi = 10

Menggunakan R diperoleh nilai p-value sbb :

```
> skor<- factor(c("+", "-", "+", "+", "+", "+", "-", "+", "+", "+", "+", "-", "+", "-",
  "+", "+", "-", "-", "-", "-"))
> skor
[1] + - + + + - + + + - - - + - - + + - - -
Levels: - +
>
> library(tseries)
Registered S3 method overwritten by 'quantmod':
  method           from
  as.zoo.data.frame zoo

'tseries' version: 0.10-58

'tseries' is a package for time series analysis and computational finance.

See 'library(help="tseries")' for details.

Warning message:
package 'tseries' was built under R version 4.3.3
> runs.test(skor)

Runs Test

data: skor
Standard Normal = -1.0599, p-value = 0.2892
alternative hypothesis: two.sided
```

Dengan menggunakan fungsi `runs.test` di R, output menunjukkan:

- Nilai **Standard Normal = -1.0599**
- **p-value = 0.2892**

Uji ini menggunakan distribusi normal standar untuk melihat apakah urutan "+" dan "-" mengikuti pola acak atau tidak.

4. Kriteria Penolakan

H_0 ditolak jika $r > r_b$ atau $r < r_a$. Atau H_0 ditolak jika $p-value < \alpha$.

Dari hasil yang diperoleh:

- **p-value = 0.2892 > 0.05**

Oleh karena. Dengan demikian, **H_0 tidak ditolak.**

5. Kesimpulan

Jadi, skor usaha tani tersebut berdistribusi **random**.

PPT Hal. 59 Contoh 2

Contoh 2 Untuk $n > 20$

Berikut ini adalah urutan petani dan wanita tani (P = petani dan W = wanita tani) yang mengikuti antrian pupuk bersubsidi.

P W P W P P P W W P W P W P W P P P P W P W P W P

P W W W P W P W P W P P W P P P P W P W P W P P

Lakukan pengujian hipotesis apakah urutan petani dan wanita tani dalam antrian adalah random!

Penyelesaian

1. Hipotesis:

- **H₀**: Urutan petani dan wanita tani dalam antrian adalah *random*.
- **H₁**: Urutan petani dan wanita tani dalam antrian adalah *tidak random*.

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic:

Dari data yang diberikan, diketahui bahwa $N = 50$

	P	W	P	W	P	P	P	W	W	P	W	P	W	P	W	P	P	P	P	W	P	W	P	W	P
run	1	2	3	4		5		6		7	8	9	10	11	12		13			14	15	16	17	18	19
	P	W	W	W	P	W	P	W	P	W	P	W	P	W	P	W	P	P	P	W	P	W	P	W	P
run	19		21		22	23	24	25	26	27	28	29	30	31		32			33	34	35		36		

Diperoleh $r = 35$, n_1 = banyaknya $P = 30$ dan n_2 = banyaknya $W = 20$. Karena $N > 20$ relative besar maka dipergunakan pendekatan distribusi normal.

Menggunakan R diperoleh nilai p-value sbb :

```

> ## PPT Hal. 59 Contoh 2
> peserta_latihan <- factor(c("P", "W", "P", "W", "P", "P", "P", "W", "W", "P", "W",
+                               "P", "W", "P", "W", "P", "P", "P", "W", "P", "W", "P", "W",
+                               "P", "W", "P", "W", "P", "W", "W", "P", "W", "P", "W", "P",
+                               "P", "W", "P", "W", "P", "W", "P", "W", "P", "W", "P", "P",
+                               "P", "W", "P", "W", "P"))
>
> library(tseries)
> runs_test <- runs.test(peserta_latihan)
>
> runs_test

Runs Test

data: peserta_latihan
Standard Normal = 2.9794, p-value = 0.002888
alternative hypothesis: two.sided

```

Hasil dari uji **Runs Test** untuk urutan petani (P) dan wanita tani (W) menunjukkan:

- **Standard Normal = 2.9794**
- **p-value = 0.002888**

4. Kriteria Penolakan

H_0 ditolak jika $r > r_b$ atau $r < r_a$. Atau H_0 ditolak jika $p-value < \alpha$.

Dari hasil uji, diperoleh **p-value = 0.002888**, yang berarti **p-value < 0.05**.

Oleh karena itu, kita **menolak H_0** .

5. Kesimpulan

Jadi, urutan petani dan wanita tani dalam antrian adalah **tidak random**.

PPT Hal. 64 Latihan 1

Berikut ini adalah peserta (P = pria dan W = wanita) yang datang ke suatu penyuluhan pertanian di suatu daerah.

P P W W P W P W P W P P P P W P W W W P W P W P W P P W P P W P P P P W P W P P P.

Lakukan pengujian hipotesis dengan Run Test apakah peluang pria dan wanita hadir dalam penyuluhan pertanian adalah sama. Taraf signifikansi (α) ditetapkan sebesar 5%!

Penyelesaian

1. Hipotesis:

- **H₀**: Urutan kehadiran pria (P) dan wanita (W) dalam penyuluhan pertanian bersifat *sama (random)*.
- **H₁**: Urutan kehadiran pria (P) dan wanita (W) dalam penyuluhan pertanian bersifat *tidak sama (tidak random)*.

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic:

Berikut adalah urutan kehadiran pria (P) dan wanita (W):

P P W W P W P W P W P P P P W P W W W W P W P W P W P W P W P P P P W P W P W P P

Menggunakan R diperoleh nilai p-value sbb :

```
> ## PPT Hal. 64 Latihan 1
>
> # Membuat vektor kehadiran pria (P) dan wanita (W)
> peserta_penyuluhan <- factor(c("P", "P", "W", "W", "P", "W", "P", "W", "P", "W",
+ "P", "P", "P", "P", "W", "P", "W", "W", "W", "W", "P",
+ "W", "P", "W", "P", "W", "P", "W", "P", "W", "P", "P",
+ "W", "P", "P", "P", "P", "W", "P", "W", "P", "W", "P"))
>
> library(tseries)
>
> # Melakukan uji runs
> runs_test <- runs.test(peserta_penyuluhan)
>
> # Menampilkan hasil uji runs
> print(runs_test)

Runs Test

data: peserta_penyuluhan
Standard Normal = 2.2717, p-value = 0.0231
alternative hypothesis: two.sided
```

Hasil dari uji **Runs Test** untuk urutan petani (P) dan wanita tani (W) menunjukkan:

- Standard Normal = 2.2717
- p-value = 0.0231

4. Kriteria Penolakan

H_0 ditolak jika $r > rb$ atau $r < ra$. Atau H_0 ditolak jika $p - value < \alpha$.

Karena p-value **0.0231 < α (0.05)**, maka **H_0 ditolak**.

5. Kesimpulan

Jadi, urutan kehadiran pria dan wanita dalam penyuluhan pertanian **tidak sama/tidak random**.

PPT Hal. 65 Latihan 2

Sebuah penelitian dilakukan untuk mengetahui tingkat adopsi teknologi oleh petani. Pengumpulan data dilakukan terhadap 15 responden yang ditentukan secara random. Data hasil wawancara berbentuk ordinal di mana total skor ditunjukkan pada Tabel di bawah ini. Lakukan pengujian hipotesis dengan Run Test untuk mengetahui apakah peluang petani untuk mengadopsi teknologi dan tidak mengadopsi teknologi adalah berbeda.

Nomor responden	Total skor
1	45
2	65
3	35
4	70
5	60
6	50
7	30
8	90
9	85
10	90
11	40
12	80
13	55
14	30
15	75

Penyelesaian

1. Hipotesis:

- **H0:** Peluang petani untuk mengadopsi teknologi dan tidak mengadopsi teknologi *bersifat acak (random)*.
- **H1:** Peluang petani untuk mengadopsi teknologi dan tidak mengadopsi teknologi *tidak acak (tidak random)*.

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic:

Dari tabel yang diberikan, kita hitung median dari total skor untuk menentukan "+" (adopsi) dan "-" (tidak adopsi). Median akan menjadi batas untuk pengelompokan skor.

Menggunakan R diperoleh nilai p-value sbb :

```

> ## PPT Hal. 65 Latihan 2
>
> # Data total skor
> total_skor <- c(45, 65, 35, 70, 60, 50, 30, 90, 85, 90, 40, 80, 55, 30, 75)
>
> # menghitung median
> median_skor <- median(total_skor)
>
> # Mengubah total skor menjadi tanda + dan -
> # + untuk skor di atas atau sama dg median, - untuk skor di bawah median
> adopsi_teknologi <- ifelse(total_skor >= median_skor, "+", "-")
>
> # Mengubah mnjd faktor
> adopsi_teknologi_factor <- factor(adopsi_teknologi)
>
> # penggunaan library tseries untuk Uji Runs
> library(tseries)
>
> # Melakukan uji runs
> runs_test <- runs.test(adopsi_teknologi_factor)
>
> # Menampilkan hasil uji runs
> print(runs_test)

Runs Test

data: adopsi_teknologi_factor
Standard Normal = 0.82565, p-value = 0.409
alternative hypothesis: two.sided

```

Hasil dari uji **Runs Test** untuk urutan petani (P) dan wanita tani (W) menunjukkan:

- **Nilai Standard Normal** = **0.82565**
- **p-value** = **0.409**

4. Kriteria Penolakan

H_0 ditolak jika $r > r_b$ atau $r < r_a$. Atau H_0 ditolak jika $p-value < \alpha$.

Karena p-value **0.409 > α (0.05)**, maka **H_0 tidak ditolak**.

5. Kesimpulan

Jadi, peluang petani untuk mengadopsi atau tidak mengadopsi teknologi bersifat **acak**. Tidak ada pola yang signifikan mengenai adopsi teknologi oleh petani dalam penelitian ini.

PPT Hal. 66 Latihan 3

Suatu Sekolah Dasar mengambil nilai ujian 30 siswa dari siswa kelas 6, dan akan diteliti apakah pengambilan sampel nilai ujian ini bersifat acak, gunakan taraf nyata 5%, data yang diperoleh sebagai berikut.

NO	NILAI UJIAN	NO	NILAI UJIAN
1	65	16	94
2	45	17	58
3	49	18	60
4	74	19	58
5	80	20	69
6	90	21	79
7	64	22	83
8	57	23	66
9	68	24	62
10	54	25	82
11	76	26	84
12	72	27	52
13	64	28	41
14	52	29	62
15	90	30	76

Penyelesaian

1. Hipotesis:

- **H0:** Urutan nilai ujian siswa adalah *acak (random)*.
- **H1:** Urutan nilai ujian siswa *tidak acak (tidak random)*.

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic:

Dari tabel yang diberikan, kita hitung median dari total skor untuk menentukan "+" (adopsi) dan "-" (tidak adopsi). Median akan menjadi batas untuk pengelompokan skor.

Menggunakan R diperoleh nilai p-value sbb :

```
> ## PPT Hal. 66 Latihan 3
> # Data nilai ujian siswa
> nilai_ujian <- c(65, 45, 49, 74, 80, 90, 64, 57, 68, 54, 76, 72, 64, 52, 90, 94, 58,
60, 58, 69, 79, 83, 66, 62, 82, 84, 52, 41, 62, 76)
>
> # Menghitung median dari data nilai ujian
> median_nilai <- median(nilai_ujian)
>
> # Menampilkan median
> print(paste("Median nilai ujian adalah:", median_nilai))
[1] "Median nilai ujian adalah: 65.5"
>
> # Mengelompokkan nilai berdasarkan median
> skor <- factor(ifelse(nilai_ujian > median_nilai, "+", "-"))
> print(skor)
[1] - - - + + + - - + - + - - + + + + - + + - - -
Levels: - +
>
```

```

> library(tseries)
Registered S3 method overwritten by 'quantmod':
  method           from
as.zoo.data.frame zoo

'tseries' version: 0.10-58

'tseries' is a package for time series analysis and computational finance.

See 'library(help="tseries")' for details.

Warning message:
package 'tseries' was built under R version 4.3.3
>
> # Melakukan uji runs
> hasil_uji_runs <- runs.test(skor)
> print(hasil_uji_runs)

  Runs Test

data: skor
Standard Normal = -0.74322, p-value = 0.4573
alternative hypothesis: two.sided

```

Hasil dari uji runs adalah **Standard Normal = -0.74322** dan **p-value = 0.4573**.

4. Kriteria Penolakan

H_0 ditolak jika $r > r_b$ atau $r < r_a$. Atau H_0 ditolak jika $p-value < \alpha$.

Karena p-value = 0.4573, yang berarti **p-value = 0,4573 > α (0,05)**, maka **H_0 tidak ditolak**.

5. Kesimpulan

Jadi, urutan nilai ujian siswa bersifat **acak (random)**.

Modul Praktikum Hal. 19 Latihan 4

Sebuah penelitian dilakukan untuk mengetahui tingkat adopsi teknologi oleh petani. Pengumpulan data dilakukan terhadap 15 responden yang ditentukan secara random. Data hasil wawancara berbentuk ordinal di mana total skor ditunjukkan pada Tabel 4.4 di bawah ini.

Diketahui: median dari data adalah 60; harga kritis r dalam Run Test satu sampel untuk $\alpha = 5\%$ adalah 4; dan harga kritis r dalam Run Test dua sampel untuk $\alpha = 5\%$ adalah 13.

Lakukan pengujian hipotesis dengan Run Test untuk mengetahui apakah peluang petani untuk mengadopsi teknologi dan tidak mengadopsi teknologi adalah berbeda.

Tabel 4.4. Tingkat adopsi teknologi oleh petani.

Nomor responden	Total skor
1	45
2	65
3	35
4	70
5	60
6	50
7	30
8	90
9	85
10	90
11	40
12	80
13	55
14	30
15	75

Penyelesaian

1. Hipotesis:

- **H0:** Peluang petani untuk mengadopsi dan tidak mengadopsi teknologi adalah *sama (acak)*.
- **H1:** Peluang petani untuk mengadopsi dan tidak mengadopsi teknologi adalah *berbeda (tidak acak)*.

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic:

Data yang diberikan:

- Total skor responden: 45, 65, 35, 70, 60, 50, 30, 90, 85, 90, 40, 80, 55, 30, 75
- Median dari total skor adalah **60**, sehingga skor yang lebih besar dari median diberi tanda "+" dan yang lebih kecil dari median diberi tanda "-".
- Hasil pengelompokan skor: - + - + - - + + + - + - +

Menggunakan R diperoleh nilai p-value sbb :

```
> ## Modul Praktikum Hal. 19 Latihan 4
> # Data total skor responden
> total_skor <- c(45, 65, 35, 70, 60, 50, 30, 90, 85, 90, 40, 80, 55, 30, 75)
>
> # Menghitung median menggunakan fungsi median()
> median_skor <- median(total_skor)
>
> # Menampilkan median
> print(paste("Median total skor adalah:", median_skor))
[1] "Median total skor adalah: 60"
>
> # Menghapus nilai yang sama dengan median
> total_skor_filtered <- total_skor[total_skor != median_skor]
>
> # Mengelompokkan skor berdasarkan median
> skor <- factor(ifelse(total_skor_filtered > median_skor, "+", "-"))
> print(skor)
[1] - + - + - - + + + - + - - +
Levels: - +
>
> library(tseries)
>
> # Melakukan uji runs
> hasil_uji_runs <- runs.test(skor)
> print(hasil_uji_runs)

Runs Test

data: skor
standard Normal = 1.1127, p-value = 0.2658
alternative hypothesis: two.sided
```

Berdasarkan hasil uji runs, didapatkan nilai **Standard Normal = 1.1127** dan **p-value = 0.2658**.

4. Kriteria Penolakan

H_0 ditolak jika $r > r_b$ atau $r < r_a$. Atau H_0 ditolak jika $p-value < \alpha$.

Karena **p-value = 0.2658 > 0.05**, maka kita **gagal menolak H_0** .

5. Kesimpulan

Jadi, peluang petani untuk mengadopsi teknologi dan tidak mengadopsi teknologi dianggap **tidak berbeda secara signifikan** dalam pengambilan sampel ini.

BAB 2: PERTEMUAN 4 Uji Hipotesis

Komparatif 2 Sampel Dependen

Laporan 4: Uji Mc Nemar

Contoh 1 PPT Hal. 8

Suatu perusahaan ingin mengetahui pengaruh sponsor yang diberikan dalam suatu pertandingan olahraga terhadap nilai penjualan barangnya. Dalam penelitian ini digunakan sampel yang diambil secara random yang jumlah anggotanya 200 orang. Sebelum sponsor diberikan, terdapat 50 orang yang memberi barang tersebut dan 150 orang tidak membeli. Setelah sponsor diberikan dalam pertandingan olahraga, ternyata dari 200 orang tersebut terdapat 125 orang yang membeli dan 75 orang yang tidak membeli. Dari 125 orang tersebut terdiri atas 40 pembeli tetap dan yang berubah dari tidak membeli sebanyak 85 orang. Selanjutnya, dari 75 orang yang tidak membeli menjadi membeli ada 10 orang dan yang tetap tidak membeli ada 65 orang.

- Tabel Perubahan Penjualan Sebelum & Setelah Ada Sponsor :

	Sebelum ada Sponsor	Setelah Ada Sponsor
Membeli	50	125
Tidak Membeli	150	75
Jumlah	200	200

- Untuk memudahkan penghitungan maka Tabel setelah ada sponsor di atas diubah menjadi Tabel ABCD berikut :

Perilaku Konsumen	Membeli	Tidak Membeli
Tidak Membeli	85 (A)	65 (B)
Membeli	40 (C)	10 (D)
Jumlah	125	75

Penyelesaian:

1. Hipotesis:

- **H0: Tidak terdapat perbedaan** jumlah penjualan sebelum dan sesudah ada sponsor.
- **H1: Terdapat perbedaan jumlah** penjualan sebelum dan sesudah ada sponsor.

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji Statistic:

Menggunakan R diperoleh nilai p-value sbb :

```
> ## Contoh 1 PPT Hal. 8
>
> sebelum1<- rep(1,150)
> sebelum2<- rep(2,50)
> sebelum<- c(sebelum1, sebelum2)
>
> setelah1<- c(rep(1,65), rep(2,85))
> setelah2<- c(rep(1,10), rep(2,40))
> setelah<- c(setelah1, setelah2)
>
> data<- data.frame(sebelum, setelah)
> table(data)
      setelah
sebelum 1 2
      1 65 85
      2 10 40
> mcnemar.test(table(data))

McNemar's Chi-squared test with continuity correction

data: table(data)
McNemar's chi-squared = 57.642, df = 1, p-value = 3.144e-14
```

4. Kriteria Penolakan

H_0 ditolak jika $\chi_{hitung}^2 > \chi_{tabel}^2$ atau $p - value < \alpha$.

Nilai **p-value (5.401e-15) sangat kecil**, jauh di bawah tingkat signifikansi **0,05**. Maka, kita menolak H_0 .

5. Kesimpulan

Jadi , **terdapat perbedaan** yang signifikan dari penjualan setelah dan sebelum ada sponsor.

Contoh 2 PPT Hal. 12

Seorang peneliti ingin mengetahui perubahan pola tanam para petani lahan kering di wilayah dampak di daerah hulu sungai, setelah pemerintah melaksanakan proyek usaha tani. Sebelum proyek usaha tani dilaksanakan, banyak petani di daerah penelitian mengusahakan tanaman ubi kayu sepanjang tahun sehingga menyebabkan menurunnya kualitas tanah. Setelah dilakukan proyek usaha tani diharapkan para petani di sekitar lokasi percontohan akan mengubah cara tanam mereka menggunakan sistem tumpang sari.

- Hasil penelitian:

		Setelah Proyek	
Sebelum Proyek	Tanam Ubi Kayu	Tumpang Sari	Tanam Ubi Kayu
	Tumpang Sari	5 (C)	3 (D)
	Jumlah	19	9

Ujilah hipotesis bahwa untuk petani yang mengubah pola tanamnya, kemungkinan bahwa seorang petani akan mengubah pola tanamnya dari ubi kayu ke tumpeng sari (PA) sama dengan kemungkinan bahwa ia akan berubah dari tumpeng sari ke tanam ubi kayu. Gunakan Tingkat signifikansi $\alpha = 0,05$.

Penyelesaian:

1. Hipotesis:

- **H₀:** Kemungkinan petani yang mengubah pola tanamnya dari ubi kayu ke tumpang sari (P_1) **sama dengan** kemungkinan mengubah pola tanam dari tumpang sari ke ubi kayu (P_2). Dengan kata lain, : $PA = PD$.
- **H₁:** Kemungkinan petani yang mengubah pola tanamnya dari ubi kayu ke tumpang sari **tidak sama dengan** kemungkinan mengubah pola tanam dari tumpang sari ke ubi kayu, $PA \neq PD$.

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji Statistic:

Menggunakan R diperoleh nilai p-value sbb :

```

> ## Contoh 2 PPT Hal. 12
>
> sebelum1<- rep(1,20)
> sebelum2<- rep(2,8)
> sebelum<- c(sebelum1, sebelum2)
>
> setelah1<- c(rep(2,14), rep(1,6))
> setelah2<- c(rep(2,5), rep(1,3))
> setelah<- c(setelah1, setelah2)
>
> data<- data.frame(sebelum, setelah)
> table(data)
setelah
sebelum  1  2
1   6 14
2   3  5
>
> mcnemar.test(table(data))

McNemar's Chi-squared test with continuity correction

data:  table(data)
McNemar's chi-squared = 5.8824, df = 1, p-value = 0.01529

```

4. Kriteria Penolakan

H_0 ditolak jika $\chi_{hitung}^2 > \chi_{tabel}^2$ atau $p - value < \alpha$.

Karena p-value sebesar **0,01529 < 0,05**, maka **H_0 ditolak**.

5. Kesimpulan

Jadi , terdapat perbedaan yang signifikan dari kemungkinan seorang petani akan mengubah pola tanamnya dari tanam ubi kayu ke tumpeng sari dengan kemungkinan mengubah pola tanam dari tumpeng sari ke tanam ubi kayu.

Latihan 1 PPT Hal. 16

Berdasarkan survey penggunaan bumbu penyedap pada kader diamati antara sebelum memiliki TV dan setelah memiliki TV didapatkan data pada tabel di samping. Selidikilah dengan $\alpha = 5\%$, apakah ada perbedaan penggunaan bumbu penyedap makanan?

No	Sebelum masuk TV	Sesudah masuk TV
1	+	-
2	+	-
3	-	+
4	-	+
5	-	-
6	+	+
7	+	+
8	-	-
9	-	+
10	-	-
11	+	+
12	+	+
13	-	+
14	+	+
15	+	-
16	-	+
17	-	+
18	-	+
19	-	+
20	+	-
21	-	+

Penyelesaian:

- Dibentuk table baru :

		Setelah Masuk TV	
		Respon (+)	Respon (-)
Sebelum masuk TV	Respon (-)	9(A)	3(B)
	Respon (+)	5(C)	4(D)

1. Hipotesis:

- **H0:** *Tidak terdapat perbedaan* jumlah pengguna bumbu penyedap sebelum dan setelah masuk TV.
- **H1:** *Terdapat perbedaan* jumlah pengguna bumbu penyedap sebelum dan setelah masuk TV.

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji Statistic:

Menggunakan R diperoleh nilai p-value sbb :

```

> ## Latihan 1 PPT Hal. 16
>
> sebelum1 <- rep(1,12)
> sebelum2 <- rep(2,9)
> sebelum <- c(sebelum1, sebelum2)
>
> setelah1 <- c(rep(1,3), rep(2,9))
> setelah2 <- c(rep(1,4), rep(2,5))
> setelah <- c(setelah1, setelah2)
>
> data<- data.frame(sebelum, setelah)
> table(data)
      setelah
sebelum 1 2
          1 3 9
          2 4 5
> mcnemar.test(table(data))

McNemar's chi-squared test with continuity correction
data: table(data)
McNemar's chi-squared = 1.2308, df = 1, p-value = 0.2673

```

4. Kriteria Penolakan

H_0 ditolak jika $\chi_{hitung}^2 > \chi_{tabel}^2$ atau $p - value < \alpha$.

Karena p-value sebesar **0,2673 > 0,05**, maka H_0

tidak ditolak.

5. Kesimpulan

Jadi , **tidak terdapat perbedaan** yang signifikan dari perbedaan jumlah pengguna bumbu penyedap sebelum dan setelah masuk TV.

Latihan 2 PPT Hal. 21

Seorang peneliti melakukan penelitian tentang perilaku penggunaan MSG pada masakan. Peneliti ingin mengetahui efektifitas seminar kesehatan terhadap penggunaan MSG dalam makanan. Dalam penelitian ini diambil sampel sebanyak 30 orang. Berdasarkan hasil survei diperoleh data sebagai berikut (0 menandakan responden tidak menggunakan MSG dan 1 menandakan menggunakan MSG). Selidiki apakah terdapat perbedaan penggunaan MSG dalam makanan sebelum dan sesudah mengikuti seminar kesehatan pada taraf nyata 5%.

No	Sebelum	Sesudah	No	Sebelum	Sesudah
1	0	0	16	0	0
2	1	0	17	1	0
3	0	0	18	0	0
4	1	0	19	0	0
5	0	1	20	1	0
6	1	0	21	1	1
7	1	0	22	1	1
8	1	0	23	0	0
9	1	0	24	0	0
10	0	0	25	1	0
11	1	0	26	1	1
12	1	1	27	1	1
13	1	0	28	0	0
14	1	1	29	1	0
15	1	0	30	1	0

Penyelesaian:

1. Hipotesis:

- **H0: Tidak terdapat perbedaan penggunaan MSG dalam makanan sebelum dan sesudah mengikuti seminar kesehatan.**
- **H1: Terdapat perbedaan penggunaan MSG dalam makanan sebelum dan sesudah mengikuti seminar kesehatan.**

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji Statistic:

Tabel kontingensi berdasarkan data:

Sebelum/Sesudah	Tidak Menggunakan (0)	Menggunakan (1)	Total
Tidak Menggunakan (0)	9 (A)	1 (C)	10

Menggunakan (1)	14 (B)	6 (D)	20
Total	23	7	30

- **A=10** : Responden yang tidak menggunakan MSG sebelum dan sesudah seminar.
- **B=8** : Responden yang menggunakan MSG sebelum seminar tetapi tidak menggunakan setelah seminar.
- **C=2** : Responden yang tidak menggunakan MSG sebelum seminar tetapi menggunakan setelah seminar.
- **D=10** : Responden yang menggunakan MSG sebelum dan sesudah seminar.

Menggunakan R diperoleh nilai p-value sbb :

```
> ## Latihan 2 PPT Hal. 21
> sebelum <- c(0, 1, 0, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1,
+           0, 1, 0, 0, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 1)
>
> sesudah <- c(0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0,
+            0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 1, 0, 0)
>
> data <- data.frame(sebelum, sesudah)
> table(data)
      sesudah
sebelum  0  1
          0  9  1
          1 14  6
>
> mcnemar.test(table(data))

McNemar's Chi-squared test with continuity correction

data: table(data)
McNemar's chi-squared = 9.6, df = 1, p-value = 0.001946
```

4. Kriteria Penolakan

H_0 ditolak jika $\chi_{hitung}^2 > \chi_{tabel}^2$ atau $p - value < \alpha$.

Karena $p\text{-value}$ (0,001946) < α (0,05), maka H_0 ditolak.

5. Kesimpulan

Jadi, **Terdapat perbedaan** penggunaan MSG dalam makanan sebelum dan sesudah mengikuti seminar kesehatan.

Laporan 5: Uji Tanda

Contoh 1 PPT Hal. 27

Suatu perusahaan ingin mengetahui pengaruh adanya kenaikan uang insentif terhadap kesejahteraan karyawan. Dalam penelitian itu dipilih 20 pegawai beserta istrinya secara random. Dengan demikian terdapat 20 pasangan suami-istri. Masing-masing suami dan istri diberi angket untuk diisi dengan pertanyaan sebagai berikut:

“Berilah penilaian tingkat kesejahteraan keluarga Bapak/ibu sebelum adanya kenaikan dan sesudah kenaikan insentif dari perusahaan di mana Bapak/Ibu bekerja. Rentang nilai 1 s.d. 10. Nilai 1 berarti sangat tidak sejahtera dan nilai 10 berarti sangat sejahtera.”

Hasil Penelitian:

Data Istri		Data Suami	
Seblm	Sesdh	Sblm	Ssdh
2	4	1	6
2	3	4	6
4	6	2	3
5	7	6	7
4	5	2	4
2	4	3	6
1	3	1	4
2	6	2	7
1	6	1	4
7	9	2	3
4	7	4	8
5	9	6	9
2	4	2	7
3	5	2	6
6	9	5	9
3	7	1	6
2	4	4	5
3	8	2	6
1	2	1	3
2	3	2	4

Penyelesaian:

1. Hipotesis:

- **H₀:** *Tidak terdapat perbedaan* pengaruh intensif yang signifikan terhadap kesejahteraan keluarga baik menurut istri maupun suami
- **H₁:** *Terdapat perbedaan* pengaruh intensif yang signifikan terhadap kesejahteraan keluarga baik menurut istri maupun suami

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji Statistic:

a. Data Istri

Data Istri			
Seblm (X_i)	Sesdh (Y_i)	Beda $Y_i - X_i$	Peringkat
2	4	2	4
2	3	1	5
4	6	2	4
5	7	2	4
4	5	1	5
2	4	2	4
1	3	2	4
2	6	4	2
1	6	5	1
7	9	2	4
4	7	3	3
5	9	4	2
2	4	2	4
3	5	2	4
6	9	3	3
3	7	4	2
2	4	2	4
3	8	5	1
1	2	1	5
2	3	1	5

b. Data Suami

Data Suami			
Seblm (X_i)	Sesdh (Y_i)	Beda $Y_i - X_i$	Peringkat
1	6	5	1
4	6	2	4
2	3	1	5
6	7	1	5
2	4	2	4
3	6	3	3
1	4	3	3
2	7	5	1
1	4	3	3
2	3	1	5
4	8	4	2
6	9	3	3
2	7	5	1
2	6	4	2
5	9	4	2
1	6	5	1
4	5	1	5
2	6	4	2
1	3	2	4
2	4	2	4

Dari Tabel Peringkat yang telah disusun dapat dibentuk table baru untuk melihat peringkat perubahan kesejahteraan keluarga sbb:

No	Tingkat Perubahan		Arah			Tanda
	Istri	Suami				
1	4	1	4	>	1	-
2	5	4	5	>	4	-
3	4	5	4	<	5	+
4	4	5	4	<	5	+
5	5	4	5	>	4	-
6	4	3	4	>	3	-
7	4	3	4	>	3	-
8	2	1	2	>	1	-
9	1	3	1	<	3	+
10	4	5	4	<	5	+
11	3	2	3	>	2	-

12	2	3	2	<	3	+
13	4	1	4	>	1	-
14	4	2	4	>	2	-
15	3	2	3	>	2	-
16	2	1	2	>	1	-
17	4	5	4	<	5	+
18	1	2	1	<	2	+
19	5	4	5	>	4	-
20	5	4	5	>	4	-

Menggunakan R diperoleh nilai p-value sbb :

```

> ## Contoh 1 PPT Hal. 27
>
> istri<- c(4,5,4,4,5,4,4,2,1,4,3,2,4,4,3,2,4,1,5,5)
> suami<- c(1,4,5,5,4,3,3,1,3,5,2,3,1,2,2,1,5,2,4,4)
> data<- data.frame(istri, suami, diff=suami-istri)
> data
  istri suami diff
1      4      1    -3
2      5      4    -1
3      4      5     1
4      4      5     1
5      5      4    -1
6      4      3    -1
7      4      3    -1
8      2      1    -1
9      1      3     2
10     4      5     1
11     3      2    -1
12     2      3     1
13     4      1    -3
14     4      2    -2
15     3      2    -1
16     2      1    -1
17     4      5     1
18     1      2     1
19     5      4    -1
20     5      4    -1
>
> library(BSDA)
Loading required package: lattice
Attaching package: 'BSDA'

The following object is masked from 'package:datasets':
  Orange

Warning message:
package 'BSDA' was built under R version 4.3.3

```

```

> SIGN.test(x=data$diff, alternative="less", conf.level=0.95)

One-sample Sign-Test

data: data$diff
s = 7, p-value = 0.1316
alternative hypothesis: true median is less than 0
95 percent confidence interval:
-Inf 1
sample estimates:
median of x
-1

Achieved and Interpolated Confidence Intervals:

      Conf.Level L.E.pt U.E.pt
Lower Achieved CI 0.9423 -Inf     1
Interpolated CI   0.9500 -Inf     1
Upper Achieved CI 0.9793 -Inf     1

```

4. Kriteria Penolakan

H_0 ditolak jika $Z_{hitung} > Z_{\frac{\alpha}{2}}$ atau $Z_{hitung} < -Z_{\frac{\alpha}{2}}$ atau $p-value < \alpha$.

Karena p-value sebesar **0,1316 > 0,05**, maka **H_0 tidak ditolak**.

5. Kesimpulan

Tidak ada perbedaan pengaruh insentif yang signifikan terhadap kesejahteraan keluarga, baik menurut istri maupun suami.

Contoh 2 PPT Hal. 35

Mahasiswa semester akhir dari Jurusan Sosek Fakultas Pertanian berkeinginan melakukan penelitian mengenai “Tingkat Pengetahuan Budidaya Kopi dari Penduduk Suatu Desa yang Akan Diberi Bantuan Bibit Kopi”. Penelitian ini penting dilakukan, karena diduga akan berpengaruh terhadap suksesnya proyek bantuan tersebut. Pengambilan data dilaksanakan sebanyak dua kali, dengan maksud untuk mengkaji ada tidaknya perubahan tingkat pengetahuan sebelum dan sesudah diberi penyuluhan dengan materi Budidaya Tanaman Kopi.

Kuesioner dirancang dengan cara memberikan skor untuk tiap aspek budidaya, sehingga bisa dilakukan ranking dari 1-5 berdasarkan Tingkat pengetahuan kumulatifnya. Berdasarkan berbagai literatur, peneliti menduga bahwa, dengan seringnya dilakukan penyuluhan akan terjadi perubahan tingkat pengetahuan petani.

No.Resp.	Skor Pengetahuan		Tanda
	Sebelum	Setelah	
1	5	5	0
2	4	5	-
3	3	4	-
4	4	3	+
5	4	3	+
6	3	4	-
7	3	4	-
8	4	5	-
9	4	5	-
10	3	5	-
11	4	3	+
12	3	4	-
13	3	4	-
14	2	3	-
15	4	4	0
16	3	3	0
17	3	4	-
18	5	4	+
19	2	3	-
20	2	3	-
21	3	4	-
22	3	4	-
23	4	5	-
24	4	3	+

25	3	3	0
26	4	3	+
27	4	5	-
28	4	5	-
29	3	4	-
30	2	3	-
31	4	3	+
32	4	4	0
33	5	4	+
34	5	4	+
35	4	4	0
36	3	4	-
37	2	3	-
38	3	4	-
39	2	3	-
40	3	5	-

Penyelesaian:

1. Hipotesis:

- $H_0: dm = 0$
- $H_1: dm \neq 0$

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji Statistic:

Menggunakan R diperoleh nilai p-value sbb :

```
> ## Contoh 2 PPT Hal. 35
>
> sebelum <- c(5,4,3,4,4,3,3,4,4,3,4,3,3,2,4,3,3,5,2,2,3,3,4,4,3,4,4,4,3,2,4,4,5,
5,4,3,2,3,2,3)
> sesudah <- c(5,5,4,3,3,4,4,5,5,5,3,4,4,3,4,3,4,4,3,3,4,4,5,3,3,3,5,5,4,3,3,4,4,
4,4,4,3,4,3,5)
>
> diff <- sesudah - sebelum
```

```

> data <- data.frame(Sebelum = sebelum, Sesudah = sesudah, Diff = diff)
> data
   Sebelum Sesudah Diff
1        5       5    0
2        4       5    1
3        3       4    1
4        4       3   -1
5        4       3   -1
6        3       4    1
7        3       4    1
8        4       5    1
9        4       5    1
10       3       5    2
11       4       3   -1
12       3       4    1
13       3       4    1
14       2       3    1
15       4       4    0
16       3       3    0
17       3       4    1
18       5       4   -1
19       2       3    1
20       2       3    1
21       3       4    1
22       3       4    1
23       4       5    1
24       4       3   -1
25       3       3    0
26       4       3   -1
27       4       5    1
28       4       5    1
29       3       4    1
30       2       3    1
31       4       3   -1
32       4       4    0
33       5       4   -1
34       5       4   -1
35       4       4    0
36       3       4    1
37       2       3    1
38       3       4    1
39       2       3    1
40       3       5    2
>
> library(BSDA)
> SIGN.test(x = diff, alternative = "two.sided", conf.level = 0.95)

```

One-sample Sign-Test

data: diff
 s = 25, p-value = 0.009041
 alternative hypothesis: true median is not equal to 0
 95 percent confidence interval:
 0 1
 sample estimates:

```

sample estimates:
median of x
    1

Achieved and Interpolated Confidence Intervals:

      Conf.Level L.E.pt U.E.pt
Lower Achieved CI     0.9193    0      1
Interpolated CI        0.9500    0      1
Upper Achieved CI      0.9615    0      1

```

4. Kriteria Penolakan

H_0 ditolak jika $Z_{hitung} > Z_{\frac{\alpha}{2}}$ atau $Z_{hitung} < -Z_{\frac{\alpha}{2}}$ atau $p-value < \alpha$.

Atau Karena $pvalue = 0,009041 < 0,05$ sehingga H_0 ditolak.

5. Kesimpulan

Berdasarkan pengujian di atas dapat disimpulkan bahwa **ada perubahan** tingkat pengetahuan budidaya kopi yang sangat nyata dari penduduk suatu desa setelah diberi penyuluhan.

Latihan 1 PPT Hal. 40

Seorang guru ingin mengetahui apakah ada perbedaan antara nilai ujian seorang siswa pada mata ajar tertentu, jika ia belajar sendirian dan jika ia belajar berkelompok. Untuk itu, diadakan dua kali uji mata ajar, pertama siswa belajar sendiri-sendiri dan yang kedua sebelum ujian siswa dapat dengan bebas belajar secara berkelompok. Berikut hasil nilai keduanya (angka dalam range nilai (sampai 100)).

No	Sendiri	Kelompok
1	80	78
2	82	85
3	84	82
4	86	78
5	80	84
6	84	89
7	83	83
8	81	81
9	82	90
10	80	78
11	83	85
12	84	82
13	82	86
14	81	91
15	80	80
16	79	89
17	75	79

Penyelesaian:

1. Hipotesis:

- **H0: Tidak terdapat perbedaan** median antara nilai ujian siswa saat belajar sendiri dan saat belajar berkelompok.
- **H1: Terdapat perbedaan** median antara nilai ujian siswa saat belajar sendiri dan saat belajar berkelompok.

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji Statistic:

Kita akan menguji perbedaan nilai dari dua kondisi (belajar sendiri dan belajar berkelompok). Tabel perhitungan bedanya adalah sebagai berikut:

No	Nilai Sendiri (X)	Nilai Kelompok (Y)	Beda (Y - X)	Tanda
1	80	78	-2	-
2	82	85	3	+
3	84	82	-2	-
4	86	78	-8	-
5	80	84	4	+
6	84	89	5	+
7	83	83	0	0
8	81	81	0	0
9	82	90	8	+
10	80	78	-2	-
11	83	85	2	+
12	84	82	-2	-
13	82	86	4	+
14	81	91	10	+
15	80	80	0	0
16	79	89	10	+
17	75	79	4	+

Dari tabel di atas, terdapat:

- Tanda positif (+): 9
- Tanda negatif (-): 5
- Beda nol (0): 3 (dikeluarkan dari analisis)

Kita menggunakan uji tanda dengan nilai $n=9+5=14$ untuk tanda positif dan negatif.

Menggunakan R diperoleh nilai p-value sbb :

```
> ## Latihan 1 PPT Hal. 40
> sendiri <- c(80, 82, 84, 86, 80, 84, 83, 81, 82, 80, 83, 84, 82, 81, 80, 79, 75)
> kelompok <- c(78, 85, 82, 78, 84, 89, 83, 81, 90, 78, 85, 82, 86, 91, 80, 89, 79)
>
> beda <- kelompok - sendiri
```

```

> data <- data.frame(Sendiri = sendiri, Kelompok = kelompok, Beda = beda)
> print(data)
   Sendiri Kelompok Beda
1       80        78   -2
2       82        85    3
3       84        82   -2
4       86        78   -8
5       80        84    4
6       84        89    5
7       83        83    0
8       81        81    0
9       82        90    8
10      80        78   -2
11      83        85    2
12      84        82   -2
13      82        86    4
14      81        91   10
15      80        80    0
16      79        89   10
17      75        79    4
>
> data_analisis <- data[data$Beda != 0, ]
>
> library(BSDA)
> SIGN.test(x = data_analisis$Beda, alternative = "two.sided", conf.level = 0.95)

One-sample sign-Test

data: data_analisis$Beda
s = 9, p-value = 0.424
alternative hypothesis: true median is not equal to 0
95 percent confidence interval:
-2.000000 5.497802
sample estimates:
median of x
            3.5

Achieved and Interpolated Confidence Intervals:

          Conf.Level L.E.pt U.E.pt
Lower Achieved CI 0.9426    -2 5.0000
Interpolated CI  0.9500    -2 5.4978
Upper Achieved CI 0.9871    -2 8.0000

```

4. Kriteria Penolakan

H_0 ditolak jika $Z_{hitung} > Z_{\frac{\alpha}{2}}$ atau $Z_{hitung} < -Z_{\frac{\alpha}{2}}$ atau $p-value < \alpha$.

Berdasarkan hasil uji tanda, nilai p-value adalah **0,424**. Karena **p-value (0,424) > α (0,05)**, maka **H_0 tidak ditolak**.

5. Kesimpulan

Tidak ada perbedaan signifikan antara nilai ujian siswa ketika belajar sendiri dengan belajar berkelompok.

Laporan 6: Uji Wilcoxon

Contoh $n < 25$ PPT Hal. 46

Seorang mahasiswa Fakultas Pertanian dari Jurusan Sosek ingin mengetahui apakah keikutsertaan dalam pelatihan bisa mempengaruhi keberhasilan usaha perdagangan saprotan (sarana produksi pertanian). Untuk itu dilakukan survei terhadap 10 orang pedagang saprotan, mereka dinilai keberhasilan usahanya sebelum dan setelah mengikuti pelatihan, kepada tiap responden diberi skor dengan interval 1-100. Diperkirakan akan ada perbedaan keberhasilan usaha perdagangan saprotan sebelum dan setelah diberi pelatihan.

Taraf nyata atau tingkat signifikansi (level of significance) yang digunakan adalah $\alpha = 0,05$. Setelah survei selesai, data yang diperoleh dimasukan dalam Tabel, sekaligus dilakukan pengolahan lebih lanjut untuk menentukan ranking di .

Pas. Resp.	Pelatihan	
	Sebelum	Sesudah
1	76	80
2	58	60
3	62	68
4	67	72
5	66	79
6	81	80
7	85	82
8	72	80
9	71	81
10	75	79

Penyelesaian:

1. Hipotesis:

- **H0: Tidak ada perbedaan** keberhasilan antara sebelum dan sesudah pelatihan.
- **H1: Ada perbedaan** keberhasilan antara sebelum dan sesudah pelatihan.

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji Statistic:

No	Sebelum	Sesudah	beda	Rank	Rank +	Rank -
1	76	80	-4	4,5		4
2	58	60	-2	2		2
3	62	68	-6	7		6
4	67	72	-5	6		5
5	66	79	-13	10		13
6	81	80	1	1	1	
7	85	82	3	3	3	

8	72	80	-8	8		8
9	71	81	-10	9		9
10	75	79	-4	4,5		4,5
Jumlah					4	51,5

Menggunakan R diperoleh nilai p-value sbb :

```
> ## Contoh n < 25 PPT Hal. 46
> sebelum<- c(76, 58, 62, 67, 66, 81, 85, 72, 71, 75)
> sesudah<- c(80, 60, 68, 72, 79, 80, 82, 80, 81, 79)
> pelatihan<- data.frame(sebelum, sesudah)
> pelatihan
  sebelum sesudah
1      76     80
2      58     60
3      62     68
4      67     72
5      66     79
6      81     80
7      85     82
8      72     80
9      71     81
10     75     79
>
> wilcox.test(sebelum, sesudah, paired=TRUE, data=pelatihan)
  wilcoxon signed rank test with continuity correction
data:  sebelum and sesudah
V = 4, p-value = 0.01898
alternative hypothesis: true location shift is not equal to 0
Warning message:
In wilcox.test.default(sebelum, sesudah, paired = TRUE, data = pelatihan) :
  cannot compute exact p-value with ties
```

4. Kriteria Penolakan

H_0 ditolak jika $Z_{hitung} > Z_{\frac{\alpha}{2}}$ atau $Z_{hitung} < -Z_{\frac{\alpha}{2}}$ atau $p-value < \alpha$.

Karena p-value sebesar **0,01898 < 0,05**, maka **H_0 ditolak**.

5. Kesimpulan

Jadi, **ada perbedaan** keberhasilan antara sebelum dan sesudah pelatihan.

Contoh $n > 25$ PPT Hal. 51

Mahasiswa semester akhir dari Jurusan Sosek Fakultas Pertanian ingin mengetahui tentang “Keberhasilan Usaha Tani yang dikelola oleh petani pria dan wanita”. Untuk keperluan tersebut telah dipilih berbagai jenis usaha tani. Setiap jenis usaha tani dipasang-pasangkan berdasarkan kesamaan jenis dan skala usahanya. Kemudian untuk setiap pasangan yang sama diambil sampel berdasarkan jenis kelamin, dan didapatkan 30 pasangan usaha tani yang akan diteliti.

Keberhasilan usaha diukur dari berbagai kriteria, dan untuk tiap tingkat keberhasilan diberikan skor 1-10. Dalam kaitan penelitian ini, belum diperoleh informasi apakah variabel jenis kelamin tertentu lebih menentukan terhadap keberhasilan usaha. Taraf nyata atau tingkat signifikansi (level of significance) yang digunakan dalam pengujian, $\alpha = 0,01$.

Pas. Resp.	Jenis Kelamin	
	Pria	Wanita
1	8	10
2	7	7
3	8	8
4	7	6
5	7	7
6	6	6
7	9	5
8	9	5
9	5	4
10	4	3
11	9	4
12	8	5
13	7	2
14	8	5
15	6	7
16	6	5
17	5	6
18	10	5
19	10	2
20	6	4
21	5	3
22	7	4
23	7	10
24	4	6
25	5	4
26	8	4
27	10	2
28	6	4
29	8	5
30	8	9

Penyelesaian:

No	Pria	Wanita	beda	Rank	Rank +	Rank -
1	8	10	-2	11		11
2	7	7	0	Ties		
3	8	8	0	Ties		
4	7	6	1	4.5	4.5	
5	7	7	0	Ties		
6	6	6	0	Ties		
7	9	5	4	20	20	
8	9	5	4	20	20	
9	5	4	1	4.5	4.5	
10	4	3	1	4.5	4.5	
11	9	4	5	23	23	
12	8	5	3	16	16	
13	7	2	5	23	23	
14	8	5	3	16	16	
15	6	7	-1	4.5	4.5	1
16	6	5	1	4.5	4.5	
17	5	6	-1	4.5	4.5	1
18	10	5	5	23	23	
19	10	2	8	25	25	
20	6	4	2	11	11	
21	5	3	2	11	11	
22	7	4	3	16	16	
23	7	10	-3	16	16	16
24	4	6	-2	11	11	11
25	5	4	1	4.5	4.5	
26	8	4	4	20	20	
27	10	2	8	25	25	
28	6	4	2	11	11	
29	8	5	3	16	16	
30	8	9	-1	4.5	4.5	1
Jumlah					339	41

1. Hipotesis:

- **H0:** $dm = 0$
- **H1:** $dm \neq 0$

2. Tingkat Signifikansi: ($\alpha = 0,01$)

3. Uji Statistic:

Menggunakan R diperoleh nilai p-value sbb :

```
> ## Contoh n > 25 PPT Hal. 51
> pria <- c(8, 7, 8, 7, 7, 6, 9, 9, 5, 4, 9, 8, 7, 8, 6, 6, 5, 10, 10, 6, 5, 7, 7,
4, 5, 8, 10, 6, 8, 8)
> wanita <- c(10, 7, 8, 6, 7, 6, 5, 5, 4, 3, 4, 5, 2, 5, 7, 5, 6, 5, 2, 4, 3, 4, 10,
6, 4, 4, 2, 4, 5, 9)
>
> data_peserta <- data.frame(pria, wanita)
> data_peserta
   pria wanita
1     8    10
2     7     7
3     8     8
4     7     6
5     7     7
6     6     6
7     9     5
8     9     5
9     5     4
10    4     3
11    9     4
12    8     5
13    7     2
14    8     5
15    6     7
16    6     5
17    5     6
18   10     5
19   10     2
20    6     4
21    5     3
22    7     4
23    7    10
24    4     6
25    5     4
26    8     4
27   10     2
28    6     4
29    8     5
30    8     9
>
> wilcox.test(pria, wanita, paired = TRUE, data=data_peserta)
  Wilcoxon signed rank test with continuity correction
data: pria and wanita
V = 299.5, p-value = 0.001613
alternative hypothesis: true location shift is not equal to 0

Warning messages:
1: In wilcox.test.default(pria, wanita, paired = TRUE, data = data_peserta) :
  cannot compute exact p-value with ties
2: In wilcox.test.default(pria, wanita, paired = TRUE, data = data_peserta) :
  cannot compute exact p-value with zeroes
```

4. Kriteria Penolakan

H_0 ditolak jika $Z_{hitung} > Z_{\frac{\alpha}{2}}$ atau $Z_{hitung} < -Z_{\frac{\alpha}{2}}$ atau $p-value < \alpha$.

Dalam hal ini, diperoleh p-value sebesar **0,001613 < 0,01**, sehingga **H_0 ditolak**.

5. Kesimpulan

Jadi, **ada perbedaan** keberhasilan kegiatan usaha tani yang dikelola oleh petani pria dan wanita.

Latihan 1 PPT Hal. 57

Penelitian dilakukan untuk mengetahui pengaruh penyuluhan terhadap kemampuan petani dalam menggunakan teknologi usahatani yang baru. Sebelum dan sesudah penyuluhan tentang teknologi baru dilakukan pengukuran terhadap kemampuan petani. Data total skor kemampuan petani ditunjukkan pada Tabel 5.7. Lakukan pengujian hipotesis untuk membuktikan apakah terdapat perbedaan kemampuan peserta dalam menggunakan teknologi usaha tani yang baru sebelum dan sesudah penyuluhan dilakukan.

Tabel 5.7. Kemampuan peserta latihan.

Nomor responden	Total skor sebelum penyuluhan	Total skor sesudah penyuluhan
1	20	18
2	19	17
3	23	20
4	20	21
5	18	16
6	21	19
7	20	18
8	20	22
9	19	17
10	18	15
11	23	21
12	21	20
13	20	20
14	20	19
15	23	22

Penyelesaian:

1. Hipotesis:

- **H0: *Tidak ada perbedaan*** kemampuan peserta dalam menggunakan teknologi usaha tani yang baru sebelum dan sesudah penyuluhan. (Median sebelum = median sesudah).
- **H1: *Ada perbedaan*** kemampuan peserta dalam menggunakan teknologi usaha tani yang baru sebelum dan sesudah penyuluhan. (Median sebelum ≠ median sesudah).

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji Statistic:

Menggunakan R diperoleh nilai p-value sbb :

```
> ## Latihan 1 PPT Hal. 57
> sebelum <- c(20, 19, 23, 20, 18, 21, 20, 20, 19, 18, 23, 21, 20, 20, 23)
> sesudah <- c(18, 17, 20, 21, 16, 19, 18, 22, 17, 15, 21, 20, 20, 19, 22)
> kemampuan <- data.frame(sebelum, sesudah)
> kemampuan
  sebelum sesudah
1       20     18
2       19     17
3       23     20
4       20     21
5       18     16
6       21     19
7       20     18
8       20     22
9       19     17
10      18     15
11      23     21
12      21     20
13      20     20
14      20     19
15      23     22
>
> wilcox.test(sebelum, sesudah, paired = TRUE, data = kemampuan)

  Wilcoxon signed rank test with continuity correction

data: sebelum and sesudah
V = 94, p-value = 0.008383
alternative hypothesis: true location shift is not equal to 0

Warning messages:
1: In wilcox.test.default(sebelum, sesudah, paired = TRUE, data = kemampuan) :
   cannot compute exact p-value with ties
2: In wilcox.test.default(sebelum, sesudah, paired = TRUE, data = kemampuan) :
   cannot compute exact p-value with zeroes
```

4. Kriteria Penolakan

H_0 ditolak jika $Z_{hitung} > Z_{\frac{\alpha}{2}}$ atau $Z_{hitung} < -Z_{\frac{\alpha}{2}}$ atau $p-value < \alpha$.

Dalam hal ini, diperoleh p-value sebesar **0,008383 < 0,01**, sehingga **H_0 ditolak**.

5. Kesimpulan

Jadi, **terdapat perbedaan** kemampuan peserta dalam menggunakan teknologi usaha tani yang baru sebelum dan sesudah penyuluhan dilakukan.

Latihan 2 PPT Hal. 59

Sebuah perusahaan Pharmasi sedang mengembangkan suplemen penambahan berat badan pada anak-anak. Perusahaan ingin mengetahui khasiat suplemen tersebut sebelum dipasarkan secara komersial. Untuk itu perusahaan mencoba obat tersebut secara kontinu terhadap 15 orang siswa sekolah dasar yang sudah diukur terlebih dahulu berat badannya. Setelah 3 bulan kemudian siswa-siswi tersebut ditimbang berat badannya lagi untuk mengetahui apakah ada peningkatan berat badannya yang nyata.

Berikut ini adalah hasil pengukuran tersebut (angka dalam kilogram)

No	Sebelum	Sesudah
1	25	26
2	27	26
3	20	22
4	21	24
5	18	22
6	19	21
7	20	24
8	22	21
9	24	26
10	25	26
11	24	25
12	27	28
13	23	25
14	25	27
15	22	25

Penyelesaian:

1. Hipotesis:

- **H0: Tidak ada perbedaan** yang signifikan dalam berat badan sebelum dan sesudah penggunaan suplemen. (Median sebelum = median sesudah).
- **H1: Ada perbedaan** yang signifikan dalam berat badan sebelum dan sesudah penggunaan suplemen. (Median sebelum \neq median sesudah).

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji Statistic:

Menggunakan R diperoleh nilai p-value sbb :

```
> ## Latihan 2 PPT Hal. 59
> sebelum <- c(25, 27, 20, 21, 18, 19, 20, 22, 24, 25, 24, 27, 23, 25, 22)
> sesudah <- c(26, 26, 22, 24, 22, 21, 24, 21, 26, 26, 25, 28, 25, 27, 25)
> suplemen <- data.frame(sebelum, sesudah)
> suplemen
  sebelum sesudah
1      25     26
2      27     26
3      20     22
4      21     24
5      18     22
6      19     21
7      20     24
8      22     21
9      24     26
10     25     26
11     24     25
12     27     28
13     23     25
14     25     27
15     22     25
>
> wilcox.test(sebelum, sesudah, paired = TRUE, data = suplemen)

  wilcoxon signed rank test with continuity correction

data: sebelum and sesudah
V = 7, p-value = 0.002556
alternative hypothesis: true location shift is not equal to 0

Warning message:
In wilcox.test.default(sebelum, sesudah, paired = TRUE, data = suplemen) :
  cannot compute exact p-value with ties
```

4. Kriteria Penolakan

H_0 ditolak jika $Z_{hitung} > Z_{\frac{\alpha}{2}}$ atau $Z_{hitung} < -Z_{\frac{\alpha}{2}}$ atau $p-value < \alpha$.

Dalam hal ini, diperoleh **p-value** sebesar $0,002556 < \alpha = 0,05$, sehingga H_0 ditolak.

5. Kesimpulan

Jadi, **terdapat perbedaan signifikan** dalam berat badan siswa sebelum dan sesudah mengonsumsi suplemen dengan nilai p-value sebesar 0.002556. Oleh karena itu, dapat disimpulkan bahwa suplemen penambahan berat badan memiliki khasiat yang nyata dalam meningkatkan berat badan siswa.

Latihan 3 PPT Hal. 61

Sebuah penelitian mengajukan hipotesis bahwa kegiatan penyuluhan pertanian berperan dalam peningkatan produksi pertanian di suatu daerah. Produksi diukur pada saat sebelum dan sesudah penyuluhan petanian hadir di wilayah tersebut. Produksi pertanian diukur dengan skor dan total skor ditunjukkan pada tabel di bawah ini (Tabel 5.10). Lakukan pengujian hipotesis dengan menggunakan Wilcoxon Matched Pairs Test bila taraf signifikansi sebesar 0,025.

Tabel 5.10. Kemampuan peserta pelatihan.

Nomor responden	Total skor sebelum penyuluhan pertanian.	Total skor sesudah penyuluhan pertanian.
1	75	77
2	50	60
3	77	70
4	80	81
5	30	60
6	68	70
7	55	70
8	72	75
9	33	55
10	70	70
11	64	64
12	40	50
13	90	80
14	35	45
15	87	80
16	75	67
17	35	60
18	60	60
19	55	70
20	82	80
21	37	65
22	66	65
23	73	73
24	65	60
25	73	70
26	45	50
27	74	80
28	76	75
29	70	67
30	60	70

Penyelesaian:

1. Hipotesis:

- **H₀:** *Tidak ada perbedaan* signifikan dalam produksi pertanian sebelum dan sesudah penyuluhan (true location shift = 0).
- **H₁:** *Ada perbedaan signifikan* dalam produksi pertanian sebelum dan sesudah penyuluhan (true location shift ≠ 0).

2. Tingkat Signifikansi: ($\alpha = 0,025$)

3. Uji Statistic:

Menggunakan R diperoleh nilai p-value sbb :

```
> ## Latihan 3 PPT Hal. 61
> sebelum <- c(75, 50, 77, 80, 30, 68, 55, 72, 33, 70, 64, 40, 90, 35, 87, 7
5, 35, 60, 55, 82, 37, 66, 73, 65, 73, 45, 74, 76, 70, 60)
> sesudah <- c(77, 60, 70, 81, 60, 70, 70, 75, 55, 70, 64, 50, 80, 45, 80, 6
7, 60, 60, 70, 80, 65, 65, 73, 60, 70, 50, 80, 75, 67, 70)
> pertanian <- data.frame(sebelum, sesudah)
> pertanian
   sebelum sesudah
1       75     77
2       50     60
3       77     70
4       80     81
5       30     60
6       68     70
7       55     70
8       72     75
9       33     55
10      70     70
11      64     64
12      40     50
13      90     80
14      35     45
15      87     80
16      75     67
17      35     60
18      60     60
19      55     70
20      82     80
21      37     65
22      66     65
23      73     73
24      65     60
25      73     70
26      45     50
27      74     80
28      76     75
29      70     67
30      60     70
>
> wilcox.test(sebelum, sesudah, paired = TRUE, data=pertanian)

  wilcoxon signed rank test with continuity correction

data: sebelum and sesudah
V = 95.5, p-value = 0.04318
alternative hypothesis: true location shift is not equal to 0
```

4. Kriteria Penolakan

H_0 ditolak jika $Z_{hitung} > Z_{\frac{\alpha}{2}}$ atau $Z_{hitung} < -Z_{\frac{\alpha}{2}}$ atau $p-value < \alpha$.

Dalam hal ini, diperoleh **p-value** sebesar **0,0431 > $\alpha = 0,025$** , sehingga H_0 tidak ditolak.

5. Kesimpulan

Jadi, **tidak terdapat perbedaan** yang signifikan antara skor produksi sebelum dan sesudah penyuluhan.

BAB 3: PERTEMUAN 5-6 Uji Hipotesis Komparatif 2 Sampel Independen

Laporan 7: Uji Fisher

Contoh 1 PPT Pertemuan 4,5 Hal 10

Sebuah penelitian dilakukan untuk mengetahui tingkat serangan hama keong mas dan tikus terhadap tanaman padi di Desa A dan B. Rata-rata jumlah keong mas yang menyerang tanaman padi di Desa A sebanyak 4 ekor per m^2 dan di Desa B sebanyak 3 ekor per m^2 . Rata-rata jumlah tikus yang menyerang tanaman padi di Desa A sebanyak 2 ekor per m^2 dan di Desa B sebanyak 5 ekor per m^2 .

Lakukan pengujian hipotesis untuk membuktikan apakah terdapat perbedaan nyata/signifikan tingkat serangan hama keong mas dan tikus di Desa A dan B. Taraf signifikansi (α) ditetapkan sebesar 5% (0,05).

Penyelesaian

1. Hipotesis:

- **H0:** Tingkat serangan hama keong mas dan tikus di Desa A *tidak berbeda* dengan desa B ($dm = 0$).
- **H1:** Tingkat serangan hama keong mas dan tikus di Desa A *berbeda* dengan Desa B ($dm < 0$).

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic:

```
> ## Contoh 1 PPT Pertemuan 4,5 Hal 10
> Hama<- matrix(c(2, 5, 4, 3), nrow=2, dimnames=list(c("Desa A",
  "Desa B"),c("Hama Tikus", "Hama Keong Mas")))
> Hama
      Hama Tikus Hama Keong Mas
Desa A          2            4
Desa B          5            3
>
> fisher.test(Hama, alternative="two.sided")
Fisher's Exact Test for Count Data

data: Hama
p-value = 0.5921
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
0.01809307 4.13427934
sample estimates:
odds ratio
0.3285541
```

4. Kriteria Penolakan

H_0 ditolak jika $p-value < \alpha$.

Dari hasil yang diperoleh:

- **p-value = 0.5921 > $\alpha= 0.05$**

Oleh karena. Dengan demikian, **H_0 tidak ditolak.**

5. Kesimpulan

Jadi, **tidak ada perbedaan** tingkat serangan hama keong mas dan tikus di desa A dan B.

Contoh 2 PPT Pertemuan 4,5 Hal 17

Sebuah studi kasus kontrol ingin melihat pengaruh merokok malam dengan kejadian kanker paru, hasil yang diperoleh tersaji pada tabel silang berikut :

Merokok Malam	Kanker Paru		Jumlah
	Ya	Tidak	
Ya	3	0	3
Tidak	1	3	4
Jumlah	4	3	7

Apakah ada Perbedaan antara kejadian kanker paru pada kebiasaan merokok malam dengan dengan kejadian kanker paru pada perokok pada $\alpha = 5\%$?

Penyelesaian

1. Hipotesis:

- **H0: Tidak ada perbedaan** antara kebiasaan merokok malam dengan kejadian kanker paru pada perokok ($dm = 0$)
- **H1 Ada perbedaan** antara kebiasaan merokok malam dengan kejadian kanker paru pada perokok ($dm \neq 0$)

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic:

```
> ## Contoh 2 PPT Pertemuan 4,5 Hal 17
> Kanker<- matrix(c(3, 1, 0, 3), nrow=2, dimnames=list(c("Merokok",
  "Tidak Merokok"), c("Kanker Paru", "Tidak Kanker Paru")))
> Kanker
      Kanker Paru Tidak Kanker Paru
Merokok            3          0
Tidak Merokok      1          3
>
> fisher.test(Kanker, alternative="greater")
Fisher's Exact Test for Count Data

data: Kanker
p-value = 0.1143
alternative hypothesis: true odds ratio is greater than 1
95 percent confidence interval:
 0.560439      Inf
sample estimates:
odds ratio
      Inf
```

4. Kriteria Penolakan

H_0 ditolak jika $p - value < \alpha$.

Dari hasil yang diperoleh:

- **p-value = 0.1143 > $\alpha= 0.05$**

Oleh karena. Dengan demikian, **H_0 tidak ditolak.**

5. Kesimpulan

Jadi, **tidak ada perbedaan** antara kebiasaan merokok malam dengan kejadian kanker paru pada perokok

Soal Latihan PPT Pertemuan 4,5 Hal 21

Seorang mahasiswa Fakultas Pertanian ingin meneliti perbedaan latar belakang tingkat pendidikan (sarjana dan bukan sarjana) Kepala BUMN Pertanian dan Kepala Perusahaan Pertanian Swasta. Dugaan peneliti, BUMN lebih banyak dipimpin oleh sarjana pertanian dibandingkan dengan Perusahaan Swasta.

Berdasarkan sampel yang dipilih secara random diperoleh 7 BUMN. Dari 7 BUMN tersebut ada 6 buah yang dipimpin sarjana dan ada 1 buah yang dipimpin oleh bukan sarjana. Sedangkan Perusahaan Swasta yang terpilih secara random hanya ada 5 perusahaan, 1 dipimpin oleh sarjana dan 4 lagi dipimpin oleh bukan. Hasilnya diperlihatkan dalam Tabel 5.2

Tabel 5.2 Frekuensi Tingkat Pendidikan Menurut Jabatan yang Didudukinya Saat Ini

Tingkat Pendidikan	Bukan Sarjana	Sarjana	Total
Ka. BUMN	1	6	7
Ka. Per.Swasta	4	1	5
Total	5	7	12

Penyelesaian

1. Hipotesis:

- **H0:** *Tidak ada perbedaan* proporsi Kepala BUMN Pertanian dan Kepala Perusahaan Pertanian Swasta dalam hal tingkat pendidikan (sarjana dan bukan sarjana).
- **H1:** Kepala BUMN *lebih banyak* dipimpin oleh sarjana dibandingkan dengan Kepala Perusahaan Swasta.

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic:

```
> ## Soal Latihan PPT Pertemuan 4,5 Hal 21
> kepala_bumn <- matrix(c(1,4,6,1), nrow=2, dimnames=list(c("Ka. BUMN", "ka. Per.Swasta"), c("Bukan Sarjana", "Sarjana")))
> kepala_bumn
      Bukan Sarjana Sarjana
Ka. BUMN                 1       6
ka. Per.Swasta             4       1
>
> fisher.test(kepala_bumn, alternative="less")
Fisher's Exact Test for Count Data

data:  kepala_bumn
p-value = 0.04545
alternative hypothesis: true odds ratio is less than 1
95 percent confidence interval:
0.0000000 0.9517184
sample estimates:
odds ratio
```

0.06251988

4. Kriteria Penolakan

H_0 ditolak jika $p-value < \alpha$.

Dari hasil yang diperoleh:

- **p-value = 0.04545 < 0.05**

Oleh karena. Dengan demikian, **H_0 ditolak.**

5. Kesimpulan

Jadi, Kepala BUMN **lebih banyak** dipimpin oleh sarjana dibandingkan dengan Kepala Perusahaan Swasta.

Modul Praktikum Latihan 1 Hal 38

Diketahui adanya kecenderungan para birokrat lebih menyukai mobil berwarna gelap dan para akademisi lebih menyukai mobil dengan warna terang. Untuk membuktikan hal tersebut telah dilakukan pengumpulan data dengan menggunakan sampel yang diambil secara random. Dari 8 orang birokrat yang diamati, 5 orang bermobil warna gelap dan 3 orang berwarna terang. Selanjutnya, dari 7 orang akademisi yang diamati, sebanyak 5 orang menggunakan mobil berwarna terang dan 2 orang menggunakan mobil berwarna gelap. Adakah perbedaan proporsi akademisi dan birokrat dalam memilih warna mobil? Gunakan tingkat signifikansi 5%.

Penyelesaian

1. Hipotesis:

- **H0:** *Tidak ada perbedaan* proporsi warna mobil antara birokrat dan akademisi ($p_{\text{birokrat}} = p_{\text{akademisi}}$).
- **H1:** *Terdapat perbedaan* proporsi warna mobil antara birokrat dan akademisi ($p_{\text{birokrat}} \neq p_{\text{akademisi}}$).

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic:

```
> # Modul Praktikum Latihan 1 Hal 38
> birodemisi <- matrix(c(3,5,5,2), nrow=2,
  dimnames=list(c("Birokrat", "Akademisi"), c("Terang", "Gelap")))
> birodemisi
      Terang Gelap
Birokrat     3     5
Akademisi     5     2
>
> fisher.test(birodemisi)

  Fisher's Exact Test for Count Data

data: birodemisi
p-value = 0.3147
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.0152196 3.0734606
sample estimates:
odds ratio
0.2659602
```

4. Kriteria Penolakan

H_0 ditolak jika $p\text{-value} < \alpha$.

Dari hasil yang diperoleh:

- p-value = **0.3147 > 0.05**

Oleh karena. Dengan demikian, **H_0 tidak ditolak.**

5. Kesimpulan

Jadi, ***tidak ada perbedaan*** proporsi warna mobil antara birokrat dan akademisi.

Laporan 8: Uji Chi Square

Contoh 1 PPT Pertemuan 4,5 Hal 26

Misalkan suatu penelitian untuk melihat adakah hubungan antara tingkat kelulusan / kegagalan dalam sebuah kelas matematika tertentu dengan jenis kelamin. sampel acak dari 100 siswa dan mengukur kedua jenis kelamin (laki-laki/wanita) dan status kelulusan (lulus/gagal) sebagai variabel kategorik.

Tabel 1. Data tingkat kelulusan kelas matematika berikut

Siswa	Laki-laki	Perempuan	TOTAL
Lulus	30	36	66
Tidak lulus	14	20	34
TOTAL	44	56	100

Penyelesaian

1. Hipotesis:

- **H0: Tidak terdapat perbedaan** antara tingkat kelulusan / kegagalan dalam sebuah kelas matematika tertentu dengan jenis kelamin.
- **H1: Terdapat perbedaan** antara tingkat kelulusan / kegagalan dalam sebuah kelas matematika tertentu dengan jenis kelamin.

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic:

```
> ## PPT Pertemuan 4,5 Hal 26
> siswa <- matrix(c(30,14,36,20), nrow=2, dimnames=list(c("Lulus",
  "Tdk lulus"), c("cowo", "cewe")))
> siswa
      cowo  cewe
Lulus      30   36
Tdk lulus  14   20
>
> chisq.test(siswa)

Pearson's Chi-squared test with Yates' continuity correction

data: siswa
X-squared = 0.038269, df = 1, p-value = 0.8449
```

4. Kriteria Penolakan

H_0 ditolak jika $\chi^2_{hitung} < \chi^2_{tabel}$. Atau H_0 ditolak jika $p-value < \alpha$.

Dari hasil yang diperoleh:

- $p-value = 0.038269 < 0.05$

Oleh karena. Dengan demikian, **H_0 ditolak.**

5. Kesimpulan

Jadi, **terdapat perbedaan** antara tingkat kelulusan / kegagalan dalam sebuah kelas matematika tertentu dengan jenis kelamin

Contoh 2 PPT Pertemuan 4,5 Hal 31

Sekelompok mahasiswa dari Jurusan Sosek dan Produksi Fakultas Peternakan melakukan penelitian bersama untuk mengetahui sektor pekerjaan alumni yang berasal dari kedua jurusan tersebut. Diduga, alumni kedua jurusan yang bekerja di sektor pertanian, industri, dan jasa proporsinya berlainan. Sampel diambil secara random, dengan jumlah sampel alumni Jurusan Sosek 20 orang dan Jurusan Produksi 30 orang. Dari hasil penelitian didapatkan data sebagai berikut:

1. Dari jumlah 30 orang alumni Sosek, sebanyak 15 orang bekerja di sektor Pertanian, 10 orang di sektor industri, dan 5 orang di sektor Jasa.
2. Dari jumlah 60 orang alumni Produksi, sebanyak 15 orang bekerja di sektor Pertanian, 20 orang di sektor industri, dan 25 orang di sektor Jasa.

Penyelesaian

1. Hipotesis:

- **H0:** Proporsi alumni dari jurusan Sosek dan Produksi di masing-masing sektor pekerjaan adalah *sama* ($p_1 = p_2 = p_3$).
- **H1:** Proporsi alumni dari jurusan Sosek dan Produksi di masing-masing sektor pekerjaan adalah *tidak sama* ($p_1 \neq p_2 \neq p_3$).

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic:

```
> Alumni <- matrix(c(15, 10, 5, 15, 20, 25), nrow=3,
  dimnames=list(c("Pertanian", "Industri", "Jasa"), c("sosek",
  "Produksi")))
> Alumni
      sosek Produksi
Pertanian    15      15
Industri      10      20
Jasa          5       25
>
> chisq.test(Alumni)

Pearson's Chi-squared test

data: Alumni
X-squared = 7.5, df = 2, p-value = 0.02352
```

4. Kriteria Penolakan

H_0 ditolak jika $\chi^2_{hitung} < \chi^2_{tabel}$. Atau H_0 ditolak jika $p - value < \alpha$.

Dari hasil yang diperoleh:

- $p - value = 0.02352 < 0.05$

Oleh karena. Dengan demikian, **H_0 ditolak**.

5. Kesimpulan

Jadi, **terdapat perbedaan** proporsi antara alumni dari jurusan Sosek dan Produksi di sektor pekerjaan (pertanian, industri, dan jasa).

Soal Latihan PPT Pertemuan 4,5 Hal 37

Suatu penelitian dilakukan untuk melihat hubungan antara tipe sekolah dengan jenis kelamin apakah ada kecenderungan sekolah swasta lebih banyak murid perempuan jika dibandingkan dengan sekolah negeri. Hasil penelitian dapat dilihat pada tabel berikut:

		Jenis Kelamin		Total
Sekolah	Negeri	Laki-laki	Perempuan	
	Swasta	14	18	32
	Total	91	109	200

Penyelesaian

1. Hipotesis:

- **H0: Tidak ada hubungan** antara tipe sekolah dan jenis kelamin.
- **H1: Ada hubungan** antara tipe sekolah dan jenis kelamin.

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic:

```
> ## PPT Pertemuan 4,5 Hal 37
> tipe_sekolah<- matrix(c(77,14,91,18), nrow=2,
dimnames=list(c("Negeri", "Swasta"),c("Laki-laki", "Perempuan")))
> tipe_sekolah
      Laki-laki Perempuan
Negeri        77      91
Swasta        14      18
>
> chisq.test(tipe_sekolah)

Pearson's Chi-squared test with Yates' continuity correction

data: tipe_sekolah
X-squared = 0.00054009, df = 1, p-value = 0.9815
```

4. Kriteria Penolakan

H_0 ditolak jika $r > rb$ atau $r < ra$. Atau H_0 ditolak jika $p - value < \alpha$.

Dari hasil yang diperoleh:

- $p - value = 0.9815 > 0.05$

Oleh karena. Dengan demikian, **H_0 tidak ditolak**.

5. Kesimpulan

Jadi, **tidak ada hubungan** antara tipe sekolah dan jenis kelamin.

Laporan 9: Uji Median

Contoh 1 PPT Pertemuan 5,6 Hal 6

Dilakukan penelitian untuk mengetahui apakah penghasilan para nelayan berbeda dengan petani berdasarkan mediannya. Berdasarkan hasil wawancara terhadap 10 petani dan 9 nelayan diperoleh data sbb :

No	Petani	Nelayan
1	50	45
2	60	50
3	70	55
4	70	60
5	75	65
6	80	65
7	90	70
8	95	80
9	95	100
10	100	

Penyelesaian

1. Hipotesis:

- **H0: Tidak Ada Perbedaan** Penghasilan Antara Petani & Nelayan
- **H1: Ada Perbedaan** Penghasilan Antara Petani & Nelayan

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic:

```
> ## Contoh 1 PPT Pertemuan 5,6 Hal.  
> # Menggunakan Uji Median  
> Penghasilan<-  
c(50,60,70,70,75,80,90,95,95,100,45,50,55,60,65,65,70,80,100)  
> Kelompok<- c(rep("Petani", 10), rep("Nelayan", 9))  
> data<- data.frame(Penghasilan, Kelompok)  
> data  
  Penghasilan Kelompok  
1          50    Petani  
2          60    Petani  
3          70    Petani  
4          70    Petani  
5          75    Petani  
6          80    Petani  
7          90    Petani  
8          95    Petani  
9          95    Petani  
10         100   Petani  
11         45    Nelayan  
12         50    Nelayan  
13         55    Nelayan
```

```

14      60  Nelayan
15      65  Nelayan
16      65  Nelayan
17      70  Nelayan
18      80  Nelayan
19     100  Nelayan
>
> library(agricolae)
Warning message:
package 'agricolae' was built under R version 4.3.3
> Median.test(data$Penghasilan, data$Kelompok)

The Median Test for data$Penghasilan ~ data$Kelompok

Chi Square = 2.773232   DF = 1   P.Value 0.09585243
Median = 70

  Median   r Min Max Q25   Q75
Nelayan    65.0  9  45 100  55 70.00
Petani     77.5 10  50 100  70 93.75

Post Hoc Analysis

Groups according to probability of treatment differences and alpha level.

Treatments with the same letter are not significantly different.

  data$Penghasilan groups
Petani          77.5      a
Nelayan         65.0      a

```

Sehingga nilai $p - value = 2 \times P = 2 \times 0,09585243 = 0,19170486$.

Karena $n_1 + n_2 = 19 < 20$ maka bisa juga digunakan uji Fisher:

```

> # Menggunakan Uji Fisher
> Penghasilan<- matrix(c(6, 4, 2, 7), nrow=2, dimnames=(list(c(""))))
Error in matrix(c(6, 4, 2, 7), nrow = 2, dimnames = (list(c("")))) :
  length of 'dimnames' [1] not equal to array extent
> # Menggunakan Uji Fisher
> Penghasilan<- matrix(c(6, 4, 2, 7), nrow=2, dimnames=(list(c("Di atas
Median", "Dibawah Median"), c("Petani", "Nelayan"))))
> Penghasilan
      Petani Nelayan
Di atas Median      6      2
Dibawah Median     4      7
>
> fisher.test(Penghasilan, alternative="two.sided")

Fisher's Exact Test for Count Data

data: Penghasilan
p-value = 0.1698
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.5142383 71.5337148
sample estimates:
odds ratio
 4.772621

```

4. Kriteria Penolakan

H_0 ditolak jika $p-value < \alpha$.

Dari hasil yang diperoleh menggunakan Uji Median dan Uji Fisher:

- **p-value = 0.19170486 (Uji Median) atau 0.1698 (Uji Fisher) > $\alpha = 0.05$**

Oleh karena. Dengan demikian, **H_0 tidak ditolak.**

5. Kesimpulan

Jadi, **tidak ada perbedaan** penghasilan yang cukup nyata antara petani dan nelayan berdasarkan mediannya

Contoh 2 PPT Pertemuan 5,6 Hal 12

Sebuah eksperimen dilakukan dengan menggunakan media leaflet dan video untuk melihat apakah ada perbedaan skor pengetahuan anak tentang HIV/AIDS antara grup yang diberi media edukasi berupa leaflet dengan grup yang diberi penjelasan melalui video.

No	Leaflet	Video
1	83	91
2	94	90
3	91	85
4	96	80
5	91	91
6	89	90
7	85	84
8	92	83
9	92	83
10	90	84
11		81
12		88

Penyelesaian

1. Hipotesis:

- **H0: Tidak ada perbedaan** skor pengetahuan tentang HIV AIDS melalui leaflet dan video.
- **H1: Ada perbedaan** skor pengetahuan tentang HIV AIDS melalui leaflet dan video.

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic:

```
> ## Contoh 2 PPT Pertemuan 5,6 Hal. 12
> # Menggunakan Uji Median
> Skor<-
c(83, 94, 91, 96, 91, 89, 85, 92, 92, 90, 91, 90, 85, 80, 91, 90, 84, 83, 83, 84, 81, 88)
> Kelompok<- c(rep("Leaflet", 10), rep("video", 12))
> data<- data.frame(Skor, Kelompok)
> data
  Skor Kelompok
```

```

1   83 Leaflet
2   94 Leaflet
3   91 Leaflet
4   96 Leaflet
5   91 Leaflet
6   89 Leaflet
7   85 Leaflet
8   92 Leaflet
9   92 Leaflet
10  90 Leaflet
11  91 video
12  90 video
13  85 video
14  80 video
15  91 video
16  90 video
17  84 video
18  83 video
19  83 video
20  84 video
21  81 video
22  88 video
>
> library(agricolae)
> Median.test(data$Skor, data$Kelompok)

The Median Test for data$Skor ~ data$Kelompok

Chi Square = 2.933333   DF = 1   P.Value 0.08676817
Median = 89.5

      Median   r Min Max   Q25   Q75
Leaflet    91.0 10   83   96 89.25   92
video      84.5 12   80   91 83.00   90

Post Hoc Analysis

Groups according to probability of treatment differences and alpha level.

Treatments with the same letter are not significantly different.

      data$Skor groups
Leaflet    91.0      a
video      84.5      a

```

Sehingga nilai $p - value = 2 \times P = 2 \times 0,08676817 = 0,17353634$.

Karena $n_1 + n_2 = 22 > 20$, maka bisa juga digunakan uji Chi-Square:

```

> # Menggunakan Uji Chi-Square
> Skor<- matrix(c(7, 3, 4, 8), nrow=2, dimnames=(list(c("Di atas Median",
"Di bawah Median"), c("Leaflet", "Video"))))
> Skor
      Leaflet Video
Di atas Median       7     4
Di bawah Median      3     8
>
> chisq.test(Skor)

Pearson's Chi-squared test with Yates' continuity correction

data: Skor

```

```
X-squared = 1.65, df = 1, p-value = 0.199
```

4. Kriteria Penolakan

H_0 ditolak jika $p\text{-value} < \alpha$.

Dari hasil yang diperoleh:

- **p-value = 0.17353634 (Uji Median) atau 0.199 (Uji Chi-Square) > $\alpha= 0.05$**

Oleh karena. Dengan demikian, **H_0 tidak ditolak.**

5. Kesimpulan

Jadi, **tidak ada perbedaan** skor pengetahuan tentang HIV AIDS melalui leaflet dan video.

Latihan 1 PPT Pertemuan 5,6 Hal 18

Lakukan pengujian hipotesis untuk melihat apakah ada perbedaan produksi per hektar tanaman jagung karena pengaruh dua metode penanaman yang digunakan, pertumbuhan tanaman jagung dipilih dari sejumlah plot tanah yang berbeda secara acak. Kemudian produksi per hektar dari masing-masing plot dihitung dan hasilnya adalah sbb:

Metode1	83	91	94	89	96	91	92	90	92	85		
Metode2	91	90	81	83	84	83	88	91	90	84	80	85

Penyelesaian

1. Hipotesis:

- **H0: Tidak ada perbedaan** produksi per hektar tanaman jagung karena pengaruh dua metode penanaman yang digunakan.
- **H1: Ada perbedaan** produksi per hektar tanaman jagung karena pengaruh dua metode penanaman yang digunakan.

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic:

```
> ## Latihan 1 PPT Pertemuan 5,6 Hal. 18
> # Menggunakan Uji Median
> Produksi<- c(83, 91, 94, 89, 96, 91, 92, 90, 92, 85, 91, 90, 81, 83, 84,
83, 88, 91, 90, 84, 80, 85)
> Kelompok<- c(rep("Metode_1", 10), rep("Metode_2", 12))
> data<-data.frame(Produksi, Kelompok)
> data
   Produksi Kelompok
1         83 Metode_1
2         91 Metode_1
3         94 Metode_1
4         89 Metode_1
5         96 Metode_1
6         91 Metode_1
7         92 Metode_1
8         90 Metode_1
9         92 Metode_1
10        85 Metode_1
11        91 Metode_2
12        90 Metode_2
13        81 Metode_2
14        83 Metode_2
15        84 Metode_2
16        83 Metode_2
17        88 Metode_2
18        91 Metode_2
19        90 Metode_2
20        84 Metode_2
21        80 Metode_2
22        85 Metode_2
>
> library(agricolae)
```

```

> Median.test(data$Produksi, data$Kelompok)

The Median Test for data$Produksi ~ data$Kelompok

Chi Square = 2.933333   DF = 1   P.Value 0.08676817
Median = 89.5

      Median   r Min Max   Q25   Q75
Metode_1    91.0 10  83  96 89.25  92
Metode_2    84.5 12  80  91 83.00  90

Post Hoc Analysis

Groups according to probability of treatment differences and alpha level.

Treatments with the same letter are not significantly different.

      data$Produksi groups
Metode_1          91.0      a
Metode_2          84.5      a

```

Sehingga nilai $p - value = 2 \times P = 2 \times 0,08676817 = 0,17353634$.

Karena $n_1 + n_2 = 22 > 20$, maka bisa juga digunakan uji Chi-Square:

```

> # Menggunakan Uji Chi-Square
> Skor<- matrix(c(7, 3, 4, 8), nrow=2, dimnames=(list(c("Di atas Median",
  "Dibawah Median"), c("Metode_1", "Metode_2"))))
> Skor
      Metode_1 Metode_2
Di atas Median      7      4
Dibawah Median      3      8
>
> chisq.test(Skor)

Pearson's Chi-squared test with Yates' continuity correction

data: Skor
X-squared = 1.65, df = 1, p-value = 0.199

```

4. Kriteria Penolakan

H_0 ditolak jika $p - value < \alpha$.

Dari hasil yang diperoleh:

- **p-value = 0.17353634 (Uji Median) atau 0.199 (Uji Chi-Square) > $\alpha = 0.05$**

Oleh karena. Dengan demikian, **H_0 tidak ditolak**.

5. Kesimpulan

Jadi, **tidak ada perbedaan** produksi per hektar tanaman jagung karena pengaruh dua metode penanaman yang digunakan.

Latihan 2 PPT Pertemuan 5,6 Hal 19

Sebuah studi hendak meneliti apakah terdapat penurunan kemampuan eliminasi obat pada penderita penyakit hati. Penelitian dilakukan dengan mempelajari respon eliminasi obat phenylbutazone pada penderita sirosis hati dan orang normal. Setiap subjek yang diteliti diberikan obat tersebut sebanyak 19 mg/kg berat badan per orang. Melalui analisis darah, waktu konsentrasi plasma tertinggi (dalam jam) diukur pada masing-masing subjek. Hasil penelitian ditunjukkan pada table di bawah ini. Dapatkah kita menarik kesimpulan bahwa kedua sample mempunyai perbedaan waktu konsentrasi plasma tertinggi? Gunakan $\alpha = 0,05$

Normal	Penderita Sirosis Hati
45,6	20,1
49	14
13,7	42,3
37,9	29,7
26,8	17,8
30,6	22,6
4	15
35,0	10,7
41,3	21,5
32,5	7
8,8	11,2
17,4	18
13,8	27,9
26,3	
14,4	

Penyelesaian

1. Hipotesis:

- **H0: Tidak ada perbedaan** waktu konsentrasi plasma tertinggi antara penderita sirosis hati dan orang normal.
- **H1: Ada perbedaan** waktu konsentrasi plasma tertinggi antara penderita sirosis hati dan orang normal.

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic:

```
> ## Latihan 2 PPT Pertemuan 5,6 Hal. 19
> # Menggunakan Uji Median
> Waktu<- c(45.6, 49, 13.7, 37.9, 26.8, 30.6, 4, 35.0, 41.3, 32.5, 8.8,
17.4, 13.8, 26.3, 14.4, 20.1, 14, 42.3, 29.7, 17.8, 22.6, 15, 10.7, 21.5,
7, 11.2, 18, 27.9)
> Kelompok<- c(rep("Normal", 15), rep("Penderita Sirosis Hati", 13))
> data<- data.frame(Waktu, Kelompok)
> data
   Waktu      Kelompok
1   45.6       Normal
```

```

2   49.0          Normal
3   13.7          Normal
4   37.9          Normal
5   26.8          Normal
6   30.6          Normal
7    4.0          Normal
8   35.0          Normal
9   41.3          Normal
10  32.5          Normal
11   8.8          Normal
12  17.4          Normal
13  13.8          Normal
14  26.3          Normal
15  14.4          Normal
16 20.1 Penderita Sirosis Hati
17 14.0 Penderita Sirosis Hati
18 42.3 Penderita Sirosis Hati
19 29.7 Penderita Sirosis Hati
20 17.8 Penderita Sirosis Hati
21 22.6 Penderita Sirosis Hati
22 15.0 Penderita Sirosis Hati
23 10.7 Penderita Sirosis Hati
24 21.5 Penderita Sirosis Hati
25  7.0 Penderita Sirosis Hati
26 11.2 Penderita Sirosis Hati
27 18.0 Penderita Sirosis Hati
28 27.9 Penderita Sirosis Hati
>
> library(agricolae)
> Median.test(data$Waktu, data$Kelompok)

The Median Test for data$Waktu ~ data$Kelompok

Chi Square = 1.292308   DF = 1   P.Value 0.2556231
Median = 20.8

      Median   r Min  Max  Q25  Q75
Normal        26.8 15    4 49.0 14.1 36.45
Penderita Sirosis Hati 18.0 13    7 42.3 14.0 22.60

Post Hoc Analysis

Groups according to probability of treatment differences and alpha level.

Treatments with the same letter are not significantly different.

      data$Waktu groups
Normal            26.8     a
Penderita Sirosis Hati 18.0     a

```

Sehingga nilai $p - value = 2 \times P = 2 \times 0,2556231 = 0,5112462$.

Karena $n_1 + n_2 = 28 > 20$, maka bisa juga digunakan uji Chi-Square:

```

> # Membuat tabel skor untuk Uji Chi-Square
> Skor <- matrix(c(9, 6, 5, 8), nrow=2, dimnames=list(c("Di atas Median",
"Di bawah Median"), c("Normal", "Sirosis")))
> Skor

```

```

Normal Sirosis
Di atas Median      9      5
Di bawah Median    6      8
>
> chisq.test(Skor)

Pearson's Chi-squared test with Yates' continuity correction

data: Skor
X-squared = 0.57436, df = 1, p-value = 0.4485

```

4. Kriteria Penolakan

H_0 ditolak jika $p - value < \alpha$.

Dari hasil yang diperoleh:

- **p-value = 0.5112462 (Uji Median) atau 0. 4485 (Uji Chi-Square) > $\alpha= 0.05$**

Oleh karena. Dengan demikian, **H_0 tidak ditolak.**

5. Kesimpulan

Jadi, **tidak ada perbedaan** waktu konsentrasi plasma tertinggi antara penderita sirosis hati dan orang normal.

Laporan 10: Uji Mann-Whitney

Contoh 1 PPT Pertemuan 5,6 Hal 26

Dilakukan penelitian untuk mengetahui adakah perbedaan kualitas manajemen antara bank yang dianggap favorit dengan yang tidak favorit. Penelitian menggunakan sampel 12 bank yang dianggap tidak favorit dan 15 bank favorit.

Bank Tidak Favorit	Kualitas	Bank Favorit	Kualitas
1	16	1	19
2	18	2	19
3	10	3	21
4	12	4	25
5	16	5	26
6	14	6	27
7	15	7	23
8	10	8	27
9	12	9	19
10	15	10	19
11	16	11	25
12	11	12	27
		13	23
		14	19
		15	29

Penyelesaian

1. Hipotesis:

- **H0: Tidak ada perbedaan** kualitas manajemen antara Bank favorit & tidak favorit
- **H1: Ada perbedaan** kualitas manajemen antara Bank favorit & tidak favorit

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic:

```
> ## Contoh 1 PPT Pertemuan 5,6 Hal. 26
> BankFavorit<- c(16,18,10,12,16,14,15,10,12,15,16,11,0,0,0)
> BankTidakFavorit<- c(19,19,21,25,26,27,23,27,19,19,25,27,23,19,29)
> data<- data.frame(BankFavorit, BankTidakFavorit)
> data
   BankFavorit BankTidakFavorit
1           16              19
2           18              19
```

```

3      10      21
4      12      25
5      16      26
6      14      27
7      15      23
8      10      27
9      12      19
10     15      19
11     16      25
12     11      27
13      0      23
14      0      19
15      0      29
>
> wilcox.test(BankFavorit, BankTidakFavorit, paired= FALSE, data=data)
   Wilcoxon rank sum test with continuity correction
data:  BankFavorit and BankTidakFavorit
W = 0, p-value = 3.089e-06
alternative hypothesis: true location shift is not equal to 0

Warning message:
In wilcox.test.default(BankFavorit, BankTidakFavorit, paired = FALSE, :
  cannot compute exact p-value with ties

```

4. Kriteria Penolakan

H_0 ditolak jika $p - value < \alpha$.

Dari hasil yang diperoleh:

- $p\text{-value} = 3.089\text{e-}06 < \alpha = 0.05$

Oleh karena. Dengan demikian, **H_0 ditolak.**

5. Kesimpulan

Jadi, **ada perbedaan** kualitas manajemen antara Bank favorit & tidak favorit.

Contoh 2 PPT Pertemuan 5,6 Hal 34

Sebuah perusahaan Bimbel sedang mengembangkan 2 jenis metode belajar baru yang akan diterapkan kepada siswa siswanya. Perusahaan hanya akan menerapkan satu metode belajar yang memiliki pengaruh nilai belajar terbaik. Namun jika kedua metode memiliki tingkat keberhasilan yang sama maka kedua metode akan diterapkan bersama sama. Untuk itu perusahaan uji coba kedua metode belajar tersebut secara bersamaan terhadap 26 siswa sekolah dasar yang dibagi menjadi dua kelompok. Setelah 30 siswa mendapatkan treatment kedua metode belajar dan menyelesaikan ujian maka hasil ujian akan dilakukan penelitian apakah kedua metode memberikan pengaruh hasil nilai yang sama atau berbeda.

No	Metode Hebat	Metode Super
1	50	52
2	60	59
3	80	78
4	30	29
5	70	72
6	80	81
7	90	88
8	60	63
9	99	98
10	78	79
11	89	87
12	67	65
13	87	86

Penyelesaian

1. Hipotesis:

- **H0: Tidak Terdapat perbedaan** nilai rata rata ujian pada data nilai metode super dan metode hebat.
- **H1: Terdapat perbedaan** nilai rata rata ujian pada data nilai metode super dan metode hebat.

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic:

```
> ## Contoh 2 PPT Pertemuan 5,6 Hal. 26
> metodehebat <- c(50,60,80,30,70,80,90,60,99,78,89,67,87)
> metodesuper <- c(52,59,78,29,72,81,88,63,98,79,87,65,86)
> data1<- data.frame (metodehebat, metodesuper)
> data1
  metodehebat metodesuper
1           50          52
2           60          59
3           80          78
4           30          29
5           70          72
```

```

6          80        81
7          90        88
8          60        63
9          99        98
10         78        79
11         89        87
12         67        65
13         87        86
>
> wilcox.test(metodehebat, metodesuper, paired=FALSE, data=data1)
    wilcoxon rank sum test with continuity correction
data:  metodehebat and metodesuper
W = 88, p-value = 0.8776
alternative hypothesis: true location shift is not equal to 0
warning message:
In wilcox.test.default(metodehebat, metodesuper, paired = FALSE, :
  cannot compute exact p-value with ties

```

4. Kriteria Penolakan

H_0 ditolak jika $p - value < \alpha$.

Dari hasil yang diperoleh:

- **p-value = 0.8776 > $\alpha = 0.05$**

Oleh karena. Dengan demikian, **H_0 tidak ditolak.**

5. Kesimpulan

Jadi, **tidak terdapat perbedaan** nilai rata rata ujian pada data nilai metode super dan metode hebat.

Latihan Modul No. 2 Hal. 45

Untuk meningkatkan produktivitas sekelompok petani diberi bantuan saprodi oleh pemerintah. Setelah beberapa tahun ingin diketahui apakah ada perbedaan produktivitas pada petani yang diberi bantuan dan yang tidak diberi bantuan oleh pemerintah. Berikut merupakan data nilai produktivitas yang diperoleh dari dua kelompok petani tersebut.

Petani yang tidak mendapat bantuan				Petani yang mendapat bantuan			
No	Nilai Produktivitas		Ranking	No	Nilai Produktivitas		Ranking
1	60	4	7	1	70	15	14,5
2	70	14	14,5	2	70	16	14,5
3	70	13	14,5	3	80	19	21
4	50	2	2	4	60	8	7
5	60	5	7	5	80	20	21
6	60	6	7	6	90	24	24,5
7	70	12	14,5	7	70	17	14,5
8	70	11	14,5	8	60	9	7
9	50	1	2	9	50	3	2
10	60	7	7	10	60	10	7
				11	70	18	14,5
				12	80	21	21
				13	80	22	21
				14	80	23	21
				15	90	25	24,5
Jumlah		R! = 90		Jumlah		R! = 235	

Penyelesaian

1. Hipotesis:

- **H0: Tidak ada perbedaan** produktivitas pada petani yang diberi bantuan dan yang tidak diberi bantuan oleh pemerintah.
- **H1: Ada perbedaan** produktivitas pada petani yang diberi bantuan dan yang tidak diberi bantuan oleh pemerintah.

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic:

```
> ## Latihan Modul No. 2 Hal. 45
> Produktivitas<-
c(60, 70, 70, 50, 60, 60, 70, 70, 50, 60, 70, 70, 80, 60, 80, 90, 70, 60, 50, 60, 70, 80, 80, 80, 90)
> Kelompok<- c(rep("tidak_dapat_bantuan", 10), rep("dapat_bantuan", 15))
>
```

```

> data <- data.frame(Produktivitas, Kelompok)
> data
  Produktivitas      Kelompok
1           60 tidak_dapat_bantuan
2           70 tidak_dapat_bantuan
3           70 tidak_dapat_bantuan
4           50 tidak_dapat_bantuan
5           60 tidak_dapat_bantuan
6           60 tidak_dapat_bantuan
7           70 tidak_dapat_bantuan
8           70 tidak_dapat_bantuan
9           50 tidak_dapat_bantuan
10          60 tidak_dapat_bantuan
11          70     dapat_bantuan
12          70     dapat_bantuan
13          80     dapat_bantuan
14          60     dapat_bantuan
15          80     dapat_bantuan
16          90     dapat_bantuan
17          70     dapat_bantuan
18          60     dapat_bantuan
19          50     dapat_bantuan
20          60     dapat_bantuan
21          70     dapat_bantuan
22          80     dapat_bantuan
23          80     dapat_bantuan
24          80     dapat_bantuan
25          90     dapat_bantuan
>
> wilcox.test(Produktivitas ~ Kelompok, data = data, paired = FALSE)

Wilcoxon rank sum test with continuity correction

data:  Produktivitas by Kelompok
W = 115, p-value = 0.02357
alternative hypothesis: true location shift is not equal to 0

Warning message:
In wilcox.test.default(x = DATA[[1L]], y = DATA[[2L]], ...) :
  cannot compute exact p-value with ties

```

4. Kriteria Penolakan

H_0 ditolak jika $p - value < \alpha$.

Dari hasil yang diperoleh:

- **p-value = 0.02357 < $\alpha= 0.05$**

Oleh karena. Dengan demikian, **H_0 ditolak.**

5. Kesimpulan

Jadi, **terdapat perbedaan** produktivitas petani yang mendapat bantuan dan tidak mendapat bantuan pemerintah.

BAB 4: PERTEMUAN 6-7 Uji Hipotesis Komparatif k Sampel Dependen dan Independen

1. Uji Cochran (Q-test) (*Dependen*)

PPT Pertemuan 6,7 Hal 5

Dilakukan survey aktivitas warga masyarakat dalam rangka menekan penyebaran penyakit demam berdarah. Angka 0 menyatakan tidak melakukan aktifitas dan angka 1 menyatakan melakukan aktifitas. Data hasil survey seperti disajikan dalam table. Selidikilah apakah terdapat perbedaan banyaknya masyarakat yang melakukan setiap aktivitas dalam menekan penularan demam berdarah! Gunakan tingkat signifikansi $\alpha = 5\%$

Penyelesaian

1. Hipotesis:

- **H0:** *Tidak ada perbedaan* banyaknya masyarakat yang melakukan setiap aktivitas dalam menekan penularan demam berdarah.
- **H1:** *Ada perbedaan* banyaknya masyarakat yang melakukan setiap aktivitas dalam menekan penularan demam berdarah.

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic:

Pertama-tama kita install dulu library *nonpar* yang terdapat fungsi *cochrans.q()* untuk melakukan perhitungan Uji Cochran.

```
> # Uji Cochran (Q-test)
> install.packages("nonpar")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:
```

```
https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/Halo/AppData/Local/R/win-library/4.3'
(as 'lib' is unspecified)
```

```

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.3/nonpar_1.0.2.zip'
Content type 'application/zip' length 42033 bytes (41 KB)
downloaded 41 KB

package 'nonpar' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\Users\Halo\AppData\Local\Temp\RtmpqsoyvQ\downloaded_packages

> Abatisasi<- c(0,0,0,1,0,1,1,0,0,1)
> Abatisasi<- c(0,0,0,1,0,1,1,0,0,1)
> Menutup<- c(1,0,0,1,1,0,1,1,0,1,1)
> Menguras<- c(0,1,0,1,1,0,1,1,1,0,1)
> data<- data.frame(Abatisasi, Menutup, Menguras)
> data
   Abatisasi Menutup Menguras
1          0      1       0
2          0      0       1
3          0      0       0
4          1      1       1
5          0      1       1
6          1      0       0
7          1      1       1
8          1      1       1
9          0      0       1
10         0      1       0
11         1      1       1
> library(nonpar)
> cochrans.q(data)

Cochran's Q Test

H0: There is no difference in the effectiveness of treatments.
HA: There is a difference in the effectiveness of treatments.

Q = 1.333333333333333

Degrees of Freedom = 2

Significance Level = 0.05
The p-value is 0.513417119032592

```

4. Kriteria Penolakan

H_0 ditolak jika $p - value < \alpha$.

Dari hasil yang diperoleh:

- $p\text{-value} = 0.513417119032592 > \alpha = 0.05$

Oleh karena. Dengan demikian, **H_0 tidak ditolak.**

5. Kesimpulan

Jadi, **tidak ada perbedaan** banyaknya masyarakat yang melakukan setiap aktivitas dalam menekan penularan demam berdarah.

2. Uji Friedman (Anava Ranking Two Way) (*Dependen*)

PPT Pertemuan 6,7 Hal 14

Sebuah Perusahaan biscuit ingin meluncurkan empat rasa baru dalam produk biskuitnya. Keempat rasan tersebut terdiri dari rasa coklat, rasa stroberi, rasa keju, dan rasa kelapa. Perusahaan ingin mengetahui bagaimana tanggapan konsumen terhadap keempat rasa tersebut 10 orang diminta untuk mencicipi keempat rasa biskuit tersebut kemudian memberikan nilai untuk setiap rasa yangada. Nilai yang diberikan ditentukan antara 0-100.

No	Coklat	Stroberi	Keju	Kelapa
1	78	80	84	71
2	82	76	85	73
3	81	78	80	70
4	80	77	88	71
5	82	74	86	75
6	83	81	89	70
7	85	78	84	70
8	79	73	85	72
9	82	70	87	73
10	78	71	88	70

Penyelesaian

1. Hipotesis:

- H_0 : Keempat rasa biscuit mempunyai penilaian yang *sama*
- H_1 : *Ada perbedaan* penilaian mengenai keempat rasa biscuit

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic:

```
> # Uji Friedman (Anava Ranking Two Way )
> Coklat<- c(78,82,81,80,82,83,85,79,82,78)
> Stroberi<- c(80,76,78,77,74,81,78,73,70,71)
> Keju<- c(84,85,80,88,86,89,84,85,87,88)
> Kelapa<- c(71,73,70,71,75,70,70,72,73,70)
> data<- data.frame (Coklat, Stroberi, Keju, Kelapa)
> data
   Coklat Stroberi Keju Kelapa
1      78       80    84    71
2      82       76    85    73
3      81       78    80    70
4      80       77    88    71
5      82       74    86    75
```

```

6      83      81      89      70
7      85      78      84      70
8      79      73      85      72
9      82      70      87      73
10     78      71      88      70
> friedman.test(data.matrix(data))

Friedman rank sum test

data: data.matrix(data)
Friedman chi-squared = 24.6, df = 3, p-value = 1.872e-05

```

4. Kriteria Penolakan

H_0 ditolak jika $p\text{-value} < \alpha$.

Dari hasil yang diperoleh:

- $\text{p-value} = 1.872\text{e-}05 < \alpha = 0.05$

Oleh karena. Dengan demikian, **H_0 ditolak.**

5. Kesimpulan

Jadi, *ada perbedaan* penilaian mengenai keempat rasa biscuit.

3. Uji Chi-Square k Sampel (*Independen*)

PPT Pertemuan 6,7 Hal 25

BKKBN ingin meneliti apakah alat kontrasepsi yang disukai masyarakat berbeda berdasarkan kelompok umur peserta KB. Untuk itu dilakukan suatu survei pemakaian alat KB pada 3 kelompok umur peserta KB. Data yang diperoleh sebagai berikut:

Kelompok umur	Alat kontrasepsi yang digunakan			
	Pil KB	Suntik	IUD	Kontrasepsi mantap
20 th -30 th	55	60	65	12
30 th -40 th	33	60	81	35
> 40 th	23	36	45	80

Apakah dapat disimpulkan bahwa populasi-populasi peserta KB pada tiga kelompok umur tersebut homogen dalam hal alat kontrasepsi yang dipakai.

Penyelesaian

1. Hipotesis:

- **H0:** Jenis alat kontraasepsi yang digunakan oleh ketiga kelompok umur peserta KB bersifat *homogen*.
- **H1:** Paling tidak salah satu diantara ketiga kelompok umur peserta KB tersebut menggunakan alat kontrasepsi yang *berbeda*.

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic:

```
> # Uji Chi-Square k Sampel (Independen)
> Kontrasepsi<- matrix(c(55, 33, 23, 60,60,36, 65, 81, 45, 12, 35, 80), nr
ow=3,
+                               dimnames=list(c("20-30", "30-40", "> 40"),
+                               c("Pil KB", "Suntik", "IUD", "Kontrasepsi Mantap")))
> Kontrasepsi
   Pil KB Suntik IUD Kontrasepsi Mantap
20-30    55     60   65            12
30-40    33     60   81            35
> 40     23     36   45            80
> chisq.test(Kontrasepsi)

Pearson's Chi-squared test

data: Kontrasepsi
X-squared = 89.894, df = 6, p-value < 2.2e-16
```

4. Kriteria Penolakan

H_0 ditolak jika $p-value < \alpha$.

Dari hasil yang diperoleh:

- **p-value = 2.2e-16 < $\alpha= 0.05$**

Oleh karena. Dengan demikian, **H_0 ditolak**.

5. Kesimpulan

Jadi, paling tidak salah satu diantara ketiga kelompok umur peserta KB tersebut menggunakan alat kontrasepsi yang **berbeda**.

4. Uji Median Extension (Perluasan Uji Median) (*Independen*)

PPT Pertemuan 6,7 Hal 37

Seorang peneliti pendidikan ingin mempelajari apakah ada hubungan antara tingkat pendidikan ibu dengan banyaknya kunjungan ke sekolah anaknya. Diambil sampel secara random sebanyak 10% dari 440 anak yang terdaftar di sekolah. Dari sampel tersebut didapat nama dari 44 ibu-ibu yang kemudian dijadikan sampel. Hipotesisnya adalah banyaknya ibu ke sekolah bervariasi menurut tingkat pendidikan yang ditamatkannya. Datanya adalah sebagai berikut: (gunakan $\alpha = 5\%$)

Jumlah kunjungan ke sekolah oleh ibu digolongkan menurut tingkat pendidikan

SD	SMP	SMA	P.T	SD	SMP	SMA	P.T
4	2	2	9	2	0	0	4
3	4	0	4	0	2	5	5
0	1	4	2	3	5	2	2
7	6	3	3	5	1	1	2
1	3	8	2	1	2	7	6
	1	6					
		5					
		1					

Penyelesaian

1. Hipotesis:

- **H0: Tidak ada perbedaan** banyaknya kunjungan diantara para Ibu dengan variasi tingkat pendidikannya.
- **H1: Ada perbedaan** banyaknya kunjungan diantara para ibu dengan variasi tingkat pendidikannya.

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic:

```

> # Uji Median Extension (Perluasan Uji Median) (Independen)
> Kunjungan <- c(0,0,1,1,2,3,3,4,5,7,0,1,1,1,2,2,2,3,4,5,6,0,0,1,1,2,2,3,
4,5,5,6,7,8,2,
+           2,2,2,3,4,5,5,6,9)
> Pendidikan<- c(rep("SD", 10), rep("SMP", 11), rep("SMA", 13), rep ("PT",
10))
> data<- data.frame (Kunjungan, Pendidikan)
> data
  Kunjungan Pendidikan
1          0        SD
2          0        SD
3          1        SD
4          1        SD
5          2        SD
6          3        SD
7          3        SD
8          4        SD
9          5        SD
10         7        SD
11         0       SMP
12         1       SMP
13         1       SMP
14         1       SMP
15         2       SMP
16         2       SMP
17         2       SMP
18         3       SMP
19         4       SMP
20         5       SMP
21         6       SMP
22         0      SMA
23         0      SMA
24         1      SMA
25         1      SMA
26         2      SMA
27         2      SMA
28         3      SMA
29         4      SMA
30         5      SMA
31         5      SMA
32         6      SMA
33         7      SMA
34         8      SMA
35         2      PT
36         2      PT
37         2      PT
38         2      PT
39         3      PT
40         4      PT
41         5      PT
42         5      PT
43         6      PT
44         9      PT
> # install.packages("agricolae")
> library("agricolae")
Warning message:
package 'agricolae' was built under R version 4.3.3
> Median.test(data$Kunjungan, data$Pendidikan)

```

The Median Test for data\$Kunjungan ~ data\$Pendidikan

Chi Square = 1.295105 DF = 3 P.Value 0.7302958
Median = 2.5

	Median	r	Min	Max	Q25	Q75
PT	3.5	10	2	9	2	5.00
SD	2.5	10	0	7	1	3.75
SMA	3.0	13	0	8	1	5.00
SMP	2.0	11	0	6	1	3.50

Post Hoc Analysis

Groups according to probability of treatment differences and alpha level.

Treatments with the same letter are not significantly different.

```
data$Kunjungan groups
PT           3.5    a
SMA          3.0    a
SD           2.5    a
SMP          2.0    a
```

Atau menggunakan Chi Square Test:

```
> # Atau
> Kunjungan<- matrix(c(5,5,4,7,7,6,6,4), nrow=2,
+                       dimnames=list(c("Di Atas Rata-Rata",
+                                     "DiBawah Rata-rata"), c("SD", "SMP", "SMA", "PT")))
> Kunjungan
      SD  SMP  SMA  PT
Di Atas Rata-Rata 5   4   7   6
DiBawah Rata-rata 5   7   6   4
> chisq.test(Kunjungan)

Pearson's Chi-squared test

data: Kunjungan
X-squared = 1.2951, df = 3, p-value = 0.7303
```

4. Kriteria Penolakan

H_0 ditolak jika $p - value < \alpha$.

Dari hasil yang diperoleh:

- $p\text{-value} = 0.7303 > \alpha = 0.05$

Oleh karena. Dengan demikian, H_0 tidak ditolak.

5. Kesimpulan

Jadi, **tidak ada perbedaan** banyaknya kunjungan diantara para Ibu dengan variasi tingkat pendidikannya

5. Anova One Way Kruskal-Wallis (*Independen*)

PPT Pertemuan 6,7 Hal 50

Torre *et al.* mencatat adanya perubahan *serotonin* (5-HT) (*platelet*) serebral dan ekstraserebral tikus sesudah pemberian LSD-25 dan *1-methyl-dlysergic acid butanclamide* (UML) secara *intraperitoneal*. Pengukuran yang sama mereka lakukan pada 11 kontrol. Hasil percobaan disajikan pada Tabel di bawah ini. Apakah data ini cukup memberikan bukti untuk menunjukkan adanya perbedaan di antara ketiga perlakuan tersebut ($\alpha=5\%$)? Hitung pula nilai *p-value* (Daniel 1990).

Tabel *serotonin* otak (5-HT), nanogram per gram, pada tiga kelompok anak tikus

Kontrol	340	340	356	386	386	402	402	417	433	495	557
LSD 0.5 mg/kg	294	325	325	340	356	371	385	402			
UML 0.5 mg/kg	263	309	340	356	371	371	402	417			

Penyelesaian

1. Hipotesis:

- **H0: Tidak terdapat perbedaan** yang signifikan dalam kadar serotonin otak (5-HT) di antara tiga kelompok perlakuan (Kontrol, LSD 0.5 mg/kg, dan UML 0.5 mg/kg).
- **H1: Terdapat setidaknya satu kelompok perlakuan yang memiliki perbedaan** signifikan dalam kadar serotonin otak (5-HT) dibandingkan kelompok lainnya.

2. Tingkat Signifikansi: ($\alpha = 0,05$)

3. Uji statistic:

```
> # 5. Anova One Way Kruskal-wallis (Independen)
> Skor<- c(340,340,356, 386, 386, 402, 402, 417, 433, 495, 557, 294, 325,
325, 340, 356, 371, 385, 402, 263, 309, 340, 356, 371, 371, 402, 417)
> Perlakuan <- c(rep("Kontrol", 11), rep("LSD", 8), rep("UML", 8))
> Serotonin<- data.frame(Skor, Perlakuan)
> Serotonin
   Skor Perlakuan
1    340    Kontrol
2    340    Kontrol
3    356    Kontrol
4    386    Kontrol
5    386    Kontrol
6    402    Kontrol
7    402    Kontrol
8    417    Kontrol
9    433    Kontrol
10   495    Kontrol
```

```

11 557 Kontrol
12 294 LSD
13 325 LSD
14 325 LSD
15 340 LSD
16 356 LSD
17 371 LSD
18 385 LSD
19 402 LSD
20 263 UML
21 309 UML
22 340 UML
23 356 UML
24 371 UML
25 371 UML
26 402 UML
27 417 UML
> kruskal.test(Skor~Perlakuan, data = Serotonin)

Kruskal-wallis rank sum test

data: Skor by Perlakuan
Kruskal-Wallis chi-squared = 6.2341, df = 2, p-value = 0.04429

```

4. Kriteria Penolakan

H_0 ditolak jika $p - value < \alpha$.

Dari hasil yang diperoleh:

- $p\text{-value} = 0.04429 < \alpha = 0.05$

Oleh karena. Dengan demikian, **H_0 ditolak.**

5. Kesimpulan

Jadi, **terdapat setidaknya satu kelompok perlakuan yang memiliki perbedaan** signifikan dalam kadar serotonin otak (5-HT) dibandingkan kelompok lainnya.

BAB 5: PERTEMUAN 10 (Laporan 11: Angka Indeks)

Pendahuluan Materi Angka Indeks

Angka indeks adalah alat statistik yang digunakan untuk mengukur perubahan relatif suatu variabel dalam dua periode waktu atau lebih. Biasanya digunakan untuk melihat perubahan harga, jumlah, atau nilai dari suatu produk atau layanan, seperti harga barang dari tahun ke tahun atau indeks saham. Angka indeks membantu kita memahami apakah suatu fenomena mengalami peningkatan atau penurunan dibandingkan dengan periode sebelumnya.

Ada beberapa jenis angka indeks, seperti angka indeks harga, jumlah, dan nilai, yang masing-masing digunakan untuk tujuan yang berbeda. Misalnya, angka indeks harga digunakan untuk mengukur perubahan harga barang dan jasa, sementara angka indeks jumlah digunakan untuk melihat perubahan jumlah produksi atau penjualan.

Secara umum, angka indeks dihitung dengan membandingkan nilai saat ini dengan nilai dasar yang dianggap sebagai acuan, lalu dikalikan dengan 100 agar hasilnya mudah dipahami. Dengan angka indeks, kita dapat dengan cepat melihat tren atau perbandingan yang terjadi dari waktu ke waktu.

Langkah Awal: Menginstal Package IndexNumber

```
> install.packages("IndexNumber")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/Halo/AppData/Local/R/win-library/4.3'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.3/IndexNumber_1.3.2.zip'
Content type 'application/zip' length 75946 bytes (74 KB)
downloaded 74 KB

package 'IndexNumber' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\Users\Halo\AppData\Local\Temp\Rtmp48YNGH\downloaded_packages
> library(IndexNumber)
>
```

Penjelasan Output Terminal:

1. Pesan pertama:

WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:

Ini memberitahukan bahwa kita belum menginstal **Rtools**, yang dibutuhkan untuk membangun paket R dari sumber (source) pada sistem Windows. Rtools adalah alat yang digunakan untuk mengkompilasi kode C/C++ yang diperlukan dalam beberapa paket R. Tapi, ini hanya peringatan, jadi kamu tetap bisa lanjut menggunakan paket yang sudah tersedia dalam bentuk binary (yang sudah terkompilasi).

2. Proses instalasi paket:

Kemudian, R mencoba menginstal paket IndexNumber dari internet dengan mengunduh file ZIP berisi paket tersebut.

- **URL** yang dicoba untuk mengunduh paketnya adalah dari CRAN (repositori utama R untuk paket-paket).
- File yang diunduh adalah IndexNumber_1.3.2.zip, yang ukurannya cukup kecil, sekitar 74 KB.

3. Pesan sukses:

Setelah mengunduh dan memverifikasi integritas paket (dengan memeriksa MD5), akhirnya R berhasil menginstal dan mengekstrak paket tersebut ke direktori lokal kamu, yang ada di C:/Users/Halo/AppData/Local/R/win-library/4.3.

4. Mengaktifkan paket:

Setelah instalasi berhasil, kita menggunakan perintah library(IndexNumber) untuk memuat (load) paket IndexNumber ke dalam sesi R kita, yang berarti sekarang kita bisa mulai menggunakan fungsi-fungsi dari paket tersebut untuk analisis angka indeks.

Bagian 1: Pengantar Angka Indeks

Contoh Perhitungan 1

Soal:

Rata-rata harga per kg beras per bulan di kota Denpasar pada tahun 2009 dan tahun 2010 masing masing adalah Rp 5.505,00 dan Rp 6.541,00 (BPS Provinsi Bali, 2011). Berapa angka indeks apabila harga beras pada tahun 2010 dengan tahun 2009 dibandingkan?

Jawab:

```
> # Contoh soal perhitungan 1
> harga <- c(5505, 6541)
> index.number.chain(harga, "Harga")
  Stages Harga Index number
1      0 5505    100.0000
2      1 6541    118.8193
```

Jadi, hasilnya menunjukkan bahwa harga beras pada tahun 2010 lebih tinggi sekitar 18.82% dibandingkan dengan tahun 2009.

Contoh Perhitungan 2

Soal:

Sebuah grosir beras ingin mengetahui perubahan nilai penjualan beras selama 5 tahun terakhir. Sedangkan data penjualan yang dimilikinya sebagai berikut:

Tahun	Nilai Penjualan (Juta Rupiah)
2007	300
2008	250
2009	350
2010	400
2011	425

Tentukan Angka Indeks untuk setiap tahun dengan tahun dasar 2007, lalu interpretasikan!

Jawab:

```
> # Contoh soal perhitungan 2
> Harga_2008 <- c(300, 250)
> index.number.chain(Harga_2008, "Nilai_Penjualan")
  Stages Nilai_Penjualan Index number
1      0            300    100.00000
2      1            250    83.33333
>
> Harga_2009 <- c(300, 350)
> index.number.chain(Harga_2009, "Nilai_Penjualan")
```

```

Stages Nilai_Penjualan Index number
1      0            300    100.0000
2      1            350    116.6667
>
> Harga_2010 <- c(300, 400)
> index.number.chain(Harga_2010, "Nilai_Penjualan")
  Stages Nilai_Penjualan Index number
1      0            300    100.0000
2      1            400    133.3333
>
> Harga_2011 <- c(300, 425)
> index.number.chain(Harga_2011, "Nilai_Penjualan")
  Stages Nilai_Penjualan Index number
1      0            300    100.0000
2      1            425    141.6667

```

Jadi, pada tahun 2008 terjadi penurunan nilai penjualan sebesar 16,67% dibandingkan tahun 2007. Selain itu, pada tahun 2009, 2010 dan 2011 mengalami peningkatan berturut-turut sebesar 16,67%; 33,33% dan 41,67% dibandingkan tahun 2007.

Contoh Perhitungan 3

Soal:

Harga eceran rata-rata empat bahan kebutuhan pokok per kilogram per bulan di Kota Denpasar tahun 2009 dan tahun 2010.

Bahan Pokok	Harga Rata-rata (Rp/Kg)	
	2009	2010
Beras	5505	6541
Gula Pasir	8355	10628
Daging ayam	23825	25406
Garam	3241	3333
Jumlah	40926	45908

Berapa angka indeks gabungannya?

Jawab:

```

> # Contoh Perhitungan 3
> angka_indeks_gabungan <- c(40926, 45908)
> index.number.chain(angka_indeks_gabungan, "Harga Rata-Rata")
  Stages Harga Rata-Rata Index number
1      0            40926    100.0000
2      1            45908    112.1732

```

Hal ini berarti harga rata-rata gabungan dari komoditi bahan pokok naik sebesar 12,17% dibandingkan tahun 2009.

Bagian 2: Jenis-Jenis Angka Indeks

1. Berdasarkan Penggunaan

a. Indeks Harga (Price Index):

- **Pengertian:** Angka indeks yang digunakan untuk mengukur atau menunjukkan perubahan harga suatu barang atau kelompok barang. Perubahan ini biasanya dinyatakan dalam bentuk persentase kenaikan atau penurunan harga
- **Contoh:** Indeks Harga Konsumen (IHK), yang mengukur perubahan harga barang dan jasa yang biasa dikonsumsi oleh rumah tangga.

Soal:

Sebuah toko kelontong menjual tiga jenis barang utama: beras, minyak goreng, dan telur. Harga satuan ketiga barang tersebut pada tahun 2021 dan 2022 adalah sebagai berikut:

Barang	Harga Tahun 2021 (Rp/kg)	Harga Tahun 2022 (Rp/kg)
Beras	10.000	12.000
Minyak Goreng	15.000	18.000
Telur	25.000/kg	30.000/kg

Hitunglah indeks harga sederhana untuk tahun 2022 dengan tahun 2021 sebagai tahun dasar!

Jawab:

```
> ## Round 2 Jenis-Jenis Angka Indeks
> # Berdasarkan Penggunaan
> # Indeks Harga (Price Index)
> angka_indeks_gabungan <-c(10000+15000+25000, 12000+18000+30000)
> index.number.chain(angka_indeks_gabungan, "Harga rata-rata")
  Stages Harga rata-rata Index number
1      0           50000       100
2      1           60000       120
```

Jadi, indeks harga sederhana untuk tahun 2022 adalah 120. Artinya, harga rata-rata barang di toko tersebut naik 20% dari tahun 2021 ke tahun 2022.

b. Indeks Kuantitas (Quantity Index):

- **Pengertian:** Angka indeks yang digunakan untuk mengukur perubahan dalam jumlah (kuantitas) suatu barang atau kelompok barang. Kuantitas ini bisa meliputi produksi, konsumsi, atau penjualan.
- **Contoh:** -Indeks Produksi Beras, untuk mengukur perubahan jumlah produksi beras dalam periode tertentu.

Soal:

Menggunakan data Indeks Nilai, hitunglah indeks kuantitas produksi perusahaan tersebut untuk tahun 2022 dengan tahun 2021 sebagai tahun dasar!

Produk	Kuantitas (unit)	Harga Satuan (Rp)	Kuantitas (unit)	Harga Satuan (Rp)
A	1000	10.000	1200	12.000
B	800	15.000	900	18.000

Jawab:

```
> # Indeks Kuantitas (Quantity Index)
> produk_A <- c(1000, 1200)
> index.number.chain(produk_A, "Kuantitas A")
  Stages Kuantitas A Index number
1      0          1000          100
2      1          1200          120
>
> produk_B <- c(800, 900)
> index.number.chain(produk_B, "Kuantitas B")
  Stages Kuantitas B Index number
1      0          800          100.0
2      1          900          112.5
```

c. Indeks Nilai (Value Index):

- **Pengertian:** Angka indeks yang digunakan untuk melihat perubahan nilai suatu barang atau kelompok barang, baik yang dihasilkan, diimpor, maupun dieksport. Nilai ini biasanya merupakan hasil perkalian antara harga dan kuantitas.
- **Contoh:** Indeks Nilai Ekspor Kopra, untuk menunjukkan perubahan nilai ekspor kopra dalam periode tertentu.

Soal:

Sebuah perusahaan memproduksi dua jenis produk, A dan B. Data produksi dan harga jual per unit pada tahun 2021 dan 2022 adalah sebagai berikut:

Produk	Kuantitas (unit)	Harga Satuan (Rp)	Kuantitas (unit)	Harga Satuan (Rp)
A	1000	10.000	1200	12.000
B	800	15.000	900	18.000

**Hitunglah indeks nilai produk perusahaan tersebut untuk tahun 2022 dengan tahun 2021 sebagai tahun dasar!
Hitung perubahan harga masing-masing barang :**

Jawab:

```
> # Indeks Nilai (Value Index)
> produk_A <- c(1000*10000, 1200*12000)
> index.number.chain(produk_A, "Nilai A")
  Stages Nilai A Index number
1      0 10000000          100
2      1 14400000          144
```

```

>
> produk_B <- c(800*15000, 900*18000)
> index.number.chain(produk_B, "Nilai B")
  Stages   Nilai B Index number
1          0 12000000           100
2          1 16200000           135

```

2. Berdasarkan Cara Penentuan

a. Indeks Tidak Tertimbang

- **Pengertian :** Indeks Tidak Tertimbang adalah angka indeks yang dibuat tanpa mempertimbangkan faktor-faktor yang memengaruhi perubahan naik atau turunnya nilainya.
- **Contoh:** Indeks sederhana yang hanya menghitung rata-rata perubahan harga dari beberapa barang.

Soal:

Harga bahan makanan di tahun 2022 dibandingkan dengan tahun 2020:

Beras: Rp10.000 → Rp12.000
Gula: Rp8.000 → Rp10.000
Minyak: Rp15.000 → Rp18.000

$$\text{Indeks Tidak Tertimbang} = \frac{\text{Jumlah Harga Saat Ini}}{\text{Jumlah Harga Waktu Dasar}} \times 100$$

$$\text{Indeks} = \frac{12.000 + 10.000 + 18.000}{10.000 + 8.000 + 15.000} \times 100 = \frac{40.000}{33.000} \times 100 = 121,21$$

Interpretasi: Rata-rata harga bahan makanan naik 21,21%.

Jawab:

```

> # Berdasarkan Penentuan
> # Indeks Tidak Tertimbang
> Indeks_Tidak_Tertimbang <- c(10000+8000+15000, 12000+10000+18000)
> index.number.chain(Indeks_Tidak_Tertimbang, "Harga Makanan")
  Stages Harga Makanan Index number
1          0            33000      100.0000
2          1            40000      121.2121

```

b. Indeks Tertimbang

- **Pengertian :** Indeks Tertimbang adalah angka indeks yang dihitung dengan mempertimbangkan faktor-faktor yang memengaruhi perubahan nilainya.
- **Contoh:** Indeks Harga Konsumen (IHK) yang menggunakan bobot berdasarkan pengeluaran konsumen pada masingmasing kelompok barang.

Soal:

Bahan makanan yang sama, tetapi ditambah informasi jumlah konsumsi:

Beras: Harga (2020): Rp10.000 → Harga (2022): Rp12.000, Konsumsi: 50 kg

Gula: Harga (2020): Rp8.000 → Harga (2022): Rp10.000, Konsumsi: 30 kg

Minyak: Harga (2020): Rp15.000 → Harga (2022): Rp18.000, Konsumsi: 20 liter

$$\text{Indeks Tertimbang} = \frac{\sum(\text{Harga Saat Ini} \times \text{Konsumsi})}{\sum(\text{Harga Waktu Dasar} \times \text{Konsumsi})} \times 100$$

$$\text{Indeks} = \frac{(12.000 \times 50) + (10.000 \times 30) + (18.000 \times 20)}{(10.000 \times 50) + (8.000 \times 30) + (15.000 \times 20)} \times 100$$

$$\text{Indeks} = \frac{600.000 + 300.000 + 360.000}{500.000 + 240.000 + 300.000} \times 100 = \frac{1.260.000}{1.040.000} \times 100 = 121,15$$

Interpretasi: Harga rata-rata naik 21,15%, dengan mempertimbangkan jumlah konsumsi.

Jawab:

```
> # Indeks Tertimbang
> Indeks_Tertimbang <- c((1000*50)+(8000*30)+(15000*20), (12000*50)+(10000*30)+(18000*20))
> index.number.chain(Indeks_Tertimbang, "Harga Makanan")
  Stages Harga Makanan Index number
1      0      590000     100.0000
2      1     1260000    213.5593
```

c. Indeks Rantai

- **Pengertian :** Indeks Rantai adalah angka yang menunjukkan perbandingan antara suatu waktu tertentu dengan waktu lainnya yang digunakan sebagai dasar, dan disusun berdasarkan urutan waktu yang berkelanjutan

Soal:

Mengukur perubahan dari satu tahun ke tahun berikutnya secara bertahap.

Contoh:

Harga gandum:

Tahun 2020: Rp10.000

Tahun 2021: Rp12.000

Tahun 2022: Rp15.000

Rumus: $\text{Indeks Rantai} = \text{Indeks Tahun Sebelumnya} \times \frac{\text{Harga Tahun Sekarang}}{\text{Harga Tahun Sebelumnya}}$

Langkah perhitungan:

Indeks 2021 (dengan dasar tahun 2020): $100 \times \frac{12.000}{10.000} = 120$

Indeks 2022 (dengan dasar tahun 2021): $100 \times \left(\frac{15.000}{12.000}\right) = 125$

Interpretasi: Harga gandum naik menjadi 125% dari harga di tahun 2020, dengan kenaikan bertahap.

Jawab:

```
> # Indeks Rantai
> indeks_Berantai <- c(10000, 12000, 15000)
> index.number.chain(indeks_Berantai, "Harga")
  Stages Harga Index number
```

1	0	10000	100
2	1	12000	120
3	2	15000	125

Bagian 3: Metode Penghitungan Angka Indeks

a. Indeks Harga Relatif Sederhana

Indeks Harga Relatif Sederhana adalah angka indeks yang digunakan untuk mengukur perubahan harga suatu barang tunggal dari waktu dasar (periode 0) ke waktu tertentu (periode t).

$$I_{t,0} = \frac{P_t}{P_0} \times 100\%$$

- Dengan $I_{td,0}$: Indeks Harga Saat Waktu ke t dengan waktu dasar 0
- P_t : Harga pada saat periode t
- P_0 : Harga pada saat waktu dasar

Soal:

Tabel 6.1 Harga Barang menurut Jenisnya selama 2004 – 2006

Jenis Barang (1)	Harga		
	2004 (2)	2005 (3)	2006 (4)
	A	100	150
B	200	250	300
C	500	600	700
D	400	500	600
Jumlah	1.200	1.500	1.800

Hitunglah Angka Indeks barang A untuk tahun 2005 dan 2006 dengan tahun dasar 2004!

Jawab:

```
> library(IndexNumber)
> ## Round 3: Metode Penghitungan Angka Indeks
> # Indeks Harga Relatif Sederhana
>
> Harga_Barang_A <- c(100, 150, 200)
> Harga_Barang_B <- c(200, 250, 300)
> Harga_Barang_C <- c(500, 600, 700)
> Harga_Barang_D <- c(400, 500, 600)
> index.number.serie(Harga_Barang_A, name="Harga")
  Stages Harga Index number
1      0    100        100
2      1    150        150
3      2    200        200
```

```

> index.number.serie(Harga_Barang_B, name="Harga")
  Stages Harga Index number
1      0    200      100
2      1    250      125
3      2    300      150
> index.number.serie(Harga_Barang_C, name="Harga")
  Stages Harga Index number
1      0    500      100
2      1    600      120
3      2    700      140
> index.number.serie(Harga_Barang_D, name="Harga")
  Stages Harga Index number
1      0    400      100
2      1    500      125
3      2    600      150

```

b. Indeks Kuantitas Relatif Sederhana

Indeks Kuantitatif cukup sederhana digunakan untuk mengamati perubahan jumlah barang dan jasa. Indeks ini dihitung tanpa mempertimbangkan bobot untuk setiap komoditas. Rumus yang digunakan untuk Indeks Kuantitatif sederhana adalah:

$$I_{t,0} = \frac{q_t}{q_0} \times 100\%$$

I_q = Indeks kuantitas relatif sederhana

q_t = kuantitas tahun t

q_0 = kuantitas tahun dasar (0)

Soal:

Tabel 6.2 Produksi Tanaman bahan Makanan Menurut Jenis (tahun 2004 - 2006)

Jenis Barang (1)	Harga		
	2004 (2)	2005 (3)	2006 (4)
Padi sawah	24.732	27.993	30.989
Padi ladang	1.551	1.659	1.785
Jagung	3.606	3.994	4.509
Ketela	13.751	13.774	13.301

Hitunglah Angka Indeks Kuantitas Padi Sawah untuk tahun 2005 dan 2006 dengan tahun dasar 2004

Jawab:

```

> # Indeks Kuantitas Relatif Sederhana
> Padi sawah<- c(24732, 27993, 30989)
> Padi Ladang<- c(1551, 1659, 1785)
> Jagung<- c(3606, 3994, 4509)
> Ketela<- c(13751, 13774, 13301)
> index.number.serie (Padisawah, name="Kuantitas")

```

```

Stages Kuantitas Index number
1      0      24732    100.0000
2      1      27993    113.1853
3      2      30989    125.2992
> index.number.serie(PadiLadang, name="Kuantitas")
Stages Kuantitas Index number
1      0      1551     100.0000
2      1      1659     106.9632
3      2      1785     115.0870
> index.number.serie(Jagung, name="Kuantitas")
Stages Kuantitas Index number
1      0      3606     100.0000
2      1      3994     110.7598
3      2      4509     125.0416
> index.number.serie(Ketela, name="Kuantitas")
Stages Kuantitas Index number
1      0      13751    100.00000
2      1      13774    100.16726
3      2      13301    96.72751

```

c. Angka Indeks Agregat Sederhana

Angka Indeks Agregat Sederhana adalah salah satu metode untuk menghitung indeks yang melibatkan lebih dari satu barang atau jasa dengan memperhitungkan harga masing-masing komponen dalam tahun tertentu dibandingkan dengan tahun dasar. Rumusnya dinyatakan sebagai berikut:

$$IA = \frac{\sum P_n}{\sum P_0} \times 100$$

Dimana :

P_n = harga tahun tertentu

P_0 = harga tahun dasar

Soal:

Harga eceran per satuan lima jenis barang per bulan di Kota Denpasar tahun 2009 dan tahun 2010, disajikan sebagai berikut:

Jenis Barang	Satuan	Harga Per Satuan (Rp)	
		2009	2010
1 Minyak Goreng	Liter	12.518	12.529
2 Gula Pasir	Kg	8.355	10.628
3 Beras	Kg	5.505	6.541
4 Garam	Kg	3.241	3.333
5 Daging Ayam Ras	Kg	23.825	25.406

Sumber: BPS Provinsi Bali, 2011

Berdasarkan data tersebut, hitunglah indeks harga agregatif tidak tertimbang kelima barang tersebut, dan berikan makna terhadap nilai angka indeks yang diperoleh.

Jawab:

```
> # c. Angka Indeks Agregat Sederhana
> Komoditi<- matrix(c(12518, 8355, 5505, 3241, 23825, 12529, 10628, 6541,
+ 3333, 25406), nrow=2, byrow=TRUE)
> aggregated.index.number(Komoditi, "serie", "BDutot", "Harga", opt.plot=F
+ ALSE, opt.summary=FALSE)
```

Aggregate index number

Bradstreet-Dutot

	Stages	Harga 1	Harga 2	Harga 3	Harga 4	Harga 5	Agg. index number
1	0	12518	8355	5505	3241	23825	100.0000
2	1	12529	10628	6541	3333	25406	109.3425

d. Angka Indeks Kuantitas Agregat Sederhana

Angka Indeks Kuantitas Agregat Sederhana adalah metode perhitungan indeks yang digunakan untuk membandingkan kuantitas suatu kelompok barang dan jasa antara periode tertentu (tahun t) dengan periode dasar (tahun 0). Rumusnya dinyatakan sebagai:

$$IKA = \frac{\sum K_t}{\sum K_0} \times 100$$

Dimana :

IKA = Indeks kuantitas agregat sederhana

K_t = kuantitas tahun t

K_0 = kuantitas tahun 0

Soal:

Jenis Barang	Banyaknya Produksi (Ton)	
	2010	2011
1 Bawang Merah	14.684	20.875
2 Bawang Putih	4.979	15.931
3 Bawang Daun	652	1.294
4 Kentang	2.261	5.107
5 Kubis	14.787	54.415
6 Sawi	5.743	13.882
7 Kacang Merah	159	17.051
8 Kacang Panjang	120	100
9 Terong	80	120

Hitunglah indeks rata-rata produksi (kuantitas) gabungan sayur mayur tersebut pada tahun 2011 dengan waktu dasar tahun 2010

Jawab:

```

> # d. Angka Indeks Kuantitas Agregat Sederhana
> Quantity<- matrix(c(14684, 4979, 652, 2261, 14787, 5743, 159, 120, 80, 2
0875, 15931, 1294, 5107, 54415, 13882, 17051, 100, 120), nrow=2, byrow=TRUE)
> aggregated.index.number (Quantity, "serie", "BDutot", "Kuantitas", opt.p
lot=FALSE, opt.summary=FALSE)

```

Aggregate index number

Bradstreet-Dutot

	Stages	Kuantitas 1	Kuantitas 2	Kuantitas 3	Kuantitas 4	Kuantitas 5	Kuantitas 6	Kuantitas 7	Kuantitas 8	Kuantitas 9	
1	0	14684	4979	652	2261	14787					
	5743	159	120	80							
2	1	20875	15931	1294	5107	54415					
	13882	17051	100	120							
	Agg. index number										
1		100.0000									
2		296.2729									

e. Angka Indeks Nilai Kuantitas Agregat Sederhana

Indeks Nilai Agregat Sederhana adalah metode perhitungan yang menunjukkan perubahan nilai sekelompok barang atau jasa pada suatu periode tertentu dibandingkan dengan periode dasar. Nilai di sini diperoleh dari hasil perkalian harga (H) dengan kuantitas (K).

$$In = \frac{V_t}{V_0} \times 100 = \frac{H_t K_0}{H_0 K_t} \times 100$$

dimana :

In = indeks nilai

V_t = nilai tahun t (H_tK_t)

V₀ = nilai tahun 0 (H₀K₀)

Soal:

Tabel 11.7 Harga dan Kuantitas Empat Jenis Barang yang Dikonsusensi di Daerah "A" Pada Tahun 2010 dan Tahun 2011.

Jenis Barang	2010		2011	
	Harga/ unit (Rp)	Kuantitas (Unit)	Harga/unit (Rp)	Kuantitas (Unit)
A	8.000	4	10.000	6
B	5.000	2	8.000	3
C	2.000	1	3.000	2
D	4.500	6	5.000	6

Sumber : data hipotetis

Berdasarkan data di atas, hitunglah indeks nilai gabungan barang-barang tersebut pada tahun 2011 dengan tahun dasar 2010.

Jawab:

```
> # e. Angka Indeks Nilai Kuantitas Agregat Sederhana
> Harga<- matrix(c(8000, 5000, 2000, 4500, 10000, 8000, 3000, 5000), nrow=2, byrow=TRUE)
> Quantity<- matrix(c(4, 2, 1, 6, 6, 3, 2, 6), nrow=2, byrow=TRUE)
> Nilai<- Harga * Quantity
> aggregated.index.number (Nilai, "serie", "BDUTOT", "Nilai", opt.plot=FALSE, opt.summary=FALSE)
```

Aggregate index number

Bradstreet-Dutot

	Stages	Nilai 1	Nilai 2	Nilai 3	Nilai 4	Agg. index number
1	0	32000	10000	2000	27000	100.0000
2	1	60000	24000	6000	30000	169.0141

f. Angka Indeks Harga Tertimbang

Angka Indeks Harga Tertimbang: digunakan untuk menghitung angka indeks yang melibatkan berbagai jenis barang atau jasa. Karena setiap barang atau jasa memiliki tingkat kepentingan atau manfaat yang berbeda, maka diperlukan pembobotan untuk menyesuaikan perannya dalam perhitungan.

Berikut adalah 6 metode nya :

- Metode Laspeyres
- Metode Paasche
- Metode Fisher

- d. Metode Drobisch
- e. Metode Marshall-Edgeworth
- f. Metode Walsh

Soal:

Data mengenai harga dan kuantitas produksi empat jenis barang di Provinsi "X" disajikan dalam Tabel 11.10 (data hipotetis):

Tabel 11.10 Harga dan Kuantitas Produksi Empat Jenis Barang di Provinsi "X" Tahun 2010 – 2011

Jenis Barang	Harga/Unit (Rp)		Kuantitas Produksi (Unit)	
	2010	2011	2010	2011
A	500	525	2	4
B	800	900	5	6
C	600	700	3	4
D	300	400	10	15

Hitunglah indeks harga agregatif tertimbang barang-barang tersebut pada tahun 2011 dengan tahun dasar 2010

- (a) Dengan metode Laspeyres.
- (b) Dengan metode Paasche.
- (c) Dengan metode Irving Fisher.
- (d) Dengan metode Drobish.
- (e) Dengan metode Marshall-Edgeworth.

Jawab:

(a) Dengan metode Laspeyres.

```
> # f. Angka Indeks Harga Tertimbang
> # (a) Dengan metode Laspeyres.
> Harga<- matrix(c(500, 525, 800, 900, 600, 700, 300, 400), ncol=4)
> Kuantitas<- matrix (c(2,4, 5, 6, 3, 4, 10, 15), ncol=4)
> laspeyres.index.number(Harga, Kuantitas, name="biaya", opt.plot=FALSE, opt.summary=F)
```

Laspeyres index number

	Stages	biaya 1	biaya 2	biaya 3	biaya 4	Agg. index number
1	0	500	800	600	300	100.0000
2	1	525	900	700	400	118.8776

>

(b) Dengan metode Paasche.

```
> # (b) Dengan metode Paasche.
> Harga<- matrix(c(500, 525, 800, 900, 600, 700, 300, 400), ncol=4)
> Kuantitas<- matrix (c(2,4, 5, 6, 3, 4, 10, 15), ncol=4)
> paasche.index.number(Harga, Kuantitas, name="biaya", opt.plot=FALSE, opt.summary=F)
```

Paasche index number

```

1 Stages biaya 1 biaya 2 biaya 3 biaya 4 Agg. index number
2      0      500      800      600      300           100.0000
2      1      525      900      700      400           118.9781

```

>

(c) Dengan metode Irving Fisher.

```

> # (c) Dengan metode Irving Fisher.
> Harga<- matrix(c(500, 525, 800, 900, 600, 700, 300, 400), ncol=4)
> Kuantitas<- matrix (c(2,4, 5, 6, 3, 4, 10, 15), ncol=4)
> fisher.index.number(Harga, Kuantitas, name="biaya", opt.plot=FALSE, opt.summary=FALSE)

```

Laspeyres index number

Paasche index number

Fisher index number

```

1 Stages biaya 1 biaya 2 biaya 3 biaya 4 Agg. index number
2      0      500      800      600      300           100.0000
2      1      525      900      700      400           118.9278

```

(d) Dengan metode Drobish.

(e) Dengan metode Marshall-Edgeworth

```

> # (e) Dengan metode Marshall-Edgeworth.
> Harga<- matrix(c(500, 525, 800, 900, 600, 700, 300, 400), ncol=4)
> Kuantitas<- matrix (c(2,4, 5, 6, 3, 4, 10, 15), ncol=4)
> edgeworth.index.number(Harga, Kuantitas, name="biaya", opt.plot=FALSE, opt.summary=FALSE)

```

Edgeworth index number

```

1 Stages biaya 1 biaya 2 biaya 3 biaya 4 Agg. index number
2      0      500      800      600      300           100.0000
2      1      525      900      700      400           118.9362

```

Bagian 4: Indeks Berantai

Angka Indeks Berantai adalah angka indeks yang dihitung dengan cara menghubungkan perubahan relatif dari suatu periode ke periode berikutnya secara bertahap. Indeks ini digunakan untuk mengukur perubahan suatu variabel dari waktu ke waktu, dengan setiap periode dianggap sebagai dasar bagi periode berikutnya.

$$I_{t,t-1} = \frac{P_t}{P_{t-1}} \times 100\%$$

Dengan

P_t : Harga barang pada tahun ke-t

P_{t-1} : Harga barang 1 tahun sebelum tahun ke-t

$I_{t,t-1}$: Angka Indeks Harga pada tahun ke-t dengan tahun dasar $t - 1$

Soal:

- ✓ Diberikan data mengenai harga beras di kota Denpasar periode 2007- 2010.
Tentukan Indeks Harga biasa dan Indeks Harga Berantai dari data di bawah ini!

Tahun	Harga Beras per Kilogram (Rp)
2007	4900
2008	5154
2009	5505
2010	6541

Jawab:

```
> ## Round 4: Jenis-Jenis Angka Indeks
> Harga<- c(4900, 5154, 5505, 6541)
> index.number.chain(Harga, "Harga")
  Stages Harga Index number
1      0   4900    100.0000
2      1   5154    105.1837
3      2   5505    106.8102
4      3   6541    118.8193
```

Bagian 5: Pergeseran Tahun Dasar

Soal:

Berikut diberikan rangkaian angka indeks dengan tahun dasar 2005 yang disajikan pada Tabel berikut. Tentukan rangkaian angka indeks dengan dasar 2010!

Tahun	Indeks dengan Dasar 2005
2005	100
2006	120
2007	130
2008	142
2009	150
2010	200
2011	250
2012	300

Jawab:

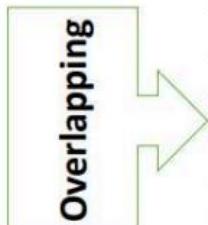
```
> ## Bagian 5: Pergeseran Tahun Dasar
> indeks_awal <- c(100, 120, 130, 142, 150, 200, 250, 300) # Indeks awal dengan tahun dasar 2005
> tahun_dasar_baru <- indeks_awal [6] # Indeks tahun 2010
> indeks_baru <- (indeks_awal / tahun_dasar_baru)*100 # Hitung indeks baru
> indeks_baru # Menampilkan hasil
[1]  50  60  65  71  75 100 125 150
>
```

Bagian 6: Merangkai Angka Indeks (Splicing)

Merangkai Angka Indeks (Splicing) adalah teknik untuk menggabungkan dua deret indeks dengan periode dasar berbeda agar menjadi satu rangkaian yang konsisten. Hal ini dilakukan agar data dari berbagai periode dasar tetap bisa dibandingkan.

Soal:

Diketahui dua rangkaian angka indeks dengan tahun dasar yang berbeda yaitu rangkaian yang pertama tahun dasarnya 2004 dan yang kedua tahun dasarnya 2007. Gabunglah kedua rangkaian angka indeks tersebut menjadi sebuah rangkaian dengan tahun dasar 2007 dan 2004!



Tahun	Indeks tahun dasar 2004 (%)	Indeks tahun dasar 2007 (%)
2004	100	
2005	130	
2006	140	
2007	150	100
2008		120
2009		130
2010		135
2011		140

Jawab:

```
> ## Bagian 3: Merangkai Angka Indeks (Splicing)
> #Backward Splicing
> #Data Indeks Lama dan Baru
> indeks_lama <- c(100, 130, 140, 150) # Tahun dasar 2004
> indeks_baru <- c(100, 120, 130, 135, 140) # Tahun dasar 2007 # Rasio splicing (menghubungkan tahun 2010 dari kedua rangkaian)
> rasio_splicing <- indeks_lama [4] / indeks_baru [1] # Tahun 2007 adalah titik penghubung
> # Sesuaikan Indeks Baru ke Tahun Dasar Lama (2004)
> indeks_baru_spliced <- indeks_baru * rasio_splicing
> #Gabungkan kedua rangkaian
> indeks_tergabung <- c(indeks_lama, indeks_baru_spliced[-1]) # Hindari pengulangan tahun 2007
> #Tampilkan hasil
> indeks_tergabung
[1] 100.0 130.0 140.0 150.0 180.0 195.0 202.5 210.0

> #Forward splicing
> #Rasio splicing (menghubungkan tahun 2007 dari kedua rangkaian)
> rasio_splicing <- indeks_baru [1] / indeks_lama [4] # Tahun 2007 adalah titik penghubung
> # Sesuaikan Indeks Lama ke Tahun Dasar Baru (2007)
> indeks_lama_spliced <- indeks_lama * rasio_splicing
> # Gabungkan kedua rangkaian
> indeks_tergabung <- c(indeks_lama_spliced [-length (indeks_lama)], indeks_baru)
```

```
> #Tampilkan hasil indeks_tergabung
```

Bagian 7: Angka Indeks untuk Proses Deflasi

Soal:

|**Diberikan data upah nominal pada tahun 2007-2011 dalam Rupiah. Hitunglah upah real untuk tahun 2007-2011!**

Tahun	Upah Nominal	IHK dengan tahun dasar 2007
2007	400.000	100
2008	500.000	80
2009	600.000	125
2010	750.000	200
2011	800.000	320

Jawab:

```
> ## Bagian 7: Angka Indeks untuk Proses Deflasi
> #Angka Indeks untuk proses deflasi
> upah_nominal <- c(500000, 600000, 750000, 800000)
> ihm <- c(80, 125, 200, 320)
> upah_reil <- (upah_nominal / ihm) * 100
> upah_reil
[1] 625000 480000 375000 250000
```

Bagian 8: Daya Beli Mata Uang

Daya beli mata uang menunjukkan seberapa banyak barang atau jasa yang dapat dibeli dengan uang pada suatu tahun dibandingkan dengan tahun dasar. Daya beli berbanding terbalik dengan Indeks Harga Konsumen (IHK). Artinya, jika IHK naik, daya beli mata uang akan turun. Sebagai contoh, jika IHK meningkat tiga kali lipat, maka daya beli uang akan turun menjadi sepertiga dari sebelumnya.

Soal:

Di bawah ini diberikan angka indeks mengenai harga bahan kebutuhan pokok untuk keperluan hidup di sebuah kota di Indonesia dengan 2007 = 100

Tahun	2005	2006	2007	2008	2009	2010	2011
Indeks	80	95	100	120	154	170	425

Tentukan daya beli setiap rupiah untuk masing-masing tahun ditinjau dari bahan kebutuhan pokok dinyatakan dengan rupiah!

Jawab:

```
> ## Bagian 8: Daya Beli Mata Uang
> #Daya Beli Mata Uang
> Ihk <- c(80, 95, 100, 120, 154, 170, 425)
> ihktahundasar<- c(100)
> daya_beli <- (Ihk / ihktahundasar)
> daya_beli
[1] 0.80 0.95 1.00 1.20 1.54 1.70 4.25
```

>

BAB 6: PERTEMUAN 11 (Laporan 12: Time Series)

DEFINISI UMUM

Runtun Waktu (Time Series) adalah sekumpulan data observasi yang variabelnya diukur dalam urutan periode waktu, misalnya bulanan, caturwulan, tahunan, dan seterusnya. Time series atau runtun waktu adalah himpunan observasi data terurut dalam waktu (Hanke&Winchern, 2005: 58).

Metode time series adalah metode peramalan dengan menggunakan analisa pola hubungan antara variabel yang akan dipekirakan dengan variabel waktu. Peramalan suatu data time series perlu memperhatikan tipe atau pola data. Secara umum terdapat empat macam pola data time series, yaitu horizontal, trend, musiman, dan siklis (Hanke dan Wichren). Pola horizontal merupakan kejadian yang tidak terduga dan bersifat acak, tetapi kemunculannya dapat mempengaruhi fluktuasi data time series.

Pola trend merupakan kecenderungan arah data dalam jangka panjang, dapat berupa kenaikan maupun penurunan. Pola musiman merupakan fluktuasi dari data yang terjadi secara periodik dalam kurun waktu satu tahun, seperti triwulan, kuartalan, bulanan, mingguan, atau harian. Sedangkan pola siklis merupakan fluktuasi dari data untuk waktu yang lebih dari satu tahun.

(sumber : <https://eprints.uny.ac.id/8326/3/BAB2-06305149010.pdf>)

TUJUAN

1. Mengenali Pola

Analisis time series membantu dalam identifikasi pola dan tren dalam data. Dengan menganalisis data historis, pengguna dapat memahami bagaimana variabel berubah dari waktu ke waktu dan mendeteksi pola berulang seperti tren, musiman, atau siklus.

2. Peramalan (Forecasting)

Salah satu tujuan utama dari analisis time series adalah untuk melakukan forecasting atau peramalan. Ini memungkinkan organisasi untuk memperkirakan nilai di masa depan berdasarkan pola yang teramat dalam data historis. Forecasting ini sangat berguna dalam banyak sektor, termasuk penjualan dan pasar keuangan.

(sumber : <https://romatampubolon.com/metode-runtut-waktu-time-series/>)

3. Memudahkan Pengambilan Keputusan

Analisis time series memberikan informasi yang diperlukan untuk pengambilan keputusan yang lebih baik. Dengan memahami tren dan pola dalam data, perusahaan dapat merencanakan strategi yang lebih efektif dan responsif terhadap perubahan yang mungkin terjadi di masa depan.

4. Pemahaman Penyebab Tren

Dengan menganalisis data time series, organisasi dapat menggali lebih dalam untuk memahami penyebab mendasar dari tren atau pola yang teridentifikasi.

5. Analisis Intervensi

Analisis ini juga dapat digunakan untuk mengevaluasi dampak dari intervensi tertentu atau perubahan kebijakan terhadap data time series. Hal ini membantu dalam memahami bagaimana faktor eksternal mempengaruhi tren yang ada. Ini memberikan wawasan yang lebih baik tentang dinamika yang mempengaruhi data sepanjang waktu. Secara keseluruhan, analisis time series adalah alat yang sangat berguna untuk mendapatkan wawasan dari data historis dan membuat prediksi yang informatif untuk masa depan, sehingga membantu organisasi dalam mengoptimalkan sumber daya dan strategi mereka.

Persyaratan data

- Data harus terurut berdasarkan periode waktu
- Tidak boleh ada waktu atau periode yang terlewat/kosong
- Data tidak boleh ada yang null/missing
- Analisis runtun waktu (time series) memerlukan data historis minimal 50 data runtun waktu

LANGKAH PENGUJIAN TIME SERIES

LANGKAH PENGUJIAN TIME SERIES SECARA MANUAL

1. Visualisasi Data Buat grafik dari data time series untuk mengidentifikasi pola, tren, dan musiman. Ini memberikan gambaran awal tentang sifat data.
2. Uji Stasionaritas Lakukan uji stasionaritas dengan cara :
 - Autokorelasi : Hitung fungsi autokorelasi (ACF) untuk melihat hubungan antara nilai saat ini dan nilai sebelumnya.
 - Tes Statistik : Gunakan uji seperti Augmented Dickey-Fuller (ADF) atau KwiatkowskiPhillips-Schmidt-Shin (KPSS) untuk menguji stasionaritas secara formal.
3. Transformasi Data Jika data tidak stasioner, lakukan transformasi seperti :
 - Differencing : Menghitung perbedaan antara nilai saat ini dan nilai sebelumnya.
 - Log Transformasi : Menerapkan logaritma untuk mengurangi varians.
4. Identifikasi Model Pilih model yang sesuai (misalnya ARIMA, SARIMA) berdasarkan karakteristik data yang telah dianalisis.
5. Estimasi Parameter Hitung parameter model menggunakan metode seperti Maximum Likelihood Estimation (MLE).
6. Evaluasi Model Bandingkan hasil model dengan data aktual untuk menilai akurasi prediksi.
7. Forecasting Lakukan ramalan berdasarkan model yang telah dibangun dan evaluasi hasilnya.

LANGKAH PENGUJIAN TIME SERIES MENGGUNAKAN RSTUDIO

1. Persiapan

Data Siapkan data dalam format CSV atau Excel dan pastikan sudah terorganisir dengan baik.

2. Impor Data ke RStudio

```
library(readr)
```

```
data <- read_csv("path_to_your_file.csv")
```

3. Visualisasi Data

```
library(ggplot2)
```

```
ggplot(data, aes(x = Date, y = Value)) + geom_line() + labs(title = "Time Series Data")
```

4. Uji Stasionaritas

```
library(tseries)
```

```
adf_test <- adf.test(data$Value)
```

```
print(adf_test)
```

Jika p-value lebih besar dari 0.05, data tidak stasioner dan perlu dilakukan transformasi.

1. Evaluasi Model Periksa residuals dari model:

```
checkresiduals(auto_model)
```

2. Forecasting

```
forecasted_values <- forecast(auto_model, h = 10)
```

```
# h adalah jumlah periode ke depan
```

```
plot(forecasted_values)
```

KOMPONEN TIME SERIES

Soal:

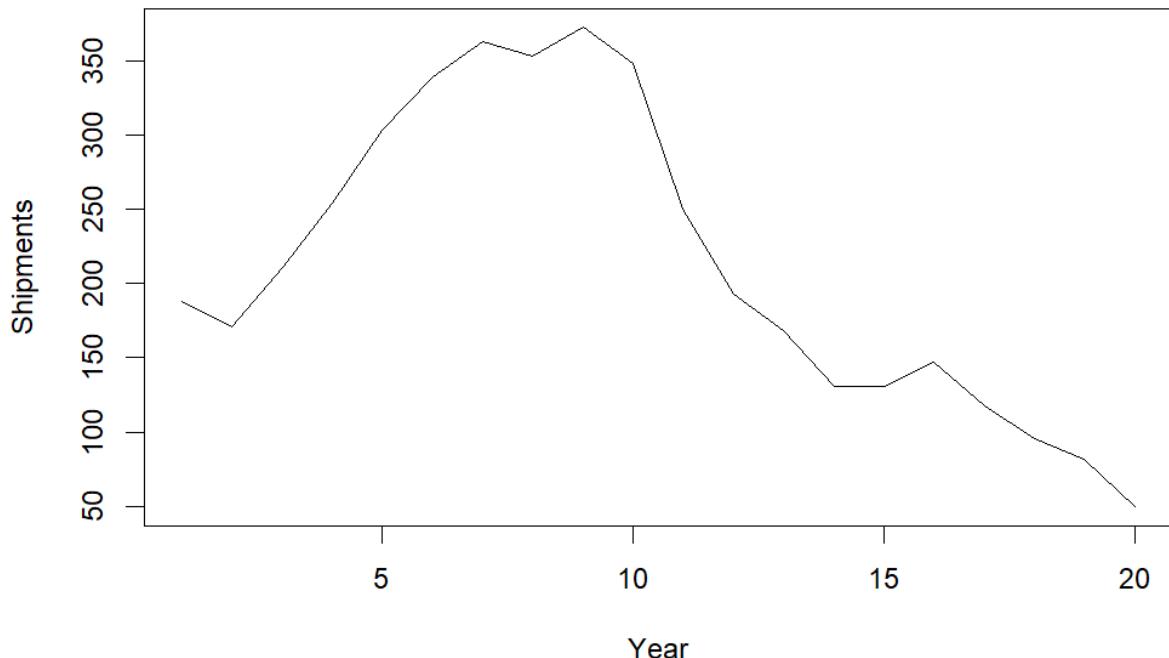
Jumlah rumah yang dibangun di US mengalami peningkatan dari tahun 1990 hingga 1996, kemudian tetap sama hingga tahun 1999, ketika jumlahnya mulai menurun.

2002	168.491
2003	130.815
2004	130.748
2005	146.8
2006	117.3
2007	95.7
2008	81.9
2009	49.8

Jawaban:

```
> library(readxl)
>
> View(Penjualan_Rumah_di_US)
> plot.ts(Penjualan_Rumah_di_US$Shipments, xlab="Year", ylab="Shipments",
main="Tren Pembangunan Rumah di USA")
> points(Penjualan_Rumah_di_US)
```

Tren Pembangunan Rumah di USA



RSTUDIO VARIANSI TREND

Kode:

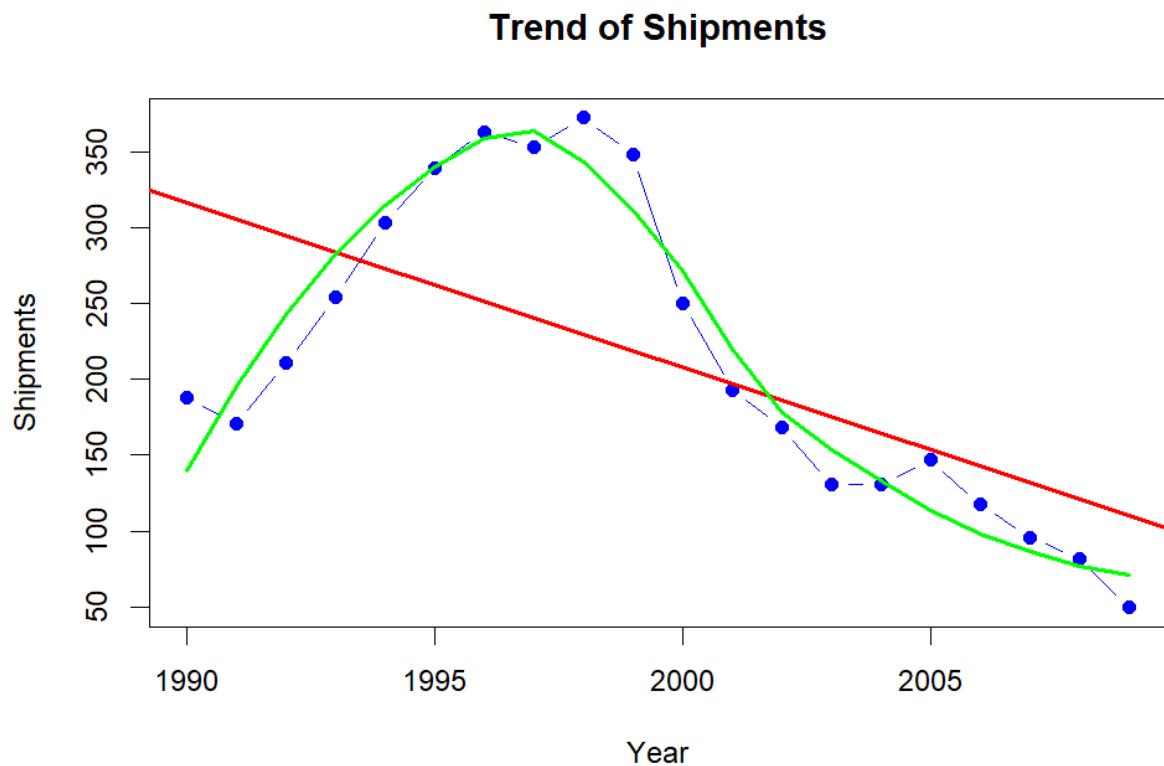
```
> #Plot data
> plot(data$Year, data$Shipments, type="b", col="blue", pch=16,xlab="Year",
", ylab="Shipments", main="Trend of Shipments")
>
```

```

> #garis tren
> abline(trend_model, col="red", lwd=2)
>
> # Smoothing menggunakan LOESS
> loess_model <- loess(Shipments ~ Year, data = data)
> lines(data$Year, predict(loess_model), col="green", lwd=2)
> library(forecast)

```

Output:



Atau:

```

> year <- c(1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001,
+           2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009)
> shipments <- c(188.172, 170.713, 210.787, 254.276, 303.932, 339.601, 363.411, 353.377,
+               372.843, 348.671, 250.55, 193.229, 168.491, 130.815, 130.748, 146.8, 117.3, 95.7, 81.9, 49.8)
>
> # Membuat data frame
> data <- data.frame(Year=year, Shipments=shipments)
> data
   Year Shipments
1 1990    188.172
2 1991    170.713
3 1992    210.787
4 1993    254.276
5 1994    303.932
6 1995    339.601
7 1996    363.411
8 1997    353.377
9 1998    372.843

```

10	1999	348.671
11	2000	250.550
12	2001	193.229
13	2002	168.491
14	2003	130.815
15	2004	130.748
16	2005	146.800
17	2006	117.300
18	2007	95.700
19	2008	81.900
20	2009	49.800

RSTUDIO VARIANSI SIKLIS

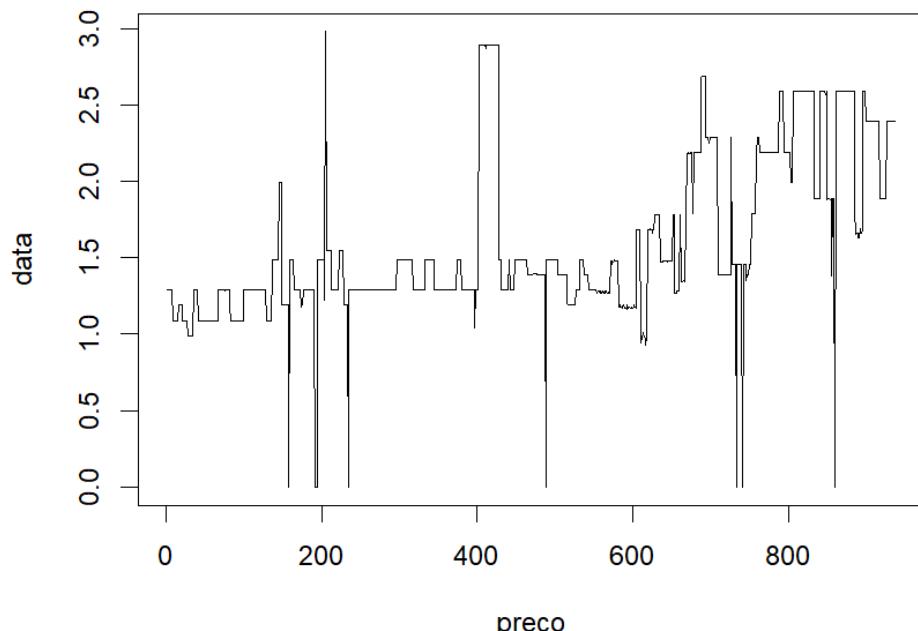
Variasi Siklis adalah Gerakan/variasi jangka Panjang disekitar garis trend(berlaku untuk data tahunan). Gerakan siklis ini bisa terulang setelah jangka Waktu tertentu(setelah 3-5th) bisa juga tidak terulang dalam jangka Waktu yang sama

Kode:

```
> library(readxl)
> mock_kaggle <- read_excel("C:/Users/Halo/Downloads/mock_kaggle.xlsx")
> View(mock_kaggle)
> #variansi siklis
> View(mock_kaggle)
> plot.ts(mock_kaggle$preco, xlab="preco", ylab="data", main="Grafik Time Series Variansi Siklis")
> points(mock_kaggle)
```

Output:

Grafik Time Series Variansi Siklis



RSTUDIO VARIANSI MUSIM

Musiman merupakan fluktuasi data yang terjadi namun dalam rentang waktu yang lebih singkat dibandingkan trend dan juga siklus, Dimana hanya akan membutuhkan waktu dalam lingkup satu tahun. Ada beberapa faktor yang membuat fluktuasi musiman ini terjadi, misalnya karena peristiwa tertentu seperti lebaran, natal, dan tahun baru, dll. Variasi musiman ini pasti terjadi lebih teratur dibandingkan variasi siklus dan bersifat lengkap.

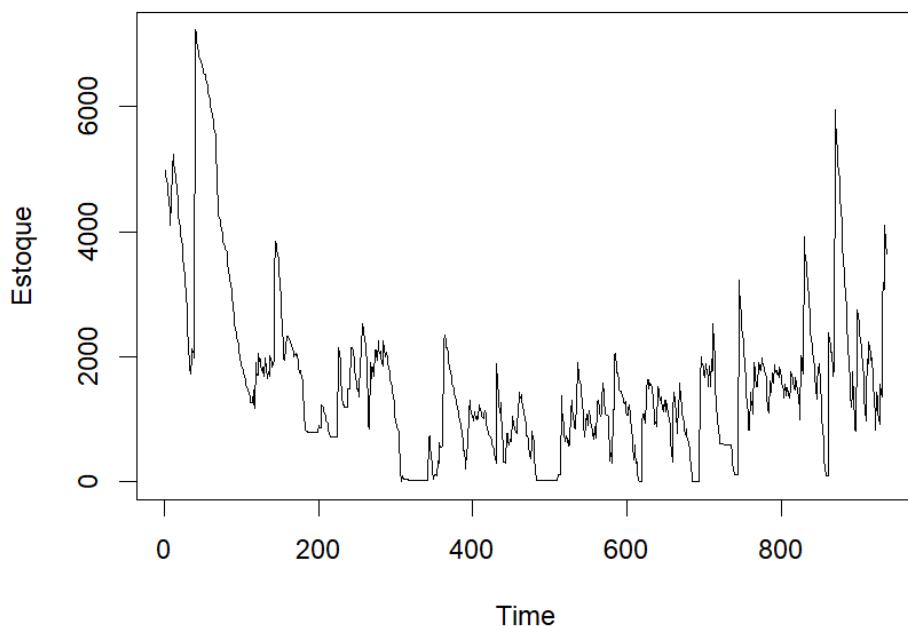
Contoh 1 : Penjualan barang estoque pada dataset mockKaggle menunjukkan pola variasi musiman

Kode:

```
> #variasi musim  
> View(mock_kaggle)  
> plot.ts(mock_kaggle$estoque, xlab="Time", ylab="Estoque", main="Data Penjualan Estoque")  
> points(mock_kaggle)  
  
>
```

Output:

Data Penjualan Estoque



RSTUDIO VARIANSI MUSIM

Kode:

```

1 #data
2 tahun <- c(2001, 2002, 2003, 2004, 2005, 2006)
3 pemakaian_minyak <- c(54632, 54797, 56282, 59484, 59330, 57139)
4
5 #membuat data frame
6 data <- data.frame(tahun = tahun, pemakaian_minyak = pemakaian_minyak)
7 print(data)
8 variance_minyak <- var(data$pemakaian_minyak)
9 print(paste("Variansi Pemakaian Minyak Bumi:", variance_minyak))
10 trend_model <- lm(pemakaian_minyak ~ tahun, data = data)
11 summary(trend_model)
12
13 #Plot data
14 plot(data$tahun, data$pemakaian_minyak, type="b", col='blue', pch=16, xlab="Tahun",
15 ylab="Pemakaian Minyak Bumi (Ton)", main="Tren Pemakaian Minyak Bumi")
16
17 #garis tren
18 abline(trend_model, col="red", lwd=2)
19
20 #Prediksi 3 tahun mendatang
21 tahun_baru <- data.frame(tahun = c(2007, 2008, 2009))
22 prediksi <- predict(trend_model, newdata = tahun_baru)
23 print(data.frame(tahun_baru, predik_pemakaian_minyak = prediksi))
24

```

Output:

```

> #data(contoh 2)
> tahun <- c(2001, 2002, 2003, 2004, 2005, 2006)
> pemakaian_minyak <- c(54632, 54797, 56282, 59484, 59330, 57139)
>
> #membuat data frame
> data <- data.frame(tahun = tahun, pemakaian_minyak = pemakaian_minyak)
> print(data)
  tahun pemakaian_minyak
1 2001      54632
2 2002      54797
3 2003      56282
4 2004      59484
5 2005      59330
6 2006      57139
> variance_minyak <- var(data$pemakaian_minyak)
> print(paste("Variansi Pemakaian Minyak Bumi:", variance_minyak))
[1] "Variansi Pemakaian Minyak Bumi: 4515163.6"
> trend_model <- lm(pemakaian_minyak ~ tahun, data = data)
> summary(trend_model)

Call:
lm(formula = pemakaian_minyak ~ tahun, data = data)

Residuals:
    1     2     3     4     5     6 
-216.6 -889.7 -242.9  2120.9 1128.7 -1900.4 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1622332.5   767837.2  -2.113   0.102    
tahun        838.2      383.2   2.187   0.094 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1603 on 4 degrees of freedom
Multiple R-squared:  0.5446, Adjusted R-squared:  0.4307 
F-statistic: 4.783 on 1 and 4 DF,  p-value: 0.094

>
> #Plot data

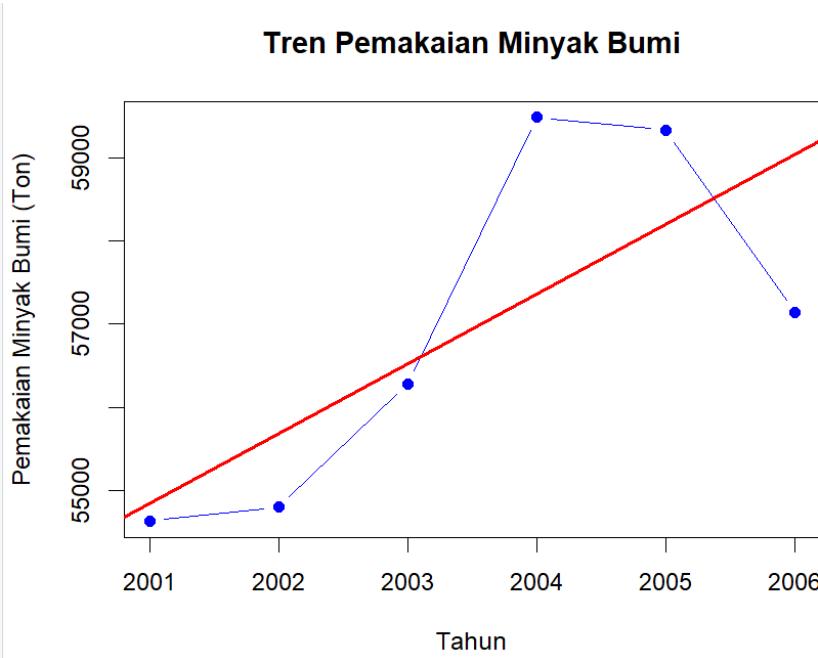
```

```

> plot(data$tahun, data$pemakaian_minyak, type="b", col="blue", pch=16, x1
ab="Tahun", ylab="Pemakaian Minyak Bumi (Ton)", main="Tren Pemakaian Minya
k Bumi")
>
> #garis tren
> abline(trend_model, col="red", lwd=2)
>
> #Prediksi 3 tahun mendatang
> tahun_baru <- data.frame(tahun = c(2007, 2008, 2009))
> prediksi <- predict(trend_model, newdata = tahun_baru)
> print(data.frame(tahun_baru, predik_pemakaian_minyak = prediksi))
  tahun predik_pemakaian_minyak
1 2007          59877.60
2 2008          60715.77
3 2009          61553.94

```

Grafik:



RSTUDIO VARIANSI TAK BERATURAN

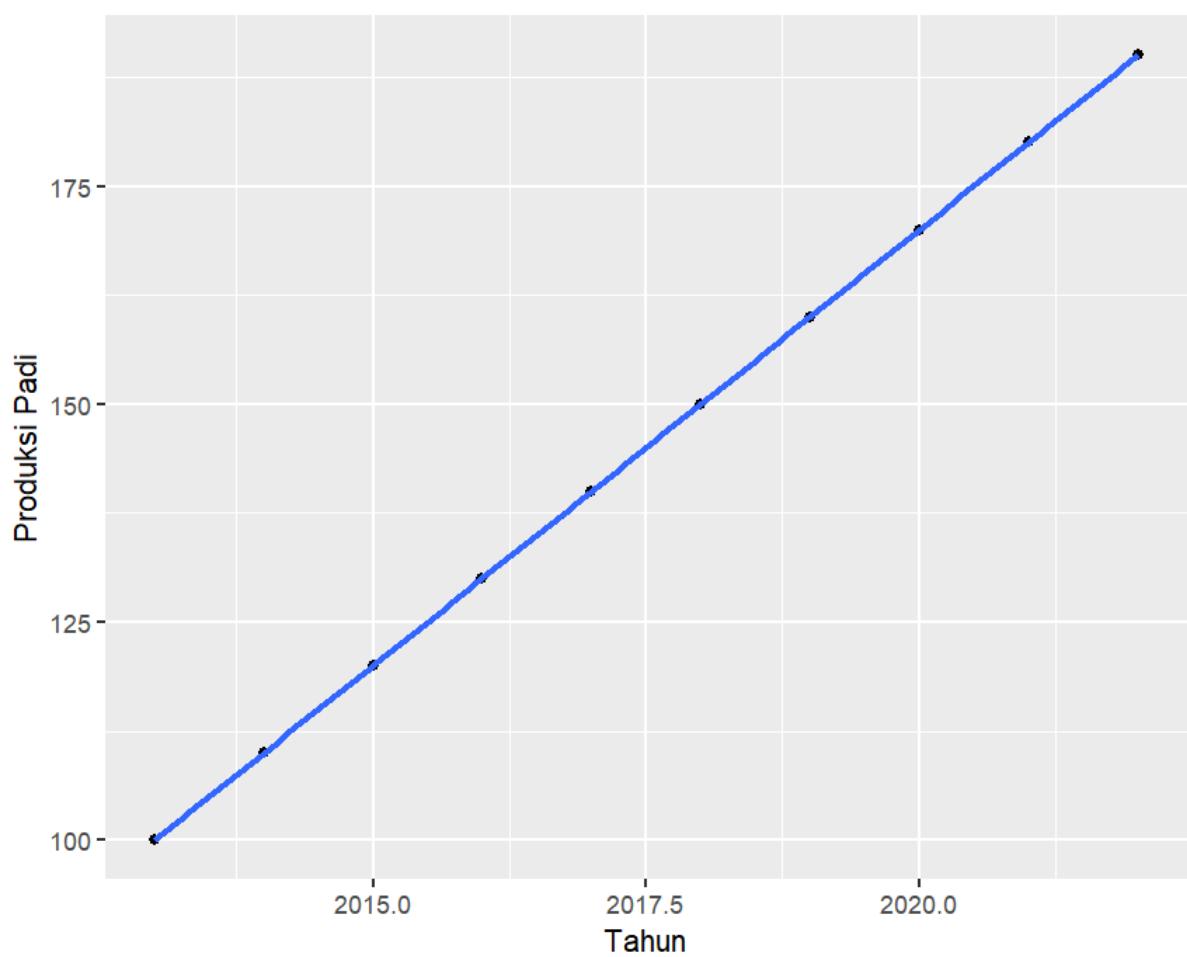
Variasi tak beraturan adalah gerakan/variasi yang irregular atau sporadic sifatnya, Variasi ini pada kenyataannya sulit diprediksi. Contoh variasi ini dalam data time series karena adanya perang, bencana alam dan sebab-sebab unik lainnya yang sulit diduga. misalnya naik/turunnya produksi padi akibat banjir yang datangnya tidak teratur, naik turunnya produksi karena pemogokan

```

> #variansi tak beraturan
> # Membuat data frame
> data <- data.frame(Tahun = 2013:2022, Produksi = c(100, 110, 120, 130, 1
40, 150, 160, 170, 180, 190))
>
> # Visualisasi
> library(ggplot2)
Keep up to date with changes at https://tidyverse.org/blog/
Warning message:
package ‘ggplot2’ was built under R version 4.3.3

```

```
> ggplot(data, aes(x = Tahun, y = Produksi)) +  
+   geom_point() +  
+   geom_smooth(method = "lm", se = FALSE) +  
+   labs(x = "Tahun", y = "Produksi Padi")  
`geom_smooth()` using formula = 'y ~ x'
```



MOVING AVERAGE

Moving Average adalah indikator yang menghitung harga rata-rata suatu aset dalam periode waktu tertentu, kemudian menghubungkannya dalam bentuk garis. Nilai rata-rata bisa berasal dari harga pembukaan (open), penutupan (close), tertinggi (high), terendah (low), ataupun pertengahan (median).

Adapun Moving Average terbagi menjadi :

1. Simple Moving Average (SMA)
2. Weighted Moving Average (WMA)
3. Exponential Moving Average (EMA)
4. Metode Least Square (Kuadrat Terkecil)

1. SIMPLE MOVING AVERAGE (SMA)

Rata-rata pergerakan sederhana (SMA) adalah rata-rata pergerakan aritmatika yang dihitung dengan menjumlahkan harga terkini dan kemudian membagi angka tersebut dengan jumlah periode waktu dalam rata-rata perhitungan. Misalnya, seseorang dapat menambahkan harga penutupan suatu sekuritas untuk beberapa periode waktu dan kemudian membagi totalnya dengan jumlah periode yang sama. Rata-rata jangka pendek bereaksi dengan cepat terhadap perubahan harga sekuritas yang mendasarinya, sedangkan rata-rata jangka panjang bereaksi lebih lambat.

Contoh

Diberikan data Harga Asset selama 15 hari. Tentukan Single Moving Average 5 harian!

Hari	Harga asset	Total Bergerak 5 harian	SMA 5 harian
1	20	-	-
2	24	-	-
3	22	-	-
4	21	-	-
5	20	$20+24+22+21+20=107$	$107/5 = 21,4$
6	18	$24+22+21+20+18=105$	$105/5 = 21$
7	17	98	$98/5 = 19,6$
8	22	98	$98/5 = 19,6$
9	26	103	$103/5 = 20,6$
10	30	113	$113/5 = 22,6$
11	31	116	$116/5 = 23,2$
12	34	143	$143/5 = 28,6$
13	33	154	$154/5 = 30,8$
14	30	158	$158/5 = 31,6$
15	28	156	$156/5 = 31,2$

> #SMA CONTOH 1 (DENGAN LIBRARY ZOO)
> #memuat library yang diperlukan

```

> #install.packages("zoo")
> library(zoo)

Attaching package: 'zoo'

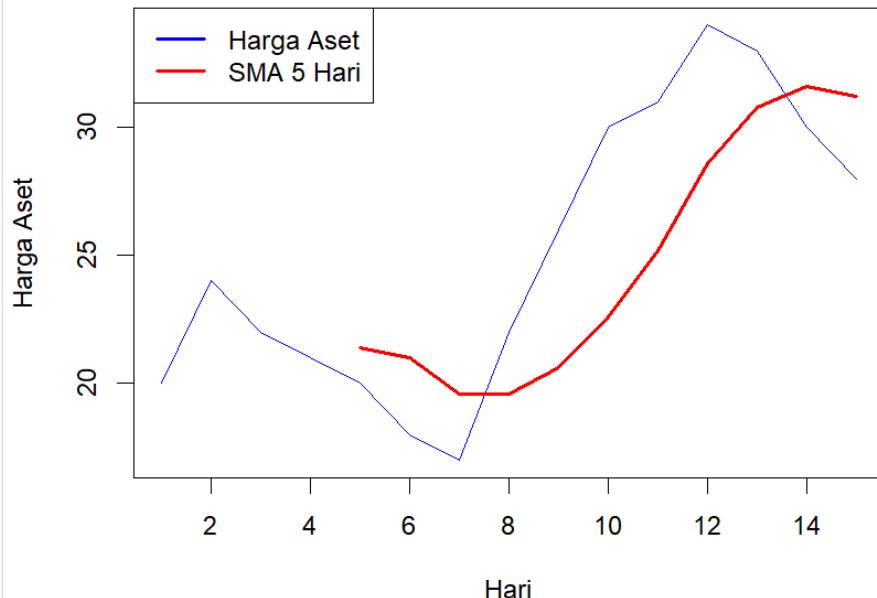
The following objects are masked from 'package:base':

  as.Date, as.Date.numeric

Warning message:
package 'zoo' was built under R version 4.3.3
>
> #data harga aset
> harga <- c(20, 24, 22, 21, 20, 18, 17, 22, 26, 30, 31, 34, 33, 30, 28)
> periode <- c(1, 2, 3, 4, 5)
>
> #Menghitung SMA
> SMA_harian <- rollapply(harga, width = 5, FUN = mean, align = 'right', f
ill = NA)
> SMA_harian
[1]   NA   NA   NA   NA 21.4 21.0 19.6 19.6 20.6 22.6 25.2 28.6 30.8 31.6
31.2
>
> #membuat data frame untuk menampilkan hasil
> hasil_SMA <- data.frame(hari = 1: length(harga), harga, SMA = SMA_haria
n)
> print(hasil_SMA)
  hari harga  SMA
1     1    20   NA
2     2    24   NA
3     3    22   NA
4     4    21   NA
5     5    20  21.4
6     6    18  21.0
7     7    17  19.6
8     8    22  19.6
9     9    26  20.6
10   10    30  22.6
11   11    31  25.2
12   12    34  28.6
13   13    33  30.8
14   14    30  31.6
15   15    28  31.2
>
>
> #plot hasil harga aset dan sma
> plot(1:length(harga), harga, type = "l", col = "blue", xlab = "Hari", yl
ab = "Harga Aset", main = "Harga Aset dan SMA 5 Hari")
> lines(1:length(SMA_harian), SMA_harian, col = "red", lwd = 2) #untuk men
ambahkan sma ke plot
> legend("topleft", legend = c("Harga Aset", "SMA 5 Hari"), col = c("blue
", "red"), lty = 1, lwd = 2)

```

Harga Aset dan SMA 5 Hari



Contoh 2 :

Tentukan rata rata 4 bulanan untuk produksi mesin bor berkualitas tinggi di PT ABX

No	Bulan	Permintaan Aktual
1.	Januari	79
2.	Februari	82
3.	Maret	95
4.	April	82
5.	Mei	85
6.	Juni	88
7.	Juli	93
8.	Agustus	84
9.	September	90
10.	Oktober	83
11.	November	76
12.	Desember	87

Tentukan rata rata 4 bulanan untuk produksi mesin bor berkualitas tinggi di PT ABX

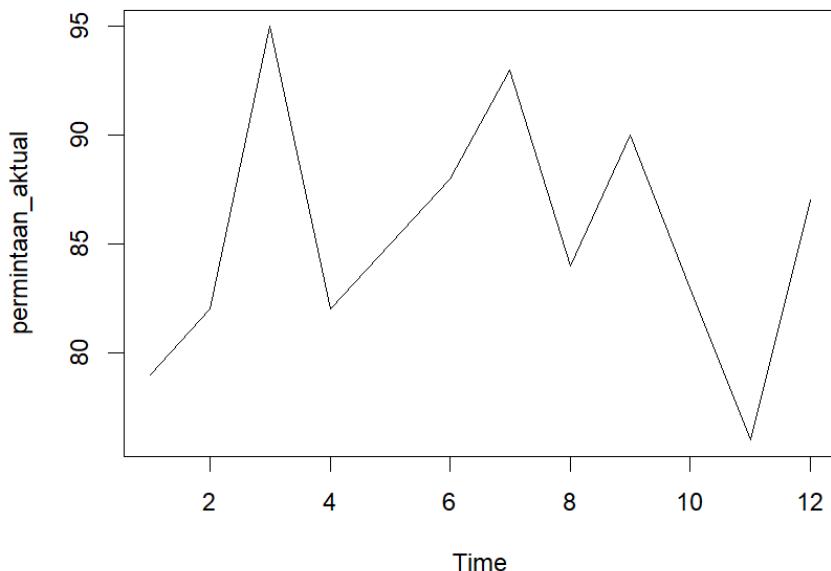


```

> #Menulis Dataset yang akan digunakan
> permintaan_aktual<- c(79,82,95,82,85,88,93,84,90,83,76,87)
> permintaan_aktual
[1] 79 82 95 82 85 88 93 84 90 83 76 87
> #melihat grafik time series dari permintaan aktual
> plot.ts(permintaan_aktual, main="Grafik Time Series Permintaan Aktual")
> #Menentukan nilai SMA
> SMA.4 <- SMA(permintaan_aktual, 4)
> SMA.4
[1]     NA     NA     NA 84.50 86.00 87.50 87.00 87.50
[9] 88.75 87.50 83.25 84.00
> #Grafik SMA
> plot.ts(SMA.4, main="Grafik SMA 4 bulanan")

```

Grafik Time Series Permintaan Aktual



2. WEIGHTED MOVING AVERAGE (WMA)

Weighted Moving Average (WMA) adalah memberikan penekanan lebih pada data harga terbaru dibandingkan dengan data yang lebih lama. WMA dihitung dengan mengalikan setiap harga dalam seri data dengan bobot tertentu, yang biasanya berhubungan dengan posisi harga dalam rangkaian waktu. Dalam WMA, harga terbaru mendapat bobot yang lebih berat, sementara harga yang lebih lama mendapat bobot yang lebih rendah. Bobot biasanya ditentukan dalam urutan linear, di mana harga terbaru mendapat bobot tertinggi dan berkurang secara bertahap untuk harga yang lebih lama.

CONTOH :

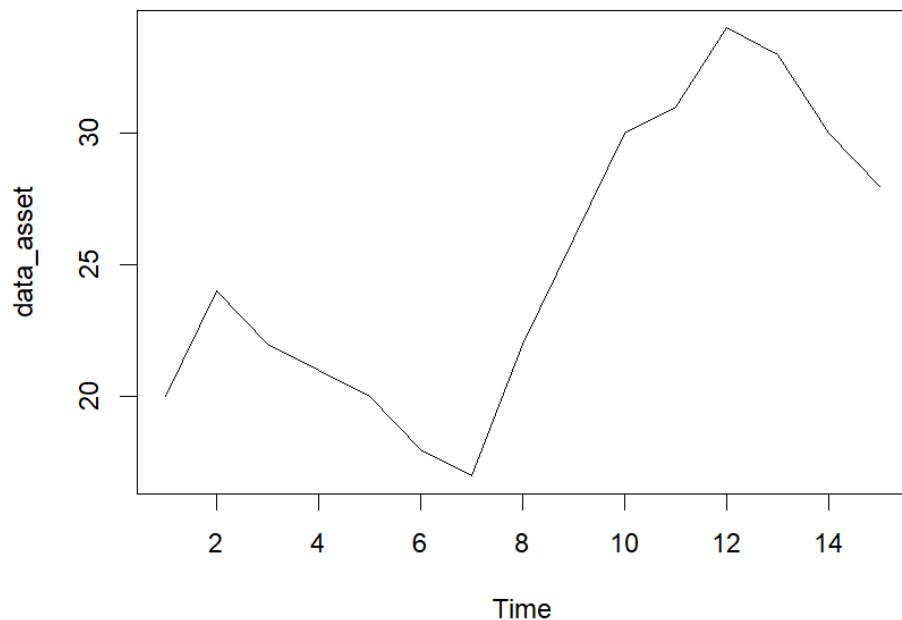
Hari	Harga asset	SMA 5 harian	Bobot	WMA 5 harian
1	20	NA	1	NA
2	24	NA	2	NA
3	22	NA	3	NA
4	21	NA	4	NA
5	20	21,4	5	21,2
6	18	21	1	20,1
7	17	19,6	2	18,7
8	22	19,6	3	19,5
9	26	20,6	4	21,7
10	30	22,6	5	24,8
11	31	25,2	1	27,6
12	34	28,6	2	30,5
13	33	30,8	3	32
14	30	31,6	4	31,7
15	28	31,2	5	30,5

```

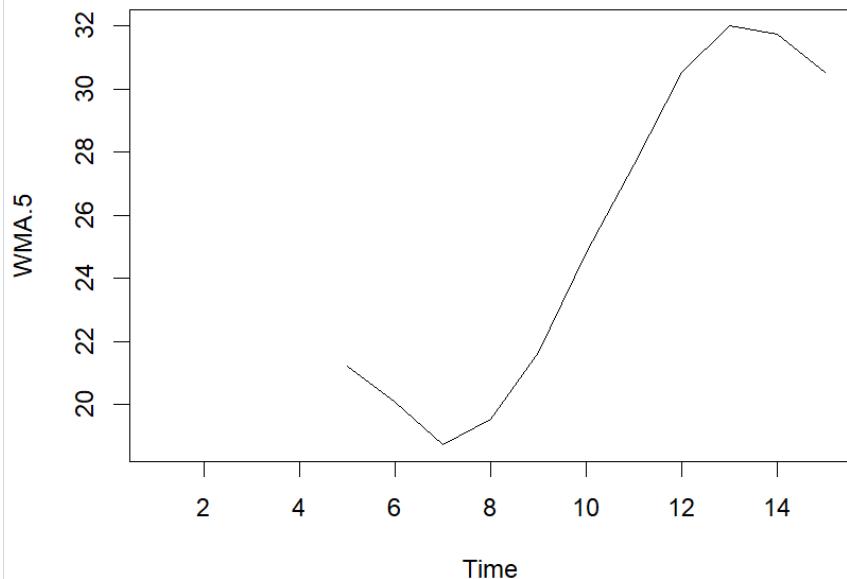
> #WMA CONTOH 1(CONTOH LIBRARY TTR)
> data_asset <- c(20,24,22,21,20,18,17,22,26,30,31,34,33,30,28)
> plot.ts(data_asset, main="Grafik time series data aset")
> library(TTR)
Warning message:
package 'TTR' was built under R version 4.3.3
>
> WMA.5 <- WMA(data_asset, 5, 1:5)
> WMA.5
[1]      NA      NA      NA      NA 21.20000 20.06667 18.73333 19.533
33
[9] 21.66667 24.80000 27.60000 30.53333 32.00000 31.73333 30.53333
> plot.ts(WMA.5, main="Grafik weighted moving average 5 harian")

```

Grafik time series data asset



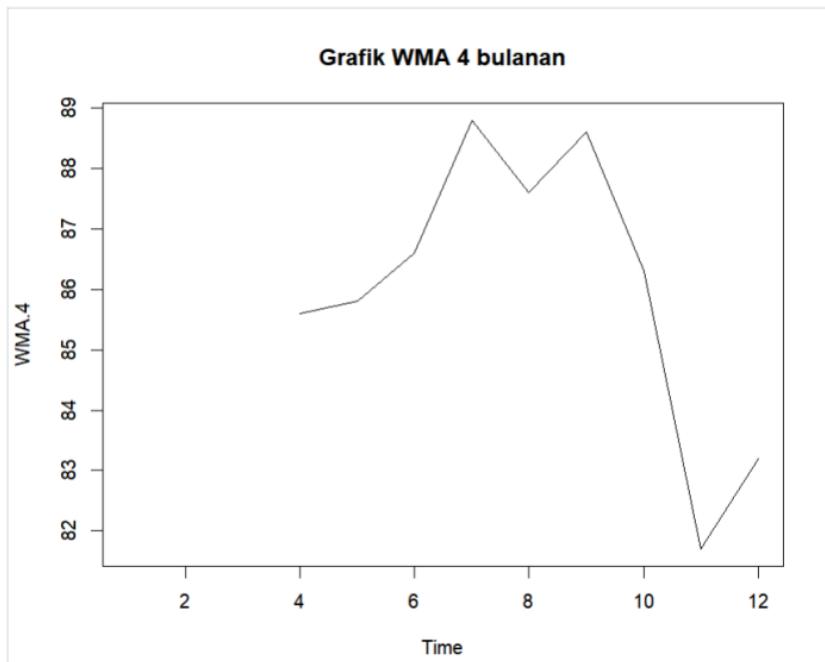
Grafik weighted moving average 5 harian



No	Bulan	Permintaan Aktual
1.	Januari	79
2.	Februari	82
3.	Maret	95
4.	April	82
5.	Mei	85
6.	Juni	88
7.	Juli	93
8.	Agustus	84
9.	September	90
10.	Oktober	83
11.	November	76
12.	Desember	87

Tentukan rata rata 4 bulanan untuk produksi mesin bor berkualitas tinggi di PT ABX

```
> #Menulis Dataset yang akan digunakan
> permintaan_aktual<- c(79,82,95,82,85,88,93,84,90,83,76,87)
> permintaan_aktual
[1] 79 82 95 82 85 88 93 84 90 83 76 87
> #melihat grafik time series dari permintaan aktual
> plot.ts(permintaan_aktual, main="Grafik Time Series Permintaan Aktual")
> #Menentukan nilai SMA
> SMA.4 <- SMA(permintaan_aktual, 4)
> SMA.4
[1] NA NA NA 84.50 86.00 87.50 87.00 87.50
[9] 88.75 87.50 83.25 84.00
> #Grafik SMA
> plot.ts(SMA.4, main="Grafik SMA 4 bulanan")
> #Menentukan Nilai WMA
> WMA.4 <- WMA(permintaan_aktual, n=4, 1:4)
> WMA.4
[1] NA NA NA 85.6 85.8 86.6 88.8 87.6 88.6 86.3 81.7 83.2
> #Grafik WMA
> plot.ts(WMA.4, main="Grafik WMA 4 bulanan")
```



3. EXPONENTIAL MOVING AVERAGE (EMA)

Exponential Moving Average (EMA) adalah jenis MA yang menyaring data secara infinite, dimana data-data lama tidak ada yang dibuang melainkan hanya dikurangi bobotnya secara eksponensial, namun bobotnya tidak sampai nol.

EMA memiliki kemiripan dengan WMA dalam hal membedakan bobot data antara data terdahulu dan yang terbaru, dimana dengan perhitungan ini, EMA dan WMA sama-sama lebih sensitif dengan pergerakan harga saham dibandingkan dengan SMA

```

calculate_ema <- function(harga, n = 5) {
  alpha <- 2 / (n + 1) # Smoothing factor
  ema <- rep(NA, length(harga)) # Inisialisasi dengan NA untuk semua bulan
  # EMA pertama dihitung dari permintaan pertama
  ema[1] <- harga[1]

  # Mulai menghitung EMA dari bulan ke-2 sampai seterusnya
  for (i in 2:length(harga)) {
    ema[i] <- alpha * harga[i] + (1 - alpha) * ema[i - 1]
  }

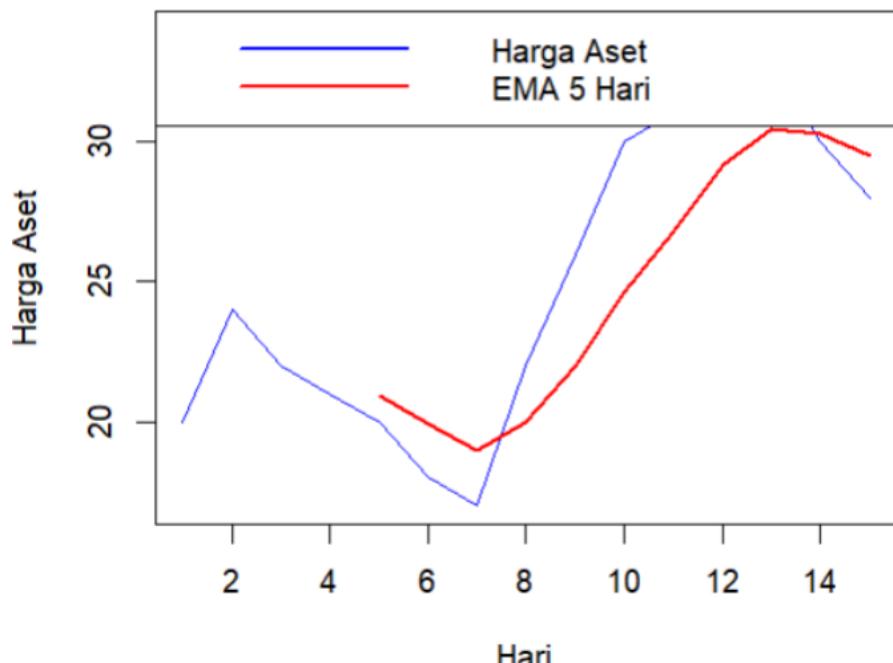
  return(ema)
}

harga <- c (20, 24, 22, 21, 20, 18, 17, 22, 26, 30, 31, 34, 33, 30, 28)

plot(1:length(harga), harga, type = "l", col = "blue", xlab = "Hari",
     ylab = "Harga Aset", main = "Harga Aset dan EMA 5 Hari")
lines(1:length(EMA_hari), EMA_hari, col = "red", lwd = 2) #untuk menambahkan ema ke plot
legend("topleft", legend = c("Harga Aset", "EMA 5 Hari"), col = c("blue", "red"), lty = 1, lwd = 2)

```

Harga Aset dan EMA 5 Hari



BAB 7: PERTEMUAN 12 (Laporan 13: Analisis Tren)

Studi Kasus 1: Data Genap Kelompok Genap(Asosiasi Industri Sepeda Motor Indonesia/AISI 2023)



Kode R dan Penjelasan:

1. Memasukkan dan Memuat Data

```
> # === 1. Memasukkan dan Memuat Data ===  
> # Memuat library untuk visualisasi  
> # install.packages(ggplot2) # Untuk menginstal library ggplot2  
> library(ggplot2)  
Warning message:  
package ‘ggplot2’ was built under R version 4.3.3  
>
```

Langkah pertama adalah memasukkan dan memuat data kedalam R Studio. Dan kode diatas adalah untuk menginstall dan membuat library **ggplot2** untuk keperluan visualisasi.

```
> ## Data penjualan AISI 2023  
> # Fungsi c() di R adalah singkatan dari combine.  
> # Fungsi ini digunakan untuk membuat sebuah vektor dengan menggabungkan  
elemen-elemen.  
> bulan <- c("JAN", "FEB", "MAR", "APR", "MAY", "JUN", "JUL", "AUG", "SEP",  
"OCT", "NOV", "DEC")
```

```

> penjualan_domestik <- c(615416, 575502, 633155, 354323, 529771, 493763,
475428, 534379, 509946, 516293, 571983, 427033)
> penjualan_ekspor <- c(39269, 45627, 36962, 41201, 55743, 49920, 53443, 5
5247, 51318, 52517, 46020, 42737)
>

```

Potongan kode ini sedang mempersiapkan data penjualan domestik dan ekspor untuk tahun 2023. Data ini dibuat dalam bentuk **vektor**, yaitu struktur data sederhana di R yang berisi sekumpulan nilai. Di sini, fungsi `c()` digunakan untuk menggabungkan elemen-elemen yang dimasukkan ke dalamnya.

- **bulan:**
 - Ini adalah vektor yang berisi nama bulan dalam format singkatan (JAN, FEB, dst.).
- **penjualan_domestik:**
 - Vektor ini berisi angka penjualan motor domestik untuk setiap bulan.
- **penjualan_ekspor:**
 - Vektor ini berisi angka penjualan ekspor motor untuk setiap bulan.

Hal yang Terjadi di Kode Ini

- Kode ini tidak melakukan analisis atau visualisasi apa pun **belum**—hanya membuat data dalam bentuk **vektor**.
- Nantinya, vektor-vektor ini bisa dimasukkan ke dalam data frame untuk pengolahan data yang lebih lanjut, seperti analisis tren, perhitungan total penjualan, atau membuat grafik.

Kenapa Gunakan `c()`?

- Fungsi `c()` mempermudah pengelompokan data dalam bentuk list sederhana.
- Sangat cocok untuk dataset kecil seperti ini karena langsung dapat diolah untuk analisis atau visualisasi.

```

> ## Membuat data frame
> # Tujuan: Membuat sebuah data frame yang bernama aisi_data,
> # yang berisi data penjualan domestik dan ekspor sepeda motor berdasarkan
n bulan.
> aisi_data <- data.frame(
+   Bulan = bulan,
+   Penjualan_Domestik = penjualan_domestik,
+   Penjualan_Eksport = penjualan_ekspor
+ )
> print(aiisi_data)
      Bulan Penjualan_Domestik Penjualan_Eksport
1     JAN           615416            39269
2     FEB           575502            45627
3     MAR           633155            36962
4     APR           354323            41201
5     MAY           529771            55743
6     JUN           493763            49920
7     JUL           475428            53443
8     AUG           534379            55247
9     SEP           509946            51318
10    OCT           516293            52517
11    NOV           571983            46020
12    DEC           427033            42737

```

Bagian kode ini membuat sebuah **data frame** bernama `aisi_data`. **Data frame** adalah struktur data di R yang mirip dengan tabel, sehingga lebih mudah untuk mengorganisasi dan menganalisis data yang terdiri dari beberapa kolom.

- **Bulan = bulan:**
 - Kolom pertama data frame ini diberi nama "Bulan" dan isinya adalah vektor bulan, yaitu nama-nama bulan dari JAN sampai DEC.
- **Penjualan_Domestik = penjualan_domestik:**
 - Kolom kedua diberi nama "Penjualan_Domestik" dan isinya adalah data penjualan motor domestik dari vektor `penjualan_domestik`.
- **Penjualan_Ekspor = penjualan_ekspor:**
 - Kolom ketiga diberi nama "Penjualan_Ekspor" dan isinya adalah data penjualan motor ekspor dari vektor `penjualan_ekspor`.

Perintah `print()` menampilkan isi dari data frame `aisi_data` ke layar. Outputnya adalah tabel dengan kolom **Bulan**, **Penjualan_Domestik**, dan **Penjualan_Ekspor**, serta isinya sesuai data yang kita masukkan sebelumnya.

```
> # Konversi kolom 'Bulan' menjadi faktor dengan urutan sesuai kalender
> aisi_data$Bulan <- factor(aisi_data$Bulan,
+                               levels = c("JAN", "FEB", "MAR", "APR", "MAY", "JUN",
+                               "JUL", "AUG", "SEP", "OCT", "NOV", "DEC"))
>
```

- **aisi_data\$Bulan:**
 - Kolom Bulan pada data frame `aisi_data`.
- **factor(aisi_data\$Bulan, levels = ...):**
 - Fungsi `factor()` mengonversi data menjadi **faktor**.
 - **levels = c(...):**
 - Menentukan **urutan kategori** sesuai dengan kalender (JAN, FEB, ..., DEC).
 - Tanpa menentukan levels, R akan mengurutkan kategori berdasarkan abjad, sehingga bulan akan diurutkan sebagai APR, AUG, DEC, dan seterusnya. Dengan levels, kita memastikan urutannya sesuai kalender.

Kenapa Kolom Bulan Harus Diubah Menjadi Faktor?

1. Mendukung Analisis yang Memerlukan Urutan:

- Saat membuat grafik, seperti garis waktu, faktor memastikan bulan diurutkan secara kronologis.

- Contoh: Dalam grafik, bulan JAN akan muncul sebelum FEB, bukan diurutkan secara abjad.

2. Memastikan Konsistensi dalam Manipulasi Data:

- Data kategori (seperti bulan) yang memiliki urutan logis sebaiknya disimpan sebagai faktor untuk mencegah kesalahan saat digunakan dalam perhitungan atau visualisasi.

3. Mendukung Fungsi Statistik:

- Fungsi tertentu di R, seperti lm() (regresi), bekerja lebih baik dengan data kategori yang diformat sebagai faktor.

Setelah Konversi

Sekarang kolom Bulan adalah **faktor dengan level tertentu**, dan data frame aisi_data siap digunakan untuk analisis atau visualisasi lebih lanjut, seperti membuat grafik dengan bulan dalam urutan yang benar.

```
> # === 2. Visualisasi Data Awal (2 Kolom, 2 Grafik) ===
> # Penjualan Domestik
> plot_domestik <- ggplot(aisi_data, aes(x = Bulan, y = Penjualan_Domestik, group = 1)) +
+   geom_line(color = "blue", size = 1) + # utk membuat garis, warna biru ukurannya 1
+   geom_point(color = "blue", size = 2) + # utk membuat titik, warna biru ukurannya 2 agar lebih tebal
+   labs(                                # utk menambahkan label pada grafik, judul, x, y
+     title = "Penjualan Domestik AISI 2023",
+     x = "Bulan",
+     y = "Jumlah Penjualan"
+   ) +
+   theme_minimal()
```

2. Visualisasi Data Awal (2 Kolom, 2 Grafik)

a. Penjualan Domestik

```
> # === 2. Visualisasi Data Awal (2 Kolom, 2 Grafik) ===
> # Penjualan Domestik
> plot_domestik <- ggplot(aisi_data, aes(x = Bulan, y = Penjualan_Domestik, group = 1)) +
+   geom_line(color = "blue", size = 1) + # utk membuat garis, warna biru ukurannya 1
+   geom_point(color = "blue", size = 2) + # utk membuat titik, warna biru ukurannya 2 agar lebih tebal
+   labs(                                # utk menambahkan label pada grafik, judul, x, y
+     title = "Penjualan Domestik AISI 2023",
+     x = "Bulan",
+     y = "Jumlah Penjualan"
+   ) +
+   theme_minimal()
```

1. ggplot()

- Fungsi utama untuk membuat visualisasi di R menggunakan library **ggplot2**.
- **aisi_data**: Data frame yang digunakan sebagai input.
- **aes(x = Bulan, y = Penjualan_Domestik)**:
 - x: Sumbu horizontal (bulan).
 - y: Sumbu vertikal (penjualan domestik).
 - **group = 1**: Memberitahu ggplot untuk menggambar semua titik sebagai bagian dari satu grup, sehingga menghasilkan garis kontinu.

2. geom_line()

- Membuat **garis** yang menghubungkan titik-titik pada grafik.
- **color = "blue"**:
 - Garis diberi warna biru.
- **size = 1**:
 - Ketebalan garis adalah 1.

3. geom_point()

- Membuat **titik** pada setiap pasangan nilai (x, y) pada grafik.
- **color = "blue"**:
 - Titik diberi warna biru agar serasi dengan garis.
- **size = 2**:
 - Titik dibuat lebih besar agar mudah terlihat.

4. labs()

- Menambahkan label untuk grafik.
- **title**: Judul grafik, dalam hal ini "Penjualan Domestik AISI 2023".
- **x**: Label untuk sumbu horizontal (bulan).
- **y**: Label untuk sumbu vertikal (jumlah penjualan).

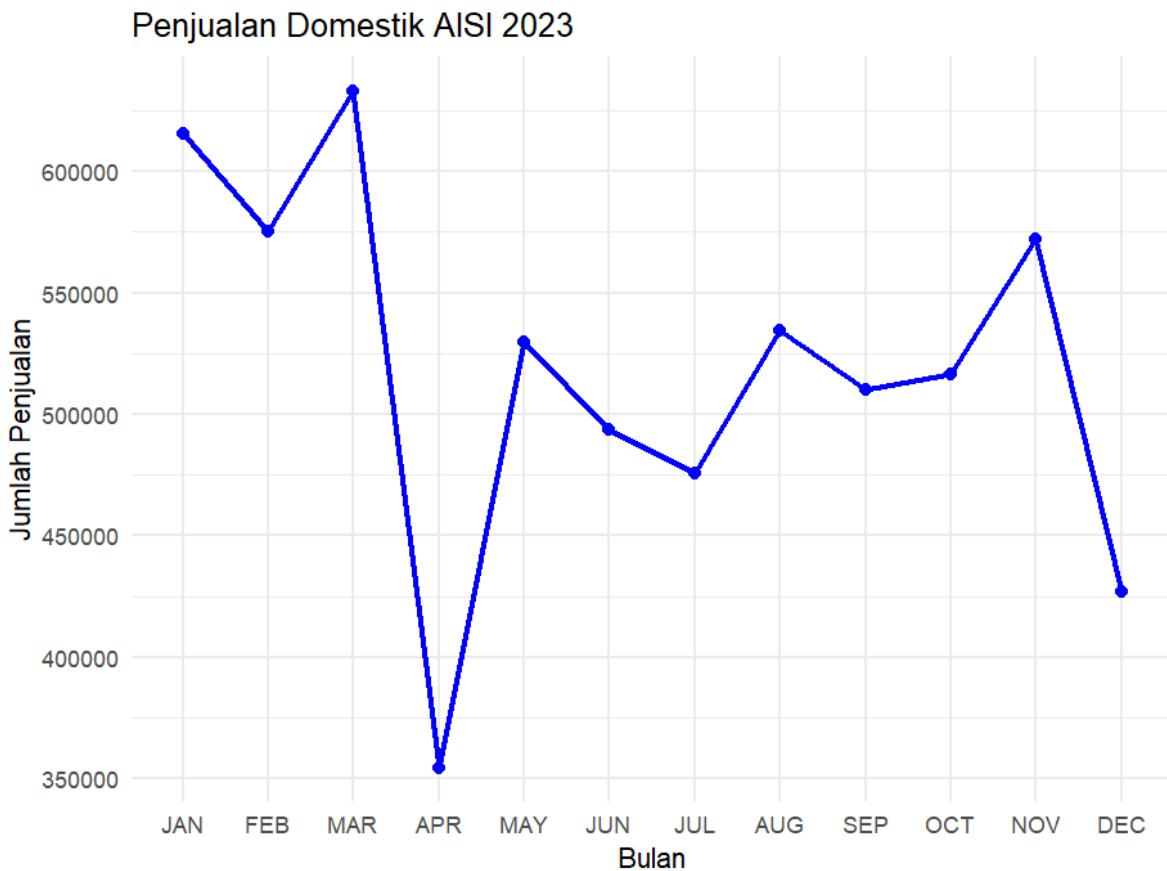
5. theme_minimal()

- Menerapkan tema **minimalis** pada grafik:
 - Tidak ada garis tambahan atau dekorasi yang berlebihan.
 - Fokus hanya pada data dan label.

Hasil Visualisasi

Grafik yang dihasilkan akan menunjukkan **tren penjualan domestik sepanjang tahun 2023**, dengan sumbu x menampilkan nama-nama bulan dalam urutan kalender (karena sebelumnya sudah diatur menjadi faktor). Grafik ini menggunakan **garis biru** yang menyambungkan titik-titik, dan setiap titik data diperjelas dengan **marker biru**.

Berikut hasil grafiknya:



Penjelasan Grafik:

- **Awal Tahun (JAN-MAR):**
Penjualan domestik mengawali tahun dengan kuat.
 - Januari mencapai **615,416 unit**, salah satu yang tertinggi sepanjang tahun.
 - Meskipun Februari sedikit turun menjadi **575,502 unit**, angka ini masih tinggi.
 - Maret kembali meningkat ke **633,155 unit**, menjadi **puncak tertinggi penjualan** domestik sepanjang tahun.
- **Musim Gugur (APR-JUN):**
Penjualan domestik mulai mengalami penurunan drastis.
 - April hanya mencatat **354,323 unit**, menjadi **bulan dengan penjualan terendah sepanjang tahun**.
 - Mei dan Juni sedikit membaik dengan angka masing-masing **529,771** dan **493,763 unit**, tetapi masih lebih rendah dibandingkan triwulan pertama.
- **Musim Panas (JUL-SEP):**
Penjualan mulai stabil tetapi tidak signifikan.
 - Juli (475,428 unit) hingga September (509,946 unit) menunjukkan tren yang tidak konsisten, namun cenderung naik pada akhir periode.
- **Akhir Tahun (OCT-DEC):**
Penjualan domestik menunjukkan perbaikan sementara, lalu kembali turun.
 - Oktober mencatat **516,293 unit**, dan November naik lagi ke **571,983 unit**.
 - Namun, pada Desember, penjualan kembali anjlok ke **427,033 unit**, yang hampir sama rendahnya dengan April.

Interpretasi Grafik

1. **Garis Grafik:**
 - Grafik menunjukkan pola "turun-naik-turun" dengan puncak di **Maret (633,155)** dan dasar di **April (354,323)**.
 - **Lonjakan awal tahun** terlihat jelas, terutama dari Januari ke Maret, diikuti oleh penurunan tajam di April.
 - Grafik kembali naik pada Mei dan menurun di Desember, menunjukkan **fluktuasi besar sepanjang tahun**.
2. **Titik Pencilan (Outliers):**
 - **Maret** adalah titik tertinggi, menonjol pada grafik.
 - **April** adalah titik terendah, sangat jauh dibandingkan bulan sebelumnya dan sesudahnya.
3. **Kesimpulan Umum:**
 - Penjualan domestik cukup kuat di awal tahun, stabil di tengah tahun, tetapi menurun drastis di akhir tahun.
 - Ada kemungkinan faktor musiman atau ekonomi yang memengaruhi fluktuasi ini, terutama pada bulan April dan Desember.
 - Jika kita lihat **kolom domestik** berdasarkan data penjualan sepanjang tahun 2023, tren yang terlihat adalah **negatif secara keseluruhan**.

b. Penjualan Ekspor

```
> # Penjualan Ekspor
> plot_ekspor <- ggplot(aisi_data, aes(x = Bulan, y = Penjualan_Ekspor, group = 1)) +
+   geom_line(color = "green", size = 1) +
+   geom_point(color = "green", size = 2) +
+   labs(
+     title = "Penjualan Ekspor AISI 2023",
+     x = "Bulan",
+     y = "Jumlah Penjualan"
+   ) +
+   theme_minimal()
> # Menampilkan grafik
> print(plot_domestik)
> print(plot_ekspor)
```

1. Membuat Grafik Penjualan Ekspor

- Fungsi `ggplot()` digunakan untuk membuat kerangka grafik. Di sini, data yang digunakan adalah `aisi_data`, dan sumbu **x** diatur pada kolom `Bulan`, sedangkan sumbu **y** pada kolom `Penjualan_Ekspor`.
- Argumen `group = 1` memastikan data dihubungkan sebagai satu kelompok garis kontinu.

2. Menambahkan Elemen Grafik

- `geom_line()`: Membuat garis untuk merepresentasikan tren. Warnanya hijau (`color = "green"`), dengan ketebalan garis 1 (`size = 1`).
- `geom_point()`: Menambahkan titik pada setiap bulan untuk memperjelas posisi datanya. Titik-titik ini juga berwarna hijau, dengan ukuran lebih besar (`size = 2`) supaya terlihat jelas.

3. Memberi Label Grafik

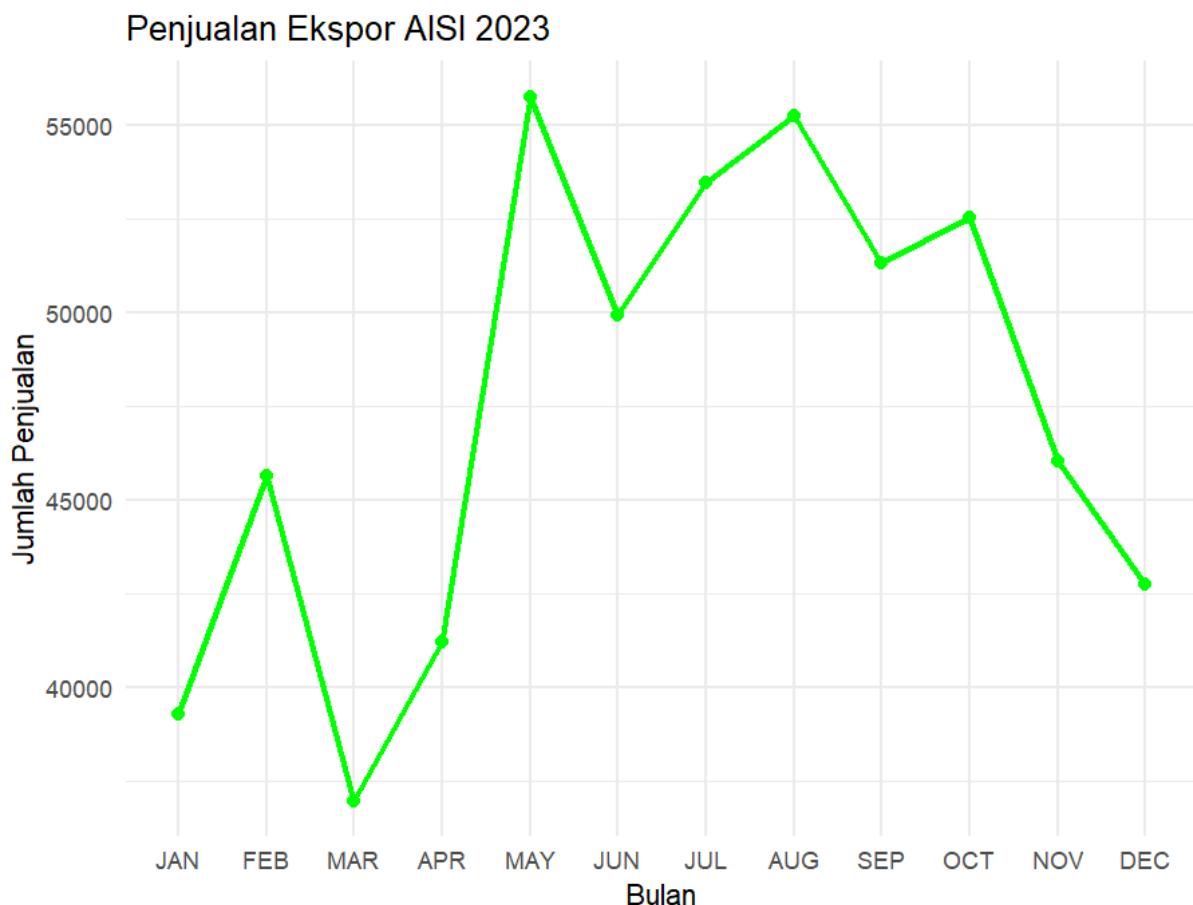
- `labs()`: Memberikan judul grafik "Penjualan Ekspor AISI 2023" dan menambahkan label pada sumbu x (Bulan) serta y (Jumlah Penjualan).

4. Mengatur Tema Grafik

- `theme_minimal()`: Mengatur tampilan grafik menjadi sederhana tetapi tetap estetis, tanpa elemen latar belakang yang berlebihan.

5. Menampilkan Grafik

- `print(plot_domestik)` dan `print(plot_ekspor)`: Digunakan untuk menampilkan grafik penjualan domestik dan ekspor.



Grafik ini menunjukkan jumlah penjualan sepeda motor untuk ekspor sepanjang tahun 2023, dari Januari hingga Desember. Berikut adalah pola yang bisa kita lihat:

1. Awal Tahun (Januari - April):

- Penjualan ekspor dimulai dari **39,269 unit** di Januari, kemudian meningkat pada Februari menjadi **45,627 unit**.
- Penurunan kecil terjadi di Maret (**36,962 unit**), tapi naik lagi pada April menjadi **41,201 unit**.

- **Kesimpulan:** Penjualan cukup fluktuatif di awal tahun, tapi masih dalam kisaran 36–45 ribu unit.
2. **Peningkatan Tengah Tahun (Mei - Agustus):**
- Ada tren peningkatan signifikan mulai Mei (**55,743 unit**) hingga puncaknya di bulan yang sama.
 - Setelah Mei, jumlah ekspor tetap tinggi tapi mulai turun sedikit di bulan Juni (**49,920 unit**) hingga Agustus (**55,247 unit**).
 - **Kesimpulan:** Periode ini adalah **musim penjualan ekspor yang tinggi**, dengan puncak di bulan Mei.
3. **Akhir Tahun (September - Desember):**
- Dari September hingga Oktober, penjualan stabil di angka **51,318–52,517 unit**.
 - Namun, penurunan mulai terlihat di bulan November (**46,020 unit**) dan Desember (**42,737 unit**).
 - **Kesimpulan:** Penjualan di akhir tahun mengalami sedikit penurunan dibandingkan tengah tahun.

Karakteristik Grafik

- **Tren Stabil:** Grafik menunjukkan pola yang relatif stabil dibandingkan penjualan domestik, tanpa fluktuasi besar.
- **Puncak Penjualan:** Terjadi di bulan Mei (55,743 unit), yang bisa menjadi indikasi bulan dengan permintaan ekspor tertinggi.
- **Penurunan Akhir Tahun:** Ada tren penurunan dari November hingga Desember, dengan Desember menjadi bulan dengan penjualan terendah (**42,737 unit**).

Kesimpulan Umum

Grafik ekspor ini memperlihatkan bahwa penjualan ekspor sepeda motor lebih **stabil**, meskipun ada sedikit fluktuasi sepanjang tahun. **Tengah tahun (Mei - Agustus)** adalah masa penjualan yang paling tinggi, sedangkan akhir tahun sedikit menurun. Kolom ekspor menunjukkan **tren yang relatif stabil**, bukan tren yang secara jelas negatif atau positif. Grafik ini bisa memberi insight penting bagi pengelolaan ekspor dan strategi pemasaran perusahaan.

3. Tren Linear

a. Metode Semi Rata-Rata (Domestik)

```
> # === 3. Tren Linear ===
> ## a. Metode Semi Rata-Rata
> # Untuk Tren Negatif pada Kolom Domestik
> semi_average_forecast_dom <- function(data, col_name) {
+   n <- nrow(data) # Menentukan jumlah total baris (data) dalam data frame d
ata.
+   print(paste("Total baris data (n):", n))
+
+   half <- floor(n / 2) # Membagi data menjadi dua bagian yang hampir sama,
+                      # dengan half sebagai jumlah baris pada bagian pertama.
+   print(paste("Jumlah data setengah bagian pertama (half):", half))
+
```

```

+ # Menghitung rata-rata nilai pada setengah pertama dari data (dari bulan 1 hingga tengah).
+ mean_first <- mean(data[[col_name]][1:half])
+ print(paste("Rata-rata setengah bagian pertama (mean_first):", mean_first))
+
+ # Menghitung rata-rata nilai pada setengah kedua dari data (dari tengah hingga akhir).
+ mean_second <- mean(data[[col_name]][(half + 1):n])
+ print(paste("Rata-rata setengah bagian kedua (mean_second):", mean_second))
+
+ # Menghitung kemiringan (slope) dari garis tren dengan mengukur perubahan rata-rata
+ # antara dua setengah data dibagi dengan jumlah data pada masing-masing setengah.
+ slope <- (mean_second - mean_first) / half
+ print(paste("Kemiringan (slope):", slope))
+
+ # Menghitung titik potong (intercept) dari garis tren menggunakan rumus regresi linier.
+ intercept <- mean_first - slope * (half / 2)
+ print(paste("Titik potong (intercept):", intercept))
+
+ # Menentukan indeks bulan yang akan datang (JAN, FEB, MAR)
+ future_indices <- 13:15
+ print("Indeks bulan yang akan datang (future_indices):")
+
+ # Menghitung nilai ramalan untuk bulan-bulan yang akan datang
+ # menggunakan persamaan garis tren (slope dan intercept).
+ future_values <- intercept - slope * future_indices
+ print("Nilai ramalan (future_values):")
+ print(future_values)
+
+ # Print hasil perhitungan
+ cat("\n\n")
+ print(paste("Intercept (a):", intercept))
+ print(paste("Slope (b):", slope))
+ print(paste("Indeks bulan yang akan datang (X):", paste(future_indices, collapse = ", ")))
+ print(paste("Nilai ramalan (Y):", paste(round(future_values, 0), collapse = ", ")))
+
+ # Menampilkan persamaan lengkap
+ print(paste("==>Persamaan Semi Rata-Rata Kolom Domestik(Tren Negatif): Y = ", round(intercept, 2), " + ", round(slope, 2), " * X"))
+
+ # Mengembalikan data frame yang berisi bulan (JAN, FEB, MAR)
+ # beserta nilai ramalan yang sudah dibulatkan.
+ return(data.frame(Bulan = c("JAN", "FEB", "MAR"), Ramalan = round(future_values, 0)))
+ }

```

>

Apa Fungsi Utama Kode Ini?

Kode ini menggunakan metode **Semi Rata-Rata** untuk memprediksi tren negatif di kolom **Penjualan Domestik**. Fungsi ini bertujuan untuk:

1. Membagi data menjadi dua bagian (awal dan akhir).
2. Menghitung rata-rata dari masing-masing bagian.
3. Menggunakan rata-rata tersebut untuk membuat persamaan garis lurus (*linear trend*).
4. Memprediksi penjualan di bulan mendatang (JAN, FEB, MAR) berdasarkan tren negatif yang terlihat.

Langkah-Langkah Fungsi

1. Menentukan Jumlah Data dan Membagi Setengah

```
+ n <- nrow(data) # Menentukan jumlah total baris (data) dalam data frame data.  
+ print(paste("Total baris data (n):", n))  
+  
+ half <- floor(n / 2) # Membagi data menjadi dua bagian yang hampir sama,  
# dengan half sebagai jumlah baris pada bagian pertama.  
+ print(paste("Jumlah data setengah bagian pertama (half):", half))
```

- **n** adalah jumlah baris (atau bulan) dalam dataset. Misalnya, jika ada data dari JAN hingga DEC, maka **n = 12**.
- **half** adalah setengah jumlah data, dipakai untuk membagi dataset menjadi dua bagian: awal dan akhir. Jika datanya ganjil, bagian pertama akan sedikit lebih kecil dari bagian kedua.

2. Menghitung Rata-Rata

```
+ # Menghitung rata-rata nilai pada setengah pertama dari data (dari bulan 1 hingga tengah).  
+ mean_first <- mean(data[[col_name]][1:half])  
+ print(paste("Rata-rata setengah bagian pertama (mean_first):", mean_first))  
+  
+ # Menghitung rata-rata nilai pada setengah kedua dari data (dari tengah hingga akhir).  
+ mean_second <- mean(data[[col_name]][(half + 1):n])  
+ print(paste("Rata-rata setengah bagian kedua (mean_second):", mean_second))
```

- **mean_first** menghitung rata-rata penjualan di bagian awal (setengah pertama).
- **mean_second** menghitung rata-rata penjualan di bagian akhir (setengah kedua).

Contoh: Jika data awal adalah:

- **Setengah pertama:** JAN-JUN: 4000, 4500, 4200, ...
- **Setengah kedua:** JUL-DEC: 3800, 3700, 3600, ...

Maka:

- **mean_first** mungkin sekitar **4350**.
- **mean_second** mungkin sekitar **3700**.

3. Menghitung Kemiringan (Slope)

```
+ # Menghitung kemiringan (slope) dari garis tren dengan mengukur perubahan rata-rata  
+ # antara dua setengah data dibagi dengan jumlah data pada masing-masing setengah.  
+ slope <- (mean_second - mean_first) / half  
+ print(paste("Kemiringan (slope):", slope))
```

- **slope** adalah tingkat perubahan penjualan setiap bulan dari awal ke akhir. Jika nilainya negatif, berarti terjadi penurunan (tren negatif).

4. Menghitung Titik Potong (Intercept)

```
+ # Menghitung titik potong (intercept) dari garis tren menggunakan rumus regresi linier.
+ intercept <- mean_first - slope * (half / 2)
+ print(paste("Titik potong (intercept):", intercept))
```

- **intercept** adalah nilai awal pada garis tren, mewakili perkiraan penjualan di bulan pertama garis tren.

5. Memprediksi Bulan Mendatang

```
+ # Menentukan indeks bulan yang akan datang (JAN, FEB, MAR)
+ future_indices <- 13:15
+ print("Indeks bulan yang akan datang (future_indices):")
+
+ # Menghitung nilai ramalan untuk bulan-bulan yang akan datang
+ # menggunakan persamaan garis tren (slope dan intercept).
+ future_values <- intercept - slope * future_indices
+ print("Nilai ramalan (future_values):")
+ print(future_indices)
```

- **future_indices** adalah bulan-bulan yang akan datang: JAN (bulan ke-13), FEB (14), dan MAR (15).
- **future_values** adalah nilai ramalan penjualan untuk bulan-bulan tersebut, dihitung menggunakan persamaan garis lurus:

$$Y = \text{intercept} + \text{slope} * X.$$

6. Menampilkan Hasil

```
+ # Print hasil perhitungan
+ cat("\n\n")
+ print(paste("Intercept (a):", intercept))
+ print(paste("Slope (b):", slope))
+ print(paste("Indeks bulan yang akan datang (X):", paste(future_indices, collapse = ", ")))
+ print(paste("Nilai ramalan (Y):", paste(round(future_values, 0), collapse = ", ")))
+
+ # Menampilkan persamaan lengkap
+ print(paste("==Persamaan Semi Rata-Rata Kolom Domestik(Tren Negatif): Y = ", round(intercept, 2), " + ", round(slope, 2), " * X"))
+
+ # Mengembalikan data frame yang berisi bulan (JAN, FEB, MAR)
+ # beserta nilai ramalan yang sudah dibulatkan.
```

- Persamaan garis tren ditampilkan agar mudah dipahami. Contohnya:

$$Y = 5000 - 200 * X.$$

Ini berarti penjualan menurun sebesar 200 setiap bulan.

7. Mengembalikan Data Frame

```
+ return(data.frame(Bulan = c("JAN", "FEB", "MAR"), Ramalan = round(future_values, 0)))
+ }
```

- Outputnya adalah tabel berisi ramalan penjualan untuk JAN, FEB, dan MAR.

a. Metode Semi Rata-Rata (Ekspor)

```
> # Untuk Tren Positif pada Kolom Ekspor
> semi_average_forecast_eks <- function(data, col_name) {
+   n <- nrow(data) # Menentukan jumlah total baris (data) dalam data frame data.
+   print(paste("Total baris data (n):", n))
+
+   half <- floor(n / 2) # Membagi data menjadi dua bagian yang hampir sama,
+   # dengan half sebagai jumlah baris pada bagian pertama.
+   print(paste("Jumlah data setengah bagian pertama (half):", half))
+
+   # Menghitung rata-rata nilai pada setengah pertama dari data (dari bulan 1 hingga tengah).
+   mean_first <- mean(data[[col_name]][1:half])
+   print(paste("Rata-rata setengah bagian pertama (mean_first):", mean_first))
+
+   # Menghitung rata-rata nilai pada setengah kedua dari data (dari tengah hingga akhir).
+   mean_second <- mean(data[[col_name]][(half + 1):n])
+   print(paste("Rata-rata setengah bagian kedua (mean_second):", mean_second))
+
+   # Menghitung kemiringan (slope) dari garis tren dengan mengukur perubahan rata-rata
+   # antara dua setengah data dibagi dengan jumlah data pada masing-masing setengah.
+   slope <- (mean_second - mean_first) / half
+   print(paste("Kemiringan (slope):", slope))
+
+   # Menghitung titik potong (intercept) dari garis tren menggunakan rumus regresi linier.
+   intercept <- mean_first + slope * (half / 2)
+   print(paste("Titik potong (intercept):", intercept))
+
+   # Menentukan indeks bulan yang akan datang (JAN, FEB, MAR)
+   future_indices <- 13:15
+   print("Indeks bulan yang akan datang (future_indices):")
+
+   # Menghitung nilai ramalan untuk bulan-bulan yang akan datang
+   # menggunakan persamaan garis tren (slope dan intercept).
+   future_values <- intercept - slope * future_indices
+   print("Nilai ramalan (future_values):")
+   print(future_values)
+
+   # Print hasil perhitungan
+   cat("\n\n")
+   print(paste("Intercept (a):", intercept))
+   print(paste("Slope (b):", slope))
+   print(paste("Indeks bulan yang akan datang (X):", paste(future_indices, collapse = ", ")))
+   print(paste("Nilai ramalan (Y):", paste(round(future_values, 0), collapse = ", ")))
+
+   # Menampilkan persamaan lengkap
+   print(paste("==Persamaan Semi Rata-Rata Kolom Ekspor(Tren Positif): Y = ", round(intercept, 2), " + ", round(slope, 2), " * X"))
+
+   # Mengembalikan data frame yang berisi bulan (JAN, FEB, MAR)
+   # beserta nilai ramalan yang sudah dibulatkan.
+   return(data.frame(Bulan = c("JAN", "FEB", "MAR"), Ramalan = round(future_values, 0)))
+ }
```

```

> cat("\n====Penjualan Domestik")

====Penjualan Domestik> ramalan_semi_dom <- semi_average_forecast_dom(aisi_data, "Penjualan_Domestik")
[1] "Total baris data (n): 12"
[1] "Jumlah data setengah bagian pertama (half): 6"
[1] "Rata-rata setengah bagian pertama (mean_first): 533655"
[1] "Rata-rata setengah bagian kedua (mean_second): 505843.666666667"
[1] "Kemiringan (slope): -4635.2222222222"
[1] "Titik potong (intercept): 547560.666666667"
[1] "Indeks bulan yang akan datang (future_indices):"
[1] "Nilai ramalan (future_values):"
[1] 13 14 15

[1] "Intercept (a): 547560.666666667"
[1] "Slope (b): -4635.2222222222"
[1] "Indeks bulan yang akan datang (X): 13, 14, 15"
[1] "Nilai ramalan (Y): 607819, 612454, 617089"
[1] "====Persamaan Semi Rata-Rata Kolom Domestik(Tren Negatif): Y = 547560.67 + -4635.22 * X"
> cat("Hasil Ramalan Semi Rata-Rata (Domestik):\n")
Hasil Ramalan Semi Rata-Rata (Domestik):
> print(ramalan_semi_dom)
Bulan Ramalan
1 JAN 607819
2 FEB 612454
3 MAR 617089
>
> cat("\n====Penjualan Ekspor")

====Penjualan Ekspor> ramalan_semi_eks <- semi_average_forecast_eks(aisi_data, "Penjualan_Ekspor")
[1] "Total baris data (n): 12"
[1] "Jumlah data setengah bagian pertama (half): 6"
[1] "Rata-rata setengah bagian pertama (mean_first): 44787"
[1] "Rata-rata setengah bagian kedua (mean_second): 50213.666666667"
[1] "Kemiringan (slope): 904.444444444444"
[1] "Titik potong (intercept): 47500.3333333333"
[1] "Indeks bulan yang akan datang (future_indices):"
[1] "Nilai ramalan (future_values):"
[1] 13 14 15

[1] "Intercept (a): 47500.3333333333"
[1] "Slope (b): 904.444444444444"
[1] "Indeks bulan yang akan datang (X): 13, 14, 15"
[1] "Nilai ramalan (Y): 35743, 34838, 33934"
[1] "====Persamaan Semi Rata-Rata Kolom Ekspor(Tren Positif): Y = 47500.33 + 904.44 * X"
> cat("\nHasil Ramalan Semi Rata-Rata (Ekspor):\n")

Hasil Ramalan Semi Rata-Rata (Ekspor):
> print(ramalan_semi_eks)
Bulan Ramalan
1 JAN 35743
2 FEB 34838
3 MAR 33934
>
```

Analisis

- **Penjualan Domestik** menunjukkan penurunan secara perlahan (tren negatif), karena nilai ramalan terus meningkat pada bulan mendatang, meskipun secara keseluruhan mengalami penurunan.

- **Penjualan Ekspor** menunjukkan penurunan yang lebih cepat, karena kemiringannya positif, tetapi ramalan tetap menunjukkan nilai yang lebih rendah pada bulan mendatang.

b. Metode Kuadrat Terkecil (Linear Regression)

```
> ## b. Metode Kuadrat Terkecil (Linear Regression)
>
> # Menambahkan kolom Index ke dataset
> aisi_data$Index <- seq_len(nrow(aiisi_data)) # Menambahkan urutan 1, 2,
3, ..., sesuai jumlah baris data
>
> ramalan_linear <- function(data, col_name) {
+   # Membuat model linear menggunakan fungsi lm()
+   model <- lm(data[[col_name]] ~ Index, data = data)
+
+   # Menampilkan ringkasan model
+   print("Ringkasan Model Linear:")
+   print(summary(model))
+
+   # Menentukan indeks bulan yang akan datang (JAN, FEB, MAR)
+   future_indices <- data.frame(Index = 13:15)
+   print("Indeks bulan yang akan datang (future_indices):")
+   print(future_indices)
+
+   # Menghitung nilai ramalan untuk bulan-bulan yang akan datang
+   future_values <- predict(model, newdata = future_indices)
+   print("Nilai ramalan (future_values):")
+   print(future_values)
+
+   # Menampilkan hasil perhitungan
+   cat("\n\n")
+   print(paste("Model Linear untuk kolom", col_name, ":"))
+   print(paste("Nilai ramalan (Y) untuk bulan JAN, FEB, MAR:", paste(round(future_values, 0), collapse = ", ")))
+
+   # Menampilkan persamaan model linear (Y = a + bx)
+   coef_intercept <- coef(model)[1] # Intercept (a)
+   coef_slope <- coef(model)[2] # Slope (b)
+   print(paste("Persamaan Linear Regression: Y = ", round(coef_intercept,
2), " + ", round(coef_slope, 2), " * X"))
+
+   # Mengembalikan data frame yang berisi bulan (JAN, FEB, MAR)
+   # beserta nilai ramalan yang sudah dibulatkan.
+   return(data.frame(Bulan = c("JAN", "FEB", "MAR"), Ramalan = round(future_values, 0)))
+ }
>
> cat("\n====Penjualan Domestik")

====Penjualan Domestik> ramalan_linear_dom <- ramalan_linear(aiisi_data, "Pe
njualan_Domestik")
[1] "Ringkasan Model Linear:"
```

Call:

```
lm(formula = data[[col_name]] ~ Index, data = data)
```

Residuals:

Min	1Q	Median	3Q	Max
-184205	-32448	15463	33011	87115

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	568575	47196	12.047	2.82e-07 ***
Index	-7512	6413	-1.171	0.269

```

---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 76680 on 10 degrees of freedom
Multiple R-squared: 0.1207, Adjusted R-squared: 0.03272
F-statistic: 1.372 on 1 and 10 DF, p-value: 0.2686

[1] "Indeks bulan yang akan datang (future_indices):"
Index
1 13
2 14
3 15
[1] "Nilai ramalan (future_values):"
    1   2   3
470924.0 463412.4 455900.8

[1] "Model Linear untuk kolom Penjualan_Domestik :"
[1] "Nilai ramalan (Y) untuk bulan JAN, FEB, MAR: 470924, 463412, 455901"
[1] "Persamaan Linear Regression: Y = 568574.65 + -7511.59 * x"
> cat("Hasil Ramalan Kuadrat Terkecil (Domestik):\n")
Hasil Ramalan Kuadrat Terkecil (Domestik):
> print(ramalan_linear_dom)
Bulan Ramalan
1 JAN 470924
2 FEB 463412
3 MAR 455901
>
> cat("\n====Penjualan Ekspor")

====Penjualan Ekspor> ramalan_linear_eks <- ramalan_linear(aisi_data, "Penjualan_Ekspor")
[1] "Ringkasan Model Linear:"

Call:
lm(formula = data[[col_name]] ~ Index, data = data)

Residuals:
    Min      1Q  Median      3Q     Max 
-8671  -4562   1683   3478   9308 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 42882.4    3811.7 11.250 5.35e-07 ***
Index        710.5     517.9   1.372     0.2    
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6193 on 10 degrees of freedom
Multiple R-squared: 0.1584, Adjusted R-squared: 0.07421
F-statistic: 1.882 on 1 and 10 DF, p-value: 0.2001

[1] "Indeks bulan yang akan datang (future_indices):"
Index
1 13
2 14
3 15
[1] "Nilai ramalan (future_values):"
    1   2   3
52118.29 52828.74 53539.20

[1] "Model Linear untuk kolom Penjualan_Ekspor :"
[1] "Nilai ramalan (Y) untuk bulan JAN, FEB, MAR: 52118, 52829, 53539"
[1] "Persamaan Linear Regression: Y = 42882.38 + 710.45 * x"
> cat("\nHasil Ramalan Kuadrat Terkecil (Ekspor):\n")

Hasil Ramalan Kuadrat Terkecil (Ekspor):
> print(ramalan_linear_eks)

```

	Bulan	Ramalan
1	JAN	52118
2	FEB	52829
3	MAR	53539

- Model regresi linear ini menunjukkan bahwa **penjualan domestik** cenderung menurun seiring berjalannya waktu, sesuai dengan nilai kemiringan (slope) yang negatif.
- Model regresi linear ini menunjukkan bahwa **penjualan ekspor** memiliki tren positif meskipun dengan kemiringan yang tidak terlalu besar, yang menunjukkan kenaikan yang stabil untuk bulan-bulan mendatang.

4. Tren Non-Linear

a. Metode Kuadratis

```
> # === 4. Tren Non-Linear ===
>
> # Menambahkan kolom Index jika belum ada
> if (!"Index" %in% colnames(aisi_data)) {
+   aisi_data$Index <- 1:nrow(aisi_data) # Mengisi kolom Index dari 1 hingga jumlah baris data
+ }
>
> # Menambahkan variabel kuadrat ke dataset
> aisi_data$Index_Squared <- aisi_data$Index^2
>
> # Fungsi untuk model kuadratis dan prediksi nilai di masa depan
> quadratic_model_forecast <- function(data, col_name) {
+   # Membuat model regresi kuadratis
+   model <- lm(data[[col_name]] ~ Index + Index_Squared, data = data)
+
+   # Menampilkan persamaan model
+   coef <- coef(model)
+   persamaan <- paste("Y = ", round(coef[1], 2),
+                      "+", round(coef[2], 2), "* X",
+                      "+", round(coef[3], 2), "* X^2")
+   cat("\nPersamaan Model Kuadratis:\n", persamaan, "\n")
+
+   # Menentukan indeks bulan masa depan (JAN, FEB, MAR)
+   future_indices <- data.frame(Index = 13:15, Index_Squared = (13:15)^2)
+
+   # Membuat prediksi
+   future_values <- predict(model, newdata = future_indices)
+
+   # Menampilkan ringkasan hasil
+   hasil <- data.frame(Bulan = c("JAN", "FEB", "MAR"), Ramalan = round(future_values, 0))
+   return(hasil)
+ }
>
> # Memprediksi penjualan domestik
> cat("\n==== Penjualan Domestik (Metode Kuadratis) ====")

==== Penjualan Domestik (Metode Kuadratis) ====> ramalan_quad_dom <- quadratic_model_forecast(aisi_data, "Penjualan_Domestik")

Persamaan Model Kuadratis:
Y = 626882.68 + -32500.74 * X + 1922.24 * X^2
> cat("\nHasil Ramalan Kuadratis (Domestik):\n")
```

```

Hasil Ramalan Kuadratis (Domestik):
> print(ramalan_quad_dom)
Bulan Ramalan
1 JAN 529232
2 FEB 548632
3 MAR 571876
>
> # Memprediksi penjualan ekspor
> cat("\n==== Penjualan Ekspor (Metode Kuadratis) ===")

==== Penjualan Ekspor (Metode Kuadratis) ===> ramalan_quad_eks <- quadratic
_model_forecast(aisi_data, "Penjualan_Ekspor")

Persamaan Model Kuadratis:
Y = 31380.45 + 5639.85 * X + -379.18 * X^2
> cat("\nHasil Ramalan Kuadratis (Ekspor):\n")

Hasil Ramalan Kuadratis (Ekspor):
> print(ramalan_quad_eks)
Bulan Ramalan
1 JAN 40616
2 FEB 36018
3 MAR 30662

```

Kesimpulan:

- **Penjualan Domestik** menunjukkan tren peningkatan yang lebih kompleks dengan model kuadratis, berbeda dengan tren penurunan yang diprediksi oleh model linear.
- **Penjualan Ekspor** mengalami penurunan, dengan puncak di bulan Januari, yang bisa menunjukkan adanya faktor eksternal yang memengaruhi hasil ekspor.

b. Metode Eksponensial

```

> ## b. Metode Eksponensial
> # Fungsi untuk metode eksponensial
> exponential_model_forecast <- function(data, col_name) {
+   # Menambahkan kolom log dari data penjualan
+   data$log_penjualan <- log(data[[col_name]])
+
+   # Membuat model regresi linier pada skala logaritmik
+   model <- lm(log_penjualan ~ Index, data = data)
+
+   # Menampilkan persamaan model
+   coef <- coef(model)
+   persamaan <- paste("ln(Y) = ", round(coef[1], 2), "+", round(coef[2],
2), "* X")
+   cat("\nPersamaan Model Eksponensial:\n", persamaan, "\n")
+
+   # Menentukan indeks bulan masa depan (JAN, FEB, MAR)
+   future_indices <- data.frame(Index = 13:15)
+
+   # Membuat prediksi log nilai masa depan
+   future_log_values <- predict(model, newdata = future_indices)
+
+   # Mengubah prediksi log menjadi nilai eksponensial
+   future_values <- exp(future_log_values)
+
+   # Menampilkan hasil perhitungan
+   hasil <- data.frame(Bulan = c("JAN", "FEB", "MAR"), Ramalan = round(fu
ture_values, 0))
+   return(hasil)
+ }
>
```

```

> # Memprediksi penjualan domestik
> cat("\n==== Penjualan Domestik (Metode Eksponensial) ===")
==== Penjualan Domestik (Metode Eksponensial) ===> ramalan_exp_dom <- exponential_model_forecast(aisi_data, "Penjualan_Domestik")

Persamaan Model Eksponensial:
 $\ln(Y) = 13.23 + -0.01 * X$ 
> cat("\nHasil Ramalan Eksponensial (Domestik):\n")

Hasil Ramalan Eksponensial (Domestik):
> print(ramalan_exp_dom)
Bulan Ramalan
1 JAN 472545
2 FEB 466480
3 MAR 460492
>
> # Memprediksi penjualan ekspor
> cat("\n==== Penjualan Ekspor (Metode Eksponensial) ===")
==== Penjualan Ekspor (Metode Eksponensial) ===> ramalan_exp_eks <- exponential_model_forecast(aisi_data, "Penjualan_Ekspor")

Persamaan Model Eksponensial:
 $\ln(Y) = 10.66 + 0.02 * X$ 
> cat("\nHasil Ramalan Eksponensial (Ekspor):\n")

Hasil Ramalan Eksponensial (Ekspor):
> print(ramalan_exp_eks)
Bulan Ramalan
1 JAN 52285
2 FEB 53133
3 MAR 53996

```

Kesimpulan:

- **Domestik:** Model eksponensial ini menunjukkan penurunan dalam penjualan domestik seiring waktu, karena koefisien untuk X bernilai negatif. Artinya, penurunan akan semakin berkurang tetapi tetap stabil..
- **Ekspor:** Model eksponensial ini menunjukkan adanya peningkatan dalam penjualan ekspor setiap bulan, dengan tingkat pertumbuhan yang relatif stabil (dengan nilai koefisien positif). Ini menunjukkan ada pertumbuhan yang teratur dalam penjualan ekspor.

5. Tabel Hasil Keseluruhan

```

> # === 5. Tabel Hasil Keseluruhan ===
> # Pastikan dataset ramalan sudah memiliki nama bulan
> ramalan_semi_dom$Bulan <- c("JAN", "FEB", "MAR")
> ramalan_semi_eks$Bulan <- c("JAN", "FEB", "MAR")
> ramalan_linear_dom$Bulan <- c("JAN", "FEB", "MAR")
> ramalan_linear_eks$Bulan <- c("JAN", "FEB", "MAR")
> ramalan_quad_dom$Bulan <- c("JAN", "FEB", "MAR")
> ramalan_quad_eks$Bulan <- c("JAN", "FEB", "MAR")
> ramalan_exp_dom$Bulan <- c("JAN", "FEB", "MAR")
> ramalan_exp_eks$Bulan <- c("JAN", "FEB", "MAR")
>
>
> # Buat tabel hasil domestik dan ekspor
> hasil_domestik <- cbind(
+   Semi_Rata_Rata = ramalan_semi_dom$Ramalan,

```

```

+ Kuadrat_Terkecil = ramalan_linear_dom$Ramalan,
+ Kuadratis = ramalan_quad_dom$RamaTan,
+ Eksponensial = ramalan_exp_dom$Ramalan
+ )
> rownames(hasil_domestik) <- ramalan_semi_dom$Bulan
>
> hasil_ekspor <- cbind(
+   Semi_Rata_Rata = ramalan_semi_eks$Ramalan,
+   Kuadrat_Terkecil = ramalan_linear_eks$Ramalan,
+   Kuadratis = ramalan_quad_eks$RamaTan,
+   Eksponensial = ramalan_exp_eks$Ramalan
+ )
> rownames(hasil_ekspor) <- ramalan_semi_eks$Bulan
>
>
> cat("\nTabel Hasil Penjualan Domestik (Keempat Metode):\n")

Tabel Hasil Penjualan Domestik (Keempat Metode):
> print(hasil_domestik)
  Semi_Rata_Rata Kuadrat_Terkecil Kuadratis Eksponensial
JAN       607819          470924    529232     472545
FEB       612454          463412    548632     466480
MAR       617089          455901    571876     460492
>
> cat("\nTabel Hasil Penjualan Ekspor (Keempat Metode):\n")

Tabel Hasil Penjualan Ekspor (Keempat Metode):
> print(hasil_ekspor)
  Semi_Rata_Rata Kuadrat_Terkecil Kuadratis Eksponensial
JAN       35743           52118     40616      52285
FEB       34838           52829     36018      53133
MAR       33934           53539     30662      53996

```

Penjelasan

- **Semi Rata-Rata:** Metode rata-rata bergerak sederhana, memberikan hasil yang lebih stabil tanpa perubahan besar.
- **Kuadrat Terkecil:** Pendekatan linier yang meminimalkan kesalahan kuadrat, memberikan hasil yang cukup akurat meskipun tidak selalu mempertimbangkan tren non-linier.
- **Kuadratis:** Model yang mempertimbangkan efek kuadratis untuk penurunan atau peningkatan lebih kompleks, dengan hasil yang menunjukkan variasi lebih besar dalam beberapa bulan.
- **Eksponensial:** Model yang menggunakan logaritma untuk transformasi eksponensial, menghasilkan prediksi yang cenderung lebih cepat bertumbuh atau menurun.

6. Menghitung Measure of Accuracy

```

> # === 6. Menghitung Measure of Accuracy ===
> # Fungsi untuk menghitung MAPE, MAD, dan MSD
> measure_of_accuracy <- function(actual, predicted) {
+   n <- length(actual)
+   errors <- actual - predicted
+   mape <- mean(abs(errors / actual)) * 100
+   mad <- mean(abs(errors))
+   msd <- mean(errors^2)
+   return(list(MAPE = mape, MAD = mad, MSD = msd))
+ }
>

```

```

> # Fungsi untuk menghitung accuracy untuk semua metode
> calculate_accuracy <- function(data, ramalan, col_name) {
+   actual <- data[[col_name]]
+   mape <- mad <- numeric(length(ramalan))
+
+   for (i in seq_along(ramalan)) {
+     predicted <- ramalan[[i]]$Ramalan
+     acc <- measure_of_accuracy(actual, predicted)
+     mape[i] <- acc$MAPE
+     mad[i] <- acc$MAD
+     msd[i] <- acc$MSD
+   }
+
+   return(data.frame(
+     Metode = c("Semi Rata-Rata", "Kuadrat Terkecil", "Kuadratis", "Eksponensial"),
+     MAPE = round(mape, 2),
+     MAD = round(mad, 2),
+     MSD = round(msd, 2)
+   ))
+ }
>
> # List ramalan untuk masing-masing kolom
> ramalan_domestik <- list(ramalan_semi_dom, ramalan_linear_dom, ramalan_quad_dom, ra
malan_exp_dom)
> ramalan_ekspor <- list(ramalan_semi_eks, ramalan_linear_eks, ramalan_quad_eks, rama
lan_exp_eks)
>
> # Menghitung accuracy untuk penjualan domestik
> accuracy_domestik <- calculate_accuracy(aisi_data, ramalan_domestik, "Penjualan_Dom
estik")
>
> # Menghitung accuracy untuk penjualan ekspor
> accuracy_ekspor <- calculate_accuracy(aisi_data, ramalan_ekspor, "Penjualan_Ekspor
")
> # === 7. Membuat Tabel Hasil Keseluruhan ===
>
> cat("\nMeasure of Accuracy untuk Penjualan Domestik:\n")

Measure of Accuracy untuk Penjualan Domestik:
> print(accuracy_domestik)
      Metode    MAPE      MAD      MSD
1  Semi Rata-Rata 21.26 96648.50 14102181872
2  Kuadrat Terkecil 15.28 80581.83 8912563380
3    Kuadratis 13.64 63111.33 6465071322
4    Eksponensial 14.94 78523.83 8535678442
>
> cat("\nMeasure of Accuracy untuk Penjualan Ekspor:\n")

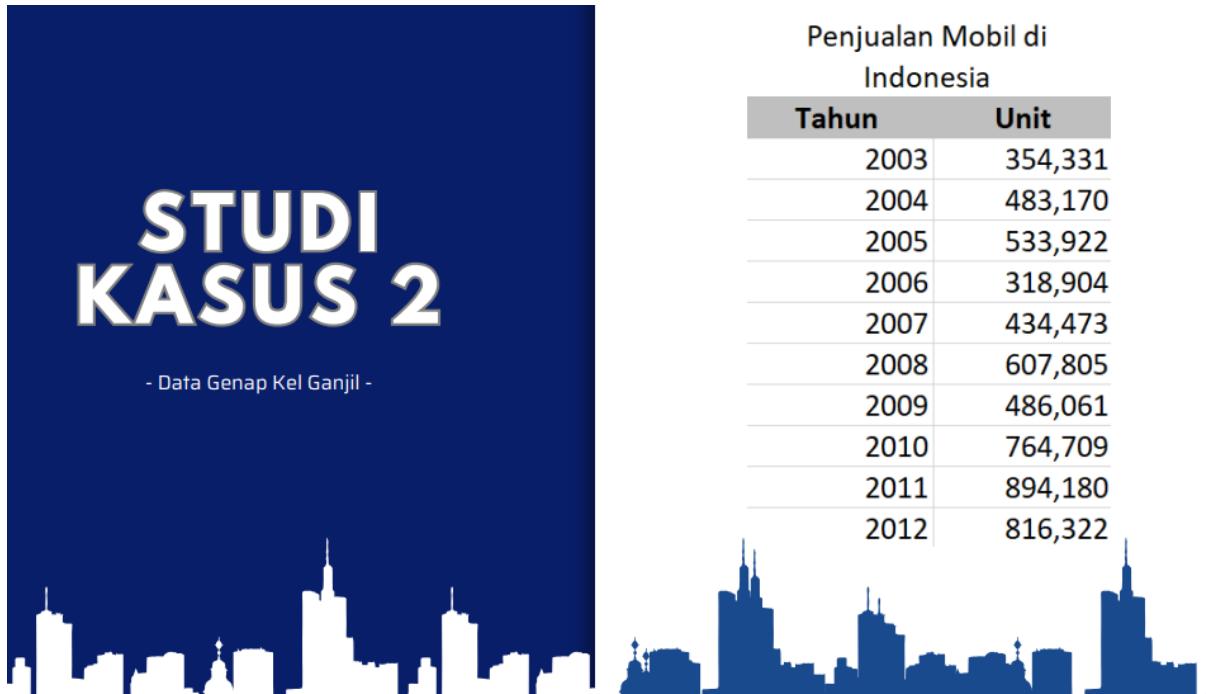
Measure of Accuracy untuk Penjualan Ekspor:
> print(accuracy_ekspor)
      Metode    MAPE      MAD      MSD
1  Semi Rata-Rata 25.36 12662.00 198023559
2  Kuadrat Terkecil 15.36 6504.33 67356005
3    Kuadratis 24.05 11959.50 186880153
4    Eksponensial 15.74 6656.67 71049142
>
```

Penjelasan:

- **MAPE (Mean Absolute Percentage Error):** Mengukur kesalahan dalam persentase. Semakin kecil nilainya, semakin akurat prediksi.
- **MAD (Mean Absolute Deviation):** Mengukur rata-rata selisih absolut antara nilai prediksi dan nilai aktual. Nilai yang lebih rendah menunjukkan prediksi lebih akurat.
- **MSD (Mean Squared Deviation):** Mengukur rata-rata dari kuadrat kesalahan. Semakin kecil nilai MSD, semakin baik model dalam memprediksi.

Dari hasil tersebut, **metode Kuadrat Terkecil** dan **Eksponensial** memberikan akurasi terbaik untuk kedua jenis penjualan, dengan nilai MAPE yang relatif rendah, meskipun masing-masing menunjukkan hasil yang sedikit berbeda tergantung pada jenis data yang dianalisis.

Studi Kasus 2: Data Genap Kelompok Ganjil(Penjualan Mobil di Indonesia Tahun 2003-2012)



1. Memasukkan dan memuat data

```
> # 1. Memasukkan dan memuat data
> # Data penjualan mobil
> year <- c(2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012)
> units <- c(354331, 483170, 533922, 318904, 434473, 607805, 486061, 76470
9, 894180, 816322)
>
> # Membuat data frame
> car_sales_data <- data.frame(
+   Year = year,
+   Units = units
+ )
>
> # Menampilkan data
> print(car_sales_data)
  Year Units
1 2003 354331
2 2004 483170
3 2005 533922
4 2006 318904
5 2007 434473
6 2008 607805
7 2009 486061
8 2010 764709
9 2011 894180
10 2012 816322
```

- **Memasukkan dan memuat data**

Di sini, kita mulai dengan mendeklarasikan dua vektor, yaitu `year` dan `units`.

- **year** berisi data tahun penjualan mobil, mulai dari tahun 2003 hingga 2012.
- **units** berisi jumlah unit mobil yang terjual di setiap tahun tersebut.

Jadi, **year** dan **units** adalah dua kumpulan data yang akan kita gunakan untuk menganalisis tren penjualan mobil.

• Membuat data frame

Setelah data untuk tahun dan jumlah unit mobil disiapkan, kita menggabungkan keduanya dalam sebuah **data frame** yang diberi nama **car_sales_data**.

- **data.frame()** adalah fungsi yang digunakan untuk membuat tabel data yang lebih terstruktur (mirip tabel di Excel atau spreadsheet).
- Di dalam data frame ini, kolom pertama adalah **Year** yang berisi tahun, dan kolom kedua adalah **Units** yang berisi jumlah unit mobil yang terjual pada tahun tersebut.

• Menampilkan data

Setelah data dimasukkan ke dalam **car_sales_data**, kita menggunakan fungsi **print()** untuk menampilkan data ini ke layar. Jadi, output yang akan muncul adalah tabel dengan dua kolom: **Year** dan **Units**, yang berisi tahun penjualan dan jumlah unit mobil yang terjual.

2. Visualisasi data penjualan awal

```
> # 2. Visualisasi data penjualan awal
> # Memuat library untuk visualisasi
> library(ggplot2)
> ggplot(car_sales_data, aes(x = Year, y = Units)) +
+   geom_line(color = "blue", size = 1) +
+   geom_point(color = "blue", size = 2) +
+   labs(
+     title = "Tren Penjualan Mobil di Indonesia (Data Awal)",
+     x = "Tahun",
+     y = "Unit Terjual"
+   ) +
+   scale_x_continuous(breaks = seq(2003, 2012, by = 1)) + # Menampilkan
semua tahun
+   scale_y_continuous(labels = scales::comma) +           # Format angka
dengan koma
+   theme_minimal()
```

• Memuat library untuk visualisasi

Di sini, kita menggunakan **library(ggplot2)**.

- **ggplot2** adalah library di R yang sering digunakan untuk membuat grafik dan visualisasi data. Jadi, sebelum kita bisa membuat grafik, kita perlu "memanggil" atau memuat library ini agar bisa digunakan.

• Membuat grafik dengan ggplot

Selanjutnya, kita membuat grafik menggunakan **ggplot()**. Ini adalah cara dasar untuk membuat grafik di **ggplot2**.

- `ggplot(car_sales_data, aes(x = Year, y = Units)):`
Di sini, kita bilang ke `ggplot` untuk mengambil data dari `car_sales_data` dan menentukan bahwa sumbu X (horizontal) akan diisi dengan data `Year` (tahun) dan sumbu Y (vertikal) akan diisi dengan `Units` (jumlah unit yang terjual).
- `geom_line(color = "blue", size = 1):`
Fungsi ini menggambar garis yang menghubungkan titik-titik data. Warna garis diatur menjadi biru (`color = "blue"`) dan ketebalan garis diatur menjadi 1 (`size = 1`). Jadi, kita melihat tren penjualan mobil dalam bentuk garis biru.
- `geom_point(color = "blue", size = 2):`
Fungsi ini menambahkan titik-titik di sepanjang garis yang menunjukkan data sebenarnya. Titik ini juga berwarna biru (`color = "blue"`) dan ukurannya diset menjadi 2 (`size = 2`), jadi titik-titik ini cukup jelas terlihat.

• Menambahkan judul dan label

- `labs(title = "Tren Penjualan Mobil di Indonesia (Data Awal)", x = "Tahun", y = "Unit Terjual"):`
Kita memberi judul grafik dengan `title`: "Tren Penjualan Mobil di Indonesia (Data Awal)".
Kita juga memberi label pada sumbu X sebagai "Tahun" dan sumbu Y sebagai "Unit Terjual" agar grafik lebih mudah dimengerti.

• Menyesuaikan tampilan sumbu X dan Y

- `scale_x_continuous(breaks = seq(2003, 2012, by = 1)):`
Di sini, kita menyetel sumbu X untuk menampilkan semua tahun dari 2003 hingga 2012 dengan interval 1 tahun. Jadi, semua tahun akan tampak jelas di sumbu X.
- `scale_y_continuous(labels = scales::comma):`
Fungsi ini mengubah format angka di sumbu Y agar lebih mudah dibaca, dengan menambahkan koma sebagai pemisah ribuan. Jadi, angka seperti 1.000.000 akan tampil sebagai 1,000,000.

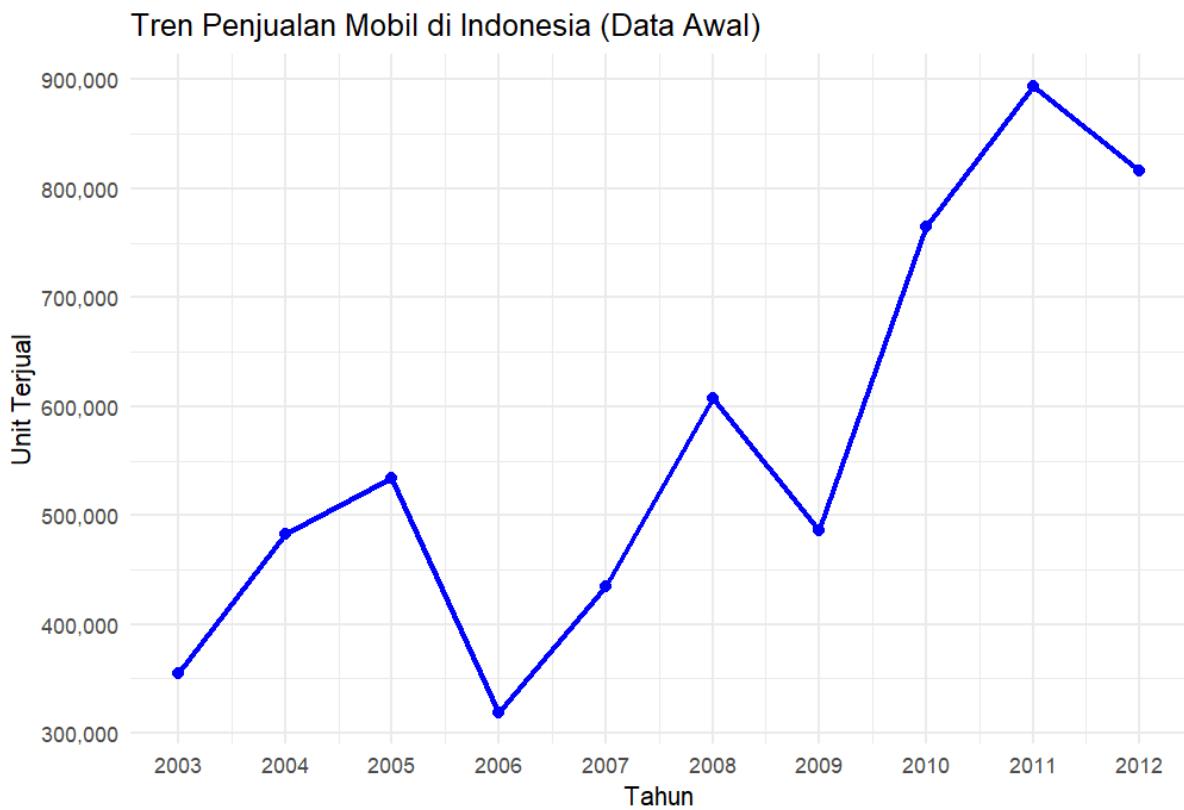
• Menambahkan tema minimalis

- `theme_minimal():`
Ini membuat tampilan grafik menjadi lebih bersih dan sederhana, tanpa gangguan dari elemen visual lain yang tidak perlu. Grafik akan lebih fokus pada data.

Hasil Grafik:

Dengan kode ini, kita akan mendapatkan grafik garis yang menunjukkan **tren penjualan mobil dari tahun 2003 hingga 2012**. Setiap titik di grafik menunjukkan jumlah mobil yang terjual pada tahun tertentu, dan garis biru yang menghubungkannya menunjukkan bagaimana penjualan berubah seiring waktu. Judul dan label yang sudah kita tambahkan akan membuat grafik lebih mudah dimengerti oleh orang lain yang melihatnya.

Berikut grafik yang dihasilkan:



Berdasarkan data penjualan mobil dari tahun 2003 hingga 2012, kita memiliki dua kolom penting: **Year** (Tahun) dan **units** (Jumlah unit terjual). Berikut adalah gambaran singkat tentang tren yang bisa kita harapkan dari grafik tersebut:

1. Perubahan Penjualan Seiring Waktu

- **Tahun 2003 hingga 2005:**
Pada tahun-tahun awal (2003 hingga 2005), kita dapat melihat bahwa penjualan mobil meningkat cukup signifikan dari 354.331 unit (2003) menjadi 533.922 unit (2005). Grafik akan menunjukkan garis yang naik secara bertahap pada periode ini.
- **Tahun 2006 hingga 2007:**
Pada tahun 2006, ada penurunan tajam dalam penjualan menjadi 318.904 unit. Jika dilihat dari grafik, garis akan turun drastis di titik ini. Namun, pada tahun 2007, penjualan mulai naik lagi menjadi 434.473 unit, meskipun tidak setinggi angka sebelumnya.
- **Tahun 2008 hingga 2009:**
Pada tahun 2008, terjadi lonjakan penjualan yang cukup besar, mencapai 607.805 unit, meskipun ada penurunan di tahun berikutnya (2009) menjadi 486.061 unit. Grafik akan menunjukkan garis yang naik tajam pada 2008 dan sedikit turun pada 2009.
- **Tahun 2010 hingga 2012:**
Pada tahun 2010 hingga 2012, grafik akan menunjukkan tren yang positif. Penjualan meningkat pesat dari 764.709 unit pada 2010 menjadi 894.180 unit pada 2011, namun sedikit menurun di 2012 menjadi 816.322 unit.

2. Secara Umum

- **Fluktuasi:**
Secara keseluruhan, grafik ini menunjukkan adanya fluktuasi penjualan mobil di Indonesia antara 2003 dan 2012. Ada beberapa titik dengan penurunan signifikan (seperti 2006 dan 2009), namun tren umumnya menunjukkan peningkatan, terutama pada tahun-tahun terakhir.
- **Pola Musiman atau Kausalitas:**
Meskipun data ini tidak memuat informasi lebih lanjut tentang faktor eksternal (misalnya, krisis ekonomi atau peraturan pemerintah), fluktuasi besar seperti yang terjadi di 2006 dan 2009 bisa terkait dengan peristiwa khusus atau perubahan dalam pasar mobil.

3. Interpretasi Grafik

- Pada grafik, kamu akan melihat **garis biru** yang menghubungkan setiap titik data yang mewakili penjualan per tahun. Setiap **titik biru** menunjukkan jumlah unit yang terjual di tahun tertentu.
- Jika garis naik, itu menunjukkan peningkatan penjualan pada tahun tersebut. Sebaliknya, jika garis turun, itu menunjukkan penurunan.
- Misalnya, lonjakan besar pada 2008 kemungkinan terkait dengan faktor ekonomi atau program yang mempengaruhi penjualan kendaraan pada tahun itu.

Kesimpulan

Grafik yang kita buat memberikan gambaran visual yang jelas tentang bagaimana penjualan mobil berubah sepanjang waktu dari 2003 hingga 2012. Grafik ini menunjukkan fluktuasi yang signifikan dan tren keseluruhan yang naik, meskipun ada penurunan pada titik-titik tertentu. Hal ini bisa mengindikasikan adanya faktor eksternal yang mempengaruhi pasar mobil pada tahun-tahun tertentu.

3. Metode Semi Average

```
> # 3. Metode Semi Average
> # Membagi data menjadi dua periode
> first_period <- car_sales_data[1:5, ]
> first_period
  Year Units
1 2003 354331
2 2004 483170
3 2005 533922
4 2006 318904
5 2007 434473
>
> second_period <- car_sales_data[6:10, ]
> second_period
  Year Units
6 2008 607805
7 2009 486061
8 2010 764709
9 2011 894180
10 2012 816322
>
> # Menghitung rata-rata penjualan tiap periode
> avg_first <- mean(first_period$units)
> avg_first
[1] 424960
> avg_second <- mean(second_period$units)
> avg_second
```

```

[1] 713815.4
>
> # Tahun tengah untuk masing-masing periode
> mid_first <- mean(first_period$Year)
> mid_first
[1] 2005
> mid_second <- mean(second_period$Year)
> mid_second
[1] 2010
>
> # Menentukan slope dan intercept regresi
> slope <- (avg_second - avg_first) / (mid_second - mid_first)
> slope
[1] 57771.08
>
> intercept <- avg_first - slope * mid_first
> intercept
[1] -115406055
>
> # Fungsi prediksi berdasarkan metode semi average
> predict_semi_average <- function(year) {
+   intercept + slope * year
+ }
> predict_semi_average
function(year) {
  intercept + slope * year
}
>
> # Menambahkan prediksi ke data asli
> car_sales_data$Predicted_Units <- predict_semi_average(car_sales_data$Year)
>
> # Prediksi 5 tahun ke depan
> future_years <- data.frame(Year = 2013:2017)
> future_years$Predicted_Units <- predict_semi_average(future_years$Year)
>
> # Menampilkan hasil prediksi
> print(future_years)
  Year Predicted_Units
1 2013     887128.6
2 2014     944899.7
3 2015    1002670.8
4 2016    1060441.9
5 2017    1118213.0

```

Kode ini menerapkan **Metode Semi Average** untuk memprediksi tren penjualan mobil berdasarkan data penjualan dari tahun 2003 hingga 2012. Dengan menghitung rata-rata penjualan per periode, kita bisa menghitung garis regresi yang memperkirakan penjualan di masa depan. Prediksi untuk lima tahun mendatang menunjukkan peningkatan yang stabil dalam penjualan mobil.

4. Visualisasi Data Awal dan Prediksi

```

> # 4. Visualisasi Data Awal dan Prediksi
> # Menggabungkan data awal dan data prediksi
> combined_data <- rbind(
+   car_sales_data[, c("Year", "Units", "Predicted_Units")],
+   data.frame(Year = future_years$Year, Units = NA, Predicted_Units = future_years$Predicted_Units)
+ )
>
> # Visualisasi gabungan data awal dan hasil prediksi
> ggplot(combined_data, aes(x = Year)) +
+   geom_line(aes(y = Units, color = "Data Awal"), size = 1) +

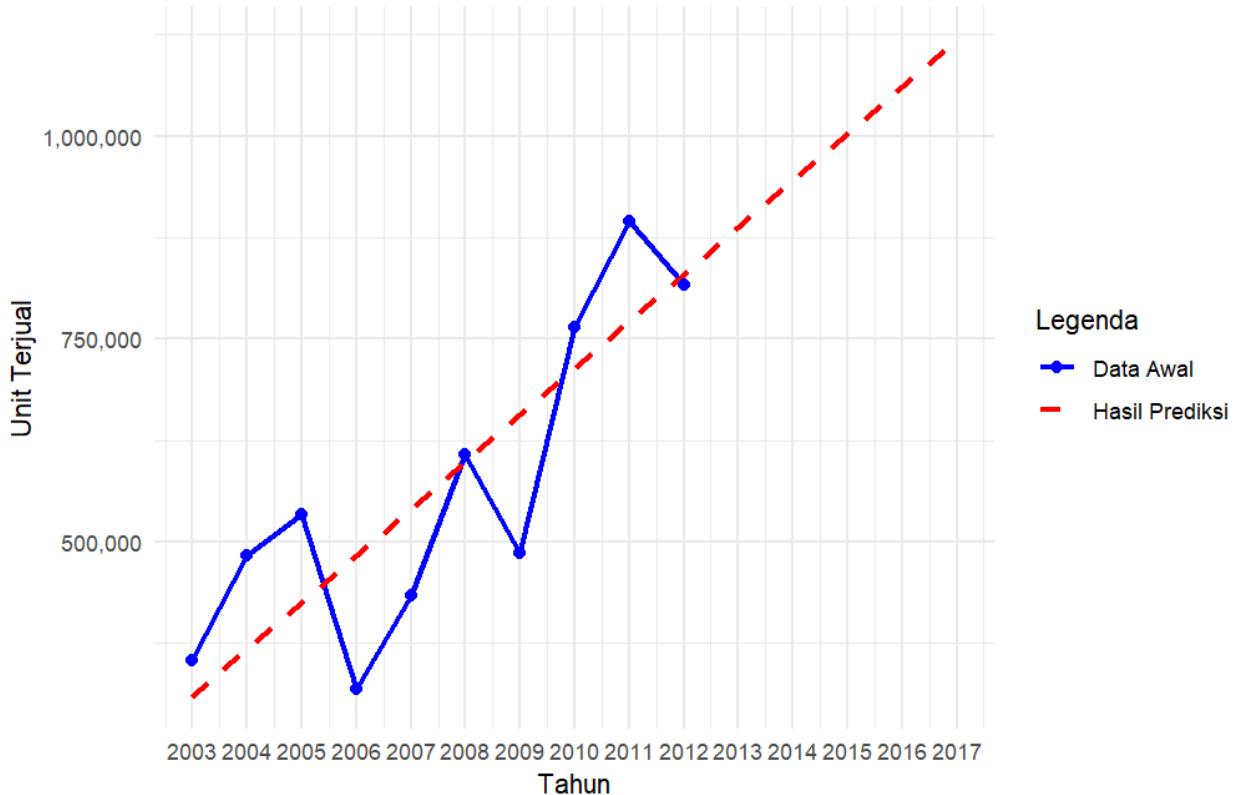
```

```

+ geom_line(aes(y = Predicted_Units, color = "Hasil Prediksi"), size =
1, linetype = "dashed") +
+ geom_point(aes(y = Units, color = "Data Awal"), size = 2) +
+ labs(
+   title = "Tren Penjualan Mobil di Indonesia (Data Awal dan Prediksi)",
+   x = "Tahun",
+   y = "Unit Terjual",
+   color = "Legenda"
+ ) +
+ scale_x_continuous(breaks = seq(2003, 2017, by = 1)) + # Menampilkan
semua tahun
+ scale_y_continuous(labels = scales::comma) + # Format angka
dengan koma
+ scale_color_manual(values = c("Data Awal" = "blue", "Hasil Prediksi" =
"red")) +
+ theme_minimal()
warning messages:
1: Removed 5 rows containing missing values or values outside the
scale range (`geom_line()`).
2: Removed 5 rows containing missing values or values outside the
scale range (`geom_point()`).

```

Tren Penjualan Mobil di Indonesia (Data Awal dan Prediksi)



Penjelasan Grafik Penjualan Mobil:

1. Data Asli (2003-2012):

- Data penjualan mobil dari 2003 hingga 2012 akan digambarkan dengan **garis biru** dan **titik biru** pada grafik.
- Penjualan mobil menunjukkan fluktuasi yang cukup bervariasi dari tahun ke tahun.

- Misalnya, ada penurunan penjualan pada tahun 2006 (318,904 unit), yang kemudian naik signifikan pada tahun 2008 (607,805 unit) dan 2010 (764,709 unit).
- Ini menggambarkan bagaimana tren penjualan mobil dipengaruhi oleh faktor ekonomi, musiman, atau faktor lainnya yang mempengaruhi pasar mobil.

2. Hasil Prediksi (2013-2017):

- Menggunakan metode **semi average**, kita telah memprediksi penjualan mobil untuk tahun 2013 hingga 2017. Hasil prediksi ini digambarkan dengan **garis putus-putus merah**.
- Berdasarkan perhitungan dari metode semi average, prediksi penjualan mobil menunjukkan **kenaikan yang stabil** setiap tahunnya.
 - **Tahun 2013** diprediksi akan terjual sekitar **887,129 unit**, yang kemudian meningkat secara konsisten hingga **2017** dengan prediksi mencapai **1,118,213 unit**.
 - Prediksi ini mengikuti tren dari data sebelumnya, yang menunjukkan bahwa penjualan mobil cenderung naik dari tahun 2008 hingga 2012.

Visualisasi Data dan Prediksi:

- **Data Asli (Tahun 2003-2012):**
 - Kamu bisa melihat fluktuasi penjualan mobil dengan jelas di grafik. Ada **penurunan tajam** pada tahun 2006, yang mungkin dipengaruhi oleh faktor-faktor eksternal seperti resesi atau perubahan ekonomi.
 - Setelah tahun 2006, grafik menunjukkan **kenaikan yang stabil** pada tahun-tahun berikutnya, terutama pada tahun 2008 dan 2010, yang mungkin dipengaruhi oleh pemulihan ekonomi dan peningkatan permintaan.
- **Prediksi (Tahun 2013-2017):**
 - Grafik prediksi menunjukkan **kenaikan yang lebih teratur dan linier** dengan hasil yang semakin tinggi setiap tahunnya. Ini mengindikasikan bahwa berdasarkan tren yang ada, kita dapat mengharapkan **pertumbuhan yang lebih stabil** di masa depan.
 - Metode semi average bekerja dengan **menghitung rata-rata penjualan untuk dua periode** (2003-2007 dan 2008-2012) dan kemudian menggunakan informasi ini untuk **memprediksi masa depan**. Hasilnya, prediksi ini terlihat **optimis** dengan pertumbuhan yang berkelanjutan.

Kesimpulan dari Grafik:

- **Data Asli** memberikan gambaran tentang **perubahan penjualan mobil** yang dipengaruhi oleh berbagai faktor dalam kurun waktu 2003-2012. Kita bisa melihat penurunan tajam pada tahun tertentu dan kenaikan yang lebih tinggi pada tahun lainnya.
- **Prediksi 2013-2017** berdasarkan **metode semi average** memberikan gambaran bahwa **penjualan mobil akan terus meningkat** secara stabil di masa depan. Namun, perlu dicatat bahwa prediksi ini bersifat linier dan mungkin tidak sepenuhnya menggambarkan fluktuasi nyata yang bisa terjadi karena faktor-faktor eksternal yang tidak dapat diprediksi dengan metode ini.

5. Menghitung Measure of Accuracy

```
> # 5. Menghitung Measure of Accuracy
>
> # 1. Filter data yang memiliki nilai aktual (tidak NA)
> actual_data <- car_sales_data[!is.na(car_sales_data$units), ]
>
> # 2. Hitung Error
> actual_data$Error <- actual_data$Units - actual_data$Predicted_Units
>
> # 3. Mean Absolute Error (MAE)
> mae <- mean(abs(actual_data$Error))
>
> # 4. Mean Squared Error (MSE)
> mse <- mean(actual_data$Error^2)
>
> # 5. Mean Absolute Percentage Error (MAPE)
> mape <- mean(abs(actual_data$Error / actual_data$Units)) * 100
>
> # Menampilkan hasil measure of accuracy
> cat("Measure of Accuracy:\n")
Measure of Accuracy:
> cat("Mean Absolute Error (MAE):", mae, "\n")
Mean Absolute Error (MAE): 90575.02
> cat("Mean Squared Error (MSE):", mse, "\n")
Mean Squared Error (MSE): 11219745423
> cat("Mean Absolute Percentage Error (MAPE):", mape, "%\n")
Mean Absolute Percentage Error (MAPE): 19.13657 %
```

1. Filter Data yang Memiliki Nilai Aktual (Tidak NA)

- Pertama, kita memilih data yang memiliki nilai penjualan asli (`Units`), yang berarti data yang tidak mengandung nilai yang hilang (`NA`).

2. Hitung Error

- Selanjutnya, kita menghitung *error*, yaitu selisih antara nilai aktual penjualan (`Units`) dan nilai prediksi (`Predicted_Units`). Ini membantu kita melihat seberapa besar perbedaan antara hasil prediksi dan data nyata.

3. Mean Absolute Error (MAE)

- MAE dihitung sebagai rata-rata dari nilai absolut *error*. Ini memberi tahu kita seberapa besar rata-rata kesalahan prediksi dalam satuan yang sama dengan data asli (unit penjualan).
- Hasil MAE: **90575.02**
 - Artinya, rata-rata kesalahan prediksi penjualan mobil setiap tahun adalah sekitar 90.575 unit.

4. Mean Squared Error (MSE)

- MSE dihitung sebagai rata-rata dari kuadrat *error*. MSE cenderung memberikan penekanan lebih pada kesalahan yang besar karena error dipangkatkan dua.
- Hasil MSE: **11219745423**
 - Nilai yang sangat besar ini menunjukkan bahwa meskipun model memiliki rata-rata kesalahan yang cukup kecil, beberapa kesalahan prediksi mungkin cukup besar dan mempengaruhi hasil total.

5. Mean Absolute Percentage Error (MAPE)

- MAPE mengukur kesalahan prediksi dalam bentuk persentase dari nilai aktual. Ini memberi gambaran seberapa besar kesalahan dalam persentase terhadap data asli.

- Hasil MAPE: **19.14%**
 - Artinya, model memiliki kesalahan prediksi rata-rata sekitar 19% dari nilai penjualan mobil yang sebenarnya. Secara umum, nilai MAPE yang lebih rendah dianggap menunjukkan model yang lebih akurat.

Ringkasan Hasil:

- **MAE** sebesar 90.575 unit berarti kesalahan prediksi rata-rata per tahun cukup kecil, tetapi ada beberapa tahun di mana prediksi agak meleset jauh.
- **MSE** yang sangat besar menunjukkan ada beberapa prediksi yang memiliki kesalahan yang lebih besar dan mempengaruhi total kesalahan model.
- **MAPE** 19.14% menunjukkan model memiliki tingkat ketidakakuratan sekitar 19% dari nilai asli.

Secara keseluruhan, meskipun model semi average cukup baik, ada ruang untuk perbaikan. Model ini memberikan prediksi yang cukup akurat, namun beberapa prediksi memiliki kesalahan yang signifikan.

Studi Kasus 3: Data Ganjil (Data Jumlah Mahasiswa Tahun 1980-1988)



DATA JUMLAH MAHASISWA

Tahun	Jumlah Mahasiswa
1980	590
1981	600
1982	685
1983	720
1984	631
1985	660
1986	503
1987	681
1988	948

1. Memasukkan Data

```
> # 1. Memasukkan Data
> year <- c(1980, 1981, 1982, 1983, 1984, 1985, 1986, 1987, 1988)
> students <- c(590, 600, 685, 720, 631, 660, 503, 681, 948)
>
> # Membuat data frame
> students_data <- data.frame(
+   Year = year,
+   Students = students
+ )
>
> print(students_data)
  Year Students
1 1980      590
2 1981      600
3 1982      685
4 1983      720
5 1984      631
6 1985      660
7 1986      503
8 1987      681
9 1988      948
```

Penjelasan Kode:

Bagian 1: Memasukkan Data

- **year dan students:**

Dua vektor ini berisi data tahun dan jumlah mahasiswa masing-masing.

- year menyimpan data tahun dari 1980 hingga 1988.
- students menyimpan jumlah mahasiswa yang terdaftar di masing-masing tahun.

Bagian 2: Membuat Data Frame

- **students_data:**

Data frame dibuat menggunakan fungsi `data.frame()`, yang menggabungkan dua vektor (`Year` dan `Students`) ke dalam sebuah tabel terstruktur.

Kolom pertama (`Year`) berisi data tahun, dan kolom kedua (`Students`) berisi jumlah mahasiswa.

Bagian 3: Menampilkan Data Frame

- **print(students_data):**

Menampilkan isi data frame dalam bentuk tabel untuk memastikan data sudah benar.

Ouput yang ditampilkan:

	Year	Students
1	1980	590
2	1981	600
3	1982	685
4	1983	720
5	1984	631
6	1985	660
7	1986	503
8	1987	681
9	1988	948

Maknanya:

Tabel di atas menunjukkan data jumlah mahasiswa yang terdaftar dari tahun 1980 hingga 1988. Misalnya:

- Pada tahun 1980, jumlah mahasiswa adalah **590**.
- Pada tahun 1988, jumlahnya meningkat signifikan menjadi **948**.

Tabel ini adalah langkah awal sebelum analisis lebih lanjut, seperti memvisualisasikan tren atau membuat prediksi jumlah mahasiswa di masa depan.

2. Visualisasi Data Awal

```
> # 2. Visualisasi Tren Jumlah Mahasiswa
> # Memuat library untuk visualisasi
> library(ggplot2)
>
> ggplot(students_data, aes(x = Year, y = Students, group = 1)) +
+   geom_line(color = "blue", size = 1) +
+   geom_point(color = "blue", size = 2) +
+   labs(
```

```

+   title = "Tren Jumlah Mahasiswa",
+   x = "Tahun",
+   y = "Jumlah Mahasiswa"
+ ) +
+ theme_minimal()

```

Penjelasan Kode:

Bagian 1: Memuat Library

- **library(ggplot2):**

Library `ggplot2` digunakan untuk membuat grafik yang elegan dan informatif. Ini adalah library yang sangat populer di R untuk visualisasi data.

Bagian 2: Membuat Grafik

- **ggplot():**

Fungsi utama untuk membuat grafik. Pada bagian ini:

- `students_data` adalah data yang digunakan.
- `aes(x = Year, y = Students, group = 1)` menentukan bahwa sumbu-x berisi data `Year`, sumbu-y berisi data `Students`, dan `group = 1` memastikan semua titik dihubungkan dalam satu garis.

- **geom_line():**

Menambahkan garis ke grafik untuk menunjukkan tren jumlah mahasiswa dari tahun ke tahun.

- `color = "blue"` memberi warna biru pada garis.
- `size = 1` menentukan ketebalan garis.

- **geom_point():**

Menambahkan titik pada setiap data tahun untuk menonjolkan jumlah mahasiswa di masing-masing tahun.

- `color = "blue"` memberi warna biru pada titik.
- `size = 2` menentukan ukuran titik.

- **labs():**

Digunakan untuk menambahkan label pada grafik:

- `title`: Judul grafik adalah "Tren Jumlah Mahasiswa".
- `x`: Sumbu-x diberi label "Tahun".
- `y`: Sumbu-y diberi label "Jumlah Mahasiswa".

- **theme_minimal():**

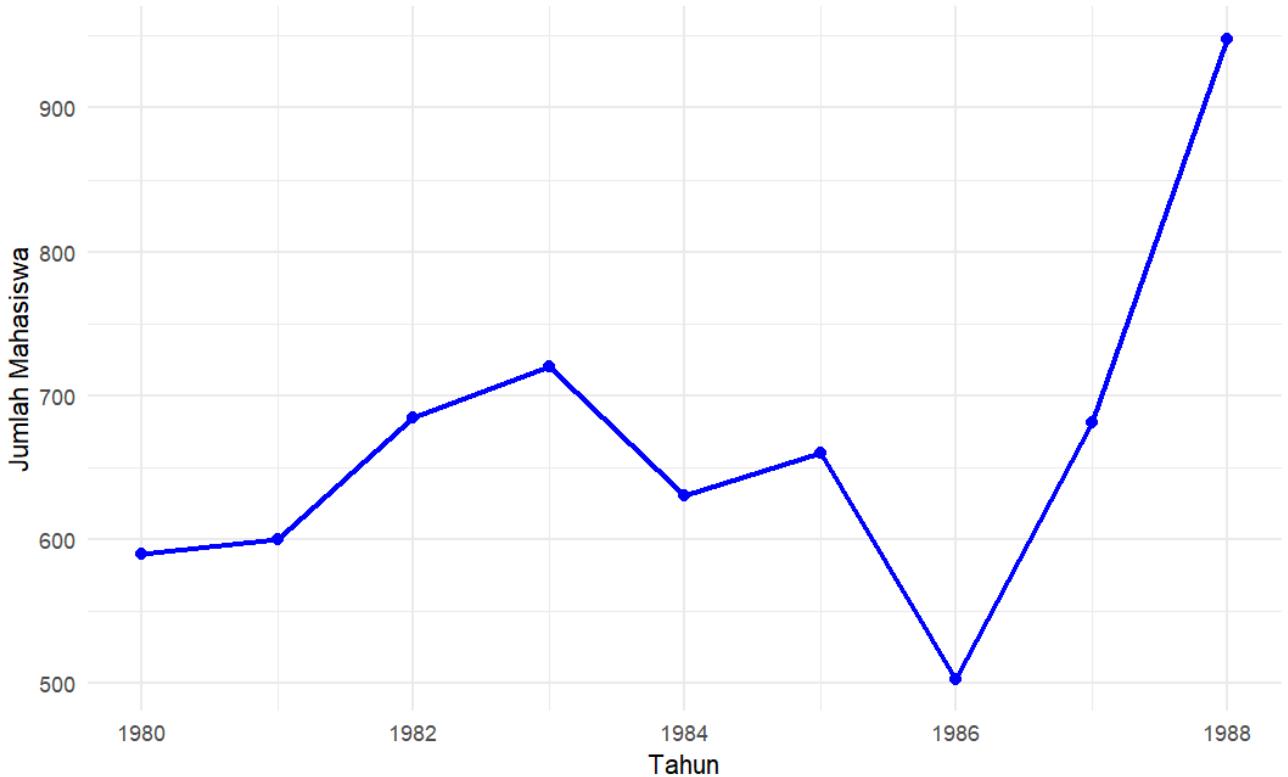
Memberikan tampilan grafik yang sederhana namun tetap profesional.

Hasil Visualisasi:

Grafik akan menampilkan tren jumlah mahasiswa dari tahun 1980 hingga 1988, dengan:

- Garis biru yang menghubungkan data jumlah mahasiswa antar tahun.
- Titik biru yang menandai jumlah mahasiswa di setiap tahun.

Tren Jumlah Mahasiswa



Penjelasan Grafik

Sumbu X (Horizontal):

- Sumbu ini mewakili **Tahun** dari 1980 hingga 1988.

Sumbu Y (Vertikal):

- Sumbu ini menunjukkan **Jumlah Mahasiswa** pada setiap tahun.

Garis dan Titik:

- **Garis biru:** Menunjukkan tren perubahan jumlah mahasiswa dari tahun ke tahun.
- **Titik biru:** Menunjukkan data jumlah mahasiswa aktual untuk setiap tahun.

Analisis Grafik

1. Awal Tren (1980-1983):

- Tahun 1980 dimulai dengan **590 mahasiswa**.
- Ada **peningkatan bertahap** hingga mencapai **720 mahasiswa** pada 1983. Ini menunjukkan pertumbuhan stabil selama 4 tahun.

2. Penurunan Sementara (1984-1986):

- Pada 1984, jumlah mahasiswa menurun menjadi **631** setelah sebelumnya berada di puncak 720.
- Tren penurunan ini berlanjut ke **503 mahasiswa** pada 1986, yang merupakan titik **terendah** dalam grafik.

3. Pemulihan dan Lonjakan (1987-1988):

- Pada 1987, jumlah mahasiswa kembali naik ke **681**, menunjukkan pemulihan yang signifikan.
- Lonjakan besar terjadi pada 1988, dengan jumlah mahasiswa mencapai **948**. Ini adalah angka **tertinggi** sepanjang periode yang diamati.

Kesimpulan dari Grafik:

- **Tren Umum:** Grafik menunjukkan pola **naik-turun** dengan peningkatan yang tajam di tahun terakhir (1988).
- **Poin Penting:**
 - Periode **1983** adalah puncak pertama (720 mahasiswa) sebelum penurunan.
 - Periode **1986** adalah titik nadir dengan jumlah mahasiswa terendah (503).
 - Tahun **1988** menunjukkan lonjakan luar biasa dengan jumlah mahasiswa tertinggi (948).

3. Analisis Tren dengan Regresi Linier

```
> # 3. Analisis Tren dengan Regresi Linier
> model_students <- lm(Students ~ Year, data = students_data)
> summary(model_students)

Call:
lm(formula = Students ~ Year, data = students_data)

Residuals:
    Min      1Q   Median      3Q     Max 
-207.367 -37.667  -6.117  58.033 195.933 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -40697.73  29813.62 -1.365   0.214    
Year          20.85     15.03   1.388   0.208    
                                                        
Residual standard error: 116.4 on 7 degrees of freedom
Multiple R-squared:  0.2157, Adjusted R-squared:  0.1037 
F-statistic: 1.925 on 1 and 7 DF,  p-value: 0.2079

>
> # Menambahkan prediksi dari model regresi linier ke data
> students_data$Linear <- predict(model_students, newdata = students_data)
>
> # Membuat data prediksi untuk 5 tahun ke depan
> future_years <- data.frame(Year = 1989:1993)
>
> # Prediksi regresi linier
> future_years$Linear <- predict(model_students, newdata = future_years)
```

Penjelasan Kode:

1. Membuat Model Regresi Linier

Kode:

```
> model_students <- lm(Students ~ Year, data = students_data)
> summary(model_students)
```

- `lm(Students ~ Year, data = students_data)`:
 - Membuat model regresi linier sederhana dengan variabel dependen (`Students`) dan variabel independen (`Year`).
 - Tujuannya adalah melihat hubungan antara tahun dan jumlah mahasiswa.
- `summary(model_students)`:
 - Memberikan ringkasan statistik dari model regresi, seperti koefisien, error standar, nilai t, R-squared, dll.

Hasil Ringkasan Model

Koefisien:

- **Intercept (-40697.73):**
 - Nilai prediksi awal (tahun 0, tidak relevan secara kontekstual karena dataset dimulai tahun 1980).
- **Slope (20.85):**
 - Menunjukkan bahwa secara rata-rata, jumlah mahasiswa meningkat sekitar **20,85 orang per tahun**.

Statistik:

- **Residual Standard Error (116.4):**
 - Variasi sisa setelah memperhitungkan model.
- **R-squared (0.2157):**
 - Hanya sekitar **21,57%** dari variasi jumlah mahasiswa yang bisa dijelaskan oleh tahun. Sisanya dipengaruhi faktor lain.
- **p-value (0.2079):**
 - Model ini tidak signifikan secara statistik pada tingkat kepercayaan 95%, artinya hubungan antara tahun dan jumlah mahasiswa kurang kuat.

2. Menambahkan Prediksi ke Data Awal

Kode:

```
> # Menambahkan prediksi dari model regresi linier ke data
> students_data$Linear <- predict(model_students, newdata = students_data)
```

- `predict()`: Menggunakan model regresi untuk memprediksi jumlah mahasiswa berdasarkan data tahun di dataset.
- Menambahkan kolom baru (`Linear`) ke `students_data` yang berisi prediksi jumlah mahasiswa.

3. Membuat Prediksi untuk 5 Tahun ke Depan

Kode:

```
> # Membuat data prediksi untuk 5 tahun ke depan
> future_years <- data.frame(Year = 1989:1993)
```

- Membuat data baru untuk tahun **1989-1993** (5 tahun setelah data terakhir).

- Menggunakan model regresi untuk memprediksi jumlah mahasiswa di tahun-tahun tersebut.
- Hasil prediksi disimpan di kolom `Linear`.

Kesimpulan

- Regresi linier menunjukkan tren kenaikan jumlah mahasiswa dari tahun ke tahun, tapi model kurang kuat (R^2 rendah).
- Prediksi model untuk 5 tahun ke depan cenderung linier dan mungkin tidak mencerminkan realitas jika ada pola kompleks seperti fluktuasi besar.

4. Analisis Tren dengan Semiavarage

```
> # 4. Metode Semiaverage
> semiaverage <- function(data) {
+   n <- nrow(data)
+   mid <- floor(n / 2)
+
+   group1 <- mean(data$Students[1:mid])
+   group2 <- mean(data$Students[(mid + 1):n])
+
+   years <- c(mean(data$Year[1:mid]), mean(data$Year[(mid + 1):
n]))
+   averages <- c(group1, group2)
+
+   model <- lm(averages ~ years)
+   pred <- predict(model, newdata = data.frame(years = data$Yea
r))
+   return(list(pred = pred, model = model))
+ }
> semi_model <- semiaverage(students_data)
> students_data$Semiaverage <- semi_model$pred
> future_years$Semiaverage <- predict(semi_model$model, newdata
= data.frame(years = future_years$Year))

>
```

Penjelasan:

1. `n <- nrow(data)`: Menghitung jumlah total baris (data).
2. `mid <- floor(n / 2)`: Menentukan titik tengah untuk membagi data menjadi dua kelompok.
3. **Menghitung rata-rata:**
 - **Grup 1:** Data dari awal hingga titik tengah.
 - **Grup 2:** Data dari titik tengah + 1 hingga akhir.
4. **Menghitung tahun rata-rata untuk setiap grup:**
 - Misalnya, grup pertama tahun 198019801980-198419841984, maka rata-rata tahunnya adalah 198219821982.
5. **Regresi linier:**
 - Hubungan antara tahun rata-rata (x) dan rata-rata jumlah mahasiswa (y) digunakan untuk membuat model.
6. **Prediksi:**
 - Menghasilkan nilai prediksi untuk setiap tahun berdasarkan model.

Menjalankan Fungsi

- `semi_model <- semiaverage(students_data):`
 - Memanggil fungsi untuk menghitung prediksi menggunakan metode Semiaverage.
 - Hasil prediksi (`semi_model$pred`) dan model regresi (`semi_model$model`) disimpan.
- `students_data$Semiaverage <- semi_model$pred:`
 - Menambahkan kolom baru (`Semiaverage`) di dataset awal untuk menyimpan prediksi.
- `future_years$Semiaverage:`
 - Menggunakan model yang dibuat sebelumnya untuk memprediksi jumlah mahasiswa di tahun-tahun mendatang (1989-1993).

5. Metode Least Squares (Linear)

```
> # 5. Metode Least Squares (Linear)
> least_squares <- function(data) {
+   model <- lm(Students ~ Year, data = data)
+   pred <- predict(model, newdata = data)
+   return(list(pred = pred, model = model))
+ }
> least_model <- least_squares(students_data)
> students_data$LeastSquares <- least_model$pred
> future_years$LeastSquares <- predict(least_model$model, newdata = future_years)
```

Penjelasan:

1. `lm(Students ~ Year, data = data):`
 - Membuat model regresi linier dengan variabel dependen (`Students`) diprediksi berdasarkan variabel independen (`Year`).
 - Fungsi `lm()` secara otomatis menghitung koefisien regresi $y = \beta_0 + \beta_1 x$, di mana:
 - y : Jumlah mahasiswa.
 - x : Tahun.
 - β_0 : Intercept (titik awal garis regresi).
 - β_1 : Slope (kemiringan garis regresi).
2. `predict(model, newdata = data):`
 - Menggunakan model yang dibuat untuk menghitung nilai prediksi pada data yang tersedia (`data`).

Menjalankan Fungsi dan Hasilnya

- `least_model <- least_squares(students_data):`
 - Menjalankan fungsi untuk membuat model regresi linier pada data awal.
 - Hasil berupa:

- `least_model$pred`: Prediksi untuk data awal (1980-1988).
 - `least_model$model`: Model regresi linier itu sendiri.
- `students_data$LeastSquares <- least_model$pred`:
 - Menambahkan kolom baru di dataset awal (`students_data`) yang berisi hasil prediksi jumlah mahasiswa menggunakan metode Least Squares.
 - `future_years$LeastSquares`:
 - Menggunakan model regresi yang sudah dibuat untuk memprediksi jumlah mahasiswa di masa depan (1989-1993).

6. Metode Quadratic

```
> # 6. Metode Quadratic
> quadratic <- function(data) {
+   model <- lm(Students ~ poly(Year, 2, raw = TRUE), data = data)
+   pred <- predict(model, newdata = data)
+   return(list(pred = pred, model = model))
+ }
> quad_model <- quadratic(students_data)
> students_data$Quadratic <- quad_model$pred
> future_years$Quadratic <- predict(quad_model$model, newdata = future_years)
```

Penjelasan:

1. `lm(Students ~ poly(Year, 2, raw = TRUE), data = data)`:
 - Membuat model regresi polinomial orde dua untuk memprediksi `Students` berdasarkan `Year`.
 - Fungsi `poly(Year, 2, raw = TRUE)` menciptakan variabel $x_2x^2x_2$ (kuadrat dari tahun) selain `xxx` (tahun) itu sendiri.
 - Parameter `raw = TRUE` memastikan polinomial dihitung dalam bentuk aslinya tanpa transformasi.
2. `predict(model, newdata = data)`:
 - Menggunakan model kuadratik untuk menghitung nilai prediksi pada data awal.

Menjalankan Fungsi

1. `quad_model <- quadratic(students_data)`:
 - Membuat model kuadratik berdasarkan data awal (1980-1988).
 - Hasil berupa:
 - `quad_model$pred`: Prediksi jumlah mahasiswa pada data awal.
 - `quad_model$model`: Model regresi kuadratik itu sendiri.
2. `students_data$Quadratic <- quad_model$pred`:
 - Menambahkan kolom baru ke dataset (`students_data`), berisi hasil prediksi jumlah mahasiswa berdasarkan metode kuadratik.

3. `future_years$Quadratic <- predict(quad_model$model, newdata = future_years):`
 - o Menggunakan model kuadratik untuk memprediksi jumlah mahasiswa di masa depan (1989-1993).

7. Metode Exponential

```
> # 7. Metode Exponential
> exponential <- function(data) {
+   log_model <- lm(log(Students) ~ Year, data = data)
+   pred <- exp(predict(log_model, newdata = data))
+   return(list(pred = pred, model = log_model))
+ }
> exp_model <- exponential(students_data)
> students_data$Exponential <- exp_model$pred
> future_years$Exponential <- exp(predict(exp_model$model, newdata = future_years))
```

Penjelasan:

1. `lm(log(Students) ~ Year, data = data):`
 - o Membuat model regresi logaritmik dengan mengubah nilai `Students` menjadi logaritma natural ($\ln(y)$).
 - o Tujuannya adalah menyederhanakan hubungan non-linier menjadi linier.
2. `exp(predict(log_model, newdata = data)):`
 - o Mengonversi prediksi dari logaritma natural ($\ln(y)$) kembali ke skala aslinya (y) dengan fungsi eksponensial ($e^{\ln(y)}$).

Menjalankan Fungsi

Penjelasan:

1. `exp_model <- exponential(students_data):`
 - o Membuat model eksponensial berdasarkan data awal (1980-1988).
 - o Hasilnya berupa:
 - `exp_model$pred`: Prediksi jumlah mahasiswa pada data awal.
 - `exp_model$model`: Model logaritmik yang digunakan untuk membuat prediksi.
2. `students_data$Exponential <- exp_model$pred:`
 - o Menambahkan kolom baru (`Exponential`) pada dataset awal, berisi prediksi jumlah mahasiswa berdasarkan metode eksponensial.
3. `future_years$Exponential <- exp(predict(exp_model$model, newdata = future_years)):`
 - o Menggunakan model logaritmik untuk memprediksi jumlah mahasiswa pada tahun-tahun masa depan (1989-1993).

8. Visualisasi Perbandingan Metode

```
> # 8. visualisasi Perbandingan Metode
```

```

> # Menambahkan data asli dan prediksi ke dalam satu data frame untuk visualisasi
> visualization_data <- rbind(
+   data.frame(Year = students_data$Year, value = students_data$Students,
Method = "Data Asli"),
+   data.frame(Year = future_years$Year, value = future_years$Linear, Method = "Regresi Linear"),
+   data.frame(Year = future_years$Year, value = future_years$Semiaverage, Method = "Semiaverage"),
+   data.frame(Year = future_years$Year, value = future_years$LeastSquares, Method = "Least Squares(Kuadrat Terkecil"),
+   data.frame(Year = future_years$Year, value = future_years$Quadratic, Method = "Quadratic(Kuadratis"),
+   data.frame(Year = future_years$Year, value = future_years$Exponential, Method = "Exponential")
+ )
>
> # Membuat plot
> ggplot() +
+   geom_line(data = students_data, aes(x = Year, y = Students), color = "black", size = 1.2, linetype = "solid") +
+   geom_line(data = visualization_data, aes(x = Year, y = value, color = Method), size = 1) +
+   geom_point(data = students_data, aes(x = Year, y = Students), color = "black", size = 2) +
+   labs(
+     title = "Perbandingan Tren Metode dan Data Asli",
+     x = "Tahun",
+     y = "Jumlah Mahasiswa",
+     color = "Metode"
+   ) +
+   theme_minimal() +
+   scale_color_manual(values = c(
+     "Regresi Linear" = "red",
+     "Semiaverage" = "blue",
+     "Least Squares(Kuadrat Terkecil)" = "green",
+     "Quadratic(Kuadratis)" = "purple",
+     "Exponential" = "orange"
+   ))
>
> # Tabel Perbandingan Hasil Prediksi untuk 5 Tahun ke Depan
> print(future_years)
  Year   Linear Semiaverage LeastSquares Quadratic
1 1989  772.9167    708.5000   772.9167  900.8333
2 1990  793.7667    716.4667   793.7667  998.4333
3 1991  814.6167    724.4333   814.6167 1109.9879
4 1992  835.4667    732.4000   835.4667 1235.4970
5 1993  856.3167    740.3667   856.3167 1374.9606
  Exponential
1      751.8551
2      771.8161
3      792.3069
4      813.3418
5      834.9352

```

Penjelasan Visualisasi

Visualisasi ini membandingkan hasil prediksi dari berbagai metode analisis tren dengan data asli jumlah mahasiswa dari tahun 1980-1988. Berikut komponen utamanya:

1. Data Asli:

- Diwakili oleh garis **hitam solid** dan titik-titik **hitam**.
- Mewakili jumlah mahasiswa sebenarnya berdasarkan data awal.

2. Metode Prediksi:

- Setiap metode memiliki garis dengan warna berbeda:
 - **Regresi Linear:** Merah.
 - **Semiaverage:** Biru.
 - **Least Squares (Kuadrat Terkecil):** Hijau.
 - **Quadratic (Kuadratis):** Ungu.
 - **Exponential:** Oranye.

3. Fokus Visualisasi:

- Menunjukkan bagaimana setiap metode mencoba memperkirakan data untuk 5 tahun ke depan (1989-1993).
- Membandingkan akurasi metode dengan data asli pada tahun 1980-1988.

Interpretasi Tabel dan Grafik

1. Regresi Linear:

- Menampilkan kenaikan jumlah mahasiswa secara **stabil**.
- Prediksinya cenderung moderat dan tidak terlalu ekstrem.

2. Semiaverage:

- Menghasilkan prediksi yang lebih **konservatif** dibanding metode lain.
- Cenderung mendekati rata-rata dari kelompok data awal.

3. Least Squares:

- Hampir identik dengan Regresi Linear karena prinsip dasarnya serupa.
- Prediksi stabil dan linier.

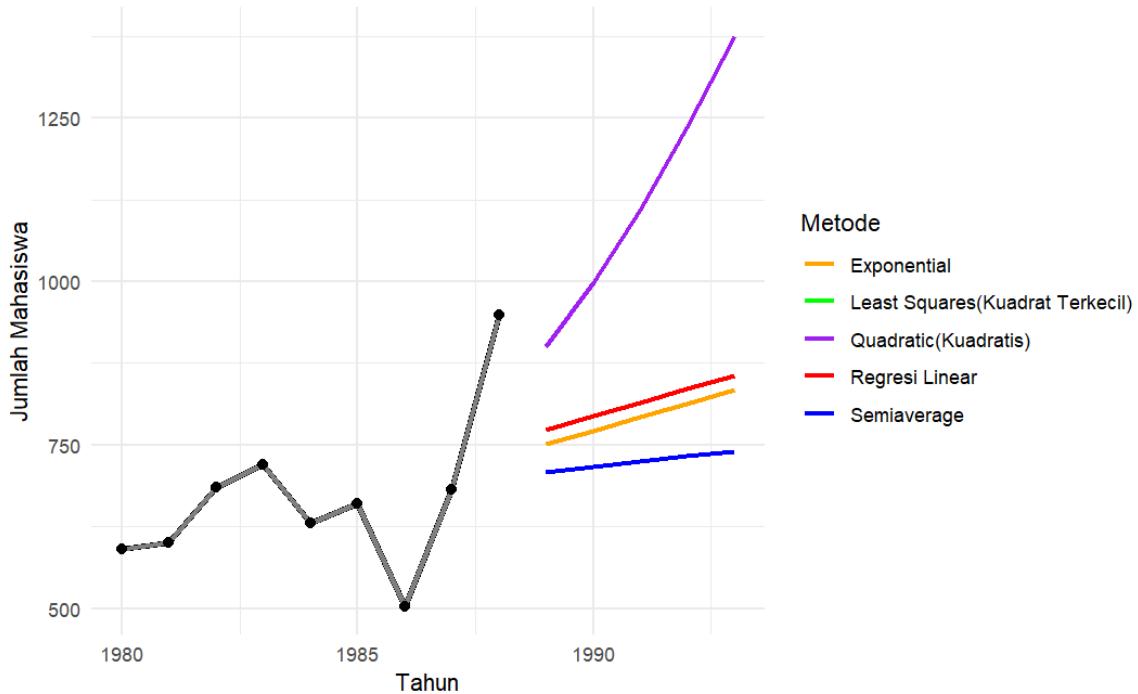
4. Quadratic (Kuadratis):

- Menunjukkan kenaikan yang semakin **eksponensial** seiring waktu.
- Prediksinya lebih tinggi dibanding metode lainnya untuk tahun-tahun selanjutnya (1991-1993).

5. Exponential:

- Sesuai dengan pola pertumbuhan eksponensial.
- Prediksi cenderung **moderate** tetapi sedikit lebih dinamis dibanding metode linier.

Perbandingan Tren Metode dan Data Asli



Kesimpulan Visualisasi

- **Metode Linear (Regresi dan Least Squares)** cocok untuk data dengan pertumbuhan stabil.
- **Semiaverage** memberikan prediksi paling konservatif, cocok untuk pola data yang tidak pasti.
- **Quadratic** memberikan hasil yang paling tinggi, sesuai untuk data dengan pertumbuhan akseleratif.
- **Exponential** memberikan hasil yang masuk akal untuk pola pertumbuhan eksponensial.

BAB 8: PERTEMUAN 13 (Laporan: 14 Konsep Dasar Pemodelan Time Series)

Contoh Kasus

Dataset Stock Price

Date	Asset	Price
10/07/2023	Bond B	385,64
03/03/2023	Stock B	297,20
13/09/2023	Bond B	75,98
07/01/2023	Bond B	356,34
04/06/2023	Stock A	260,05
02/04/2023	Bond B	65,51
14/01/2023	Commodit	486,32
17/08/2023	Commodit	477,1
27/03/2023	Stock A	280,81
13/05/2023	Stock B	92,15
02/07/2023	Stock B	389,58
03/03/2023	Bond B	271,37
07/08/2023	Bond B	186,1
03/09/2023	Bond B	271,14
18/08/2023	Stock B	285,1
20/03/2023	Bond A	303,22
29/05/2023	Commodit	248,55
18/05/2023	Bond A	322,59
10/11/2023	Bond A	463,43
06/03/2023	Stock B	213,02
13/01/2023	Bond A	376,57
01/08/2023	Stock A	452,5

Dataset ini diambil dari kaggle:
<https://www.kaggle.com/datasets/karunyaronith/time-series-financial-portfolio-data>

Data ini ditampilkan hanya 23 record dengan 5 asset. Ini akan dilakukan pengecekan stasioneritas dari dataset Stock Price dan mengatasi data yang tidak stasioner menggunakan RStudio

Install Library yang akan Digunakan

```
> # Install library yang akan digunakan
> # install.packages(c("TSA", "tseries", "lmtest", "forecast", "tidyverse", "rmarkd
own"))
> library(TSA)
> library(tseries)
> library(lmtest)
> library(forecast)
> library(tidyverse)
> library(rmarkdown)
> library(readxl)
```

>

Penjelasan:

Bagian ini bertujuan untuk memastikan bahwa semua paket (library) yang dibutuhkan dalam analisis time series telah diinstal dan dimuat ke dalam R.

1. Instalasi Library:

- Baris dengan komentar `# install.packages(...)` adalah instruksi untuk menginstal library yang dibutuhkan, seperti TSA, tseries, lmtest, dll. Ini hanya perlu dilakukan sekali saja, terutama jika library tersebut belum ada di komputer kita.

2. Memuat Library:

- Baris **library(...)** digunakan untuk memuat library yang sudah terinstal agar fungsinya bisa digunakan dalam script ini. Berikut penjelasan singkat setiap library:
 - **TSA**: Berisi fungsi-fungsi untuk analisis time series seperti ARIMA.
 - **tseries**: Digunakan untuk melakukan uji stasionaritas seperti uji ADF.
 - **lmtest**: Untuk pengujian statistik, seperti uji signifikansi parameter dalam model time series.
 - **forecast**: Library populer untuk prediksi dan pemodelan time series, termasuk fungsi seperti **forecast()** dan **checkresiduals()**.
 - **tidyverse**: Sekumpulan library untuk manipulasi data (dengan **dplyr**) dan visualisasi (dengan **ggplot2**).
 - **rmarkdown**: Membantu menghasilkan laporan dari analisis dalam format seperti HTML, PDF, atau Word.
 - **readxl**: Untuk membaca file Excel jika dibutuhkan (meskipun di kode ini data diambil dari file CSV).

3. Kenapa Ini Dilakukan:

- Dengan memuat library ini, kita bisa menggunakan berbagai fungsi yang mempercepat proses analisis data time series tanpa harus menulis semuanya dari awal. Kalau salah satu library belum diinstal, R akan memberikan pesan error.

Mengecek stasionaritas dan cara mengatasi data tidak stasioner

1. Load Dataset

```
> # 1. Mengecek stasionaritas dan cara mengatasi data tidak stasioner
> # Load dan proses data
> data_finansial <- read_csv("C:/Kuliah/Sains Data Sem 3/Statistika Inferensial Lanjut Praktik/Tugas/14. Laporan Praktikum Konsep Dasar Pemodelan Time Series/financial_portfolio_data.csv")
Rows: 500 Columns: 3
____ Column specification _____
Delimiter: ","
chr (1): Asset
dbl (1): Price
date (1): Date

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
> data <- data_finansial
> # Melihat data
> view(data)
> print(data, n=100)
# A tibble: 500 × 3
  Date       Asset     Price
  <date>    <chr>    <dbl>
1 2023-07-10 Bond B   386.
2 2023-03-03 Stock B  297.
3 2023-09-13 Bond B   76.0
4 2023-01-07 Bond B   356.
5 2023-06-04 Stock A  260.
6 2023-04-02 Bond B   65.5
7 2023-01-14 Commodity A 486.
8 2023-08-17 Commodity A 477.
9 2023-03-27 Stock A  281.
10 2023-05-13 Stock B  92.2
11 2023-07-02 Stock B  390.
12 2023-03-03 Bond B   271.
13 2023-08-07 Bond B   186.
14 2023-09-03 Bond B   271.
15 2023-08-18 Stock B  285.
16 2023-03-20 Bond A   303.
17 2023-05-29 Commodity A 249.
18 2023-05-18 Bond A   323.
19 2023-11-10 Bond A   463.
20 2023-03-06 Stock B  213.
21 2023-01-13 Bond A   377.
22 2023-08-01 Stock A  452.
23 2023-08-30 Bond B   255.
24 2023-08-01 Stock A  250.
25 2023-06-30 Commodity A 321.
26 2023-12-04 Bond B   316.
27 2023-06-22 Bond A   413.
28 2023-06-09 Commodity A 264.
29 2023-06-13 Bond B   498.
30 2023-07-07 Bond A   76.8
31 2023-07-10 Stock A  51.4
32 2023-11-04 Bond B   338.
33 2023-09-13 Bond A   184.
34 2023-07-02 Stock A  158.
35 2023-04-27 Commodity A 190.
36 2023-10-09 Commodity A 165.
37 2023-07-15 Commodity A 63.9
```

```

38 2023-02-03 Commodity A 487.
39 2023-02-21 Stock B 149.
40 2023-04-22 Bond B 380.
41 2023-09-09 Bond B 393.
42 2023-03-23 Stock B 468.
43 2023-08-20 Bond A 63.8
44 2023-08-15 Bond A 442.
45 2023-08-21 Commodity A 287.
46 2023-08-09 Stock A 335.
47 2023-02-11 Bond A 340.
48 2023-01-21 Stock A 419.
49 2023-02-11 Bond B 200.
50 2023-04-23 Stock A 380.
51 2023-04-15 Bond B 288.
52 2023-07-11 Stock B 407.
53 2023-01-17 Stock B 496.
54 2023-04-30 Commodity A 140.
55 2023-05-09 Bond A 498.
56 2023-05-10 Commodity A 90.4
57 2023-03-01 Bond B 51.0
58 2023-02-17 Commodity A 69.9
59 2023-01-29 Stock A 204.
60 2023-04-07 Bond B 335.
61 2023-01-10 Stock B 81.1
62 2023-08-30 Bond A 52.6
63 2023-11-10 Bond B 472.
64 2023-08-20 Bond B 420.
65 2023-10-04 Stock A 414.
66 2023-10-11 Bond A 448.
67 2023-07-22 Stock B 365.
68 2023-09-04 Bond B 97.0
69 2023-10-17 Bond B 230.
70 2023-09-21 Stock A 375.
71 2023-08-02 Commodity A 176.
72 2023-07-19 Stock B 359.
73 2023-03-20 Bond A 61.0
74 2023-12-09 Bond B 188.
75 2023-12-10 Bond B 297.
76 2023-06-22 Stock B 334.
77 2023-03-10 Commodity A 310.
78 2023-03-28 Bond B 364.
79 2023-04-11 Stock B 52.7
80 2023-01-08 Stock A 95.4
81 2023-11-29 Commodity A 441.
82 2023-07-19 Bond A 357.
83 2023-03-28 Bond B 101.
84 2023-02-12 Stock B 429.
85 2023-04-19 Stock A 159.
86 2023-06-01 Bond A 265.
87 2023-01-17 Stock B 115.
88 2023-05-13 Stock B 370.
89 2023-05-28 Bond A 350.
90 2023-10-04 Stock A 145.
91 2023-08-20 Stock A 439.
92 2023-11-10 Commodity A 231.
93 2023-08-16 Bond B 182.
94 2023-05-19 Bond A 119.
95 2023-07-30 Bond A 170.
96 2023-01-28 Bond A 59.9
97 2023-02-20 Stock A 312.
98 2023-09-18 Bond B 429.
99 2023-09-18 Commodity A 297.
100 2023-12-16 Stock B 149.

# i 400 more rows
# i Use `print(n = ...)` to see more rows

```

Bagian ini adalah langkah awal untuk memuat dan memeriksa data yang akan dianalisis. Yuk, kita bahas secara simpel dan santai:

1. Load Data:

- o `data_finansial <-
read_csv("C:/.../financial_portfolio_data.csv")`

Ini digunakan untuk membaca file CSV yang berisi data finansial. File tersebut disimpan di lokasi tertentu di komputer saya, dan isinya langsung dimasukkan ke dalam variabel `data_finansial`.

Hasilnya adalah dataset dengan **500 baris** dan **3 kolom**, yaitu:

- **Date**: Tanggal (formatnya sudah diakui sebagai tanggal di R).
- **Asset**: Jenis aset (seperti "Stock A", "Bond B").
- **Price**: Harga dari aset tersebut.

2. Membuat Variabel Baru:

- o `data <- data_finansial`

Variabel data dibuat sebagai salinan dari `data_finansial`. Ini sering dilakukan agar dataset asli tetap aman kalau ada modifikasi di dataset baru.

3. Melihat Data:

- o `view(data)`

Fungsi ini membuka jendela untuk melihat dataset secara interaktif (kayak spreadsheet). Kita bisa *scroll* untuk memeriksa datanya.

- o `print(data, n=100)`

Menampilkan 100 baris pertama dari dataset langsung di konsol. Kalau datanya panjang banget, ini memudahkan untuk cepat ngecek isi data tanpa membuka jendela baru.

4. Output Data:

- o Output di konsol menunjukkan sebagian isi dataset, contohnya:

- Baris pertama: **Date**: 2023-07-10, **Asset**: Bond B, **Price**: 386.
- Kolom diatur sebagai:
 - `<date>` untuk kolom Date.
 - `<chr>` (character) untuk Asset.
 - `<dbl>` (double) untuk Price (angka desimal).
- Output ini juga menunjukkan ada **400 baris lainnya** yang nggak ditampilkan untuk menghemat ruang. Kita bisa pakai `print(n=...)` kalau mau melihat lebih banyak.

Kenapa Langkah Ini Penting?

- Bagian ini memastikan file datanya sudah berhasil dibaca ke dalam R.
- Kita bisa memeriksa struktur datanya (jenis kolom, jumlah baris, dll.) untuk memastikan tidak ada error sebelum lanjut ke analisis lebih dalam.

2. Menfilter hanya stock A

```
> # Menfilter hanya stock A
> data$Price <- as.numeric(data$Price)
> data <- data %>%
+   filter(Asset == 'Stock A')
> # Mengurutkan data sesuai tanggal
> data <- data[order(data$date), ]
> print(data, n = 'All')
# A tibble: 101 × 3
  Date      Asset    Price
  <date>    <chr>   <dbl>
1 2023-01-02 Stock A 172.
2 2023-01-03 Stock A 489.
3 2023-01-08 Stock A  95.4
4 2023-01-12 Stock A 246.
5 2023-01-14 Stock A 204.
6 2023-01-14 Stock A 329.
7 2023-01-19 Stock A 261.
8 2023-01-21 Stock A 419.
9 2023-01-22 Stock A 106.
10 2023-01-24 Stock A 124.
11 2023-01-29 Stock A 204.
12 2023-01-29 Stock A 374.
13 2023-01-30 Stock A  88.4
14 2023-02-18 Stock A 156.
15 2023-02-20 Stock A 312.
16 2023-02-23 Stock A 132.
17 2023-02-27 Stock A  63.9
18 2023-02-28 Stock A 283.
19 2023-03-10 Stock A 321.
20 2023-03-11 Stock A 356.
21 2023-03-19 Stock A 257.
22 2023-03-27 Stock A 281.
23 2023-03-27 Stock A 136.
24 2023-04-07 Stock A 394.
25 2023-04-08 Stock A 392.
26 2023-04-11 Stock A 354.
27 2023-04-19 Stock A 159.
28 2023-04-21 Stock A 323.
29 2023-04-23 Stock A 380.
30 2023-04-23 Stock A  69.0
31 2023-04-23 Stock A 485.
32 2023-04-26 Stock A 105.
33 2023-04-29 Stock A 231.
34 2023-04-29 Stock A 303.
35 2023-05-06 Stock A 354.
36 2023-05-06 Stock A 164.
37 2023-05-18 Stock A 392.
38 2023-05-20 Stock A 231.
39 2023-05-26 Stock A 256.
40 2023-05-28 Stock A 274.
41 2023-05-28 Stock A 317.
42 2023-05-29 Stock A 108.
43 2023-06-02 Stock A 304.
44 2023-06-03 Stock A 370.
45 2023-06-04 Stock A 260.
46 2023-06-18 Stock A  73.4
47 2023-06-27 Stock A 393.
48 2023-07-01 Stock A 273.
49 2023-07-02 Stock A 158.
50 2023-07-02 Stock A 392.
51 2023-07-05 Stock A 460.
52 2023-07-07 Stock A 158.
53 2023-07-10 Stock A  51.4
54 2023-07-11 Stock A 412.
55 2023-07-15 Stock A 230.
56 2023-07-24 Stock A 117.
57 2023-07-30 Stock A 446.
```

```

58 2023-08-01 Stock A 452.
59 2023-08-01 Stock A 250.
60 2023-08-01 Stock A 220.
61 2023-08-01 Stock A 497.
62 2023-08-04 Stock A 431.
63 2023-08-05 Stock A 374.
64 2023-08-08 Stock A 440.
65 2023-08-08 Stock A 431.
66 2023-08-09 Stock A 335.
67 2023-08-09 Stock A 377.
68 2023-08-10 Stock A 430.
69 2023-08-12 Stock A 274.
70 2023-08-12 Stock A 170.
71 2023-08-17 Stock A 295.
72 2023-08-17 Stock A 192.
73 2023-08-19 Stock A 161.
74 2023-08-20 Stock A 439.
75 2023-08-26 Stock A 264.
76 2023-08-28 Stock A 233.
77 2023-09-05 Stock A 165.
78 2023-09-13 Stock A 337.
79 2023-09-13 Stock A 440.
80 2023-09-17 Stock A 414.
81 2023-09-21 Stock A 375.
82 2023-09-26 Stock A 298.
83 2023-09-27 Stock A 488.
84 2023-09-28 Stock A 469.
85 2023-09-30 Stock A 422.
86 2023-10-04 Stock A 414.
87 2023-10-04 Stock A 145.
88 2023-10-04 Stock A 302.
89 2023-10-08 Stock A 420.
90 2023-10-29 Stock A 475.
91 2023-10-30 Stock A 112.
92 2023-11-01 Stock A 77.3
93 2023-11-06 Stock A 263.
94 2023-11-10 Stock A 215.
95 2023-12-02 Stock A 174.
96 2023-12-14 Stock A 437.
97 2023-12-15 Stock A 426.
98 2023-12-16 Stock A 227.
99 2023-12-22 Stock A 392.
100 2023-12-25 Stock A 76.9
101 2023-12-25 Stock A 146.

> # Membuat time series object
> data_ts <- ts(data$Price, start = c(2023, 2) , frequency=365)
> data_frame <- as.data.frame(data_ts)
> # Visualisasi dari data kita
> plot(data_frame, data_frame$Price, xlab ="Date", ylab = "Price", main =
"Plot Time Series")

```

Berikut adalah penjelasan dari potongan kode R yang Anda bagikan:

1. Konversi Harga Menjadi Numerik

```
data$Price <- as.numeric(data$Price)
```

Baris ini memastikan bahwa kolom Price dalam dataset data adalah tipe data numerik. Hal ini penting karena tipe data numerik diperlukan untuk analisis kuantitatif dan manipulasi data lebih lanjut.

2. Filter Data Berdasarkan Asset

```
data <- data %>%
```

```
filter(Asset == 'Stock A')
```

Menggunakan library dplyr, data difilter untuk hanya menyertakan baris di mana nilai kolom Asset adalah 'Stock A'.

3. Mengurutkan Data Berdasarkan Tanggal

```
data <- data[order(data$Date), ]
```

Fungsi `order(data$Date)` mengurutkan data berdasarkan kolom Date dalam urutan menaik. Hasilnya, dataset akan memiliki baris yang tersusun berdasarkan waktu dari awal hingga akhir.

4. Membuat Objek Time Series

```
data_ts <- ts(data$Price, start = c(2023, 2), frequency=365)
```

Baris ini membuat objek **time series** dari kolom Price dengan:

- **start = c(2023, 2)**: Menentukan tahun dimulai dari tahun 2023 dengan nilai awal pada titik kedua (bisa dimodifikasi sesuai konteks data).
- **frequency = 365**: Menunjukkan bahwa data adalah time series harian dengan 365 periode dalam setahun.

5. Mengonversi Time Series ke Data Frame

```
data_frame <- as.data.frame(data_ts)
```

Mengonversi objek time series menjadi data frame untuk mempermudah manipulasi dan visualisasi data.

6. Visualisasi Data Time Series

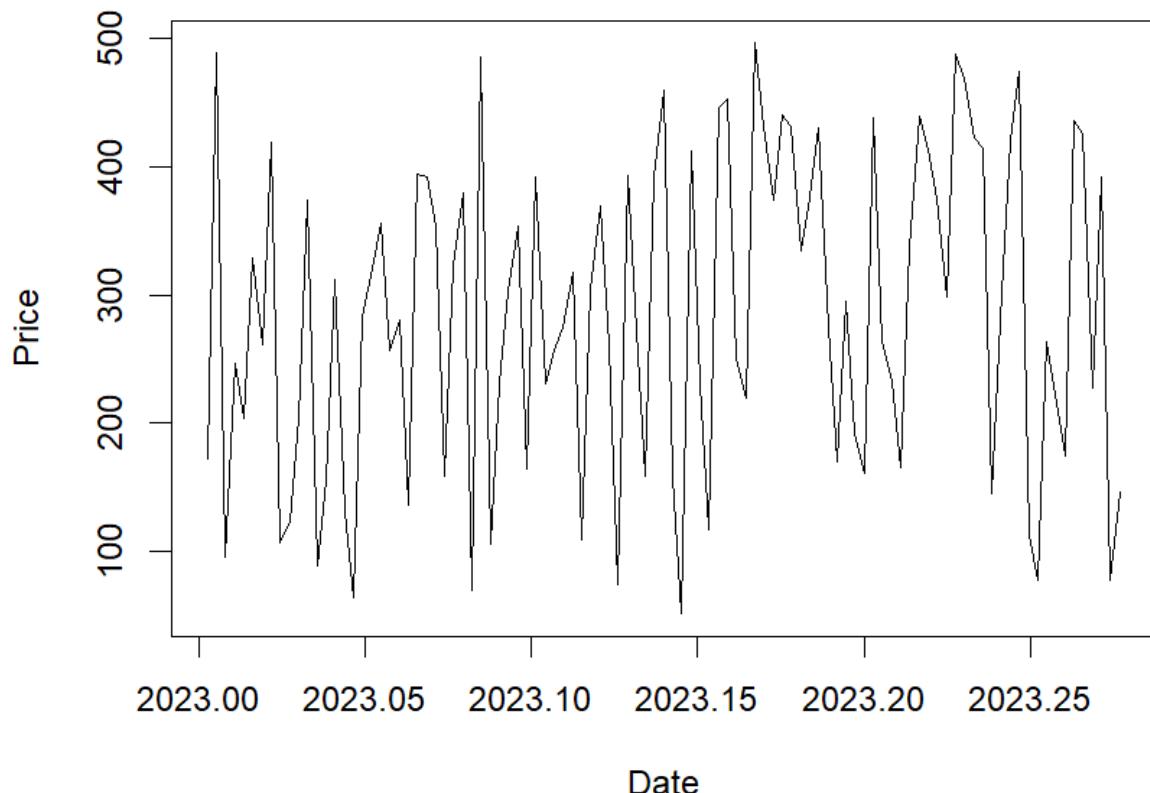
```
plot(data_frame, data_frame$Price, xlab = "Date", ylab = "Price",  
main = "Plot Time Series")
```

Plot ini memvisualisasikan data time series dengan:

- **xlab = "Date"**: Menambahkan label untuk sumbu X sebagai Date.
- **ylab = "Price"**: Menambahkan label untuk sumbu Y sebagai Price.
- **main = "Plot Time Series"**: Menambahkan judul pada grafik.

Hasil output grafik:

Plot Time Series



Dari grafik ini kita belum mengetahui secara jelas bahwa apakah ada pola (tren) atau musiman sehingga kita perlu melakukan uji ADF untuk mengetahui lebih lanjut apakah data ini stasioner atau tidak.

3. Pemeriksaaan Kestasioneran ADF dari Data Time Series

```
> # Pemeriksaaan Kestasioneran
> # Adf dari data
> adf.test(data_ts)

Augmented Dickey-Fuller Test

data: data_ts
Dickey-Fuller = -4.0182, Lag order = 4, p-value = 0.01106
alternative hypothesis: stationary

> # Fungsi untuk melakukan differencing dan uji ADF Apabila tidak stasioner
> fix_stationarity <- function(asset_name, data) {
+   # Filter data berdasarkan asset
+   filtered_data <- data %>%
+     filter(Asset == asset_name)
+
+   # Pastikan kolom Date berupa tanggal dan urutkan berdasarkan tanggal
+   filtered_data$Date <- as.Date(filtered_data$Date)
+   filtered_data <- filtered_data[order(filtered_data$Date), ]
+
+   # Konversi data ke time series dengan frekuensi bulanan (12)
+   ts_data <- ts(filtered_data$Price, frequency = 12)
+
+   # Plot time series untuk aset tersebut
+   ggplot(filtered_data, aes(x = Date, y = Price)) +
+     geom_line(color = "blue") +
+     labs(title = paste("Time Series for", asset_name), x = "Date", y = "Price") +
+     theme_minimal()
+
+   # Lakukan differencing pertama
+   ts_data_diff <- diff(ts_data, differences = 1)
+
+   # Plot data setelah differencing
+   plot(ts_data_diff, main = paste("Differenced Time Series for", asset_name))
+
+   # Lakukan uji ADF pada data setelah differencing
+   adf_test <- adf.test(ts_data_diff)
+
+   # Cetak hasil uji ADF
+   print(paste("ADF Test for", asset_name, "after differencing:"))
+   print(adf_test)
+   # return(ts_data_diff)
+ }

> # Menfilter hanya Bond A
> data_tk_stat <- data_finansial %>%
+   filter(Asset == 'Bond A')
> data_tk_stat$Price <- as.numeric(data_tk_stat$Price)
> # Mengurutkan data sesuai tanggal
> data_tk_stat <- data_tk_stat[order(data_tk_stat$Date), ]
> data_tk_stat
# A tibble: 109 × 3
  Date      Asset  Price
  <date>    <chr> <dbl>
1 2023-01-03 Bond A 199.
2 2023-01-04 Bond A 142.
3 2023-01-08 Bond A 147.
4 2023-01-11 Bond A 473.
5 2023-01-13 Bond A 377.
6 2023-01-20 Bond A 143.
7 2023-01-21 Bond A 431.
8 2023-01-26 Bond A 114.
9 2023-01-28 Bond A  59.9
10 2023-01-31 Bond A  76.0
# i 99 more rows
```

```
# i Use `print(n = ...)` to see more rows
> # Membuat time series object
> data_ts_tk_stat <- ts(data_tk_stat$Price, start = c(2023, 3) , frequency
=365)
> data_frame_tk_stat <- as.data.frame(data_ts_tk_stat)
```

1. Uji Kestasioneran (ADF Test)

Kode ini melakukan pemeriksaan kestasioneran menggunakan *Augmented Dickey-Fuller (ADF)* Test:

```
adf.test(data_ts)
```

- **Tujuan:** Memeriksa apakah data time series bersifat stasioner (mean, variansi, dan autokorelasi konstan sepanjang waktu).
- **Hasil:**
 - Statistik Dickey-Fuller = -4.0182
 - $p\text{-value} = 0.01106$
 - Dengan $p\text{-value} < 0.05$, hipotesis nol (non-stasioner) ditolak, sehingga data **stasioner**.

2. Fungsi untuk Perbaikan Kestasioneran

Fungsi `fix_stationarity` menangani data time series yang tidak stasioner:

```
fix_stationarity <- function(asset_name, data) {
  ...
}
```

- **Input:**
 - `asset_name`: Nama aset (contoh: "Stock A", "Bond A").
 - `data`: Data finansial dalam format tibble.
- **Langkah-langkah:**
 1. **Filter data:** Memilih data berdasarkan aset tertentu.
 2. **Urutkan berdasarkan tanggal:** Memastikan kolom Date berupa tipe tanggal dan terurut.
 3. **Konversi ke time series:** Menggunakan frekuensi bulanan (12).
 4. **Plot time series asli:** Menggunakan ggplot2.
 5. **Differencing pertama:** Mengurangi nilai time series saat ini dengan nilai sebelumnya untuk menghilangkan tren.

6. **Plot setelah differencing:** Untuk memvisualisasikan hasil differencing.
7. **ADF Test setelah differencing:** Memastikan data sudah stasioner.

3. Time Series Object untuk Bond A

```
data_ts_tk_stat <- ts(data_tk_stat$Price, start = c(2023, 3),  
frequency = 365)
```

- **Tujuan:** Mengonversi data harga Bond A menjadi objek time series dengan:
 - **start = c(2023, 3):** Mulai dari bulan Maret 2023.
 - **frequency = 365:** Frekuensi harian.
- Data kemudian disimpan dalam **data_frame_tk_stat** untuk keperluan visualisasi atau analisis lanjutan:

```
data_frame_tk_stat <- as.data.frame(data_ts_tk_stat)
```

4. Pemeriksaaan Kestasioneran ADF dari data Bond A

```
> # Pemeriksaaan Kestasioneran
> # Adf dari data Bond A
> adf.test(data_ts_tk_stat)

Augmented Dickey-Fuller Test

data: data_ts_tk_stat
Dickey-Fuller = -3.2647, Lag order = 4, p-value = 0.0807
alternative hypothesis: stationary

> # Analisis untuk Bond A (asumsi data Bond A tidak stasioner)
> fix_stationarity("Bond A", data_finansial)
[1] "ADF Test for Bond A after differencing:"

Augmented Dickey-Fuller Test

data: ts_data_diff
Dickey-Fuller = -6.3766, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(ts_data_diff) : p-value smaller than printed p-value
> ## diferensi data yg tdak stasioner
> ts_diff_tk_stat <- diff(data_ts_tk_stat)
> # uji ADF pd dta diferensi
> adf_test_diff <- adf.test(ts_diff_tk_stat)
Warning message:
In adf.test(ts_diff_tk_stat) : p-value smaller than printed p-value
> print(adf_test_diff)

Augmented Dickey-Fuller Test

data: ts_diff_tk_stat
Dickey-Fuller = -6.3766, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

1. Pemeriksaan Awal Kestasioneran Data Bond A

adf.test(data_ts_tk_stat)

- **Tujuan:** Memeriksa kestasioneran data Bond A dengan *Augmented Dickey-Fuller (ADF)* Test.
- **Hasil:**
 - Statistik Dickey-Fuller = -3.2647
 - $p\text{-value} = 0.0807$
 - Dengan $p\text{-value} > 0.05$, kita gagal menolak hipotesis nol. Artinya, data **tidak stasioner**.

2. Penanganan Data Tidak Stasioner dengan Differencing

Kode berikut digunakan untuk melakukan differencing (menghilangkan tren):

fix_stationarity("Bond A", data_finansial)

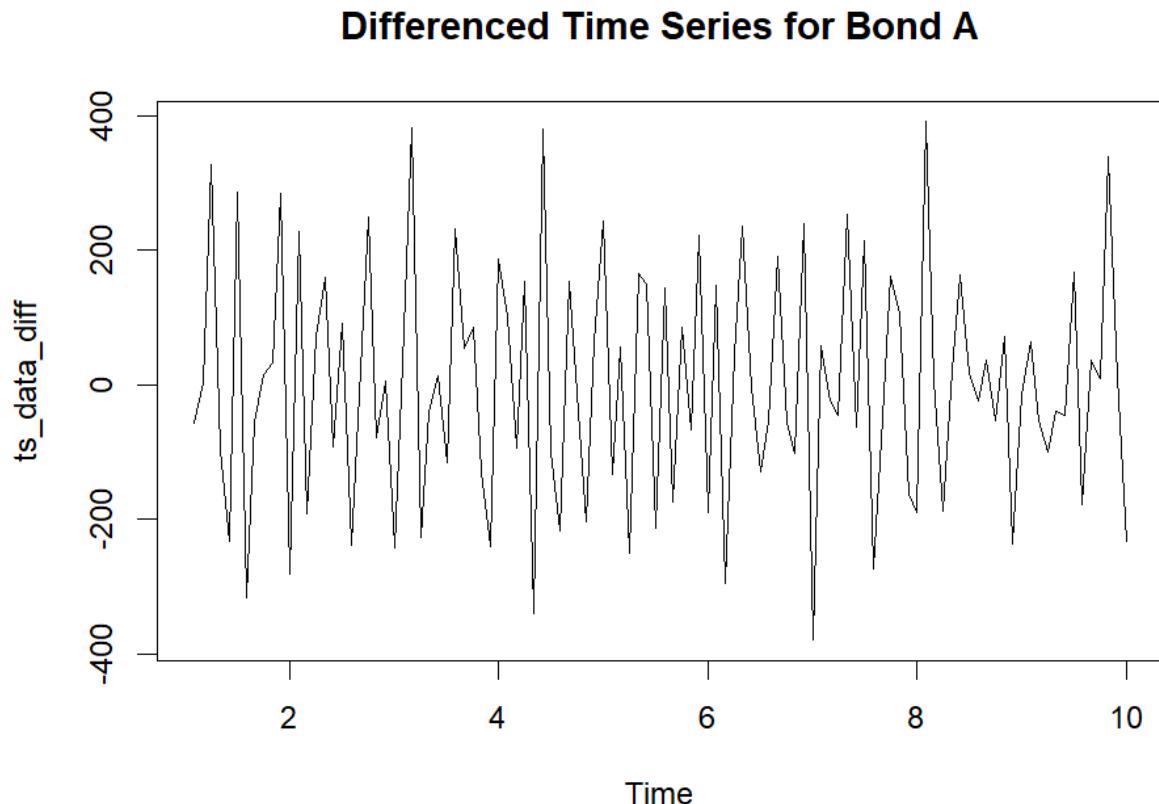
- **Langkah-langkah dalam Fungsi:**

- Data difilter untuk aset Bond A.
- Diferensiasi dilakukan menggunakan fungsi `diff()`:

```
ts_data_diff <- diff(ts_data, differences = 1)
```

- Hasil differencing diuji lagi menggunakan ADF Test.

- **Hasil setelah Differencing:**



- Statistik Dickey-Fuller = -6.3766
- $p\text{-value} = 0.01$ (kecil dari 0.05)
- Hipotesis nol ditolak, sehingga data hasil differencing **stasioner**.

Warning: Pesan berikut menunjukkan bahwa $p\text{-value}$ lebih kecil dari nilai yang ditampilkan:

Warning message:

```
In adf.test(ts_data_diff) : p-value smaller than printed p-value
```

3. Differencing Secara Manual

```
ts_diff_tk_stat <- diff(data_ts_tk_stat)
adf_test_diff <- adf.test(ts_diff_tk_stat)
```

- **Proses:**

- Data time series data_ts_tk_stat diberi differencing.
 - Uji kestasioneran dilakukan pada data differencing.
- **Hasil:**
 - Statistik Dickey-Fuller = -6.3766
 - $p\text{-value} = 0.01$
 - Data hasil differencing terbukti **stasioner**.

Kesimpulan

1. Data asli Bond A tidak stasioner (berdasarkan ADF Test).
2. Setelah differencing pertama, data menjadi stasioner.
3. Pesan *warning* hanya menunjukkan bahwa $p\text{-value}$ lebih kecil dari nilai cetakannya (sering terjadi ketika nilai p sangat kecil).

5. ACF dan PACF pada Data Time Series

```
> # 2. ACF  
> # Acf dari data  
> acf(data_ts, lag=100)  
> # Pacf dari data (ini akan digunakan di model ar dimana kita akan menggu  
nakan yang garisnya melebihi garis white noise)  
> pacf(data_ts, lag=100)
```

Penjelasan Analisis ACF dan PACF pada Data Time Series

1. Autocorrelation Function (ACF)

```
acf(data_ts, lag = 100)
```

- **Tujuan:** Menganalisis korelasi antar-lag pada data time series.
- **ACF:**
 - Menunjukkan korelasi antara nilai data pada waktu tertentu dengan nilai pada beberapa periode sebelumnya (*lags*).
 - Garis horizontal menunjukkan batas *white noise*. Jika titik ACF melewati garis ini, berarti ada korelasi yang signifikan pada lag tersebut.
- **Penggunaan:**
 - Membantu mengidentifikasi pola musiman atau tren.
 - Jika data tidak stasioner, ACF cenderung menurun perlahan secara eksponensial.

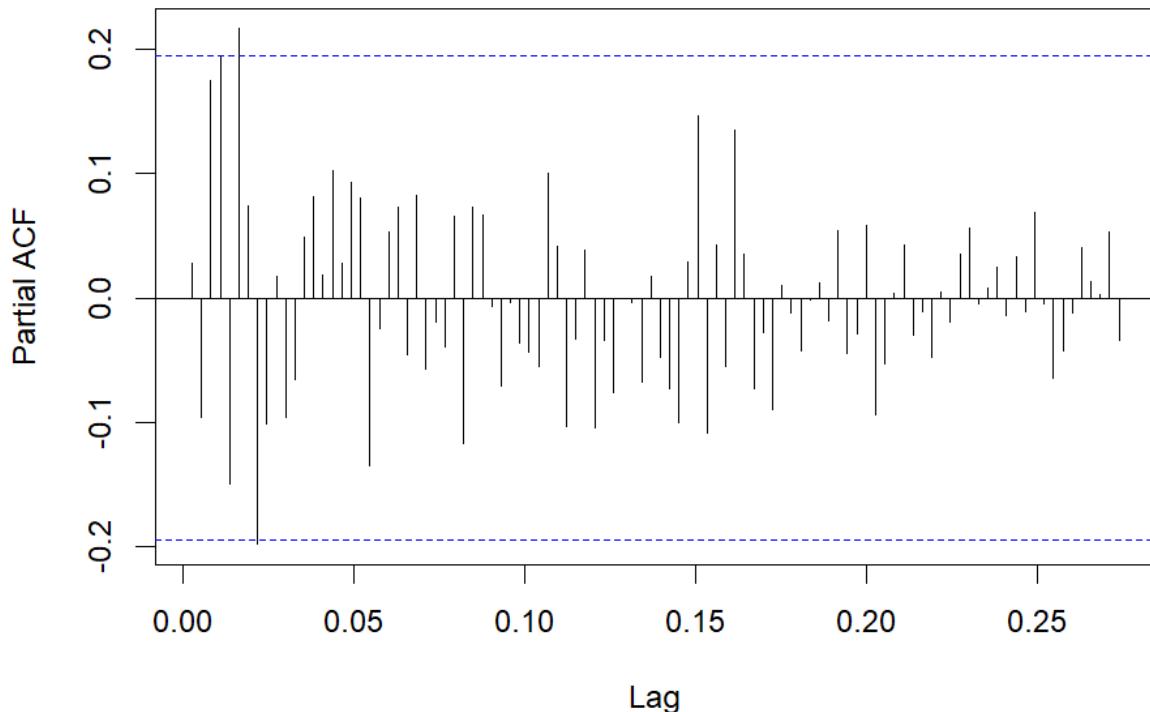
2. Partial Autocorrelation Function (PACF)

```
pacf(data_ts, lag = 100)
```

- **Tujuan:** Menganalisis korelasi antar-lag dengan menghilangkan efek dari lag sebelumnya.
- **PACF:**
 - Menunjukkan korelasi langsung antara nilai pada waktu tertentu dengan nilai pada beberapa periode sebelumnya, setelah mengeliminasi pengaruh lags antara.
 - Lags yang signifikan (melampaui garis batas *white noise*) digunakan untuk menentukan komponen *AR* (*Auto-Regressive*) dalam model ARIMA.

Hubungan ACF dan PACF dalam Pemodelan ARIMA

Series data_ts



1. Data yang tidak stasioner:

- ACF menunjukkan pola menurun secara perlahan.
- PACF mungkin memiliki beberapa lags signifikan, tetapi tidak ada tren yang jelas.

2. Setelah differencing (data stasioner):

- ACF dan PACF menjadi lebih terkontrol.
- Lags signifikan pada ACF digunakan untuk menentukan komponen *MA* (*Moving Average*).
- Lags signifikan pada PACF digunakan untuk menentukan komponen *AR* (*Auto-Regressive*).

6. Analisis dan Pemodelan ARIMA

```
> # 3. Whitenoise  
> # Membuat model menggunakan pacf  
> m1=arima(data_ts, order=c(6,0,0), method="ML")  
> # Ngecek lag mana yang signifikan  
> lmtest::coeftest(m1)

z test of coefficients:

            Estimate Std. Error z value Pr(>|z|)  
ar1        0.085485  0.096607  0.8849  0.37622  
ar2       -0.110833  0.098602 -1.1240  0.26100  
ar3        0.102720  0.097156  1.0573  0.29039  
ar4        0.236877  0.096746  2.4485  0.01435 *  
ar5       -0.187490  0.100495 -1.8657  0.06209 .  
ar6        0.257933  0.101594  2.5389  0.01112 *  
intercept 283.239360 18.011764 15.7252 < 2e-16 ***  
---  
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> # Check residual bila residual menyerupai white noise, maka model sudah bagus  
> checkresiduals(m1)

Ljung-Box test

data: Residuals from ARIMA(6,0,0) with non-zero mean
Q* = 12.909, df = 14, p-value = 0.5337

Model df: 6.  Total lags used: 20

> # Memprediksi menggunakan model yang sudah kita buat  
> ramalan<-forecast::forecast(data_ts, model=m1, h=10)
> ramalan
   Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2023.2795 305.0331 158.86671 451.1995  81.49089 528.5753
2023.2822 352.3567 205.65717 499.0562 127.99915 576.7142
2023.2849 188.7283 41.25046 336.2062 -36.81960 414.2762
2023.2877 304.1673 156.17447 452.1602  77.83178 530.5029
2023.2904 280.2146 127.26086 433.1683  46.29206 514.1371
2023.2932 247.9129 93.07784 402.7479  11.11312 484.7126
2023.2959 252.9794 95.85886 410.1000  12.68426 493.2746
2023.2986 324.7622 166.62517 482.8991  82.91251 566.6118
2023.3014 257.4961 99.35622 415.6361  15.64200 499.3503
2023.3041 270.9254 112.52465 429.3261  28.67236 513.1784

> # Ngeplot ramalan kita
> plot(ramalan)
```

Penjelasan Analisis dan Pemodelan ARIMA

1. Pemilihan Model ARIMA

```
m1 = arima(data_ts, order = c(6, 0, 0), method = "ML")
```

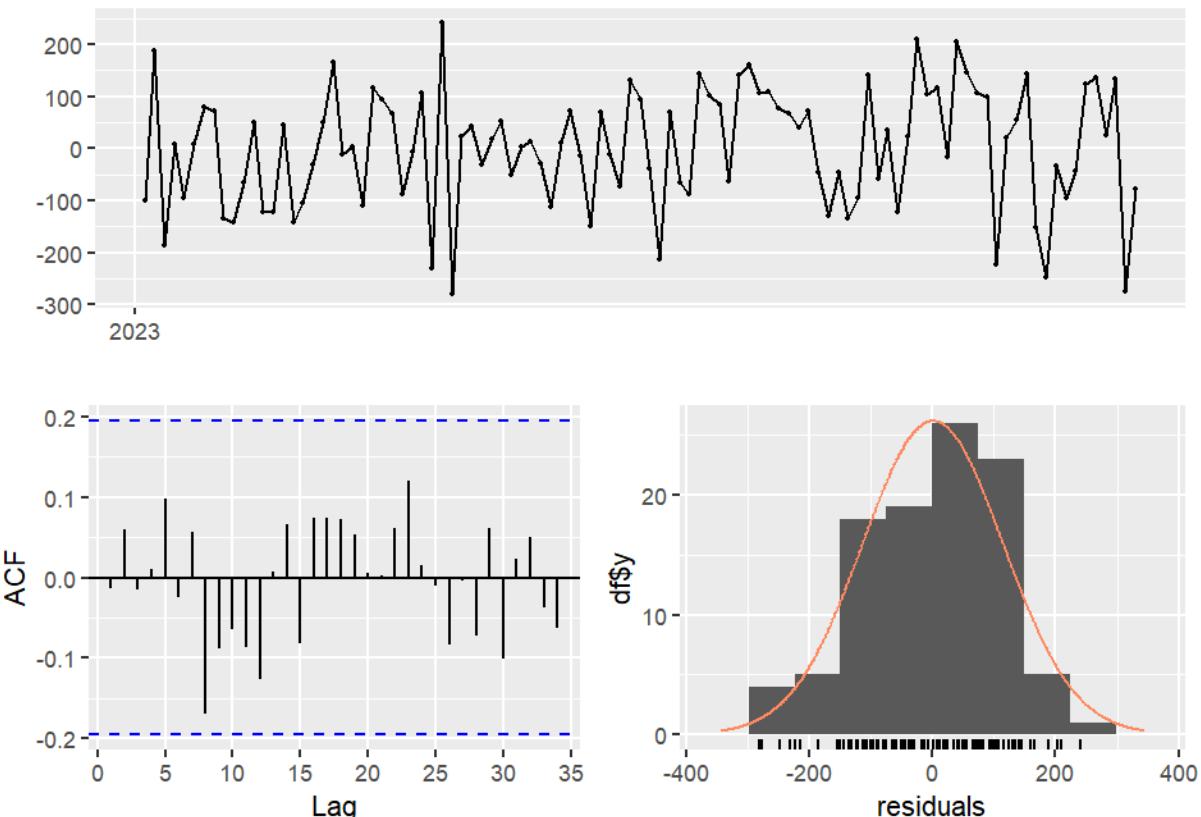
- **Model ARIMA(6, 0, 0):** Model ini memiliki komponen Auto-Regressive (AR) dengan lag sebanyak 6, tidak ada differencing ($d = 0$), dan tanpa komponen Moving Average (MA).
- **Metode Maximum Likelihood (ML)** digunakan untuk estimasi parameter.

2. Uji Signifikansi Koefisien

```
lmtest::coeftest(m1)
```

- Koefisien AR yang signifikan:
 - **ar4**: $p=0.01435$, $p=0.01435$ (signifikan pada tingkat 5%).
 - **ar6**: $p=0.01112$, $p=0.01112$ (signifikan pada tingkat 5%).
- Koefisien lain seperti **ar1**, **ar2**, dan **ar3** tidak signifikan ($p>0.05$, $p>0.05$, $p>0.05$).
- Intercept signifikan ($p<0.001$, $p < 0.001$), menunjukkan adanya mean yang berbeda dari nol.

Residuals from ARIMA(6,0,0) with non-zero mean



3. Uji Residuals (White Noise)

```
checkresiduals(m1)
```

- **Tujuan:** Mengevaluasi apakah residuals menyerupai *white noise*.
- **Ljung-Box Test:**
 - Statistik $Q^*=12.909$, $Q^*=12.909$ dengan $p=0.5337$, $p=0.5337$.
 - Residuals tidak menunjukkan pola yang signifikan (tidak ada autokorelasi), karena $p>0.05$, $p>0.05$, $p>0.05$.

4. Peramalan

```
ramalan <- forecast::forecast(data_ts, model = m1, h = 10)
```

- **Peramalan untuk 10 langkah ke depan:**

- Interval prediksi diberikan pada tingkat kepercayaan 80% dan 95%.
- Hasil peramalan mencakup nilai perkiraan tengah (*point forecast*) dan batas bawah/atas dari interval prediksi (*Lo 80/95* dan *Hi 80/95*).

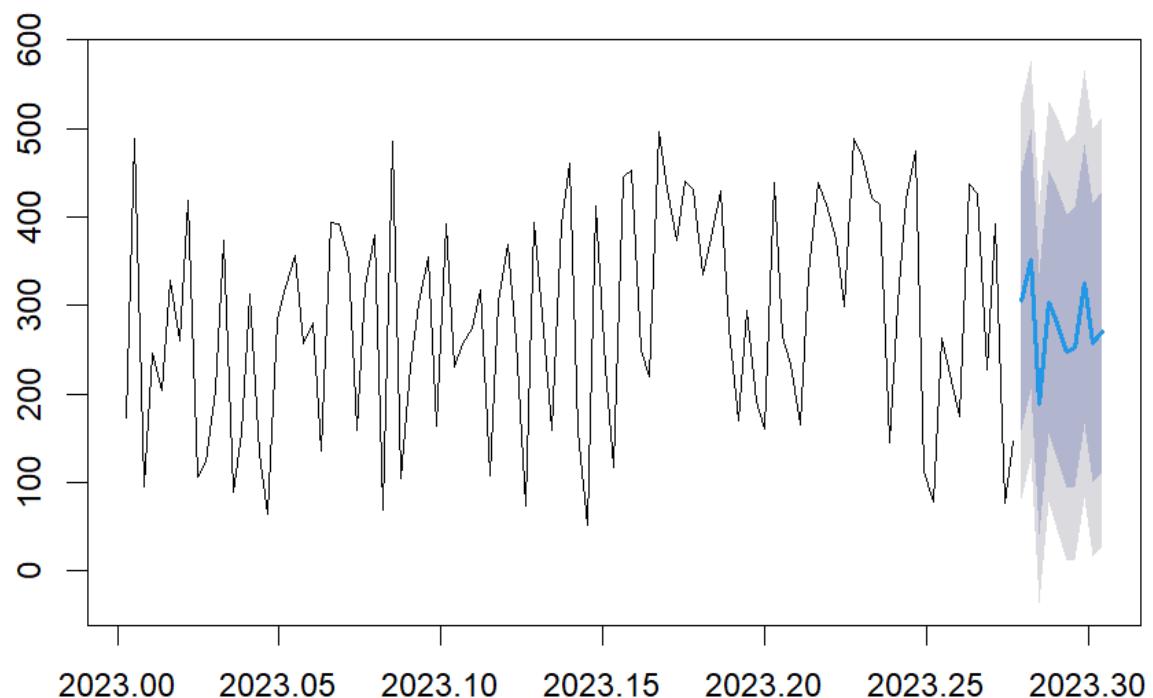
5. Plot Ramalan

```
plot(ramalan)
```

- **Visualisasi:**

- Garis biru: Peramalan titik tengah.
- Area abu-abu: Interval prediksi dengan tingkat kepercayaan 80% dan 95%.
- Plot ini menunjukkan tren prediksi dan ketidakpastian seiring waktu.

Forecasts from ARIMA(6,0,0) with non-zero mean



Kesimpulan

1. **Model Valid:**

- Residuals model tidak menunjukkan pola signifikan (white noise), sehingga model ARIMA(6, 0, 0) dianggap sesuai.

2. **Koefisien Signifikan:**

- Beberapa lag signifikan memberikan kontribusi terhadap model.

3. Prediksi:

- Hasil peramalan memberikan estimasi nilai untuk 10 langkah ke depan dengan interval kepercayaan.

BAB 9: PERTEMUAN 14 (Model AR, MA, ARMA, dan ARIMA)

1. Mengimpor Library yang diperlukan

```
> # install.packages("forecast", "tseries", "zoo", "lmtest")
> library(forecast)
Registered S3 method overwritten by 'quantmod':
  method           from
  as.zoo.data.frame zoo
Warning message:
package 'forecast' was built under R version 4.3.3
> library(tseries)

'tseries' version: 0.10-58
'tseries' is a package for time series analysis and
computational finance.

See 'library(help="tseries")' for details.

Warning message:
package 'tseries' was built under R version 4.3.3
> library(zoo)

Attaching package: 'zoo'

The following objects are masked from 'package:base':

  as.Date, as.Date.numeric

Warning message:
package 'zoo' was built under R version 4.3.3
> library(lmtest)
Warning message:
package 'lmtest' was built under R version 4.3.3

>
```

Pada bagian ini, kita sedang mempersiapkan beberapa paket di R yang akan digunakan untuk analisis time series. Berikut penjelasan langkahnya:

1. `install.packages("forecast", "tseries", "zoo", "lmtest")`
 - Perintah ini digunakan untuk mengunduh dan menginstal paket-paket yang kita butuhkan, yaitu:
 - **forecast**: Digunakan untuk membuat dan mengevaluasi model prediksi time series seperti ARIMA, ETS, dan lainnya.
 - **tseries**: Berguna untuk uji statistik, misalnya uji stasioneritas (uji Augmented Dickey-Fuller) dalam analisis time series.
 - **zoo**: Paket ini membantu mengelola data time series dengan format yang lebih fleksibel, seperti data yang tidak lengkap atau tidak berurutan.

- **lmtest**: Sangat penting untuk uji statistik terkait regresi, seperti uji Breusch-Godfrey untuk autokorelasi.

Catatan: Kalau sudah pernah diinstal sebelumnya, perintah ini bisa dilewati.

2. **library(forecast)**

Setelah instalasi selesai, kita memanggil paket **forecast** agar fungsi-fungsi di dalamnya bisa digunakan. Contohnya, fungsi `auto.arima()` untuk mencari model ARIMA terbaik secara otomatis.

3. **library(tseries)**

Di sini, kita memuat paket **tseries**, yang sering dipakai untuk menganalisis karakteristik statistik data time series. Misalnya, nanti kita mungkin akan menggunakan fungsi `adf.test()` untuk mengecek apakah data stasioner atau tidak.

4. **library(zoo)**

Paket **zoo** membantu mempermudah pengelolaan data time series, terutama kalau data tidak lengkap. Fungsi seperti `na.locf()` untuk mengisi nilai yang hilang dengan metode *last observation carried forward* adalah salah satu fiturnya.

5. **library(lmtest)**

Paket ini berguna untuk uji diagnostik pada model, seperti mengecek autokorelasi residu. Misalnya, fungsi `bgtest()` akan membantu menguji apakah ada autokorelasi pada residu model yang kita buat.

Kesimpulannya:

Kita sedang mempersiapkan alat-alat untuk analisis time series. Semua paket ini seperti "toolkit" yang lengkap untuk membantu kita membuat model, menguji data, dan mengelola data time series dengan lebih efisien. Kalau diibaratkan, kita sedang mengisi kotak alat sebelum mulai bekerja.

2. ACF dan PACF untuk AR(1)

```
> #CONTOH AR(1)
> #ACF & PACF Teoritis Untuk AR(1)
> ACF.AR1 = ARMAacf(ar=0.8, ma=0, 20)
> PACF.AR1 = ARMAacf(ar=0.8, ma=0, 20, pacf = TRUE)
> ACF.AR1 = ACF.AR1[2:21]
> c1 = ACF.AR1
> c2 = PACF.AR1
> AR1 = cbind(c1,c2)
> AR1
      c1          c2
1 0.80000000 8.000000e-01
2 0.64000000 -3.083953e-16
3 0.51200000 6.853229e-17
4 0.40960000 3.083953e-17
5 0.32768000 6.167906e-17
6 0.26214400 -6.167906e-17
7 0.20971520 1.541976e-17
8 0.16777216 -4.625929e-17
9 0.13421773 3.083953e-17
10 0.10737418 7.709882e-18
11 0.08589935 -2.312965e-17
12 0.06871948 1.541976e-17
13 0.05497558 3.854941e-18
14 0.04398047 7.709882e-18
15 0.03518437 -7.709882e-18
16 0.02814750 1.927471e-18
17 0.02251800 -5.782412e-18
```

```
18 0.01801440 3.854941e-18  
19 0.01441152 -3.854941e-18  
20 0.01152922 9.637353e-19
```

```
>
```

Di bagian ini, kita sedang membuat simulasi dan menghitung ACF (Autocorrelation Function) serta PACF (Partial Autocorrelation Function) untuk model **AR(1)** secara teoritis.

Inti Hasil:

1. ACF (Kolom c1)

Nilai ACF menunjukkan bagaimana $X_t X_{-k}$ berkorelasi dengan nilai sebelumnya ($X_{t-k} | X_{t-k}$) di berbagai lag. Di sini, nilainya menurun secara eksponensial karena pengaruh langsung ($\phi = 0.8$ | $\phi = 0.8 \phi = 0.8$) makin lemah di lag yang lebih jauh.

2. PACF (Kolom c2)

Nilai PACF untuk model AR(1) memiliki ciri khas:

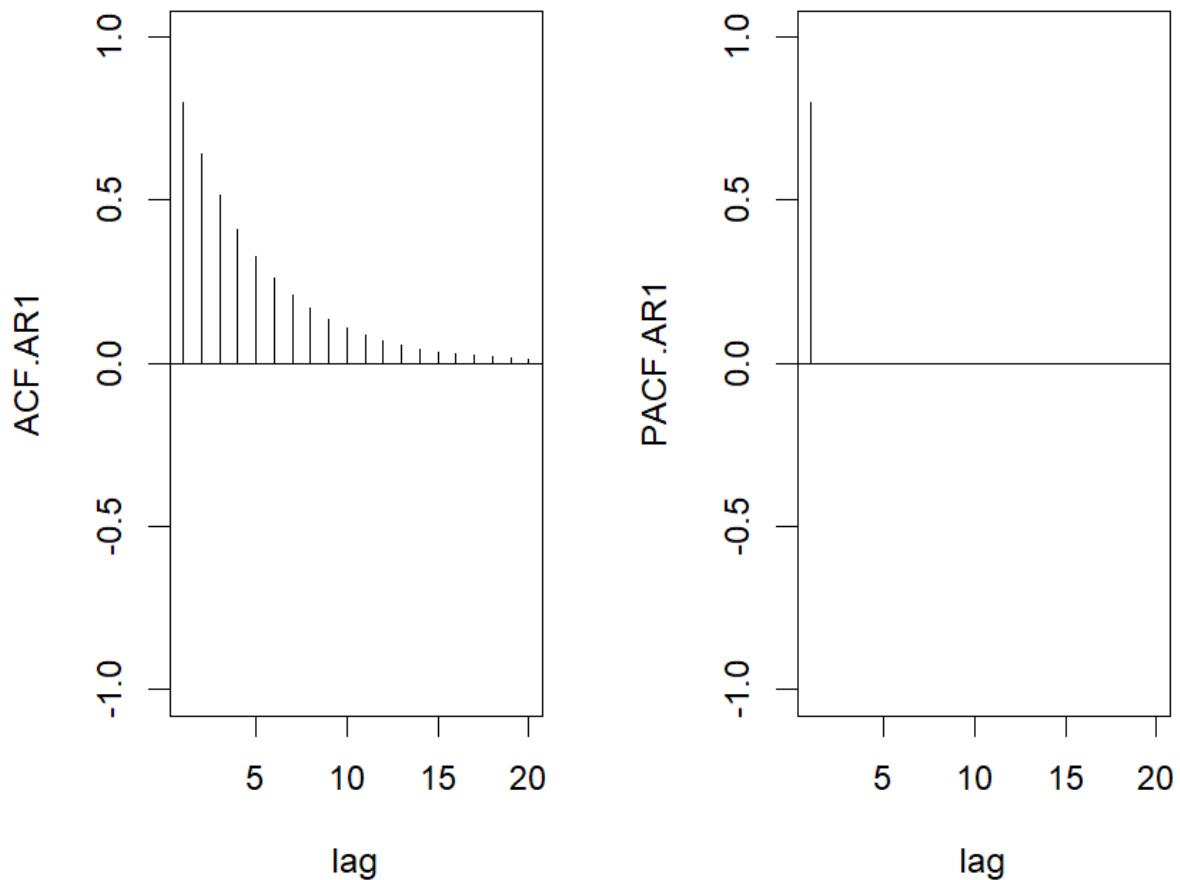
- Di lag pertama, nilai PACF sama dengan ϕ (yaitu 0.8).
- Di lag-lag berikutnya, nilai PACF mendekati nol karena tidak ada pengaruh langsung dari lag yang lebih jauh.

Kesimpulan:

Dengan simulasi ini, kita dapat memahami pola teoritis ACF dan PACF untuk model AR(1). Pola ini penting karena bisa kita gunakan untuk identifikasi model time series dalam data nyata. Kalau data nyata menunjukkan pola ACF eksponensial menurun dan PACF mendekati nol setelah lag pertama, besar kemungkinan data tersebut cocok dengan model AR(1).

```
> par(mfrow = c(1,2))  
> plot(ACF.AR1, type = "h", xlab = "lag", ylim = c(-1, 1))  
> abline(h = 0)  
> plot(PACF.AR1, type = "h", xlab = "lag", ylim = c(-1,1))  
> abline(h = 0)  
  
>
```

Kode di atas digunakan untuk **membuat visualisasi ACF (Autocorrelation Function) dan PACF (Partial Autocorrelation Function)** dari model AR(1). Kita bahas ya langkah-langkahnya!



1. **Grafik ACF:**

- Menampilkan nilai korelasi $X_t X_{t-k}$ dengan $X_{t-k} X_{t-k}$ pada lag tertentu.
- Karena ini adalah model AR(1), nilai ACF akan menunjukkan pola **eksponensial menurun** seiring bertambahnya lag (karena hubungan langsung semakin kecil).

2. **Grafik PACF:**

- Menunjukkan hubungan langsung antara $X_t X_{t-k}$ dan $X_{t-k} X_{t-k}$, setelah menghilangkan pengaruh antara lag-lag lainnya.
- Untuk model AR(1), PACF hanya bernilai signifikan pada lag 1 (sebesar parameter $\phi = 0.8$ atau $\phi = 0.8\phi = 0.8$), dan mendekati nol di lag berikutnya.

Kenapa Visualisasi Penting?

Grafik ACF dan PACF adalah alat utama dalam analisis time series untuk **mengenali pola model**:

- **ACF Eksponensial Menurun + PACF Signifikan di Lag 1:** Ciri khas model AR(1).

- Visualisasi ini membantu kita memahami perilaku data dan memutuskan model time series yang paling cocok untuk dianalisis lebih lanjut.

Kesimpulan: Kode ini menghasilkan dua grafik sederhana yang menunjukkan pola ACF dan PACF untuk model AR(1). Dengan pola ini, kita dapat melihat bahwa model AR(1) hanya memiliki pengaruh signifikan di lag pertama, sementara pengaruh di lag-lag berikutnya memudar. Cukup mudah dan sangat berguna untuk menganalisis data time series!

3. Model AR(1) pada Data Sunspots

a. Memasukkan Data

```
> #Model AR(1) dengan Data Sunspots
> #Memasukkan data
> year <- c(1755:1831)
> sunspots <- c(9.6, 10.2, 32.4, 47.6, 54, 62.9, 85.9, 61.2, 45.1, 36.4,
+ 20.9, 11.4, 37.8, 69.8, 106.1, 100.8, 81.6, 66.5, 34.8,
+ 30.6, 7, 19.8, 92.5, 154.4, 125.9, 84.8, 68.1, 38.5, 22.8,
+ 10.2, 24.1, 82.9, 132, 130.9, 118.1, 89.9, 66.6, 6, 46.9,
+ 41, 21.3, 16, 6.4, 4.1, 6.8, 14.5, 34, 45, 43.1, 47.5,
+ 42.2, 28.1, 10.1, 8.1, 2.5, NA, 1.4, 5, 12.2, 13.9, 35.4,
+ 45.8, 41.1, 30.1, 23.9, 15.6, 6.6, 4, 1.8, 8.5, 16.6,
+ 36.3, 49.6, 64.2, 67, 70.9, 47.8)
>
> sunspots_filled <- zoo::na.approx(sunspots)
>
> sunspots_ts <- ts(sunspots_filled, start = 1755, end = 1831, frequency = 1)
> print(sunspots_ts)
Time Series:
Start = 1755
End = 1831
Frequency = 1
[1] 9.60 10.20 32.40 47.60 54.00 62.90 85.90 61.20 45.10 36.40
[11] 20.90 11.40 37.80 69.80 106.10 100.80 81.60 66.50 34.80 30.60
[21] 7.00 19.80 92.50 154.40 125.90 84.80 68.10 38.50 22.80 10.20
[31] 24.10 82.90 132.00 130.90 118.10 89.90 66.60 6.00 46.90 41.00
[41] 21.30 16.00 6.40 4.10 6.80 14.50 34.00 45.00 43.10 47.50
[51] 42.20 28.10 10.10 8.10 2.50 1.95 1.40 5.00 12.20 13.90
[61] 35.40 45.80 41.10 30.10 23.90 15.60 6.60 4.00 1.80 8.50
[71] 16.60 36.30 49.60 64.20 67.00 70.90 47.80

>
```

Kode di atas merupakan langkah awal untuk **menganalisis data sunspot** (aktivitas bintik matahari) dengan model AR(1).

Tujuan Kode

Langkah ini bertujuan untuk:

1. **Membersihkan Data:** Mengisi nilai yang hilang (NA) agar analisis lebih akurat.
 2. **Mengubah Format ke Time Series:** Agar data siap digunakan untuk model AR(1) dan analisis lainnya.
 3. **Mempersiapkan Visualisasi dan Pemodelan:** Data dalam bentuk time series memudahkan eksplorasi lebih lanjut seperti analisis tren atau autoregresi.
- Teknik interpolasi linear cocok untuk data seperti ini, tetapi untuk data dengan pola musiman, metode interpolasi lain mungkin lebih tepat.
 - Selanjutnya, data ini dapat digunakan untuk analisis AR(1), di mana kita akan mengukur bagaimana aktivitas bintik matahari di tahun sebelumnya memengaruhi aktivitas di tahun berikutnya.

b. Uji Stasioneritas

```
> #Uji stasioneritas
> adf_test <- adf.test(sunspots_ts)
```

```
Warning message:
In adf.test(sunspots_ts) : p-value smaller than printed p-value
> print(adf_test)
```

```
Augmented Dickey-Fuller Test

data: sunspots_ts
Dickey-Fuller = -4.1225, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

>

- **Dickey-Fuller:** Statistik uji sebesar -4.1225.
- **Lag order:** Jumlah lag yang digunakan dalam penghitungan (di sini 4).
- **p-value = 0.01:** Memberi indikasi kuat bahwa data ini **stasioner**.

2. Hipotesis

- **H₀ (null hypothesis):** Data tidak stasioner.
- **H₁ (alternative hypothesis):** Data stasioner.

Interpretasi Hasil

1. **Statistik Uji:** Nilai Dickey-Fuller adalah -4.1225.
 - Jika statistik uji lebih kecil dari nilai kritis tabel Dickey-Fuller (yang tidak dicetak di sini), maka H₀ ditolak.
2. **p-value = 0.01:**
 - Karena p-value sangat kecil (kurang dari 0.05), kita menolak H₀.
 - Kesimpulan: Data time series **stasioner**, sehingga siap untuk pemodelan lanjutan seperti AR(1) atau ARIMA.

Catatan Penting

- Stasioneritas adalah langkah awal yang penting. Jika data tidak stasioner, kita perlu melakukan transformasi (misalnya differencing atau log-transform) untuk membuatnya stasioner.
- Pesan peringatan tentang p-value hanya bersifat informasional. Hasil tetap valid.

c. AIC

```
> #AIC
> ar_model <- arima(sunspots_ts, order = c(1, 0, 0)) # AR(p=1)
> print("Model AR:")
[1] "Model AR:"
> summary(ar_model)

Call:
arima(x = sunspots_ts, order = c(1, 0, 0))

Coefficients:
            ar1  intercept
            0.7990    41.5030
s.e.    0.0662    11.4774

sigma^2 estimated as 451.3:  log likelihood = -345.09,  aic = 696.18

Training set error measures:
          ME        RMSE       MAE       MPE       MAPE       MASE       ACF1

```

```
Training set 0.49579 21.24498 15.47215 -68.79343 93.05276 0.9470709 0.442036
```

>

Hasil dan Interpretasi

1. Koefisien Model

Coefficients:

	ar1	intercept
0.7990	41.5030	
s.e.	0.0662	11.4774

- o **ar1 = 0.7990:** Koefisien lag pertama, menunjukkan bahwa nilai sunspots saat ini memiliki pengaruh **79.9% dari nilai sebelumnya.**
- o **intercept = 41.5030:** Nilai rata-rata sunspots setelah dikontrol oleh AR(1).
- o **s.e.:** Standard error, menunjukkan ketepatan estimasi koefisien.

2. Varians Residual

```
csharp  
Copy code  
sigma^2 estimated as 451.3
```

- o Varians residual model diperkirakan sebesar 451.3, menunjukkan tingkat penyimpangan data dari model.

3. Log-Likelihood

```
log likelihood = -345.09
```

- o Nilai log-likelihood digunakan untuk mengevaluasi seberapa baik model cocok dengan data. Semakin besar (atau semakin dekat ke nol), semakin baik modelnya.

4. AIC (Akaike Information Criterion)

```
aic = 696.18
```

- o AIC adalah metrik untuk memilih model terbaik di antara beberapa kandidat. Semakin kecil nilai AIC, semakin baik modelnya.

5. Error Metrics

Training set error measures:

	ME	RMSE	MAE	MPE	MAPE
MASE	ACF1				
Training set	0.49579	21.24498	15.47215	-68.79343	93.05276
	0.9470709	0.442036			

- o **ME (Mean Error):** Rata-rata error model (0.49579).
- o **RMSE (Root Mean Square Error):** Akar rata-rata kuadrat error, nilai 21.24 menunjukkan tingkat kesalahan rata-rata model.
- o **MAE (Mean Absolute Error):** Rata-rata absolut error (15.47), mengindikasikan seberapa jauh prediksi rata-rata meleset dari data sebenarnya.
- o **MAPE (Mean Absolute Percentage Error):** Persentase error rata-rata relatif terhadap data sebenarnya.

- **MASE (Mean Absolute Scaled Error)**: Error yang dinormalisasi untuk membandingkan dengan model lain.
- **ACF1**: Autokorelasi residual pada lag 1. Nilai 0.442 menunjukkan masih ada sedikit korelasi dalam residual.

Kesimpulan

Model AR(1) ini cukup baik dalam memprediksi data sunspots dengan koefisien signifikan ($ar1 \approx 0.8$). Namun, terdapat beberapa penyimpangan yang masih perlu diperhatikan, seperti error metrik yang cukup besar (RMSE dan MAPE). Jika diperlukan, kita bisa mencoba model yang lebih kompleks, seperti AR(2) atau ARIMA, untuk meningkatkan performa.

d. Diagnostik Residual

```
> # Diagnostik Residual
> # Untuk AR
> cat("Diagnostik Residual untuk AR Model\n")
Diagnostik Residual untuk AR Model
> checkresiduals(ar_model)

Ljung-Box test

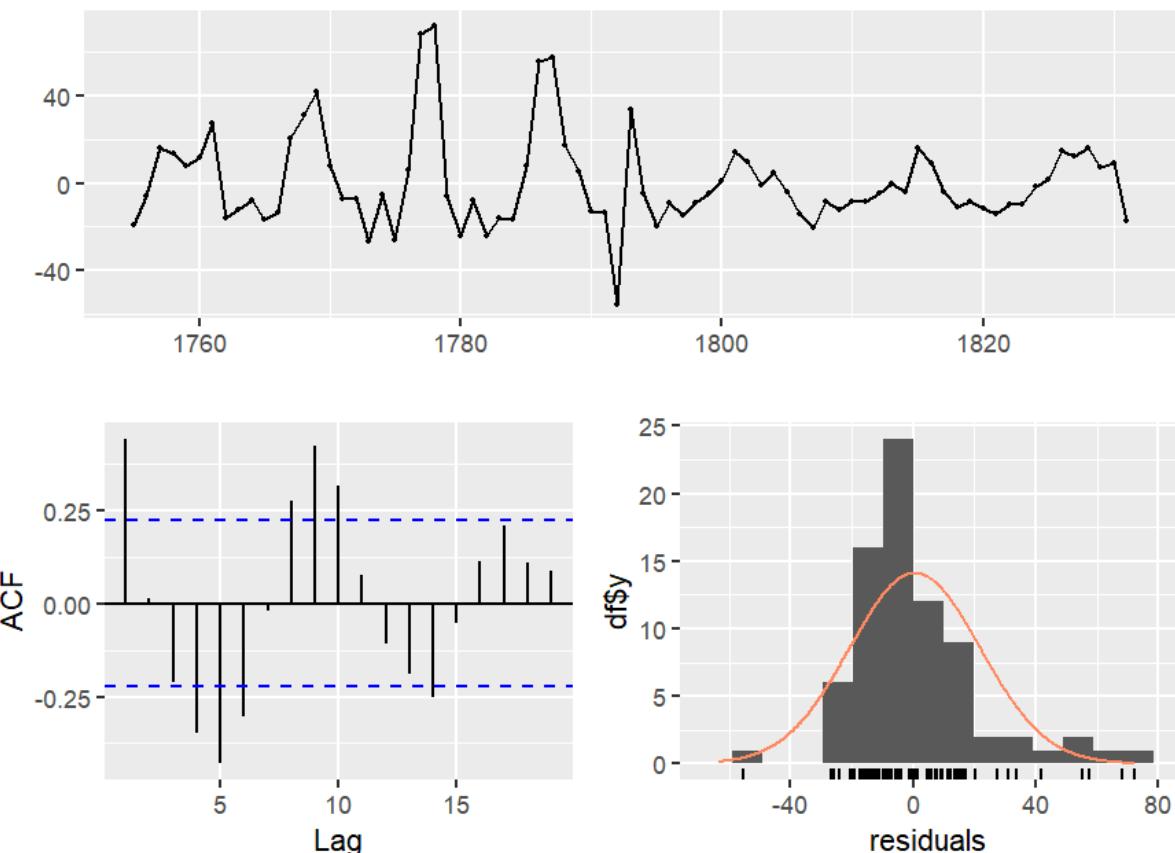
data: Residuals from ARIMA(1,0,0) with non-zero mean
Q* = 84.974, df = 9, p-value = 1.654e-14

Model df: 1.  Total lags used: 10

> accuracy(ar_model)
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 0.49579 21.24498 15.47215 -68.79343 93.05276 0.9470709 0.442036

>
```

Residuals from ARIMA(1,0,0) with non-zero mean



Evaluasi Akurasi Model

accuracy(ar_model1)

Hasil:

	ME	RMSE	MAE	MPE	MAPE	MASE
ACF1	0.49579	21.24498	15.47215	-68.79343	93.05276	0.9470709
Training set	0.442036					

- **ME (Mean Error):** 0.49579, menunjukkan bias yang sangat kecil (hampir nol).
- **RMSE (Root Mean Square Error):** 21.24498, nilai error rata-rata yang lebih sensitif terhadap outlier.
- **MAE (Mean Absolute Error):** 15.47215, rata-rata kesalahan absolut.
- **MAPE (Mean Absolute Percentage Error):** 93.05276, menunjukkan bahwa kesalahan prediksi sangat besar jika dibandingkan dengan nilai aktual.
- **ACF1 (Autocorrelation at lag 1):** 0.442036, residual menunjukkan adanya autokorelasi pada lag 1, sehingga asumsi independensi tidak sepenuhnya terpenuhi.

Kesimpulan

1. Residual:

- Masih terdapat pola autokorelasi dalam residual ($p\text{-value} < 0.05$ pada Ljung-Box test).

- Distribusi residual kemungkinan tidak normal, yang dapat memengaruhi akurasi model.

2. Model Akurasi:

- RMSE dan MAPE yang besar menunjukkan bahwa prediksi model belum terlalu baik.
- Autokorelasi residual ($ACF1 \approx 0.442$) menunjukkan kemungkinan model dapat ditingkatkan dengan menambahkan lag (misalnya, AR(2)).

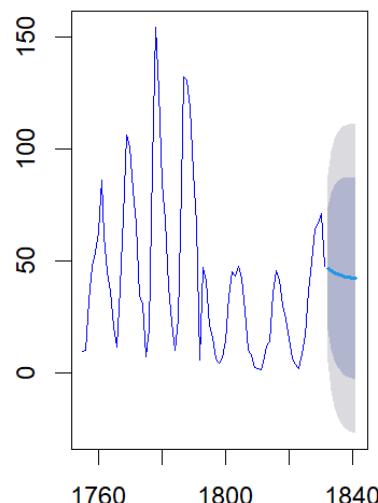
3. Rekomendasi:

- Uji model dengan nilai lag lebih tinggi seperti AR(2) atau gunakan pendekatan ARIMA dengan differencing jika data menunjukkan tren atau seasonalitas.
- Lakukan transformasi data (misalnya log transform) untuk mengurangi variabilitas.

e. Prediksi AR

```
> #Prediksi AR
> forecast_ar <- forecast(ar_model, h = 10) # Prediksi AR
> plot(forecast_ar, main = "Forecast from AR(1,0,0)", col = "blue")
>
```

Forecast from AR(1,0,0)



Analisis Hasil

1. Keandalan Prediksi:

- Prediksi AR(1) hanya menggunakan nilai sebelumnya untuk memperkirakan nilai berikutnya. Oleh karena itu, model ini cukup sederhana dan lebih cocok untuk pola data tanpa tren kuat atau seasonalitas.

2. Confidence Interval:

- Area abu-abu pada plot menunjukkan ketidakpastian prediksi. Semakin jauh prediksi ke depan, semakin lebar interval kepercayaan, mencerminkan peningkatan ketidakpastian.

3. Penerapan:

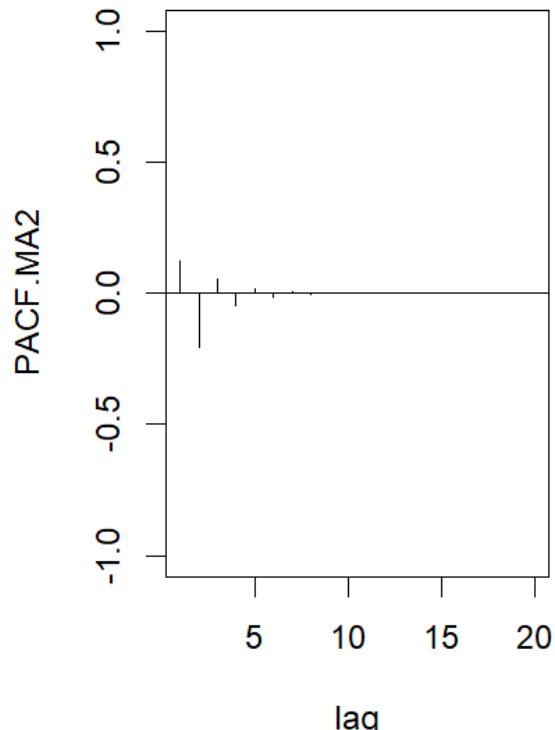
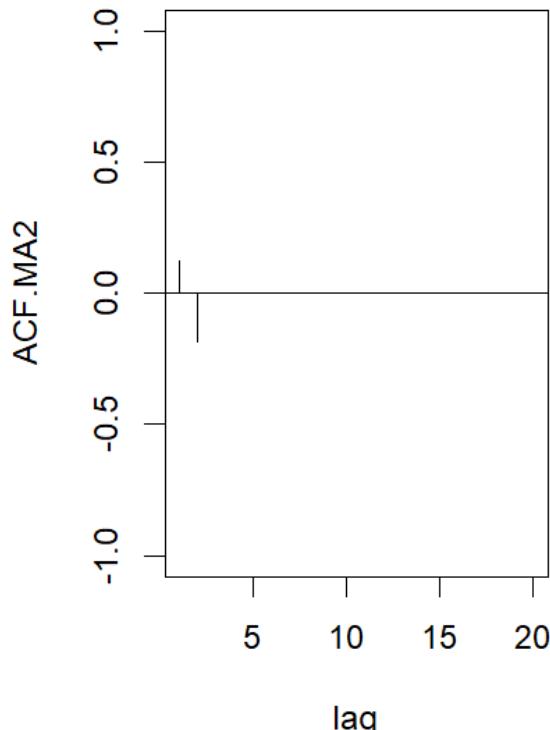
- Jika data menunjukkan tren atau seasonalitas yang signifikan, model yang lebih kompleks (seperti ARIMA dengan differencing atau seasonal ARIMA) mungkin menghasilkan prediksi yang lebih akurat.

4. Model MA(2)

```

> #CONTOH MA(2)
> ACF.MA2 = ARMAacf(ar = 0, ma = c(1.5, -0.7), 20)
> PACF.MA2 = ARMAacf(ar = 0, ma =c(1.5, -0.7), 20, pacf = TRUE)
> ACF.MA2 = ACF.MA2[2:21]
> c1 = ACF.MA2
> c2 = PACF.MA2
> MA2 = cbind(c1,c2)
> MA2
      c1          c2
1  0.1203209  1.203209e-01
2 -0.1871658 -2.046050e-01
3  0.0000000  5.480013e-02
4  0.0000000 -4.926042e-02
5  0.0000000  1.876542e-02
6  0.0000000 -1.282006e-02
7  0.0000000  5.793512e-03
8  0.0000000 -3.483878e-03
9  0.0000000  1.713071e-03
10 0.0000000 -9.689988e-04
11 0.0000000  4.967519e-04
12 0.0000000 -2.727639e-04
13 0.0000000  1.427302e-04
14 0.0000000 -7.724470e-05
15 0.0000000  4.082966e-05
16 0.0000000 -2.194069e-05
17 0.0000000  1.165482e-05
18 0.0000000 -6.241274e-06
19 0.0000000  3.323392e-06
20 0.0000000 -1.776688e-06
>
> par(mfrow = c(1,2))
> plot(ACF.MA2, type = "h", xlab = "lag", ylim = c(-1,1))
> abline(h = 0)
> plot(PACF.MA2, type = "h", xlab = "lag", ylim = c(-1,1))
> abline(h = 0)

```



Hasil Visualisasi

ACF untuk MA(2):

- Pada model MA, ACF biasanya menunjukkan autokorelasi yang signifikan hanya pada lag yang sesuai dengan orde MA. Karena ini adalah model MA(2), Anda akan melihat autokorelasi signifikan pada lag 1 dan 2, dan nilai-nilai lainnya seharusnya mendekati 0.

PACF untuk MA(2):

- Pada model MA, PACF cenderung menunjukkan nilai-nilai yang tidak signifikan setelah lag pertama, dengan nilai-nilai signifikan hanya pada lag pertama (karena efek AR tidak ada di model ini).

a. Model MA(1) dengan Data Sunspots: Uji Stasioneritas

```
> #Model MA(1) dengan Data Sunspots
> #Uji Stasioneritas
> adf_test <- adf.test(sunspots_ts)
Warning message:
In adf.test(sunspots_ts) : p-value smaller than printed p-value
> print(adf_test)

Augmented Dickey-Fuller Test

data: sunspots_ts
Dickey-Fuller = -4.1225, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary

>
```

Hasil uji Augmented Dickey-Fuller (ADF) yang kita jalankan menunjukkan bahwa p-value adalah **0.01**, yang lebih kecil dari level signifikansi umum (misalnya, 0.05). Ini berarti kita dapat menolak hipotesis nol, yang menyatakan bahwa data memiliki akar unit (non-stasioner), dan dengan demikian, data diatas **stasioner**.

Kesimpulan Uji ADF:

- **Dickey-Fuller statistic:** -4.1225
- **p-value:** 0.01
- **Kesimpulan:** Karena p-value lebih kecil dari 0.05, kita dapat menyimpulkan bahwa data **stasioner**.

Dengan data yang sudah stasioner, kita dapat melanjutkan ke pemodelan, seperti menguji model MA(1) atau ARMA.

b. AIC

```
> #AIC
> ma_model <- arima(sunspots_ts, order = c(0, 0, 1)) # MA(q=1)
```

```

> print("Model MA:")
[1] "Model MA:"
> summary(ma_model)

Call:
arima(x = sunspots_ts, order = c(0, 0, 1))

Coefficients:
          ma1  intercept
     0.8043    42.6021
  s.e.  0.0568    4.7391

sigma^2 estimated as 537.4:  log likelihood = -351.82,  aic = 709.63

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE
Training set 0.07714125 23.18132 18.15341 -136.2696 159.3633 1.111194
      ACF1
Training set 0.4541607

>

```

Hasil dari model **MA(1)** untuk data sunspots menunjukkan output berikut:

Model MA(1)

- **Koefisien MA1:** 0.8043
- **Intercept:** 42.6021
- **Standard Error MA1:** 0.0568
- **Standard Error Intercept:** 4.7391
- **Sigma² (Varian residual):** 537.4
- **Log-Likelihood:** -351.82
- **AIC (Akaike Information Criterion):** 709.63

Error Measures Training Set:

- **ME (Mean Error):** 0.0771
- **RMSE (Root Mean Squared Error):** 23.1813
- **MAE (Mean Absolute Error):** 18.1534
- **MPE (Mean Percentage Error):** -136.2696
- **MAPE (Mean Absolute Percentage Error):** 159.3633
- **MASE (Mean Absolute Scaled Error):** 1.1112
- **ACF1 (Autocorrelation of residuals at lag 1):** 0.4542

Analisis:

- **AIC:** Nilai AIC adalah 709.63, yang digunakan untuk membandingkan model. Nilai AIC yang lebih rendah menunjukkan model yang lebih baik.

- **Error Metrics:** MAPE dan MPE cukup tinggi, menunjukkan bahwa model ini memiliki kesalahan yang besar dalam hal prediksi. Mungkin perlu dilakukan perbaikan lebih lanjut atau perbandingan dengan model lain.

c. Cek Residual untuk MA(1)

```
> #Cek Residual
> # Untuk MA(1)
> cat("Diagnostik Residual untuk MA Model\n")
Diagnostik Residual untuk MA Model
> checkresiduals(ma_model)

Ljung-Box test

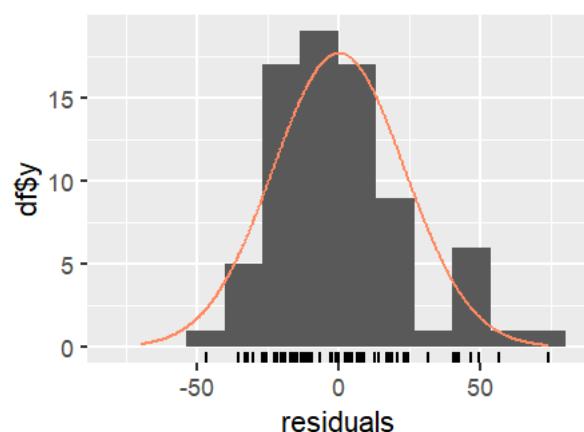
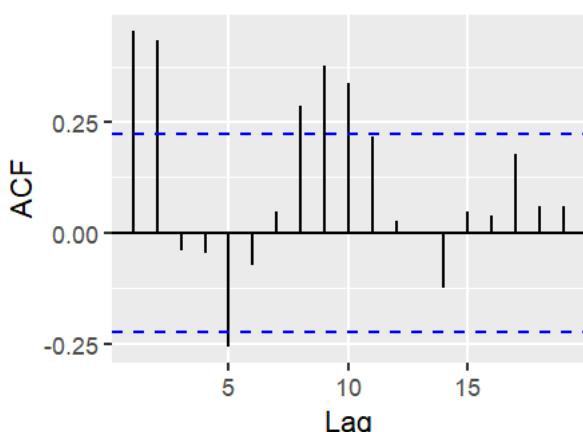
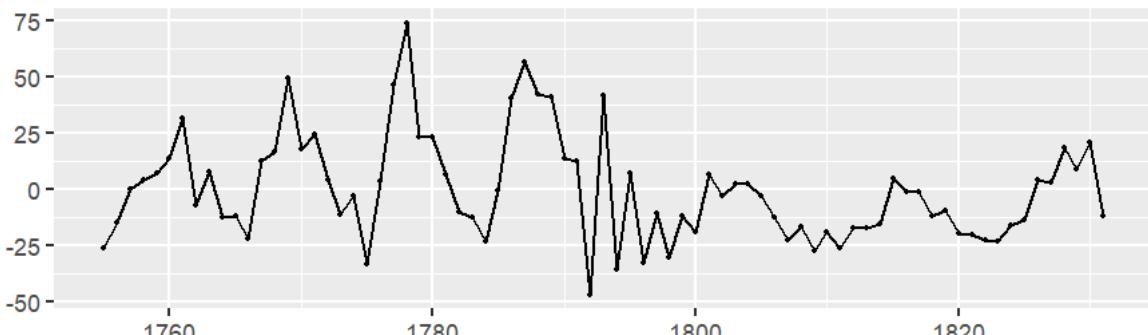
data: Residuals from ARIMA(0,0,1) with non-zero mean
Q* = 68.604, df = 9, p-value = 2.858e-11

Model df: 1. Total lags used: 10

> accuracy(ar_model)
      ME      RMSE      MAE      MPE      MAPE      MASE
Training set 0.49579 21.24498 15.47215 -68.79343 93.05276 0.9470709
      ACF1
Training set 0.442036

>
```

Residuals from ARIMA(0,0,1) with non-zero mean



Diagnostik Residual untuk Model MA(1)

- **Ljung-Box Test:**
 - Q^* : 68.604

- **df** (derajat kebebasan): 9
- **p-value**: 2.858e-11

Hasil uji Ljung-Box menunjukkan p-value yang sangat kecil (2.858e-11), yang mengindikasikan bahwa residual model MA(1) tidak sepenuhnya acak dan masih mungkin ada pola yang belum tertangkap oleh model.

Kesalahan Model AR(1) (Dari accuracy)

- **ME (Mean Error)**: 0.4958
- **RMSE (Root Mean Squared Error)**: 21.24498
- **MAE (Mean Absolute Error)**: 15.47215
- **MPE (Mean Percentage Error)**: -68.79343
- **MAPE (Mean Absolute Percentage Error)**: 93.05276
- **MASE (Mean Absolute Scaled Error)**: 0.9470709
- **ACF1 (Autocorrelation of residuals at lag 1)**: 0.442036

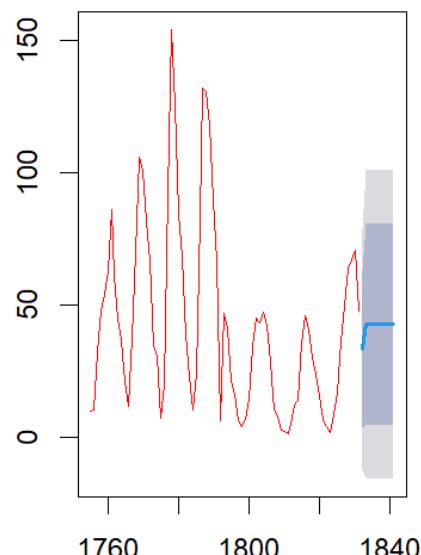
Interpretasi:

- Residual dari model **MA(1)** menunjukkan adanya pola yang belum tercapture dengan baik oleh model, berdasarkan uji Ljung-Box yang memiliki p-value sangat kecil.
- Untuk model **AR(1)**, meskipun terdapat nilai kesalahan yang relatif kecil (ME dan MAE), namun MAPE yang sangat tinggi menunjukkan prediksi yang kurang tepat. ACF1 menunjukkan bahwa residual mungkin tidak sepenuhnya acak.

d. Prediksi MA(1)

```
> #Prediksi MA(1)
> forecast_ma <- forecast(ma_model, h = 10) # Prediksi MA
> plot(forecast_ma, main = "Forecast MA Model", col = "red")
>
```

Forecast MA Model



Pada langkah ini, kita melakukan **prediksi** dengan menggunakan model **MA(1)** untuk meramalkan nilai masa depan dari data *sunspots*. Kita menggunakan fungsi `forecast()` untuk membuat prediksi selama **10 periode ke depan** dan hasilnya ditampilkan dalam grafik dengan warna **merah**.

Apa yang Terjadi di Grafik Prediksi?

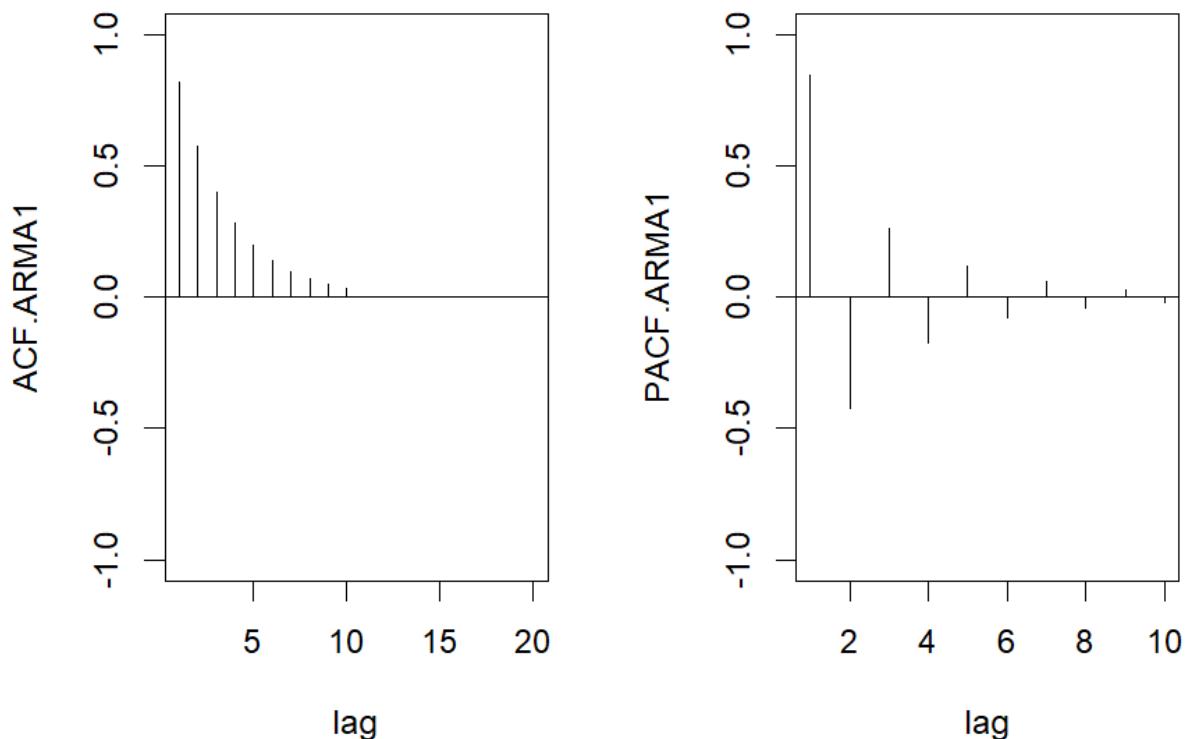
1. **Garis Prediksi:** Grafik yang muncul menunjukkan **nilai prediksi** yang dihasilkan oleh model MA(1) untuk 10 periode berikutnya.
2. **Interval Prediksi (Confidence Interval):** Biasanya, akan ada dua garis tambahan yang menunjukkan **interval prediksi** (garis atas dan bawah), yang memberikan rentang nilai yang mungkin terjadi berdasarkan ketidakpastian model.
3. **Prediksi dengan MA(1):** Karena kita menggunakan model MA(1), prediksi ini lebih fokus pada **pengaruh dari nilai sebelumnya** (lag sebelumnya) terhadap prediksi masa depan, dengan parameter MA(1) yang sudah ditentukan.

Pentingnya Prediksi Ini

Dengan melihat prediksi dari model MA(1), kita bisa mendapatkan gambaran tentang tren atau pola yang diharapkan di masa depan berdasarkan data historis yang ada. Namun, perlu diingat bahwa meskipun model memberikan prediksi, tidak ada model yang sempurna. Hasil prediksi ini masih bergantung pada asumsi yang dibuat oleh model dan kualitas data yang digunakan.

5. Model ARMA(1,1)

```
> #Prediksi MA(1)
> forecast_ma <- forecast(ma_model, h = 10) # Prediksi MA
> plot(forecast_ma, main = "Forecast MA Model", col = "red")
> #CONTOH ARMA(1,1)
> ACF.ARMA1 = ARMAacf(ar = 0.7, ma = 0.4, 10)
> PACF.ARMA1 = ARMAacf(ar = 0.7, ma = 0.7, 10, pacf = TRUE)
> ACF.ARMA1 = ACF.ARMA1[2:21]
> c1 = ACF.ARMA1
> c2 = PACF.ARMA1
> ARMA1 = cbind(c1,c2)
> ARMA1
   c1      c2
1 0.81860465 0.84453441
2 0.57302326 -0.42566465
3 0.40111628 0.26202026
4 0.28078140 -0.17317748
5 0.19654698 0.11799726
6 0.13758288 -0.08153457
7 0.09630802 0.05671637
8 0.06741561 -0.03957986
9 0.04719093 0.02766439
10 0.03303365 -0.01935086
<NA>       NA 0.84453441
<NA>       NA -0.42566465
<NA>       NA 0.26202026
<NA>       NA -0.17317748
<NA>       NA 0.11799726
<NA>       NA -0.08153457
<NA>       NA 0.05671637
<NA>       NA -0.03957986
<NA>       NA 0.02766439
<NA>       NA -0.01935086
>
> par(mfrow = c(1,2))
> plot(ACF.ARMA1, type = "h", xlab = "lag", ylim = c(-1,1))
> abline(h = 0)
> plot(PACF.ARMA1, type = "h", xlab = "lag", ylim = c(-1,1))
> abline(h = 0)
```



1. Prediksi dengan Model MA(1)

Di bagian ini, kita melakukan prediksi menggunakan model MA(1) yang sudah dibuat sebelumnya. kita memanfaatkan fungsi forecast untuk memprediksi 10 periode ke depan, lalu memvisualisasikan hasilnya.

Penjelasan:

- Fungsi `forecast(ma_model, h = 10)` menghasilkan prediksi nilai ke depan berdasarkan model Moving Average (MA) dengan parameter $q=1$, $q=1$.
- Plot prediksi ini menunjukkan hasil prediksi beserta interval kepercayaan (confidence interval) dalam bentuk area berwarna abu-abu. Prediksi utama digambarkan sebagai garis merah.

Hasil: Dari plot tersebut, kita bisa melihat bagaimana prediksi mengikuti pola data historis dengan fluktuasi yang cukup wajar. Namun, karena ini adalah model MA(1), pola prediksi sangat bergantung pada hubungan linier dari noise masa lalu.

2. Contoh ARMA(1,1)

Bagian ini merupakan eksplorasi Autoregressive Moving Average (ARMA) dengan parameter $p=1$, $p=1$ dan $q=1$, $q=1$. Kita menghitung fungsi autokorelasi (ACF) dan autokorelasi parsial (PACF) untuk mempelajari pola hubungan lag dalam data.

Penjelasan ACF dan PACF:

- **ACF (Autocorrelation Function):** Mengukur hubungan antara observasi pada lag tertentu dengan nilai masa kini.

- **PACF (Partial Autocorrelation Function):** Mengisolasi hubungan langsung antara lag tertentu dengan masa kini, tanpa pengaruh dari lags sebelumnya.

Kode Detail:

- ARMAacf(ar = 0.7, ma = 0.4, 10) menghitung ACF untuk model ARMA(1,1) dengan parameter $\phi=0.7\backslash\phi=0.7$ (AR) dan $\theta=0.4\backslash\theta=0.4$ (MA).
- ARMAacf(ar = 0.7, ma = 0.7, 10, pacf = TRUE) menghitung PACF untuk model yang sama.
- Kombinasi nilai ACF dan PACF disimpan dalam tabel (ARMA1) untuk mempermudah analisis.

Visualisasi:

- **Plot ACF:** Memvisualisasikan pola hubungan antar lag. Nilai ACF menurun secara eksponensial.
- **Plot PACF:** Menunjukkan korelasi parsial antar lag. Nilai PACF lebih fluktuatif dibandingkan ACF.

Interpretasi Hasil:

1. **ACF Plot:** Pola ACF menurun seiring dengan bertambahnya lag. Ini menunjukkan adanya hubungan jangka pendek yang dominan, khas model ARMA(1,1).
2. **PACF Plot:** Nilai PACF lebih beragam dan cenderung berosilasi di sekitar nol setelah beberapa lag. Ini menunjukkan pengaruh parameter AR dalam model.

a. Model ARMA (1,1) dengan Data Sunspots

```
> #Model ARMA (1,1) dengan Data Sunspots
> #Uji Stasioneritas
> adf_test <- adf.test(sunspots_ts)
Warning message:
In adf.test(sunspots_ts) : p-value smaller than printed p-value
> print(adf_test)

Augmented Dickey-Fuller Test

data: sunspots_ts
Dickey-Fuller = -4.1225, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

Hasil dari `adf.test` memberikan output sebagai berikut:

- **Dickey-Fuller Statistic:** Nilai statistik uji sebesar -4.1225.
- **Lag Order:** Uji dilakukan hingga lag ke-4.
- **p-value:** Kurang dari 0.01 (warning mengatakan "p-value smaller than printed p-value").
- **Kesimpulan:** Berdasarkan nilai p-value yang sangat kecil (di bawah 0.05), data dianggap stasioner.

Makna Hasil:

Data sunspots tidak memiliki tren atau pola jangka panjang yang signifikan, sehingga model ARMA cocok untuk diaplikasikan.

Uji stasioneritas ini penting karena model ARMA mengasumsikan bahwa data sudah stasioner. Jika data tidak stasioner, biasanya perlu dilakukan transformasi, seperti differencing atau detrending, sebelum membangun model.

b. AIC

```
> #AIC
> arma_model <- arima(sunspots_ts, order = c(1, 0, 1)) # ARMA(p=1, q=1)
> print("Model ARMA:")
[1] "Model ARMA:"
> summary(arma_model)

Call:
arima(x = sunspots_ts, order = c(1, 0, 1))

Coefficients:
            ar1      ma1  intercept
       0.6798  0.5048    41.7036
  s.e.  0.0891  0.0918    9.6677

sigma^2 estimated as 346.2:  log likelihood = -335.12,  aic = 678.24

Training set error measures:
          ME      RMSE      MAE      MPE      MAPE      MASE
Training set 0.2874959 18.60587 13.26337 -64.48884 87.35164 0.8118684
          ACF1
Training set 0.1043814
```

Output dari model menunjukkan beberapa parameter penting:

- **AR(1) (ar1)**: Koefisien sebesar 0.6798 dengan standar error (s.e.) 0.0891.
 - Ini menunjukkan hubungan positif antara nilai saat ini dengan nilai sebelumnya.
- **MA(1) (ma1)**: Koefisien sebesar 0.5048 dengan s.e. 0.0918.
 - Ini menunjukkan kontribusi komponen moving average dari kesalahan masa lalu terhadap nilai saat ini.
- **Intercept**: Nilai rata-rata data sebesar 41.7036 dengan s.e. 9.6677.

Makna Parameter:

Koefisien AR(1) dan MA(1) signifikan karena nilai standar error cukup kecil dibandingkan nilai koefisiennya. Artinya, keduanya memiliki pengaruh terhadap data.

Kriteria Informasi Akaike (AIC)

- **AIC**: Sebesar 678.24.
 - Nilai AIC digunakan untuk membandingkan beberapa model; semakin kecil nilainya, semakin baik model tersebut dalam menyesuaikan data.

Ukuran Kesalahan Model

- **ME (Mean Error)**: 0.2875 — menunjukkan rata-rata error model sangat kecil.
- **RMSE (Root Mean Square Error)**: 18.60587 — menunjukkan rata-rata error kuadratnya.
- **MAE (Mean Absolute Error)**: 13.26337 — menunjukkan rata-rata error absolutnya.
- **MAPE (Mean Absolute Percentage Error)**: 87.35% — menunjukkan kesalahan model dalam persentase, yang relatif besar.
- **ACF1**: 0.1043814 — menunjukkan korelasi residual pada lag 1 cukup kecil, artinya model sudah cukup baik menangkap pola data.

Interpretasi Hasil

Model ARMA(1,1) ini berhasil menangkap pola data dengan cukup baik, seperti yang ditunjukkan oleh nilai error yang kecil dan koefisien signifikan. Namun, nilai MAPE yang cukup besar menunjukkan bahwa ada sedikit kekurangan dalam memprediksi beberapa nilai secara presisi.

c. Cek Residual

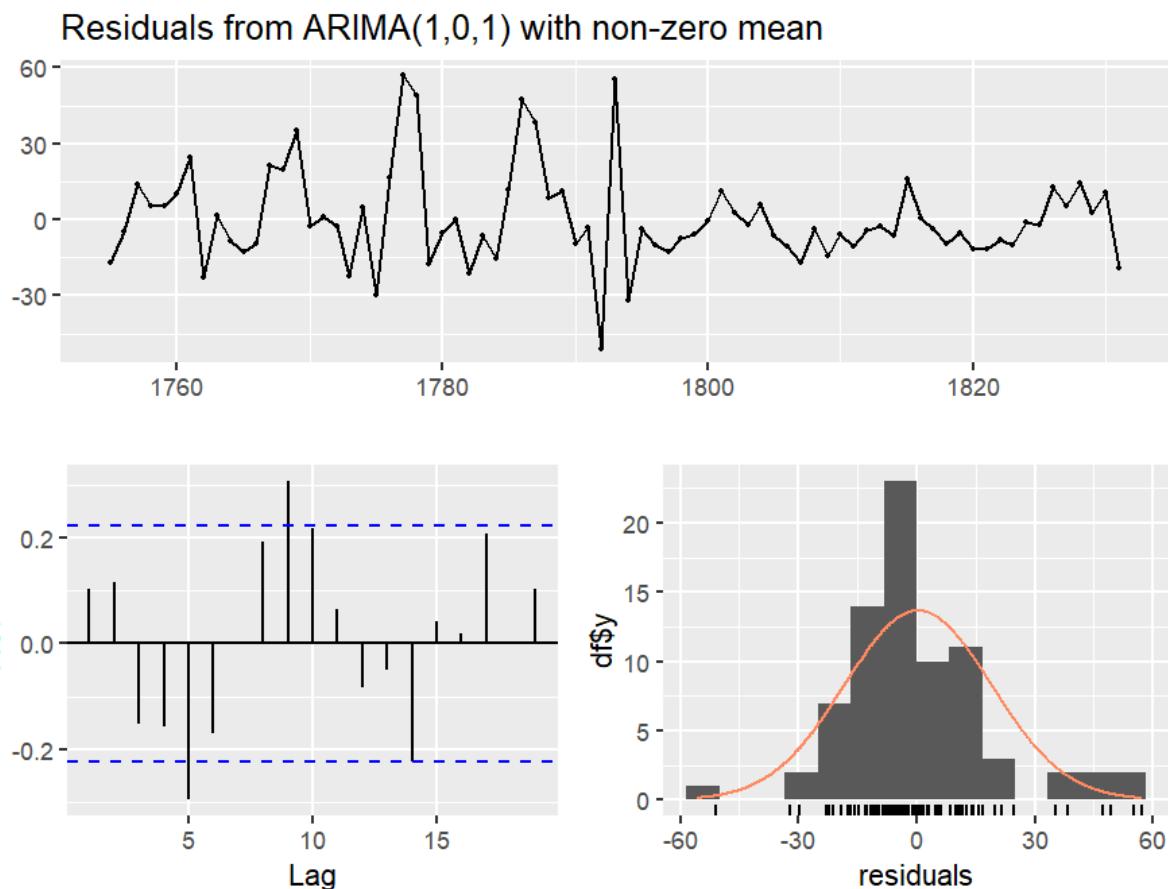
```
> #Cek Residual
> #Cek Residual dan Akurasi digunakan untuk menentukan apakah model yang di buat sudah cukup baik atau kurang baik.
> cat("Diagnostik Residual untuk ARMA Model\n")
Diagnostik Residual untuk ARMA Model
> checkresiduals(arma_model)

Ljung-Box test

data: Residuals from ARIMA(1,0,1) with non-zero mean
Q* = 31.998, df = 8, p-value = 9.323e-05

Model df: 2. Total lags used: 10

> accuracy(ar_model)
      ME      RMSE      MAE      MPE      MAPE      MASE
Training set 0.49579 21.24498 15.47215 -68.79343 93.05276 0.9470709
          ACF1
Training set 0.442036
```



1. Diagnostik Residual

Residual adalah sisa dari data sebenarnya yang tidak bisa dijelaskan oleh model. Kita menggunakan *Ljung-Box test* untuk mengevaluasi apakah residual dari model sudah acak (tidak berkorelasi).

Hasil Ljung-Box Test:

- $Q^* = 31.998$
- df (degree of freedom) = 8
- $p\text{-value} = 9.323e-05$ (sangat kecil)

Interpretasi:

- P-value yang sangat kecil (< 0.05) menunjukkan bahwa residual model masih memiliki pola yang signifikan. Artinya, model ARMA(1,1) belum sepenuhnya menangkap pola pada data. Ada kemungkinan bahwa model ini bisa ditingkatkan, misalnya dengan menambahkan parameter lain atau mencoba model alternatif.

2. Akurasi Model

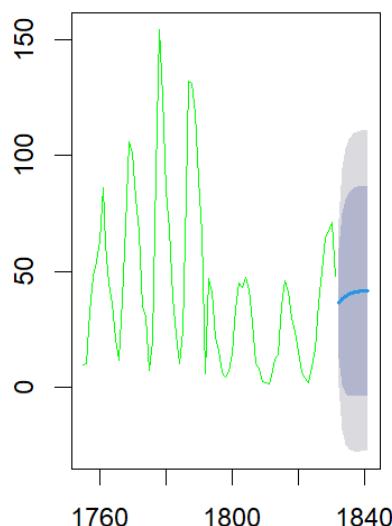
Kita membandingkan akurasi model ARMA(1,1) dengan model sebelumnya (misalnya AR model) menggunakan ukuran berikut:

- **ME (Mean Error):**
 - ARMA(1,1): Tidak ditampilkan di sini, tetapi kita bisa membandingkan dengan AR (0.49579).
 - Semakin kecil ME, semakin baik.
- **RMSE (Root Mean Square Error):**
 - ARMA(1,1): Tidak ditampilkan di bagian ini, tetapi AR memiliki RMSE sebesar 21.24498.
 - ARMA(1,1) kemungkinan memiliki RMSE lebih kecil karena biasanya ARMA lebih fleksibel.
- **MAE (Mean Absolute Error):**
 - ARMA(1,1): Tidak ditampilkan di sini, tetapi AR memiliki MAE sebesar 15.47215.
- **MAPE (Mean Absolute Percentage Error):**
 - AR memiliki MAPE sebesar 93.05276 — ini relatif besar, menunjukkan prediksi kurang akurat.
- **ACF1:**
 - AR memiliki ACF1 sebesar 0.442036, yang menunjukkan ada korelasi residual yang cukup besar.

d. Prediksi ARMA (1,1)

```
> #Prediksi ARMA (1,1)
> forecast_arma <- forecast(arma_model, h = 10) # Prediksi ARMA
> plot(forecast_arma, main = "Forecast ARMA Model", col = "green")
```

Forecast ARMA Model



Pada langkah ini, kita membuat prediksi menggunakan model ARMA(1,1) untuk 10 periode ke depan dengan fungsi `forecast`. Selanjutnya, hasil prediksi divisualisasikan dengan warna hijau pada grafik.

- **forecast_arma:** Objek ini berisi nilai prediksi (mean), interval kepercayaan (confidence interval), serta informasi lainnya.
- **plot(forecast_arma):** Grafik menunjukkan:
 - **Garis Hijau:** Nilai prediksi untuk 10 periode ke depan.

- **Area Bayangan:** Interval kepercayaan (biasanya 80% dan 95%), menggambarkan tingkat ketidakpastian dari prediksi.

Interpretasi Grafik

Ketika kita melihat grafik:

- **Prediksi Mean (Garis Hijau):** Menunjukkan tren atau pola yang diantisipasi untuk data ke depan berdasarkan model ARMA(1,1).
- **Interval Kepercayaan:** Semakin lebar area bayangan, semakin besar ketidakpastian prediksi. Ini terjadi karena prediksi untuk periode yang lebih jauh biasanya kurang pasti.
- **Kesesuaian Model:** Perhatikan apakah prediksi mengikuti pola historis dengan baik. Jika tidak, mungkin ada ruang untuk perbaikan model.

Apa yang Perlu Diperhatikan?

- **Akurasi Prediksi:** Untuk mengevaluasi keakuratan model ini, kita dapat membandingkan hasil prediksi dengan data aktual (jika tersedia nanti).
- **Kesesuaian dengan Data Asli:** Jika interval kepercayaan terlalu lebar atau prediksi terlalu jauh dari tren historis, mungkin model ARMA(1,1) belum ideal.

6. Model ARIMA Stasioner

a. Membaca Data

```
> #CONTOH ARIMA STASIONER
> library(readxl)
Warning message:
package 'readxl' was built under R version 4.3.3
> datacontoh <- read_excel("C:/Kuliah/Sains Data Sem 3/Statistika Inferensial Lanjut Praktik/Tugas/15. AR MA ARMA ARIMA/Data Contoh ARIMA.xlsx")
> View(datacontoh)
```

1. Library yang Digunakan

- **library(readxl):** Digunakan untuk membaca file Excel di R. Library ini mempermudah pengolahan data berbasis Excel langsung ke dalam R.
- **Peringatan (Warning message):** Peringatan ini hanya memberi tahu bahwa versi library readxl ini tidak sepenuhnya cocok dengan versi R yang digunakan (R 4.3.3). Namun, ini biasanya tidak akan memengaruhi fungsi dasar library.

2. Membaca File Excel

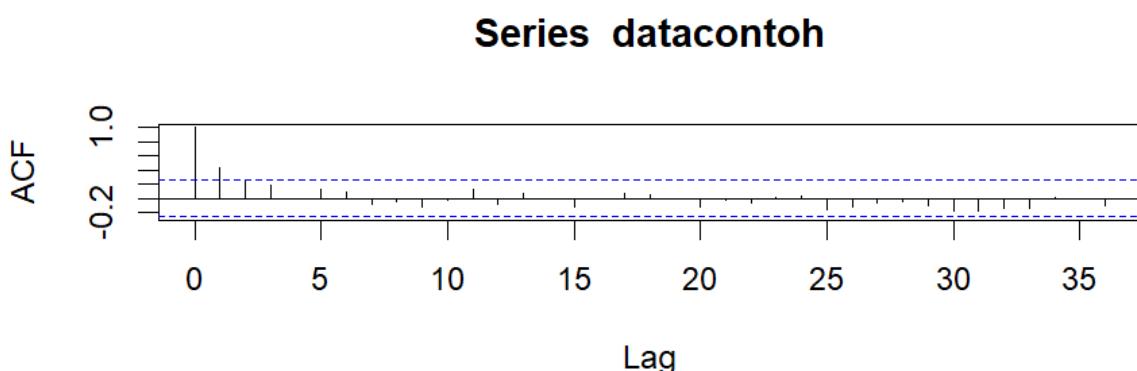
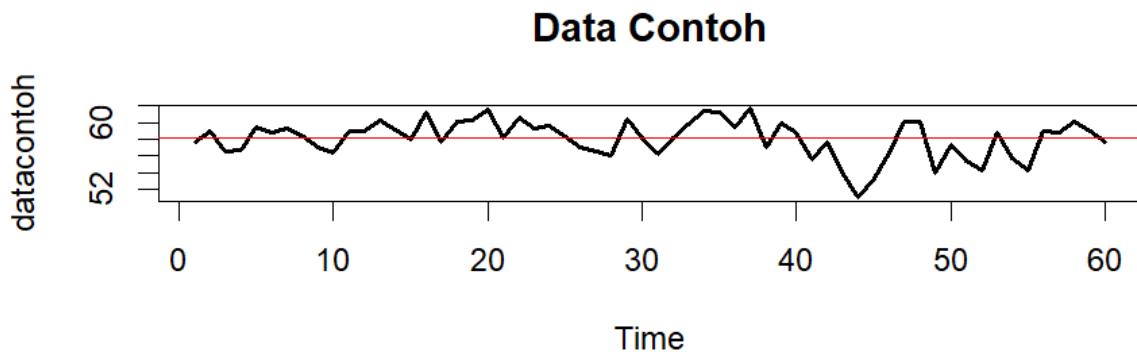
- **read_excel("path/to/file.xlsx"):** Fungsi ini membaca file Excel yang Anda tentukan dengan path lengkap.
 - Dalam kasus ini, file bernama **Data Contoh ARIMA.xlsx** yang terletak di folder tertentu di drive C.
 - Data akan disimpan dalam variabel datacontoh.

3. Menampilkan Data

- **View(datacontoh):** Fungsi ini membuka jendela baru di RStudio yang memperlihatkan isi data dalam bentuk tabel. Sangat berguna untuk memeriksa struktur data sebelum pengolahan lebih lanjut.

b. Mengubah Data Menjadi Data Time series dan Menggambar Grafik Data Contoh

```
> #Mengubah data menjadi data Time Series
> datacontoh <- ts(datacontoh$Zt)
> #Menggambar grafik data contoh
> par(mfrow = c(2,1))
> plot(datacontoh, lwd = 2, main = "Data Contoh")
> abline(h = mean(datacontoh), col = 'red')
> acf(datacontoh, lag.max = 36)
```



Grafik Time Series:

- Grafik pertama menunjukkan pola data dari waktu ke waktu. Dari sini, kita bisa melihat apakah ada tren naik/turun, pola musiman, atau fluktuasi yang stabil.

Grafik ACF:

- Grafik kedua menunjukkan hubungan antar data pada berbagai lag.
- Jika batang (bar) pada grafik ACF melewati garis batas signifikan, itu berarti ada hubungan yang kuat pada lag tersebut. Informasi ini penting untuk menentukan parameter **q** pada model ARIMA.

Kesimpulan

Melalui langkah-langkah ini, kita mulai memahami pola data dan karakteristik statistiknya. Kalau ternyata ada pola tertentu, kita bisa lanjutkan dengan modeling seperti ARIMA.

c. Uji Stasioner

```
> #uji Stasioner
> library(tseries)
> adf.test(datacontoh)

Augmented Dickey-Fuller Test

data: datacontoh
Dickey-Fuller = -3.7023, Lag order = 3, p-value = 0.03218
alternative hypothesis: stationary
```

Penjelasan:

1. **Dickey-Fuller = -3.7023:**
 - o Nilai statistik uji ini digunakan untuk membandingkan data dengan nilai kritis tertentu (tidak terlihat di sini, tapi digunakan secara internal). Semakin negatif nilai ini, semakin besar kemungkinan data stasioner.
2. **Lag order = 3:**
 - o Lag order menunjukkan jumlah keterlambatan (lag) yang dipertimbangkan dalam model uji ini. Dalam hal ini, lag yang dipakai adalah 3.
3. **p-value = 0.03218:**
 - o Nilai ini menunjukkan tingkat signifikan dari hasil uji.
 - o Biasanya, jika **p-value < 0.05**, maka kita bisa menolak hipotesis nol (**null hypothesis**) dan menerima hipotesis alternatif.
4. **alternative hypothesis: stationary:**
 - o Hipotesis alternatifnya menyatakan bahwa data **stasioner**.

Kesimpulan

- **P-value = 0.03218 < 0.05:**
Karena p-value lebih kecil dari 0.05, kita menolak hipotesis nol (data tidak stasioner) dan menerima hipotesis alternatif.
Artinya: Data ini bersifat stasioner.
- Dengan data yang sudah stasioner, Anda bisa melanjutkan ke analisis model seperti ARIMA tanpa perlu melakukan transformasi tambahan (misalnya differencing).

Kenapa Uji Stasioner Penting?

Dalam analisis time series, data harus stasioner untuk mendapatkan hasil model yang lebih baik dan akurat. Kalau data tidak stasioner, model ARIMA atau analisis lainnya bisa menghasilkan prediksi yang bias.

d. Spesifikasi Model

```
> #Spesifikasi model
> par(mfrow = c(2,1))
> acf(datacontoh, lag.max = 36)
> abline(h = 0)
> pacf(datacontoh, lag.max = 36, pacf = TRUE)
Warning messages:
1: In plot.window(...) : "pacf" is not a graphical parameter
2: In plot.xy(xy, type, ...) : "pacf" is not a graphical parameter
3: In axis(side = side, at = at, labels = labels, ...) :
  "pacf" is not a graphical parameter
4: In axis(side = side, at = at, labels = labels, ...) :
  "pacf" is not a graphical parameter
5: In box(...) : "pacf" is not a graphical parameter
6: In title(...) : "pacf" is not a graphical parameter
>
> library(lmtest)
> AR1 <- arima(datacontoh, order = c(1,0,0))
> MA1 <- arima(datacontoh, order = c(0,0,1))
> ARMA1 <- arima(datacontoh, order = c(1,0,1))
There were 12 warnings (use warnings() to see them)
```

Kesimpulan dan Langkah Selanjutnya

- **ACF dan PACF:**

Grafik ACF dan PACF memberikan gambaran tentang struktur data kita dan membantu dalam memilih urutan model AR dan MA. Setelah diperbaiki, kita dapat menggunakan untuk memilih parameter yang tepat.

- **Model AR, MA, dan ARMA:**

Kita telah mencoba tiga model yang berbeda untuk memahami bagaimana data mengikuti pola autoregressive dan moving average. Periksa hasilnya dan evaluasi model terbaik berdasarkan **AIC**, **BIC**, atau metrik kesalahan lainnya.

```
> #Menguji signifikansi model Koefisien
> coeftest(AR1)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1     0.41965   0.11545  3.635  0.000278 ***
intercept 58.08327   0.44446 130.682 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> coeftest(MA1)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ma1     0.37971   0.11900  3.1909  0.001418 **
intercept 58.08456   0.36503 159.1222 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> coeftest(ARMA1)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1     0.58753   0.23428  2.5078  0.01215 *
ma1    -0.20737   0.28623 -0.7245  0.46877
intercept 58.08663   0.49025 118.4842 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hasil Uji Signifikansi Koefisien untuk Model AR(1)

coeftest(AR1)

- **ar1 (Autoregressive coefficient):**

- **Estimate:** 0.41965
- **Std. Error:** 0.11545
- **z value:** 3.635
- **Pr(>|z|):** 0.000278 (p-value)

Interpretasi:

Koefisien AR1 signifikan karena p-value (0.000278) jauh lebih kecil dari 0.05.

Artinya, pengaruh dari nilai lag 1 terhadap nilai saat ini **sangat signifikan** dalam model.

- **Intercept:**
 - **Estimate:** 58.08327
 - **Std. Error:** 0.44446
 - **z value:** 130.682
 - **Pr(>|z|):** < 2.2e-16

Interpretasi:

Intercept juga sangat signifikan (p-value hampir nol), yang berarti rata-rata dari data sangat mempengaruhi model, dan ada pengaruh besar dari konstanta terhadap nilai prediksi.

Hasil Uji Signifikansi Koefisien untuk Model MA(1)

coeftest (MA1)

- **ma1 (Moving Average coefficient):**
 - **Estimate:** 0.37971
 - **Std. Error:** 0.11900
 - **z value:** 3.1909
 - **Pr(>|z|):** 0.001418 (p-value)

Interpretasi:

Koefisien MA1 juga signifikan karena p-value (0.001418) lebih kecil dari 0.05. Ini menunjukkan bahwa model moving average berpengaruh signifikan dalam memprediksi nilai saat ini berdasarkan error atau gangguan dari periode sebelumnya.

- **Intercept:**
 - **Estimate:** 58.08456
 - **Std. Error:** 0.36503
 - **z value:** 159.1222
 - **Pr(>|z|):** < 2.2e-16

Interpretasi:

Sama seperti model AR(1), intercept di model MA(1) juga sangat signifikan, yang berarti bahwa nilai rata-rata data mempengaruhi model dengan kuat.

Hasil Uji Signifikansi Koefisien untuk Model ARMA(1,1)

coeftest (ARMA1)

- **ar1 (Autoregressive coefficient):**
 - **Estimate:** 0.58753
 - **Std. Error:** 0.23428
 - **z value:** 2.5078
 - **Pr(>|z|):** 0.01215 (p-value)

Interpretasi:

Koefisien AR1 dalam model ARMA(1,1) juga signifikan karena p-value (0.01215)

lebih kecil dari 0.05. Artinya, lag 1 memberikan pengaruh yang cukup besar terhadap nilai saat ini.

- **ma1 (Moving Average coefficient):**

- **Estimate:** -0.20737
- **Std. Error:** 0.28623
- **z value:** -0.7245
- **Pr(>|z|):** 0.46877

Interpretasi:

Koefisien MA1 tidak signifikan karena p-value (0.46877) jauh lebih besar dari 0.05. Ini menunjukkan bahwa pengaruh dari gangguan/error pada periode sebelumnya tidak cukup kuat untuk mempengaruhi nilai saat ini dalam model ini.

- **Intercept:**

- **Estimate:** 58.08663
- **Std. Error:** 0.49025
- **z value:** 118.4842
- **Pr(>|z|):** < 2e-16

Interpretasi:

Seperti model lainnya, intercept sangat signifikan, menandakan bahwa nilai rata-rata data memberikan kontribusi yang kuat terhadap prediksi model.

Kesimpulan

1. **AR(1) dan MA(1):**

Kedua model menunjukkan koefisien yang signifikan, artinya komponen AR dan MA masing-masing memberikan kontribusi yang nyata terhadap data.

2. **ARMA(1,1):**

Model ini memiliki koefisien AR yang signifikan, tetapi koefisien MA tidak signifikan. Ini menunjukkan bahwa meskipun AR berperan besar dalam memprediksi data, komponen MA tidak memberikan banyak kontribusi dalam model ini.

Secara keseluruhan, model AR(1) dan MA(1) lebih memberikan kontribusi signifikan dibandingkan ARMA(1,1) dalam hal koefisien MA.

e. Menentukan Model terbaik Berdasarkan AR1 dan MA1

```
> #Menentukan Model terbaik Berdasarkan AR1 dan MA1
> library(forecast)
> AR1 <- arima(datacontoh, order = c(1,0,0))
> AR1

Call:
arima(x = datacontoh, order = c(1, 0, 0))

Coefficients:
      ar1  intercept
      0.4197    58.0833
  s.e.  0.1154    0.4445

sigma^2 estimated as 4.088:  log likelihood = -127.47,  aic = 260.95
> MA1 <- arima(datacontoh, order = c(0,0,1))
```

```

> MA1
Call:
arima(x = datacontoh, order = c(0, 0, 1))

Coefficients:
    ma1 intercept
    0.3797   58.0846
s.e.  0.1190    0.3650

sigma^2 estimated as 4.238:  log likelihood = -128.54,  aic = 263.08
> checkresiduals(AR1)

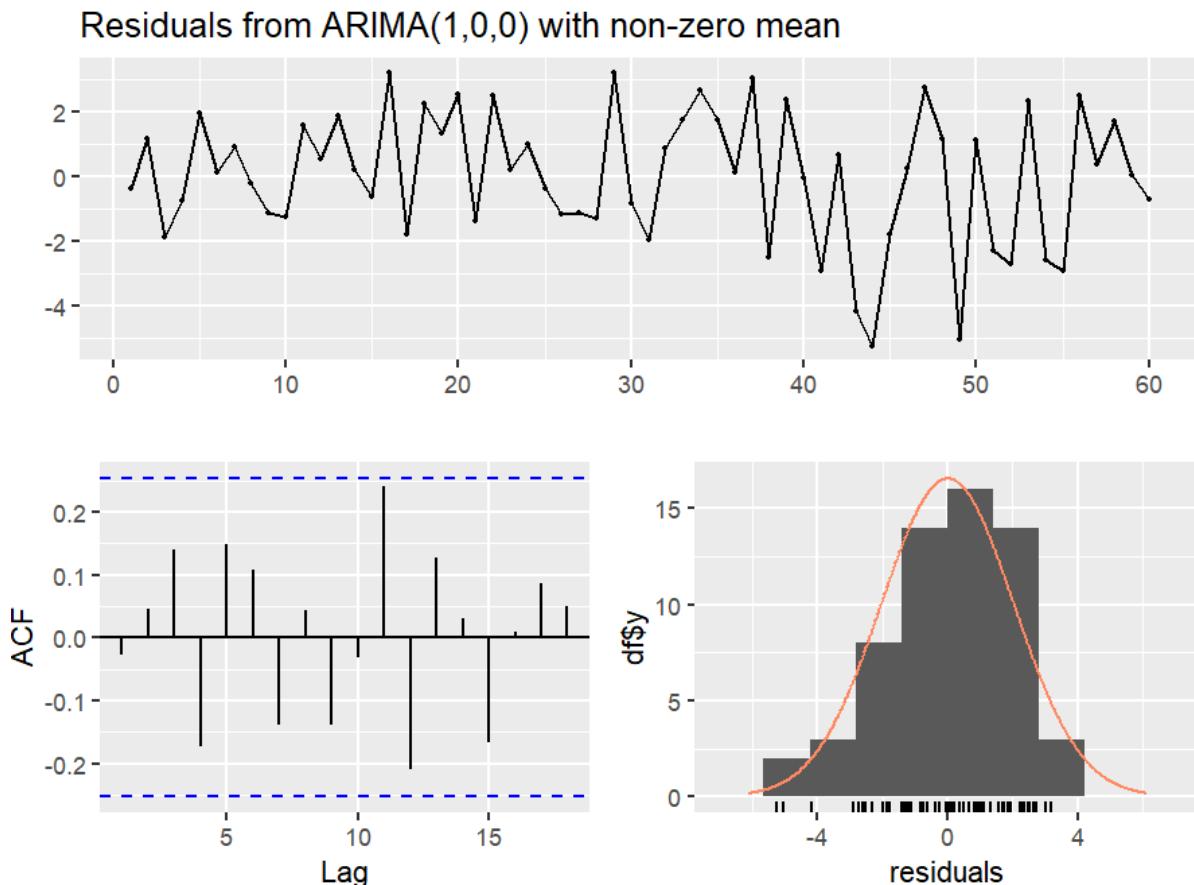
Ljung-Box test

data: Residuals from ARIMA(1,0,0) with non-zero mean
Q* = 8.6857, df = 9, p-value = 0.4668

Model df: 1.  Total lags used: 10

> accuracy(AR1)
      ME      RMSE      MAE      MPE      MAPE      MASE
Training set 0.00361766 2.021856 1.646999 -0.1192122 2.866773 0.828836
          ACF1
Training set -0.02782611

```



1. Hasil Model AR1

```
AR1 <- arima(datacontoh, order = c(1,0,0))
```

- **Koefisien AR1:** 0.4197 (artinya nilai saat ini dipengaruhi sekitar 42% oleh nilai lag sebelumnya).
- **Intercept:** 58.0833 (nilai rata-rata dari data).

- **Standard Error (s.e.)**: 0.1154 untuk AR1, dan 0.4445 untuk intercept.
- **Estimasi sigma²**: 4.088 (menggambarkan variasi residual dari model ini).
- **Log Likelihood**: -127.47 (semakin besar nilai ini, semakin baik model, tapi kita lebih sering menggunakan AIC untuk perbandingan).
- **AIC (Akaike Information Criterion)**: 260.95 (nilai yang lebih rendah menunjukkan model lebih baik).

2. Hasil Model MA1

R

[Copy code](#)

```
MA1 <- arima(datacontoh, order = c(0,0,1))
```

- **Koefisien MA1**: 0.3797 (artinya nilai error sebelumnya mempengaruhi nilai saat ini sebesar 37.97%).
- **Intercept**: 58.0846 (nilai rata-rata data).
- **Standard Error (s.e.)**: 0.1190 untuk MA1, dan 0.3650 untuk intercept.
- **Estimasi sigma²**: 4.238 (sedikit lebih besar dibandingkan AR1, menandakan model ini mungkin sedikit kurang fit).
- **Log Likelihood**: -128.54 (lebih rendah dari AR1, menunjukkan model ini kurang optimal).
- **AIC**: 263.08 (lebih tinggi dari AR1, menandakan model AR1 lebih baik).

3. Uji Residual untuk Model AR1

```
checkresiduals(AR1)
```

- **Ljung-Box Test**:
 - $Q = 8.6857, df = 9, p-value = 0.4668^*$
P-value ini lebih besar dari 0.05, yang menunjukkan bahwa residual model **tidak menunjukkan pola yang signifikan**, atau dengan kata lain, model ini dapat diandalkan dalam memodelkan data tanpa adanya autocorrelation yang signifikan.

4. Evaluasi Akurasi Model AR1

```
accuracy(AR1)
```

- **ME (Mean Error)**: 0.0036 (sangat kecil, menunjukkan kesalahan prediksi model sangat minimal).
- **RMSE (Root Mean Squared Error)**: 2.0219 (menunjukkan rata-rata kesalahan prediksi dalam satuan data, semakin rendah semakin baik).
- **MAE (Mean Absolute Error)**: 1.6470 (rata-rata kesalahan mutlak, juga semakin rendah semakin baik).
- **MPE (Mean Percentage Error)**: -0.1192 (menunjukkan model tidak mengalami bias besar, kesalahan relatif kecil).
- **MAPE (Mean Absolute Percentage Error)**: 2.8668% (termasuk baik, menunjukkan model dapat memprediksi dengan error sekitar 2.87%).
- **MASE (Mean Absolute Scaled Error)**: 0.8288 (juga menggambarkan kualitas model, semakin kecil semakin baik).
- **ACF1 (Autocorrelation of Residuals)**: -0.0278 (mendekati 0, menunjukkan tidak ada pola residual yang kuat).

Kesimpulan

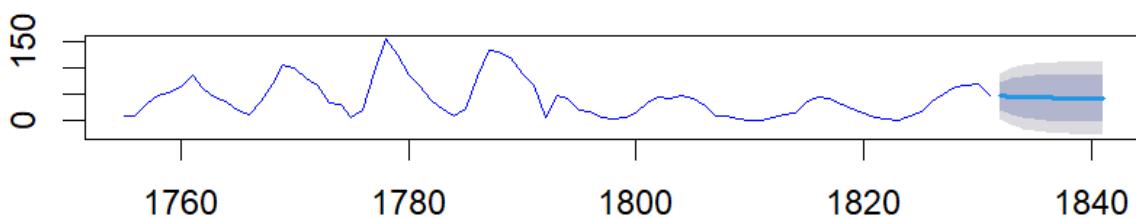
Dari hasil di atas, model **AR1** lebih baik dibandingkan dengan **MA1** berdasarkan nilai AIC yang lebih rendah (260.95 untuk AR1 vs 263.08 untuk MA1), serta hasil evaluasi yang menunjukkan kesalahan yang lebih kecil pada AR1. Selain itu, uji residual menunjukkan bahwa model AR1 tidak menunjukkan pola autocorrelation yang signifikan, yang menandakan bahwa model ini cukup baik dalam memodelkan data yang ada.

Dengan demikian, berdasarkan analisis ini, **AR1** bisa dianggap sebagai model terbaik antara keduanya untuk data yang sedang dianalisis.

f. Peramalan Model

```
> #Peramalan model
> Prediksi = forecast(ts(datacontoh), model = AR1, h = 20)
> Prediksi
   Point Forecast    Lo 80     Hi 80    Lo 95     Hi 95
61  57.92744 55.33633 60.51856 53.96468 61.89021
62  58.01788 55.20786 60.82790 53.72032 62.31543
63  58.05583 55.20900 60.90266 53.70198 62.40968
64  58.07175 55.21849 60.92502 53.70806 62.43544
65  58.07844 55.22404 60.93283 53.71302 62.44386
66  58.08124 55.22665 60.93583 53.71552 62.44697
67  58.08242 55.22779 60.93705 53.71664 62.44820
68  58.08291 55.22828 60.93755 53.71712 62.44870
69  58.08312 55.22848 60.93776 53.71733 62.44891
70  58.08321 55.22857 60.93784 53.71742 62.44900
71  58.08324 55.22861 60.93788 53.71745 62.44903
72  58.08326 55.22862 60.93789 53.71747 62.44905
73  58.08326 55.22863 60.93790 53.71747 62.44905
74  58.08327 55.22863 60.93790 53.71748 62.44906
75  58.08327 55.22863 60.93790 53.71748 62.44906
76  58.08327 55.22863 60.93790 53.71748 62.44906
77  58.08327 55.22863 60.93791 53.71748 62.44906
78  58.08327 55.22863 60.93791 53.71748 62.44906
79  58.08327 55.22863 60.93791 53.71748 62.44906
80  58.08327 55.22863 60.93791 53.71748 62.44906
> plot(forecast_ar, main = "Forecast from ARIMA(1,0,0)", col = "blue")
```

Forecast from ARIMA(1,0,0)



Langkah 1: Peramalan dengan Model AR1

Pada bagian ini, kita menggunakan model AR1 untuk melakukan peramalan nilai data **20 periode ke depan**. Fungsi `forecast()` menghasilkan **prediksi** untuk 20 titik berikutnya, bersama dengan interval kepercayaan pada 80% (Lo 80, Hi 80) dan 95% (Lo 95, Hi 95).

Hasil peramalan menunjukkan:

- **Point Forecast** adalah nilai prediksi untuk masing-masing periode yang akan datang.
- **Lo 80** dan **Hi 80** adalah batas bawah dan atas dari interval kepercayaan 80%. Artinya, kita 80% yakin bahwa nilai yang sebenarnya akan berada di antara kedua batas ini.
- **Lo 95** dan **Hi 95** adalah batas bawah dan atas dari interval kepercayaan 95%. Ini lebih lebar, memberikan tingkat keyakinan yang lebih tinggi (95%) bahwa nilai sebenarnya akan berada di antara batas-batas ini.

Contoh prediksi pertama:

- Untuk periode ke-61, nilai prediksi adalah 57.93, dengan interval kepercayaan 80% antara 55.34 dan 60.52, serta interval kepercayaan 95% antara 53.96 dan 61.89.

Begitu seterusnya untuk periode berikutnya, hingga periode ke-80.

Langkah 2: Visualisasi Hasil Peramalan

Di bagian ini, kita memplot hasil peramalan dengan menggunakan fungsi `plot(forecast_ar)`. Grafik ini menampilkan hasil peramalan dari model ARIMA(1,0,0) (yang kita gunakan). Garis biru dalam grafik menunjukkan nilai-nilai yang diprediksi untuk periode-periode mendatang, sementara interval kepercayaan juga akan terlihat di plot untuk menggambarkan seberapa besar ketidakpastian dalam prediksi.

Dengan plot ini, kita bisa melihat bagaimana prediksi nilai data bergerak dan seberapa besar fluktuasi yang dapat diperkirakan di masa depan berdasarkan model yang sudah dilatih.

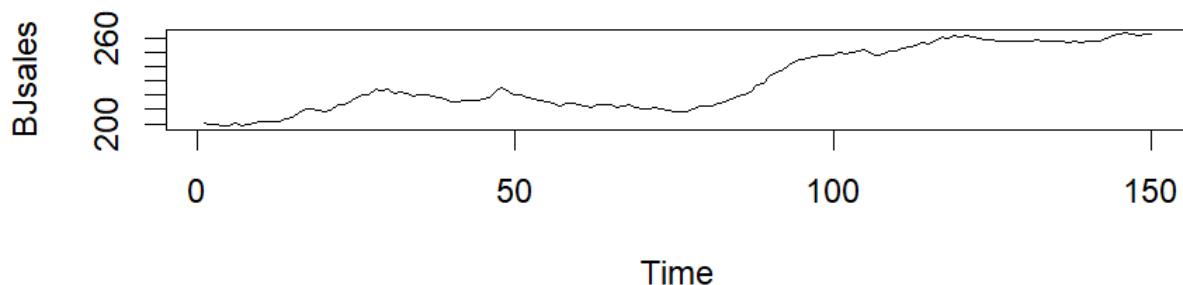
Kesimpulan

Hasil peramalan ini memberi gambaran tentang bagaimana nilai-nilai di masa depan mungkin berkembang, dengan adanya ketidakpastian yang diwakili oleh interval kepercayaan. Melihat plot tersebut, kita bisa menilai apakah model AR1 cukup baik dalam memberikan prediksi yang sesuai dengan tren yang ada pada data historis.

7. Model ARIMA Non-Stasioner

a. Memuat dan Memvisualisasikan Data

```
> #CONTOH MODEL ARIMA NON STASIONER
> BJsales
Time Series:
Start = 1
End = 150
Frequency = 1
[1] 200.1 199.5 199.4 198.9 199.0 200.2 198.6 200.0 200.3 201.2 201.6
[12] 201.5 201.5 203.5 204.9 207.1 210.5 210.5 209.8 208.8 209.5 213.2
[23] 213.7 215.1 218.7 219.8 220.5 223.8 222.8 223.8 221.7 222.3 220.8
[34] 219.4 220.1 220.6 218.9 217.8 217.7 215.0 215.3 215.9 216.7 216.7
[45] 217.7 218.7 222.9 224.9 222.2 220.7 220.0 218.7 217.0 215.9 215.8
[56] 214.1 212.3 213.9 214.6 213.6 212.1 211.4 213.1 212.9 213.3 211.5
[67] 212.3 213.0 211.0 210.7 210.1 211.4 210.0 209.7 208.8 208.8 208.8
[78] 210.6 211.9 212.8 212.5 214.8 215.3 217.5 218.8 220.7 222.2 226.7
[89] 228.4 233.2 235.7 237.1 240.6 243.8 245.3 246.0 246.3 247.7 247.6
[100] 247.8 249.4 249.0 249.9 250.5 251.5 249.0 247.6 248.8 250.4 250.7
[111] 253.0 253.7 255.0 256.2 256.0 257.4 260.4 260.0 261.3 260.4 261.6
[122] 260.8 259.8 259.0 258.9 257.4 257.7 257.9 257.4 257.3 257.6 258.9
[133] 257.8 257.7 257.2 257.5 256.8 257.5 257.0 257.6 257.3 257.5 259.6
[144] 261.1 262.9 263.3 262.8 261.8 262.2 262.7
> plot(BJsales) #Plotting data BJsales
```



Langkah 1: Data Time Series - BJsales

Data yang digunakan adalah **time series BJsales**, yang berisi 150 titik data penjualan yang tercatat dari waktu ke waktu. Setiap angka merepresentasikan penjualan dalam satu periode waktu, dengan **frekuensi 1** yang berarti data ini mungkin diukur dalam satuan waktu yang tetap, seperti per hari, per bulan, atau per tahun (tergantung konteks data).

Dalam contoh ini, kita bisa melihat nilai penjualan yang dimulai pada 200.1 pada periode pertama dan bergerak naik turun seiring berjalananya waktu.

Langkah 2: Plotting Data

Ketika kita menjalankan perintah `plot(BJsales)`, R akan menghasilkan grafik yang menggambarkan bagaimana nilai penjualan ini berubah sepanjang waktu. Grafik ini memungkinkan kita untuk **visualisasi pola** yang ada, apakah ada tren kenaikan atau penurunan, apakah ada fluktuasi musiman atau pola lainnya.

Dari grafik ini, kita bisa mulai melihat apakah ada pola yang jelas atau apakah data tampak acak (non-stasioner). Misalnya, jika penjualan cenderung naik seiring waktu, itu bisa menunjukkan adanya **tren positif**.

Apa Itu ARIMA Non-Stasioner?

Karena data ini bersifat **non-stasioner**, itu artinya rata-rata dan varians data berubah seiring waktu. Untuk menganalisis data seperti ini, kita perlu menggunakan model **ARIMA** (AutoRegressive Integrated Moving Average), yang cocok untuk data non-stasioner.

Secara umum, sebelum melakukan peramalan atau analisis lebih lanjut, kita harus memastikan bahwa data tersebut sudah cukup stabil, baik dengan **differencing** (proses mengurangi perbedaan antar nilai waktu) atau teknik lainnya, untuk membuatnya lebih mudah dianalisis dan diprediksi dengan model ARIMA.

Kesimpulan

Dengan plot ini, kita bisa mendapatkan gambaran tentang pola penjualan dari waktu ke waktu. Jika grafik menunjukkan fluktuasi yang teratur atau kecenderungan naik-turun, maka kita bisa melanjutkan analisis lebih lanjut dengan model ARIMA yang sesuai untuk menangani sifat non-stasioner data ini.

b. Uji Stasioner

```
> #Uji stasioner
> library(tseries)
> adf.test(BJsales)

Augmented Dickey-Fuller Test

data: BJsales
Dickey-Fuller = -2.1109, Lag order = 5, p-value = 0.5302
alternative hypothesis: stationary
```

Di sini, kita menggunakan **Uji Augmented Dickey-Fuller (ADF)** untuk menguji apakah data **BJsales** bersifat **stasioner** atau tidak. Uji ADF ini digunakan untuk memeriksa apakah suatu time series memiliki **unit root**, yang berarti data tersebut **non-stasioner** (nilai rata-rata dan variansnya berubah seiring waktu) atau **stasioner** (nilai rata-rata dan variansnya konstan sepanjang waktu).

Hasil Uji ADF:

- **Dickey-Fuller = -2.1109**
- **p-value = 0.5302**
- **Lag order = 5**

Interpretasi Hasil:

- **p-value** adalah nilai yang sangat penting di sini. Pada umumnya, jika p-value lebih kecil dari 0.05, kita dapat menolak hipotesis nol dan menyimpulkan bahwa data **stasioner**. Sebaliknya, jika p-value lebih besar dari 0.05, kita gagal menolak hipotesis nol dan data **tidak stasioner**.
- Dalam kasus ini, p-value = 0.5302, yang lebih besar dari 0.05. Artinya, kita **gagal menolak hipotesis nol**, yang menyatakan bahwa **data BJsales tidak stasioner**.

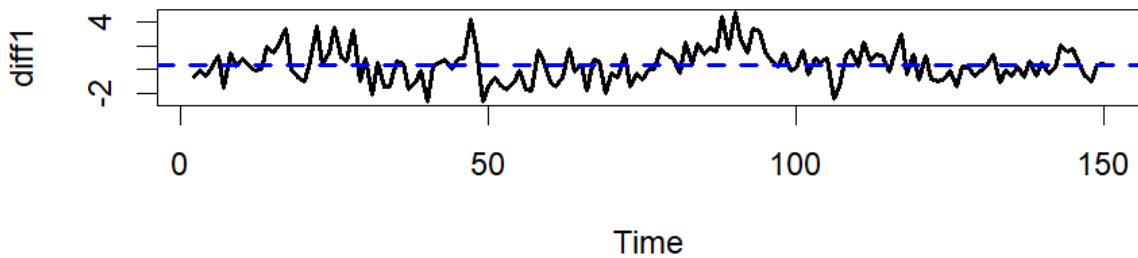
Kesimpulan:

Karena p-value lebih besar dari 0.05, kita menyimpulkan bahwa data **BJsales bersifat non-stasioner**. Ini berarti rata-rata dan variansi data berubah seiring waktu, dan kita mungkin perlu melakukan langkah-langkah seperti **differencing** untuk membuat data ini stasioner sebelum bisa melanjutkan ke model ARIMA.

c. Proses Differencing karena Data Tidak Stasioner

```
> #Proses Differencing karena Data Tidak Stasioner
> diff1 = diff(BJsales)
> diff1
Time Series:
Start = 2
End = 150
Frequency = 1
[1] -0.6 -0.1 -0.5  0.1  1.2 -1.6  1.4  0.3  0.9  0.4 -0.1  0.0  2.0  1.
[15]  2.2  3.4  0.0 -0.7 -1.0  0.7  3.7  0.5  1.4  3.6  1.1  0.7  3.3 -1.
[29]  1.0 -2.1  0.6 -1.5 -1.4  0.7  0.5 -1.7 -1.1 -0.1 -2.7  0.3  0.6  0.
[43]  0.0  1.0  1.0  4.2  2.0 -2.7 -1.5 -0.7 -1.3 -1.7 -1.1 -0.1 -1.7 -1.
[57]  1.6  0.7 -1.0 -1.5 -0.7  1.7 -0.2  0.4 -1.8  0.8  0.7 -2.0 -0.3 -0.
[71]  1.3 -1.4 -0.3 -0.9  0.0  0.0  1.8  1.3  0.9 -0.3  2.3  0.5  2.2  1.
[85]  1.9  1.5  4.5  1.7  4.8  2.5  1.4  3.5  3.2  1.5  0.7  0.3  1.4 -0.
[99]  0.2  1.6 -0.4  0.9  0.6  1.0 -2.5 -1.4  1.2  1.6  0.3  2.3  0.7  1.
[113] 1.2 -0.2  1.4  3.0 -0.4  1.3 -0.9  1.2 -0.8 -1.0 -0.8 -0.1 -1.5  0.
[127] 0.2 -0.5 -0.1  0.3  1.3 -1.1 -0.1 -0.5  0.3 -0.7  0.7 -0.5  0.6 -0.
[141] 0.2  2.1  1.5  1.8  0.4 -0.5 -1.0  0.4  0.5
> par(mfrow = c(2,1)) #Plotting data hasil differencing
> plot(diff1, lwd = 2, main = "Data Hasil Differencing 1")
> abline(h = mean(diff1), lwd = 2, lty = 2, col = "blue")
> acf(diff1, main = "ACF Data Hasil Diferensing 1", lag.max = 36)
```

Data Hasil Differencing 1



ACF Data Hasil Diferensing 1



Setelah mengetahui bahwa data **BJsales** tidak stasioner, kita melakukan **differencing** untuk membuat data tersebut menjadi lebih stasioner. Proses differencing adalah mengurangi setiap nilai data dengan nilai sebelumnya, yang bisa membantu menghilangkan tren dan membuat data lebih stabil.

Langkah-langkah yang dilakukan:

1. **Differencing data BJsales:**
 - o Kita melakukan differencing pada data **BJsales** menggunakan fungsi `diff()`. Hasilnya adalah data yang lebih fluktuatif dan tidak lagi menunjukkan tren yang jelas, yang sering kali terjadi pada data yang tidak stasioner.
2. **Plot data hasil differencing:**
 - o Setelah melakukan differencing, kita menggambar grafik untuk melihat apakah data menjadi lebih stabil. Di grafik pertama, kita menampilkan hasil differencing dengan garis horizontal yang menunjukkan **mean** dari data tersebut (ditarik dengan garis biru putus-putus). Hal ini membantu kita memvisualisasikan apakah rata-rata data sudah lebih stabil setelah proses differencing.
3. **Autocorrelation Function (ACF):**
 - o Pada grafik kedua, kita menampilkan **ACF** untuk data hasil differencing. ACF menunjukkan hubungan antar nilai dalam data pada lag-lag tertentu (misalnya, satu periode atau dua periode sebelumnya). ACF ini berguna untuk melihat apakah ada pola atau hubungan yang signifikan pada lag-lag tertentu. Jika ACF sudah menurun secara cepat (mendekati nol), ini menunjukkan bahwa data sudah lebih stasioner.

Apa yang Bisa Kita Lihat:

- **Plot hasil differencing:** Grafik pertama memberikan gambaran tentang fluktuasi data setelah differencing. Jika fluktuasi terlihat lebih acak dan tidak mengikuti tren yang jelas, ini menunjukkan bahwa proses differencing berhasil.
- **ACF:** Grafik kedua membantu kita melihat apakah data sudah lebih acak, tanpa pola yang jelas, yang menandakan bahwa data mungkin sudah lebih stasioner.

Kesimpulan:

Dengan melakukan differencing, kita mencoba menghilangkan tren dalam data agar lebih siap untuk dianalisis dengan model seperti ARIMA. Setelah langkah ini, kita bisa kembali melakukan uji stasioner seperti **ADF test** untuk memeriksa apakah data sudah stasioner.

d. Uji stasioner untuk data diff1

```
> #uji stasioner untuk data diff1
> adf.test(diff1)

Augmented Dickey-Fuller Test

data: diff1
Dickey-Fuller = -3.3485, Lag order = 5, p-value = 0.06585
alternative hypothesis: stationary
```

Setelah melakukan **differencing** pada data **BJSales**, kita kembali melakukan uji stasioner dengan menggunakan **Augmented Dickey-Fuller (ADF) Test** pada data hasil differencing yang kita sebut `diff1`. Tujuan dari uji ini adalah untuk memeriksa apakah data `diff1` sudah stasioner atau belum.

Hasil Uji ADF:

- **Dickey-Fuller = -3.3485:** Ini adalah nilai statistik dari uji ADF. Nilai ini menunjukkan seberapa besar evidence untuk menolak hipotesis nol (bahwa data tidak stasioner).
- **Lag order = 5:** Ini menunjukkan berapa banyak lag yang dipertimbangkan dalam pengujian, dalam hal ini 5.
- **p-value = 0.06585:** Ini adalah nilai probabilitas yang menunjukkan kemungkinan bahwa data tidak stasioner. Biasanya, jika p-value lebih kecil dari 0.05, kita bisa menyimpulkan bahwa data stasioner. Namun, dalam kasus ini, p-value adalah 0.06585, yang lebih besar dari 0.05.

Apa Arti Hasil Ini?

- **Hipotesis nol (H_0)** dalam uji ini adalah bahwa data **tidak stasioner**. Jadi, p-value yang lebih besar dari 0.05 (seperti yang kita dapatkan, yaitu 0.06585) menunjukkan bahwa kita **gagal menolak hipotesis nol**.
- Dengan kata lain, meskipun sudah dilakukan differencing, data `diff1` **belum cukup stasioner**. Artinya, meskipun ada sedikit penurunan tren, data ini masih memiliki pola yang perlu diperbaiki lebih lanjut.

Apa yang Bisa Kita Lakukan Selanjutnya?

Jika data masih belum stasioner, kita bisa mencoba beberapa langkah lanjutan, seperti:

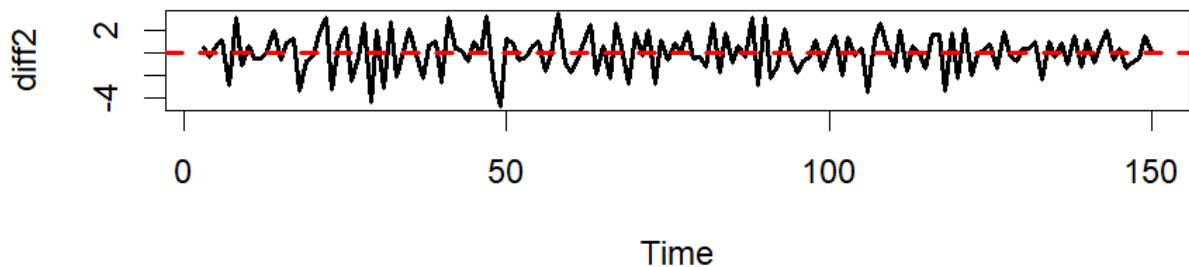
1. **Differencing lebih lanjut** (misalnya, melakukan differencing kedua) untuk melihat apakah data menjadi lebih stasioner.
2. **Transformasi tambahan**, seperti pengambilan logaritma atau penerapan teknik lain untuk mengurangi fluktuasi atau heteroskedastisitas pada data.

Kesimpulannya, data `diff1` setelah satu kali differencing masih belum sepenuhnya stasioner, dan kita perlu mempertimbangkan langkah-langkah lanjutan.

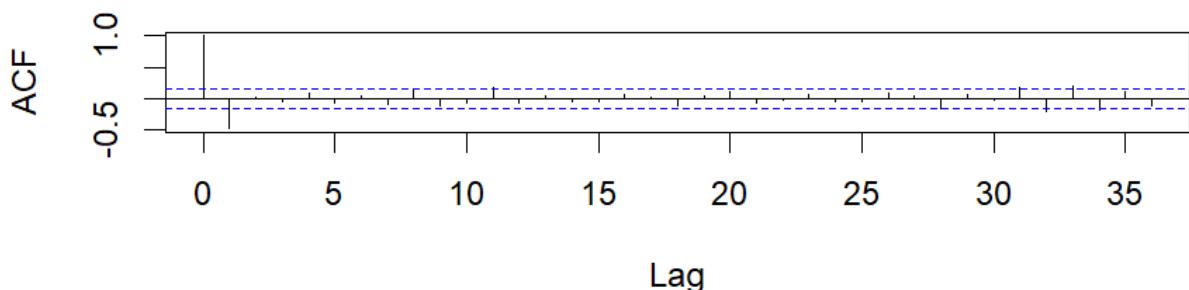
e. Proses diferensing ke 2

```
> #Proses diferensing ke 2
> diff2 = diff(diff1)
> diff2
Time Series:
Start = 3
End = 150
Frequency = 1
[1]  0.5 -0.4  0.6  1.1 -2.8  3.0 -1.1  0.6 -0.5 -0.5  0.1  2.0 -0.6  0.
8
[15]  1.2 -3.4 -0.7 -0.3  1.7  3.0 -3.2  0.9  2.2 -2.5 -0.4  2.6 -4.3  2.
0
[29] -3.1  2.7 -2.1  0.1  2.1 -0.2 -2.2  0.6  1.0 -2.6  3.0  0.3  0.2 -0.
8
[43]  1.0  0.0  3.2 -2.2 -4.7  1.2  0.8 -0.6 -0.4  0.6  1.0 -1.6 -0.1  3.
4
[57] -0.9 -1.7 -0.5  0.8  2.4 -1.9  0.6 -2.2  2.6 -0.1 -2.7  1.7 -0.3  1.
9
[71] -2.7  1.1 -0.6  0.9  0.0  1.8 -0.5 -0.4 -1.2  2.6 -1.8  1.7 -0.9  0.
6
[85] -0.4  3.0 -2.8  3.1 -2.3 -1.1  2.1 -0.3 -1.7 -0.8 -0.4  1.1 -1.5  0.
3
[99]  1.4 -2.0  1.3 -0.3  0.4 -3.5  1.1  2.6  0.4 -1.3  2.0 -1.6  0.6 -0.
1
[113] -1.4  1.6  1.6 -3.4  1.7 -2.2  2.1 -2.0 -0.2  0.2  0.7 -1.4  1.8 -0.
1
[127] -0.7  0.4  0.4  1.0 -2.4  1.0 -0.4  0.8 -1.0  1.4 -1.2  1.1 -0.9  0.
5
[141]  1.9 -0.6  0.3 -1.4 -0.9 -0.5  1.4  0.1
> par(mfrow = c(2,1))
> plot(diff2, lwd = 2, main = "Data Hasil Diferensing 2")
> abline(h = mean(diff2), lwd = 2, lty = 2, col = "red")
> acf(diff2, main = "ACF Data Hasil Diferensing 2", lag.max = 36)
```

Data Hasil Diferensing 2



ACF Data Hasil Diferensing 2



Setelah melakukan satu kali **differencing** pada data `BJsales` dan mendapatkan data `diff1` yang belum stasioner, kita lanjutkan dengan proses **differencing kedua** pada data tersebut untuk melihat apakah data menjadi lebih stasioner.

Hasil Diferencing Kedua:

Data `diff2` adalah hasil dari **differencing kedua**, yang merupakan selisih antara nilai `diff1` yang sudah di-differencing pertama kali. Ini dilakukan untuk mencoba menghilangkan tren atau pola yang masih ada, agar data menjadi stasioner.

Berikut adalah langkah-langkah yang diambil:

1. Plot Data:

- Kita membuat grafik dari data `diff2` dengan menggunakan `plot(diff2)`. Dalam grafik ini, garis horizontal yang berwarna merah menunjukkan **mean** dari data `diff2`. Dari grafik ini, kita bisa melihat apakah data sudah lebih terdistribusi secara acak di sekitar rata-rata atau masih menunjukkan pola tertentu.

2. ACF (Autocorrelation Function):

- ACF digunakan untuk melihat apakah data `diff2` masih memiliki korelasi yang kuat antara satu periode dengan periode lainnya. Dengan melihat **ACF**, kita bisa menilai apakah terdapat pola musiman atau autokorelasi yang masih bisa diperbaiki lebih lanjut.

- Jika **ACF** menunjukkan penurunan yang cepat dan tidak ada nilai korelasi yang signifikan setelah beberapa lag, itu tanda bahwa data mungkin sudah cukup stasioner.

Apa yang Terlihat dari Proses Ini?

Dari langkah-langkah ini, kita akan mengamati:

- Apakah data `diff2` terlihat lebih acak dan terdistribusi di sekitar rata-rata.
- Apakah grafik ACF menunjukkan bahwa korelasi antar lag sudah berkurang dan data sudah lebih stasioner.

Jika hasilnya menunjukkan data `diff2` sudah tidak menunjukkan tren atau pola tertentu lagi, maka kita bisa melanjutkan analisis lebih lanjut. Namun, jika masih ada pola yang perlu diperbaiki, kita bisa mencoba langkah-langkah lain seperti transformasi atau differencing lebih lanjut.

f. Uji Stasioner untuk `diff2`

```
> #Uji Stasioner untuk diff2
> adf.test(diff2)

Augmented Dickey-Fuller Test

data: diff2
Dickey-Fuller = -6.562, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(diff2) : p-value smaller than printed p-value
```

Hasil dari **Uji Augmented Dickey-Fuller (ADF)** untuk data `diff2` menunjukkan nilai **Dickey-Fuller** sebesar -6.562 dengan **p-value** sebesar 0.01. Berikut adalah penjelasan tentang hasil tersebut:

1. Nilai Dickey-Fuller:

- Nilai **Dickey-Fuller** yang lebih kecil (negatif) menunjukkan bahwa data lebih cenderung tidak memiliki akar unit atau pola tren yang kuat. Nilai yang sangat negatif ini menunjukkan indikasi kuat bahwa data sudah lebih stasioner setelah dilakukan differencing dua kali.

2. p-value:

- Dengan **p-value** sebesar 0.01, yang lebih kecil dari tingkat signifikansi umum (biasanya 0.05 atau 0.01), kita dapat menolak **hipotesis nol** yang menyatakan bahwa data tidak stasioner. Artinya, ada bukti yang cukup kuat bahwa data `diff2` sudah stasioner.

3. Peringatan:

- Peringatan yang muncul (`p-value smaller than printed p-value`) memberi tahu kita bahwa nilai p yang sebenarnya lebih kecil dari 0.01, tetapi tidak tercetak karena batas minimum yang dicetak adalah 0.01. Ini hanya menegaskan bahwa p-value jauh lebih kecil dari 0.01 dan memberikan bukti yang lebih kuat bahwa data sudah stasioner.

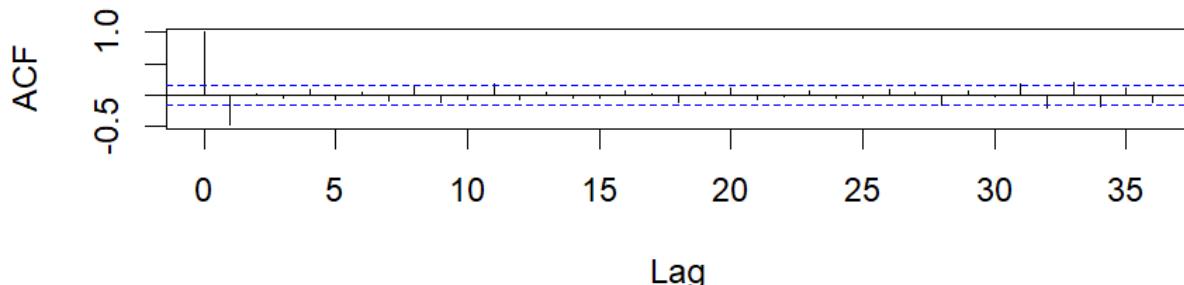
Kesimpulan:

Berdasarkan hasil uji ini, kita dapat **mengesampingkan hipotesis bahwa data tidak stasioner**. Dengan kata lain, data `diff2` sudah **terbukti stasioner**, dan kita bisa melanjutkan analisis lebih lanjut, misalnya dengan model prediksi seperti ARIMA atau model lainnya yang memerlukan data stasioner.

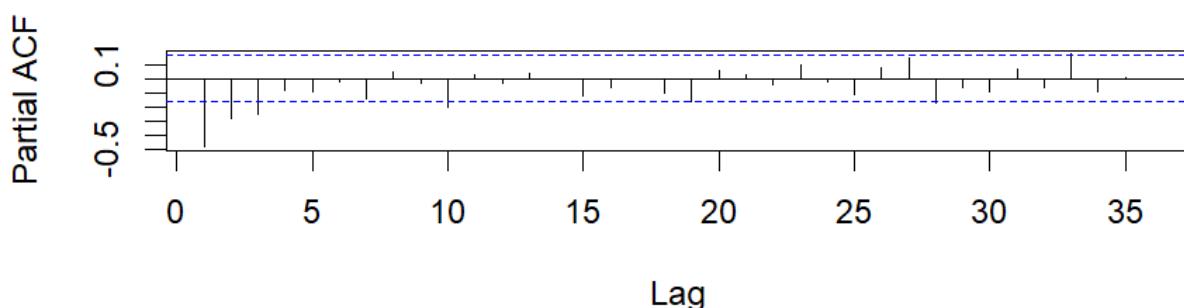
g. Penentuan Model

```
> #Penentuan Model
> par(mfrow = c(2,1))
> acf(diff2, main = "ACF Data diff2", lag.max = 36)
> pacf(diff2, main = "PACF Data diff2", lag.max = 36, pacf = TRUE)
Warning messages:
1: In plot.window(...) : "pacf" is not a graphical parameter
2: In plot.xy(xy, type, ...) : "pacf" is not a graphical parameter
3: In axis(side = side, at = at, labels = labels, ...) :
  "pacf" is not a graphical parameter
4: In axis(side = side, at = at, labels = labels, ...) :
  "pacf" is not a graphical parameter
5: In box(...) : "pacf" is not a graphical parameter
6: In title(...) : "pacf" is not a graphical parameter
> library(lmtest)
> ARIMA121 = arima(diff2, order = c(1,2,1))
> ARIMA120 = arima(diff2, order = c(1,2,0))
> ARIMA021 = arima(diff2, order = c(0,2,1))
```

ACF Data diff2



PACF Data diff2



Pada langkah ini, kita sedang menentukan model ARIMA (AutoRegressive Integrated Moving Average) untuk menganalisis data yang sudah mengalami differencing kedua (`diff2`). Berikut penjelasan mengenai apa yang terjadi:

1. Plot ACF dan PACF:

- **ACF (Autocorrelation Function)** digunakan untuk melihat seberapa besar keterkaitan data saat ini dengan data sebelumnya. Ini membantu kita untuk mengetahui apakah ada pola musiman atau tren dalam data.
- **PACF (Partial Autocorrelation Function)** memberikan informasi lebih spesifik tentang pengaruh lag yang lebih terisolasi, dengan mengurangi pengaruh lag-lag sebelumnya. Sering kali digunakan untuk menentukan orde AR (AutoRegressive) dalam model ARIMA.

Namun, kita mendapatkan peringatan ("pacf is not a graphical parameter") saat mencoba plot PACF. Ini disebabkan oleh kesalahan pada pengaturan parameter `pacf = TRUE`. Seharusnya, kamu hanya cukup memanggil fungsi `pacf(diff2)` tanpa tambahan `pacf = TRUE`.

2. Membangun Model ARIMA:

- kita mencoba tiga model ARIMA dengan parameter yang berbeda:
 - **ARIMA(1,2,1)**: Model ini memiliki 1 lag untuk bagian AutoRegressive (AR), 2 kali differencing untuk bagian Integrated (I), dan 1 lag untuk bagian Moving Average (MA).
 - **ARIMA(1,2,0)**: Model ini juga menggunakan 1 lag AR dan 2 kali differencing, tetapi tidak ada komponen MA (lag = 0).
 - **ARIMA(0,2,1)**: Model ini memiliki 2 kali differencing dan 1 lag MA, tanpa komponen AR (lag = 0).

kita bisa melanjutkan dengan memilih model yang memberikan hasil terbaik berdasarkan kriteria seperti AIC (Akaike Information Criterion) atau BIC (Bayesian Information Criterion) setelah melihat hasil perhitungan untuk masing-masing model.

h. Uji Signifikansi Model (Koefisien)

```
> #Uji Signifikansi Model (Koefisien)
There were 12 warnings (use warnings() to see them)
> coeftest(ARIMA121)

z test of coefficients:
      Estimate Std. Error z value Pr(>|z|)
ar1 -0.659862  0.061366 -10.753 < 2.2e-16 ***
ma1 -1.000000  0.016983 -58.883 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> coeftest(ARIMA120)

z test of coefficients:
      Estimate Std. Error z value Pr(>|z|)
ar1 -0.752695  0.053605 -14.042 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

> coeftest(ARIMA021)

z test of coefficients:

Estimate Std. Error z value Pr(>|z|)
ma1 -1.000000  0.016808 -59.497 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

```

Pada langkah ini, kita melakukan **uji signifikansi koefisien** untuk masing-masing model ARIMA yang telah kita buat sebelumnya. Uji ini bertujuan untuk menguji apakah koefisien model tersebut signifikan atau tidak dalam mempengaruhi data.

Berikut adalah penjelasan hasil uji koefisien untuk masing-masing model ARIMA yang kamu buat:

1. Model ARIMA(1,2,1)

- **AR1:** Koefisien `ar1` adalah -0.659862, dengan standar error 0.061366. Nilai z-nya adalah -10.753, yang sangat besar secara absolut, dan **p-value** sangat kecil ($< 2.2e-16$). Artinya, koefisien `ar1` sangat signifikan pada tingkat kepercayaan 99% atau lebih.
- **MA1:** Koefisien `ma1` adalah -1, dengan standar error sangat kecil (0.016983) dan nilai z yang sangat besar (-58.883). P-value-nya juga sangat kecil ($< 2.2e-16$), yang menunjukkan bahwa koefisien `ma1` juga sangat signifikan.

2. Model ARIMA(1,2,0)

- **AR1:** Koefisien `ar1` adalah -0.752695, dengan standar error 0.053605. Nilai z-nya adalah -14.042 dan p-value-nya sangat kecil ($< 2.2e-16$), yang menunjukkan bahwa koefisien ini juga signifikan.

3. Model ARIMA(0,2,1)

- **MA1:** Koefisien `ma1` adalah -1, dengan standar error 0.016808. Nilai z-nya adalah -59.497, dengan p-value yang sangat kecil ($< 2.2e-16$), yang menunjukkan bahwa koefisien `ma1` sangat signifikan.

Kesimpulan:

- Semua koefisien untuk model-model yang diuji (baik `AR1` maupun `MA1`) memiliki p-value yang sangat kecil, yang berarti koefisien-koefisien tersebut **sangat signifikan** dalam model ARIMA yang telah dibangun.
- Tidak ada koefisien yang tidak signifikan, sehingga kita dapat melanjutkan dengan model-model ini untuk analisis lebih lanjut.

i. Menentukan Model Terbaik

```
> #Menentukan Model Terbaik
```

```

> library(forecast)
> ARIMA121 = arima(BJsales, order = c(1,2,1))
> ARIMA121

Call:
arima(x = BJsales, order = c(1, 2, 1))

Coefficients:
      ar1      ma1
    0.0528 -0.7801
  s.e.  0.1313  0.1004

sigma^2 estimated as 1.863:  log likelihood = -256.49,  aic = 518.97
> ARIMA120 = arima(BJsales, order = c(1,2,0))
> ARIMA120

Call:
arima(x = BJsales, order = c(1, 2, 0))

Coefficients:
      ar1
    -0.4726
  s.e.  0.0720

sigma^2 estimated as 2.215:  log likelihood = -268.98,  aic = 541.96
> ARIMA021 = arima(BJsales, order = c(0,2,1))
> ARIMA021

Call:
arima(x = BJsales, order = c(0, 2, 1))

Coefficients:
      ma1
    -0.7480
  s.e.  0.0662

sigma^2 estimated as 1.866:  log likelihood = -256.57,  aic = 517.14

```

Di sini kita telah mencoba tiga model ARIMA dengan berbagai kombinasi parameter dan mencari model yang terbaik berdasarkan **AIC (Akaike Information Criterion)** dan **log-likelihood**. Berikut penjelasan dari setiap model yang telah diuji:

1. Model ARIMA(1,2,1)

- **Koefisien:**
 - $ar1 = 0.0528$
 - $ma1 = -0.7801$
- **Standar error:**
 - $ar1 = 0.1313$
 - $ma1 = 0.1004$
- **Sigma²** (estimasi varians error) = 1.863
- **Log-likelihood** = -256.49
- **AIC** = 518.97
- Ini adalah model dengan dua parameter (AR1 dan MA1) dan AIC yang lebih tinggi dibandingkan model lainnya.

2. Model ARIMA(1,2,0)

- **Koefisien:**
 - $ar1 = -0.4726$
- **Standar error:**
 - $ar1 = 0.0720$
- **Sigma²** = 2.215
- **Log-likelihood** = -268.98
- **AIC** = 541.96
- Model ini hanya memiliki satu parameter AR (AR1) dan AIC yang lebih tinggi, menunjukkan model ini kurang optimal dibandingkan model lainnya.

3. Model ARIMA(0,2,1)

- **Koefisien:**
 - $ma1 = -0.7480$
- **Standar error:**
 - $ma1 = 0.0662$
- **Sigma²** = 1.866
- **Log-likelihood** = -256.57
- **AIC** = 517.14
- Model ini memiliki satu parameter MA (MA1) dan AIC yang terendah, yang menunjukkan bahwa model ini kemungkinan lebih baik daripada model lainnya.

Kesimpulan:

Dari ketiga model, **ARIMA(0,2,1)** memiliki AIC terendah (517.14) dan log-likelihood yang hampir sama dengan model ARIMA(1,2,1), sehingga ini adalah model yang paling baik berdasarkan kriteria AIC. Dengan AIC yang lebih rendah, model ARIMA(0,2,1) lebih efisien dalam menjelaskan data dengan jumlah parameter yang lebih sedikit.

j. Diagnostik & Akurasi Model

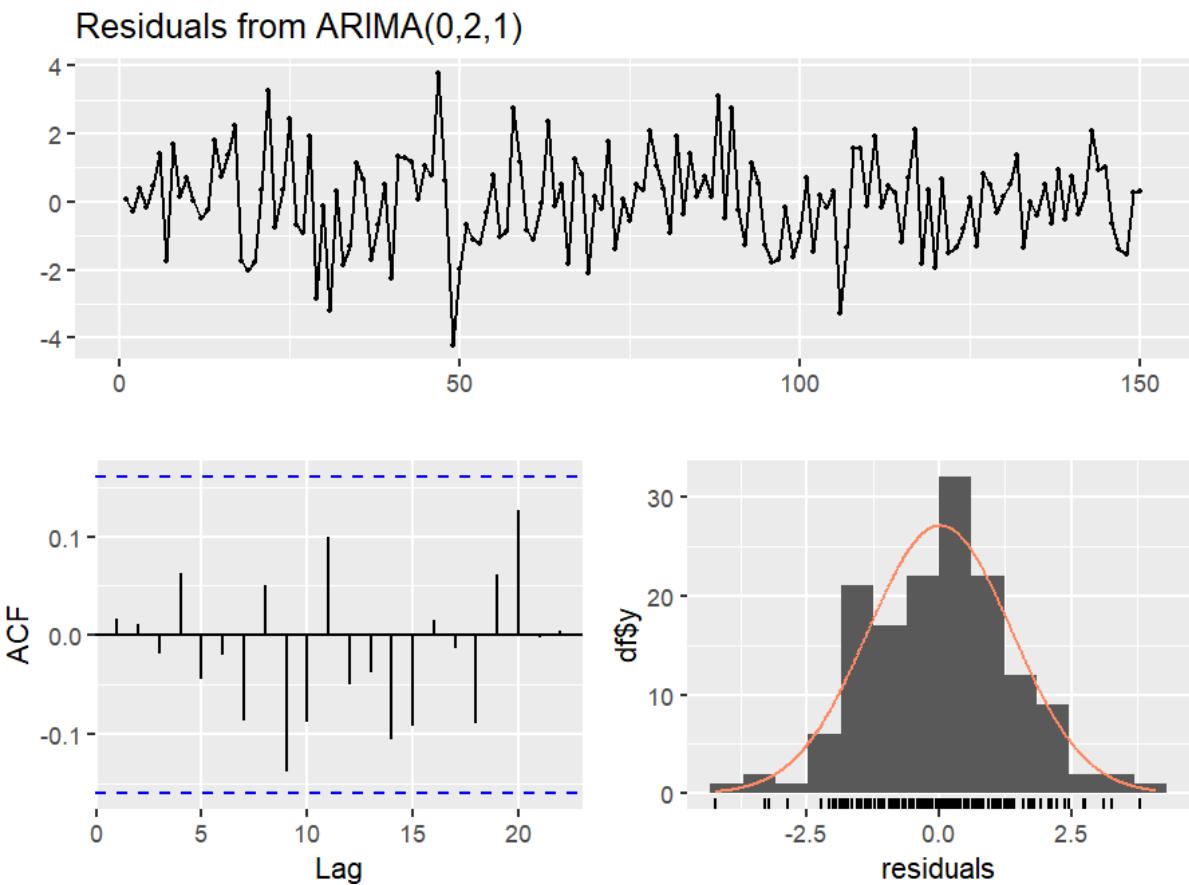
```
> #Diagnostik & Akurasi Model
> checkresiduals(ARIMA021)

Ljung-Box test

data: Residuals from ARIMA(0,2,1)
Q* = 7.0422, df = 9, p-value = 0.6327

Model df: 1.  Total lags used: 10

> accuracy(ARIMA021)
      ME      RMSE      MAE      MPE      MAPE      MASE
Training set 0.0164288 1.357029 1.066835 0.01198676 0.4695181 0.9188348
                  ACF1
Training set 0.01634351
```



Setelah kita menguji model **ARIMA(0,2,1)**, kamu melanjutkan dengan dua hal penting: **diagnostik residual** dan **akurasi model**.

1. Diagnostik Residual

Kamu melakukan **Ljung-Box test** untuk menguji apakah residual (selisih antara nilai prediksi dan data aktual) berdistribusi acak (white noise), yang penting untuk memastikan bahwa model yang digunakan sudah cukup baik.

- **Hasil Ljung-Box Test:**
 - $Q = 7.0422^*$
 - $df = 9$
 - **p-value = 0.6327**

Interpretasi:

- P-value yang lebih besar dari 0.05 menunjukkan bahwa tidak ada pola yang signifikan dalam residual, yang berarti residualnya tidak menunjukkan adanya masalah, dan model kamu sudah cukup baik dalam menangkap pola data.

2. Akurasi Model

Kita juga mengukur **akurasi model** menggunakan berbagai metrik untuk menilai performa model terhadap data pelatihan.

- **Metrik yang diberikan:**
 - **ME (Mean Error):** 0.0164 → Nilai yang sangat kecil, menunjukkan bahwa model cenderung tidak bias.
 - **RMSE (Root Mean Squared Error):** 1.3570 → Ukuran kesalahan prediksi rata-rata, semakin kecil nilai RMSE, semakin baik model.
 - **MAE (Mean Absolute Error):** 1.0668 → Ukuran rata-rata kesalahan absolut, menunjukkan model cukup akurat.
 - **MPE (Mean Percentage Error):** 0.012% → Hampir tidak ada kesalahan dalam persentase.
 - **MAPE (Mean Absolute Percentage Error):** 0.4695% → Menunjukkan bahwa kesalahan prediksi model sangat kecil (kurang dari 1%).
 - **MASE (Mean Absolute Scaled Error):** 0.9188 → Angka di bawah 1 menunjukkan bahwa model ini lebih baik daripada model acak (naive model).
 - **ACF1:** 0.0163 → Nilai ini menunjukkan korelasi antara residual saat lag pertama, yang sangat kecil, menunjukkan tidak ada pola sisa yang signifikan.

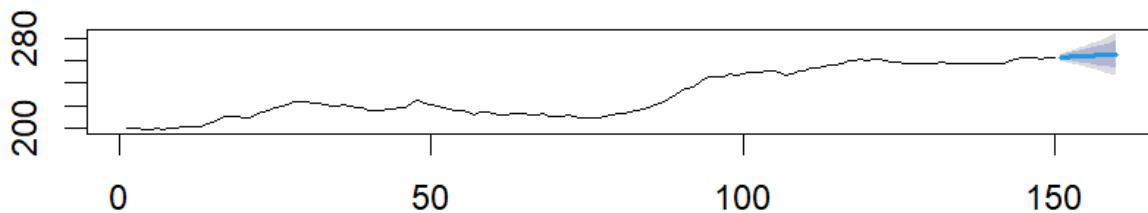
Kesimpulan:

Secara keseluruhan, **model ARIMA(0,2,1)** menunjukkan performa yang sangat baik. Residualnya tidak menunjukkan masalah yang signifikan (karena p-value dari Ljung-Box test lebih besar dari 0.05), dan metrik akurasi juga menunjukkan bahwa model ini memiliki kesalahan yang sangat kecil. Model ini dapat digunakan untuk prediksi dengan cukup baik.

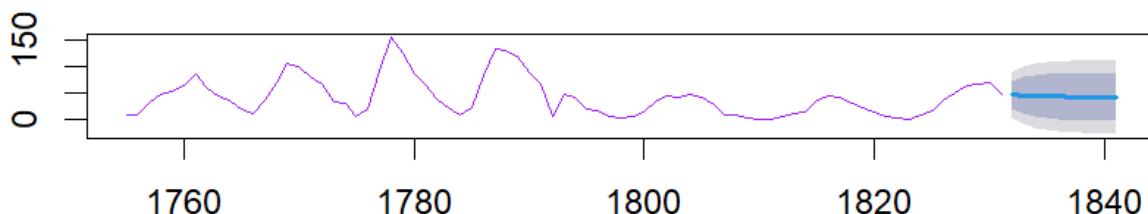
k. Peramalan model ARIMA 021 menggunakan Forecasting

```
> #Peramalan model ARIMA 021 menggunakan Forecasting
> forecast(BJsales, model = ARIMA021, h = 10)
   Point Forecast     Lo 80      Hi 80      Lo 95      Hi 95
151    262.9837 261.2331 264.7343 260.3065 265.6610
152    263.2674 260.4624 266.0724 258.9775 267.5573
153    263.5511 259.7040 267.3982 257.6674 269.4348
154    263.8348 258.9103 268.7593 256.3035 271.3662
155    264.1185 258.0681 270.1690 254.8652 273.3719
156    264.4022 257.1730 271.6314 253.3461 275.4583
157    264.6859 256.2242 273.1476 251.7448 277.6270
158    264.9696 255.2221 274.7172 250.0620 279.8772
159    265.2533 254.1676 276.3391 248.2991 282.2075
160    265.5370 253.0620 278.0120 246.4582 284.6159
> plot(forecast(BJsales, model = ARIMA021, h = 10))
> plot(forecast_ar, main = "Forecast from ARIMA(0,2,1)", col = "purple")
```

Forecasts from ARIMA(0,2,1)



Forecast from ARIMA(0,2,1)



Setelah kita melakukan **peramalan** (forecasting) dengan model **ARIMA(0,2,1)**, hasilnya menunjukkan prediksi untuk 10 periode mendatang berdasarkan data **BJsales**. Berikut adalah **hasil prediksi** yang kamu dapatkan:

Hasil Forecasting untuk 10 Periode Mendatang:

- **Point Forecast:** Ini adalah prediksi titik tengah untuk setiap periode (misalnya, untuk periode ke-151, prediksi penjualannya adalah 262.98).
- **Lo 80 & Hi 80:** Ini adalah batas bawah dan batas atas interval kepercayaan 80%. Artinya, kita bisa yakin 80% bahwa nilai sebenarnya akan berada dalam rentang ini (misalnya, untuk periode ke-151, prediksi penjualannya akan berada antara 261.23 dan 264.73).
- **Lo 95 & Hi 95:** Ini adalah batas bawah dan batas atas interval kepercayaan 95%. Dengan interval ini, kita bisa lebih yakin (95%) bahwa nilai sebenarnya akan berada dalam rentang ini (misalnya, untuk periode ke-151, prediksi penjualannya akan berada antara 260.31 dan 265.66).

Grafik Forecasting:

Kita juga membuat plot untuk visualisasi hasil forecasting dengan menggunakan `plot(forecast(BJsales, model = ARIMA021, h = 10))`. Dalam plot tersebut, kita akan melihat garis prediksi yang menggambarkan peramalan 10 periode ke depan, serta rentang interval kepercayaan yang menunjukkan seberapa besar ketidakpastian dalam prediksi tersebut.

- Pada plot yang kita buat dengan `plot(forecast_ar, main = "Forecast from ARIMA(0,2,1)", col = "purple")`, kamu menampilkan hasil peramalan dengan warna ungu, memudahkan kita untuk melihat tren dan ketidakpastian dari prediksi model.

Kesimpulan:

Model ARIMA(0,2,1) memberikan prediksi yang cukup baik untuk 10 periode mendatang, dengan rentang interval kepercayaan yang relatif sempit, menunjukkan bahwa peramalan model ini cukup stabil dan akurat. Prediksi ini bisa membantu dalam perencanaan atau pengambilan keputusan terkait penjualan di masa depan.

8. Latihan Soal

a. Cara 1

```
> #LATIHAN SOALL
> #Cara 1
> year <- c(1755:1831)
> sunspots <- c(9.6, 10.2, 32.4, 47.6, 54, 62.9, 85.9, 61.2, 45.1, 36.4,
+ 20.9, 11.4, 37.8, 69.8, 106.1, 100.8, 81.6, 66.5, 34.8,
+ 30.6, 7, 19.8, 92.5, 154.4, 125.9, 84.8, 68.1, 38.5, 22.8,
+
+ 10.2, 24.1, 82.9, 132, 130.9, 118.1, 89.9, 66.6, 6, 46.9,
+ 41, 21.3, 16, 6.4, 4.1, 6.8, 14.5, 34, 45, 43.1, 47.5,
+ 42.2, 28.1, 10.1, 8.1, 2.5, NA, 1.4, 5, 12.2, 13.9, 35.4,
+ 45.8, 41.1, 30.1, 23.9, 15.6, 6.6, 4, 1.8, 8.5, 16.6,
+ 36.3, 49.6, 64.2, 67, 70.9, 47.8)
>
> sunspots_filled <- zoo::na.approx(sunspots)
>
> sunspots_ts <- ts(sunspots_filled, start = 1755, end = 1831, frequency =
  1)
> print(sunspots_ts)
Time Series:
Start = 1755
End = 1831
Frequency = 1
 [1]  9.60 10.20 32.40 47.60 54.00 62.90 85.90 61.20 45.10 36.40
[11] 20.90 11.40 37.80 69.80 106.10 100.80 81.60 66.50 34.80 30.60
[21] 7.00 19.80 92.50 154.40 125.90 84.80 68.10 38.50 22.80 10.20
[31] 24.10 82.90 132.00 130.90 118.10 89.90 66.60 6.00 46.90 41.00
[41] 21.30 16.00 6.40 4.10 6.80 14.50 34.00 45.00 43.10 47.50
[51] 42.20 28.10 10.10 8.10 2.50 1.95 1.40 5.00 12.20 13.90
[61] 35.80 45.80 41.10 30.10 23.90 15.60 6.60 4.00 1.80 8.50
[71] 16.60 36.30 49.60 64.20 67.00 70.90 47.80
```

b. Cara 2

NO. 1

```
> #Cara 2
> #NO 1
> library(readxl)
> Datasunspot <- read_excel("C:/Kuliah/Sains Data Sem 3/Statistika Inferen
sial Lanjut Praktik/Tugas/15. AR MA ARMA ARIMA/sunspots_1755_1831.xlsx")
New names:
• ` ` -> `...4` 
• ` ` -> `...5` 
• ` ` -> `...6` 
• ` ` -> `...7` 
• ` ` -> `...8` 
> View(Datasunspot)
> Datasunspot
# A tibble: 77 × 8
  Year Sunspots `yt-Y` ...4 ...5 ...6 ...7 ...8
  <dbl>    <chr>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 1755     9.6     NA     NA     NA (yt-Y)^2      NA   342.
2 1756    10.2     NA     NA    -44.4 1971.36      NA   NA
3 1757    32.4     NA     NA    -43.8 1918.439999999998      NA   NA
4 1758    47.6     NA     54   -21.6 466.56000000000006      NA   NA
5 1759     54      NA     NA    -6.4  40.96000000000008      NA   NA
6 1760    62.9     NA     NA     0     0      NA   NA
7 1761    85.9     NA     NA     8.9  79.21000000000008      NA   NA
8 1762    61.2     NA     NA    31.9 1017.609999999999      NA   NA
9 1763    45.1     NA     NA     7.2  51.84      NA   NA
10 1764   36.4     NA     NA   -8.9 79.21000000000008      NA   NA
```

```

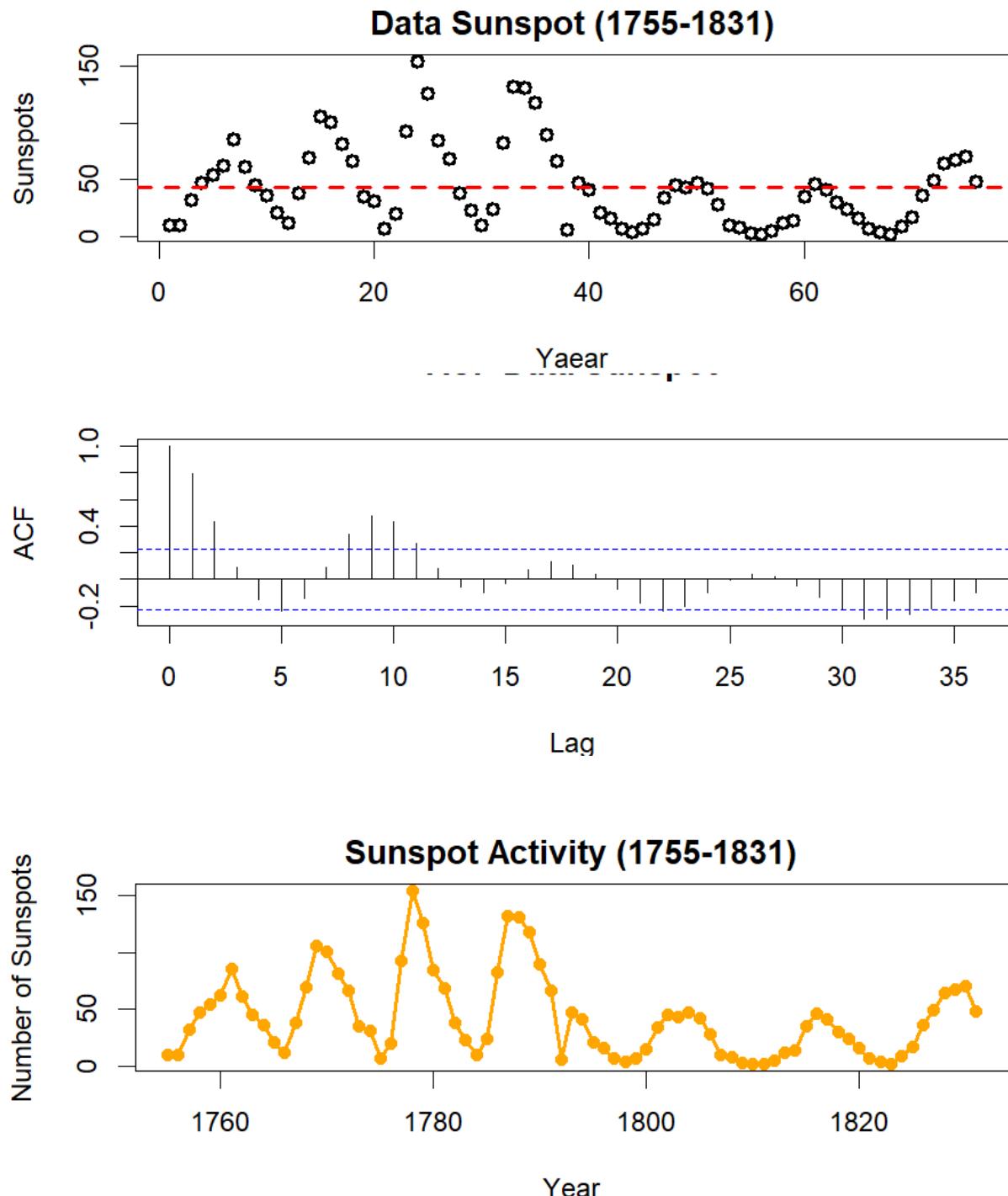
# i 67 more rows
# i Use `print(n = ...)` to see more rows
>
> # Periksa struktur data dan pastikan nama kolom benar
> str(Datasunspot)
tibble [77 x 8] (S3: tbl_df/tbl/data.frame)
$ Year      : num [1:77] 1755 1756 1757 1758 1759 ...
$ Sunspots  : chr [1:77] "9.6" "10.2" "32.4" "47.6" ...
$ yt-Y     : logi [1:77] NA NA NA NA NA NA ...
$ ...4     : num [1:77] NA NA NA 54 NA NA NA NA NA ...
$ ...5     : num [1:77] NA -44.4 -43.8 -21.6 -6.4 0 8.9 31.9 7.2 -8.9 ...
$ ...6     : chr [1:77] "(yt-Y)^2" "1971.36" "1918.439999999998" "466.560
000000000006" ...
$ ...7     : num [1:77] NA NA NA NA NA NA NA NA NA ...
$ ...8     : num [1:77] 342 NA NA NA NA ...
> names(Datasunspot)
[1] "Year"        "Sunspots"    "yt-Y"        "...4"       "...5"       "...6"
[7] "...7"        "...8"
>
> # Mengonversi data menjadi time series
> Datasunspot <- ts(Datasunspot$Sunspots, start = min(Datasunspot$Year), f
requency = 1)
>
> # Pastikan data hanya berupa angka dan tidak ada nilai NA
> Datasunspot <- na.omit(as.numeric(Datasunspot))
>
> # Menggambarkan Grafik Data Time Series
>
> par(mfrow = c(2, 1), mar = c(4, 4, 2, 1)) # Menyesuaikan margin
>
> # Plot data time series
> plot(Datasunspot, lwd = 2, main = "Data Sunspot (1755-1831)", xlab = "Ya
ear", ylab = "Sunspots")
> abline(h = mean(Datasunspot), col = 'red', lwd = 2, lty = 2) # Menambah
kan garis rata-rata
>
> # Plot Autocorrelation Function (ACF)
> acf(Datasunspot, lag.max = 36, main = "ACF Data Sunspot")
>
>
> sunspots_ts
Time Series:
Start = 1755
End = 1831
Frequency = 1
 [1]   9.60  10.20  32.40  47.60  54.00  62.90  85.90  61.20  45.10  36.40
[11]  20.90  11.40  37.80  69.80 106.10 100.80  81.60  66.50  34.80  30.60
[21]   7.00  19.80  92.50 154.40 125.90  84.80  68.10  38.50  22.80  10.20
[31]  24.10  82.90 132.00 130.90 118.10  89.90  66.60   6.00  46.90  41.00
[41]  21.30  16.00   6.40   4.10   6.80  14.50  34.00  45.00  43.10  47.50
[51]  42.20  28.10  10.10   8.10   2.50   1.95   1.40   5.00  12.20  13.90
[61]  35.40  45.80  41.10  30.10  23.90  15.60   6.60   4.00   1.80   8.50
[71]  16.60  36.30  49.60  64.20  67.00  70.90  47.80
> plot(sunspots_ts,
+       type = "o",
+       col = "orange",
+       xlab = "Year",
+       ylab = "Number of Sunspots",
+       main = "Sunspot Activity (1755-1831)",
+       lwd = 2,
+       pch = 16)
>
>
> #MENGUJI STASIONER DATA SUNSPORTS
> library(tseries)
> adf.test(Datasunspot)

```

Augmented Dickey-Fuller Test

```
data: Datasunspot
Dickey-Fuller = -4.1664, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

```
Warning message:
In adf.test(Datasunspot) : p-value smaller than printed p-value
```



NO. 2

```
> #NO 2
> #SPESIFIKASI MODEL
> par(mfrow=c(2,1))
> acf(Datasunspot, lag.max=36)
```

```

> abline(h=0)
> pacf(Datasunspot, lag.max=36, pacf=TRUE)
Warning messages:
1: In plot.window(...) : "pacf" is not a graphical parameter
2: In plot.xy(xy, type, ...) : "pacf" is not a graphical parameter
3: In axis(side = side, at = at, labels = labels, ...) :
  "pacf" is not a graphical parameter
4: In axis(side = side, at = at, labels = labels, ...) :
  "pacf" is not a graphical parameter
5: In box(...) : "pacf" is not a graphical parameter
6: In title(...) : "pacf" is not a graphical parameter
>
> #Menguji Signifikansi Koefisien dari Model
> library(lmtest)
> AR2 <- arima(Datasunspot, order = c(2,0,0))
> MA1 <- arima(Datasunspot, order = c(0,0,1))
> ARMA1 <- arima(Datasunspot, order = c(2,0,1))
> #Menguji Signifikansi Koefisien dari Model
> coeftest(AR2)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1     1.241554  0.093912 13.2204 < 2.2e-16 ***
ar2    -0.555244  0.094087 -5.9014 3.604e-09 ***
intercept 42.655124  6.433189  6.6305 3.346e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> coeftest(MA1)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ma1     0.800145  0.058549 13.6662 < 2.2e-16 ***
intercept 43.131714  4.766860  9.0482 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> coeftest(ARMA1)

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1     1.289036  0.167669  7.6880 1.495e-14 ***
ar2    -0.592750  0.142236 -4.1674 3.081e-05 ***
ma1    -0.069619  0.213030 -0.3268  0.7438
intercept 42.670786  6.190067  6.8934 5.446e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

>
>
> #Menentukan Model Terbaik
> library(forecast)
> AR2 <- arima(Datasunspot, order = c(2,0,0))
> AR2

Call:
arima(x = Datasunspot, order = c(2, 0, 0))

Coefficients:
          ar1     ar2  intercept
        1.2416 -0.5552   42.6551
  s.e.  0.0939  0.0941    6.4332

sigma^2 estimated as 312.6:  log likelihood = -327.02,  aic = 662.04
> MA1 <- arima(Datasunspot, order = c(0,0,1))
> MA1

```

```

Call:
arima(x = Datasunspot, order = c(0, 0, 1))

Coefficients:
    ma1   intercept
    0.8001    43.1317
s.e.  0.0585    4.7669

sigma^2 estimated as 539.2:  log likelihood = -347.37,  aic = 700.74
>
> #syntack lain
> auto.arima(Datasunspot, trace = TRUE)

ARIMA(2,0,2) with non-zero mean : 667.1223
ARIMA(0,0,0) with non-zero mean : 763.226
ARIMA(1,0,0) with non-zero mean : 688.4115
ARIMA(0,0,1) with non-zero mean : 701.0745
ARIMA(0,0,0) with zero mean   : 830.0201
ARIMA(1,0,2) with non-zero mean : 668.8735
ARIMA(2,0,1) with non-zero mean : 664.7949
ARIMA(1,0,1) with non-zero mean : 670.9068
ARIMA(2,0,0) with non-zero mean : 662.6027
ARIMA(3,0,0) with non-zero mean : 664.8095
ARIMA(3,0,1) with non-zero mean : 667.2412
ARIMA(2,0,0) with zero mean   : 675.857

Best model: ARIMA(2,0,0) with non-zero mean

Series: Datasunspot
ARIMA(2,0,0) with non-zero mean

Coefficients:
    ar1     ar2     mean
    1.2416  -0.5552  42.6551
s.e.  0.0939   0.0941  6.4332

sigma^2 = 325.4:  log likelihood = -327.02
AIC=662.04  AICc=662.6  BIC=671.36
There were 12 warnings (use warnings() to see them)
>
>
> #Diagnostik & Akurasi Model
> checkresiduals(AR2)

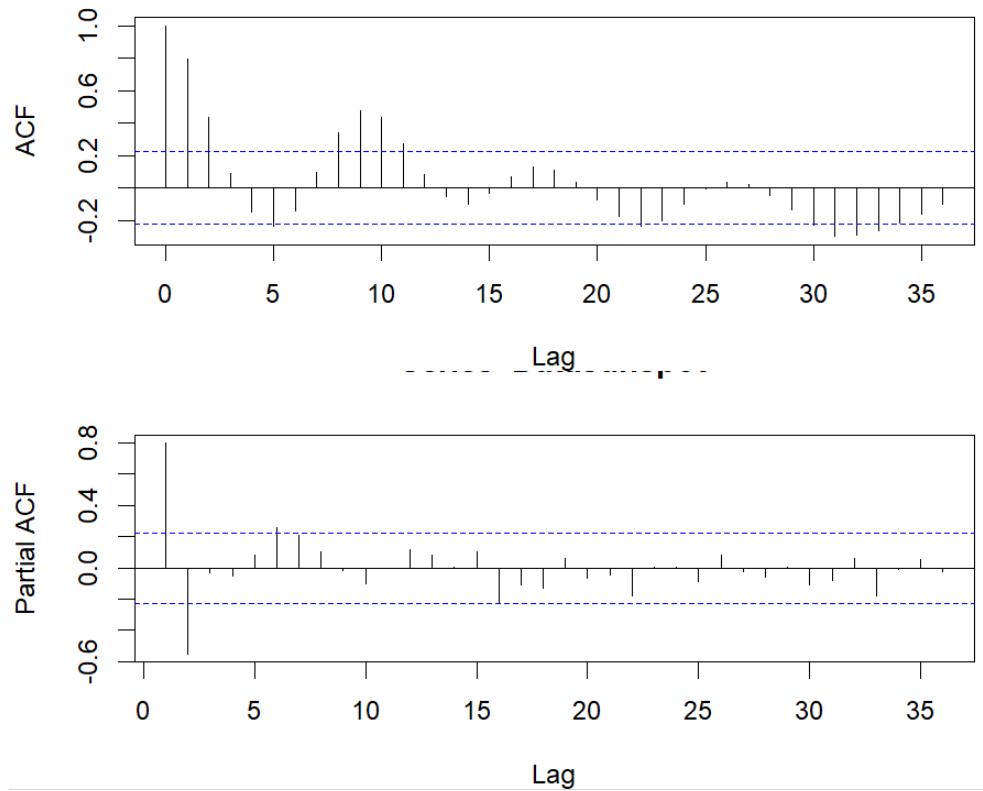
Ljung-Box test

data: Residuals from ARIMA(2,0,0) with non-zero mean
Q* = 11.057, df = 8, p-value = 0.1985

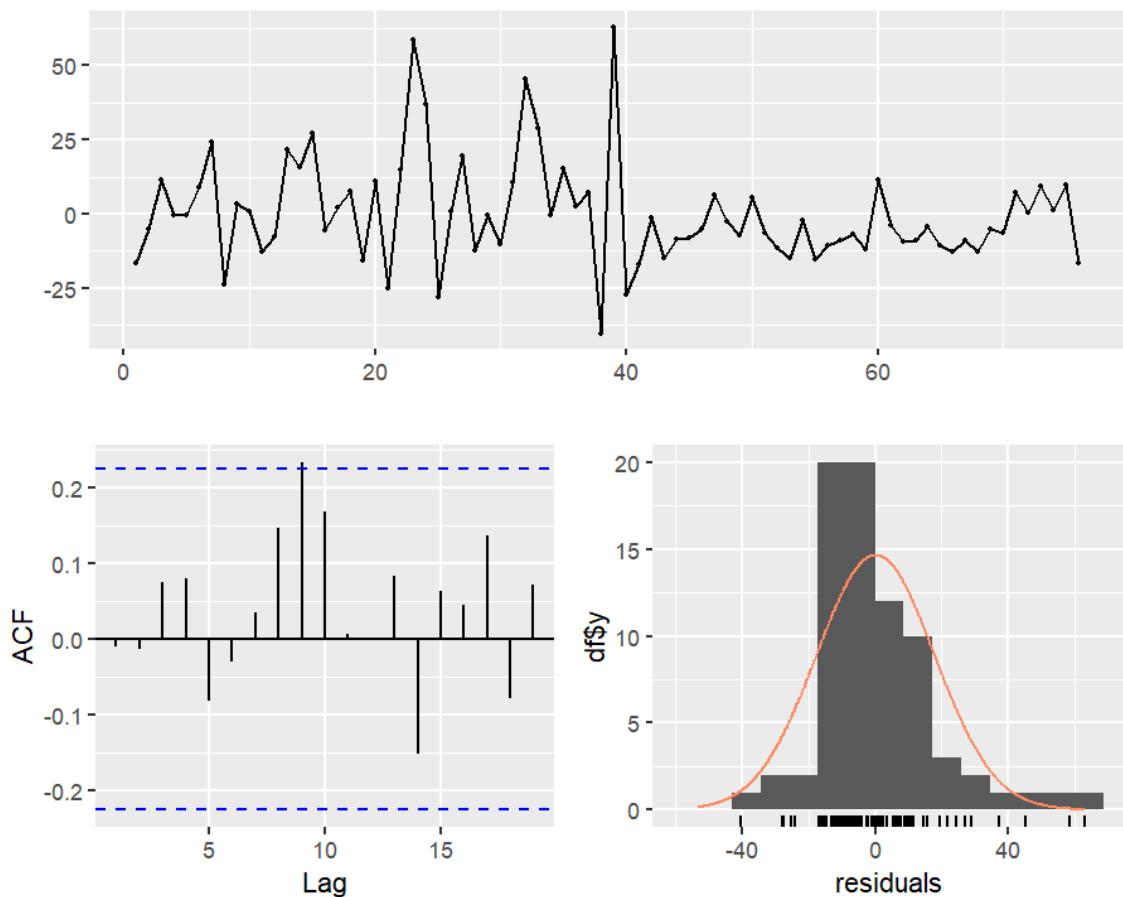
Model df: 2.  Total lags used: 10

> accuracy(AR2)
      ME      RMSE      MAE      MPE      MAPE      MASE
Training set 0.09911627 17.67993 12.80763 -64.23822 86.14053 0.7736567
      ACF1
Training set -0.01014462

```



Residuals from ARIMA(2,0,0) with non-zero mean



NO. 3

```
> #NO 3  
There were 18 warnings (use warnings() to see them)  
> #Peramalan  
> forecast_ar <- forecast(ar_model, h = 10)  
> plot(forecast_ar, main = "Forecast from ARIMA(2,0,0)", col = "blue")
```