BAHAN AJAR MATA KULIAH STATISTIKA

PENYUSUN

Retno Tri Vulandari, S.Si, M.Si **0613038801**

Dibiayai oleh :

Direktorat Riset dan Pengabdian Masyarakat

Direktorat Jenderal Penguatan Riset dan Pengembangan

Kementerian Riset, Teknologi, dan Pendidikan Tinggi

Sesuai dengan Kontrak Penelitian Tahun Anggaran 2018

Nomor : 059 / K6 / KM / SP2H / PENELITIAN / 2018

SEKOLAH TINGGI MANAJEMEN INFOMATIKA DAN KOMPUTER

(STMIK) SINAR NUSANTARA

SURAKARTA

2018

SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
STMIK SINAR NUSANTARA
Jl. KH. Samanhudi 84-86 Surakarta 57142 Telp/Fax 0271-716500
Http://www.sinus.ac.id E-mail: sekretariat@sinus.ac.id

Kode Dok. : F-PRG-002.05
Revisi : 0

LEMBAR PENGESAHAN
BAHAN AJAR

MATA KULIAH	: STATISTIKA
SKS	: 3
PROGRAM STUDI	: TEKNIK INFORMATIKA
JENJANG PENDIDIKAN	: DIPLOMA TIGA

Surakarta, 10 Juli 2018

Penyusun,	Mengesahkan,
	Ketua Prodi D3 Teknik Informatika
Retno Tri Vulandari, S.Si., M.Si.	Dwi Remawati, S.Kom., M.Kom.
NIK 110 000 093	NIP. 197303062005012002

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT, yang telah memberikan rahmatNya sehingga bab kuliah **Statistika** ini dapat terselesaikan dengan baik. Bab ini disusun sebagai media pembelajaran sehingga mahasiswa mampu memahami materi dengan lebih mudah dan dapat dimanfaatkan sebagai penunjang kelancaran pembelajaran. Bab ini terdiri dari 8 bab yang dimaksudkan untuk dua belas pertemuan.

Bahan Ajar ini berisikan mengenai ringkasan-ringkasan dari mulai materi pengenalan statistika secara umum, pengetahuan mengenai data dan variabel, proses pemeriksaan data, statistika deskripsi, dan statistika inferensial. Selain materi, bahan ajar ini juga memberikan contoh soal dan soal latihan sebagai bahan latihan sehingga dapat membantu mahasiswa menguasai materi.

Akhir kata penulis mengharapkan koreksi dan saran yang membangun dari semua pihak demi kesempurnaan bahan ajar ini. Penulis berharap semoga bahan ajar **Statistika** ini dapat berguna dan bermanfaat bagi para pembaca.

.

Surakarta, Juli 2018

Penulis

TINJAUAN UMUM

Sekarang ini, statistika adalah ilmu yang memiliki peran penting di segala bidang. Statistika adalah ilmu yang memperlajari bagaimana merencanakan, mengumpulkan, menginterpretasi, dan mempresentasikan data. Sering rancu mengenai pengertian antara statistika dan statistik, statistika adalah ilmu yang berhubungan dengan data sedangkan statistik adalah data, informasi, atau hasil penerapan algoritma statistika pada suatu data. Berdasarkan kumpulan data tersebut, statistika dapat digunakan untuk menyimpulkan disebut statistika inferensial atau untuk mendeskripsikan data disebut statistika deskriptif. Statistika banyak diterapkan dalam berbagai disiplin ilmu, baik ilmu maupun ilmu sosial. Statistika juga digunakan dalam pemerintahan untuk berbagai macam tujuan misal sensus penduduk. Penerapan statistika lainnya adalah prosedur jajak pendapat atau *polling* serta hitung cepat atau *quick count*. Di bidang komputasi, statistika dapat pula diterapkan dalam pengenalan pola maupun kecerdasan buatan.

Dengan adanya bab ini diharapkan mahasiswa dapat menjelaskan konsep statistika deskriptif dan statistika inferensial serta mampu mengaplikasikannya untuk kepentingan pengolahan, analisis data, dan pengujian hipotesis dalam bidang penelitian.

Pada akhir pembelajaran mata kuliah Statistika, diharapkan mahasiswa dapat menjelaskan konsep dasar, tujuan dan kegunaan statistika. Mahasiswa dapat menjelaskan konsep statistik deskriptif dan statistik inferensial. Mahasiswa dapat menyebutkan jenis-jenis skala pengukuran dan pemanfaatannya. Mahasiswa dapat menyusun distribusi frekuensi data kualitatif dan data kuantitatif. Mahasiswa dapat menggambarkan grafik frekuensi, frekuensi relatif, dan frekuensi kumulatif. Mahasiswa dapat menjelaskan manfaat dari berbagai ukuran pemusatan dan ukuran sebaran. Mahasiswa dapat memahami distribusi data dan normalitas distribusi. Mahasiswa dapat melakukan uji persyaratan statistik. Mahasiswa dapat melakukan uji perbedaan dua rata-rata. Mahasiswa dapat menggunakan Software komputer untuk pengolahan data statistik.

BAB I
KONSEP DASAR STATISTIKA

Capaian Pembelajaran	: menjelaskan peran statistika dalam penelitian
Kemampuan Akhir Pembelajaran	: mahasiswa dapat 1. menjelaskan arti data, statistik, dan statistika 2. menentukan jenis variabel dalam penelitian dan memberikan contoh 3. menentukan jenis skala pengukuran dari suatu variabel 4. menjelaskan syarat-syarat data yang baik 5. membedakan data menurut jenisnya 6. menggambarkan proses dan metode yang digunakan dalam pengumpulan dan pengolahan data

A. PENDAHULUAN KONSEP DASAR STATISTIKA

Statistika adalah suatu disiplin ilmu yang penting pada dewasa ini, antara lain untuk memperbaiki teori-teori statistika yang sudah ada, ataupun member gambaran tentang hasil suatu penyelidikan/percobaan. Statistika berkaitan dengan pengumpulan informasi atau keterangan, penyajian dalam bentuk daftar, diagram, atau grafik sehingga memudahkan untuk dianalisa , yang selanjutnya disimpulkan dan diambil kesimpulan. Setiap informasi atau keterangan yang diperoleh disebut datum, dalam bentuk jamak adalah data.

B. MATEMATIKA, STATISTIK, DAN STATISTIKA

Matematika adalah ilmu yang hasil bersifat deterministik (jelas dan pasti). Statistik menyatakan kumpulan data, bilangan maupun non bilangan yang disusun dalam bentuk tabel atau diagram, yang menggambarkan suatu permasalahan. Statistika adalah ilmu pengetahuan yang berhubungan dengan cara-cara pengumpulan data, pengolahan atau analisa dan penarikan kesimpulan berdasarkan kumpulan data dan penganalisisan yang dilakukan. Misal 40% siswa nilai matematika kurang dari 6.5 maka nilai 40% ini dinamakan statistik. Nilai statistik bersifat probabilistik, berhubungan dengan peluang atau kemungkinan.

Jenis statistika ada 2, yaitu:

1. Statistika deskriptif adalah tahap statistika hanya berusaha melukiskan dan menganalisa kelompok data tanpa menarik kesimpulan.

   Contoh:

   - Diagram batang, diagram titik, diagram garis
   - Tabel distribusi frekuensi
   - Diagram batang Daun

2. Statistika inferensial atau induktif adalah tahap statistika yang berkaitan dengan kondisi suatu kesimpulan diambil.

   Contoh:

   - Analisis Korelasi
   - Analisis Regresi
   - Analisis Variansi

C. POPULASI DAN SAMPEL

Perhatikan kalimat-kalimat berikut ini :

a. Lima juta penduduk Indonesia mahir 3 bahasa asing.
b. Delapan puluh persen penduduk pulau Jawa menggunakan motor merk X.
c. Baterai XYZ tahan lebih lama.

Kalimat di atas menyangkut himpunan yang universal, yaitu semua penduduk Indonesia, penduduk pulau Jawa, dan semua baterai. Dalam statistika, himpunan universal (semesta) dengan karakteristik tertentu disebut **populasi.**

Untuk keperluan itu, kita dapat menggunakan atau mengambil contoh yang dipilih dari populasi, yang disebut **sampel.** Jadi, **sampel** adalah himpunan bagian dari populasi. Metode statistika tentang cara mengambil sampel yang tepat disebut teknik sampling. Nilai-nilai yang diperoleh dari sampel disebut **statistik.** Statistik inilah yang digunakan untuk men-duga populasi. Nilai-nilai populasi disebut **parameter.**

D. JENIS DATA

Berdasarkan **wujudnya** data dibagi menjadi 2, yaitu:

1. Data Kuantitatif, yaitu data berupa kumpulan angka. Ditinjau dari cara memperolehnya, data kuantitatif dapat dibedakan menjadi dua macam, yaitu
   a. Data Diskrit, adalah data yang diperoleh dengan cara mencacah, membilang, atau menghitung banyak objek. Contoh: jumlah mahasiswa, jumlah , dan sebagainya.
   b. Data Kontinu, adalah data yang diperoleh dengan cara mengukur besaran objek. Sebagai contoh data tentang luas petak sawah dan data tentang berat padi gabah kering.
2. Data Kualitatif, yaitu data yang diamati berdasarkan atribut, data yang bukan berupa angka (non-numerik) misalnya pendapat siswa terhadap pelajaran Matematika, seperti amat senang – senang – kurang senang – tidak senang.

Untuk keperluan perhitungan maupun analisis, sering dikehendaki data kuantitatif dalam bentuk yang lebih sederhana. Untuk menyederhanakan bilangan-bilangan, diadakan aturan pembulatan sebagai berikut :

a. Aturan umum, yaitu jika kurang dari 0,5 dihilangkan dan jika sama atau lebih dari 0,5 menjadi 1,

   Misal : 3,48 dibulatkan menjadi 3
   2,5 dibulatkan menjadi 3

b. Aturan genap terdekat, yaitu kurang dari 0,5 dihilagkan, lebih dari 0,5 menjadi 1, dan sama dengan 0,5 dihilangkan jika angka yang mendahului genap atau menjadi 1 jika angka yang mendahului ganjil, Misal :6,948 dibulatkan menjadi 6,9 (sampai satu tempat desimal)

   Sebelum data diolah lebih lanjut, perlu diadakan pemeriksaan data kembali. Hal ini untuk menghindari kekeliruan dalam analisa maupun kesimpulan yang diambil. Beberapa data yang dipandang meragukan hendaknya diyakini kebenarannya. Kemungkinan kesalahan terjadi pada alat ukur, kesalahan mengukur, kekeliruan mencatat, instruksi yang tidak jelas, atau kecerobohan dalam mengumpilkan data. Semua kesalahan itu perlu diperhatikan agar diperoleh data yang akurat.

Berdasarkan **cara memperolehnya**, data dibagi menjadi 2, yaitu:

1. Data primer, adalah data yang diperoleh langsung dari sumbernya, missal melalui wawancara, penyebaran kuesioner, pengukuran langsung.
2. Data sekunder, adalah data yang diambil atau disadur dari pihak lain, misal diambil dari Koran, jurnal, penelitian/publikasi pihak lain.

E. SKALA PENGUKURAN

Skala pengukuran : cara mengukur suatu variabel.

Terdapat 4 jenis skala pengukuran :

1. Skala Nominal : angka yang diberikan pada objek/variabel pengukuran hanya memiliki arti sebagai label saja (asal dapat dibedakan). Tidak memiliki tingkatan.
2. Skala Ordinal : angka yang diberikan pada objek/variabel pengukuran mengandung pengertian tingkatan.
3. Skala Interval : angka yang diberikan pada objek/variabel pengukuran mengandung sifat ordinal ditambah sifat jarak/ interval.
4. Skala Rasio : angka yang diberikan pada objek/variabel pengukuran mengandung sifat interval ditambah sifat yang mampu memberikan keterangan tentang nilai absolut variabel yang diukur. Artinya apabila menunjuk angka 0 (nol), maka berarti benar – benar nol, tidak ada, atau kosong.

F. SOAL

1. Berdasarkan tabel berikut, sebut dan jelaskan informasi yang akan Anda peroleh!

Nagari	Pendidikan (orang)						
	Belum sekolah	Tidak tamat SD	SD	SLTP	SLTA	D3	S1
Maninjau	271	462	691	1049	850	11	5
Bayur	345	588	881	1337	1083	14	6
II Koto	378	644	966	1466	1187	13	7
Koto Kaciak	298	507	760	1153	934	12	5
III Koto	387	660	990	1502	1217	15	7
Tanjung Sani	470	801	1200	1822	1477	18	8
Sungai Batang	325	555	832	1262	1023	12	6
Jumlah	2.474	4.217	6.320	9.591	7.771	95	44
Persentase (%)	8.11	13.82	20.71	31.43	25.47	0.31	0.14

Sumber: Diolah dari BPS Kabupaten Agam, (2005) dan Kec. Tanjung Raya dalam Angka (2005)

2. Berdasarkan tabel berikut, sebut dan jelaskan informasi yang akan Anda peroleh!

Bagian	Kerusakan	Kerugian	Total
Rumah Tangga	13,9	1,4	15,3
Sosial	3,9	0,2	4,1
Produksi	4,3	4,7	9,2
Infrastruktur	0,6	0,3	0,6
Total	22,8	6,3	29,1

3. Berdasarkan tabel berikut, sebut dan jelaskan informasi yang akan Anda peroleh!

Tabel. Data hasil pengamatan produksi kedelai (ton/ha) menurut jenis varietas, daerah panen, dan jenis tanah

Varietas Kedelai	Tangkiling		Kalampangan		Total
	Tanah Gambut	Tanah Berpasir	Tanah Gambut	Tanah Berpasir	
Wilis	62	55	50	56	223
Sindoro	66	58	55	62	241
Slamet	70	64	59	65	258
Galunggung	73	68	64	72	277
Orba	78	72	71	75	296
Total	349	317	299	330	1295

BAB II
PEMERIKSAAN DATA

Capaian Pembelajaran	menjelaskan peran statistika dalam penelitian
Kemampuan Akhir Pembelajaran	mahasiswa dapat menggambarkan proses dan metode yang digunakan dalam pengumpulan dan pengolahan data

A. PENDAHULUAN PEMERIKSAAN DATA

Dalam sebuah kuisoner akan diperiksa kevalidan untuk tiap-tiap pertanyaan sehingga data dapat digunakan untuk proses analisa selanjutnya. Sebelum data diolah lebih lanjut, perlu diadakan pemeriksaan data kembali. Hal ini untuk menghindari kekeliruan dalam analisa maupun kesimpulan yang diambil. Beberapa data yang dipandang meragukan hendaknya diyakini kebenarannya. Kemungkinan kesalahan terjadi pada alat ukur, kesalahan mengukur, kekeliruan mencatat, instruksi yang tidak jelas, atau kecerobohan dalam mengumpilkan data. Semua kesalahan itu perlu diperhatikan agar diperoleh data yang akurat.

B. VALIDITAS DATA

Usaha untuk memperoleh informasi yang objektif merupakan langkah yang penting dalam suatu penyelidikan (observasi). Hal ini berkaitan dengan tujuan penyelidikan itu sendiri. Sesuai dengan tujuan penyelidikan, maka pengumpulan data dapat dilakukan dengan metode :

1. **Pengamatan (observasi),** yaitu cara pengumpulan data dengnan mengamati secara langsung subjek yang diteliti.
2. **Penelusuran literatur**, yaitu cara pengumpulan data dengan menggunakan sebagian atau seluruh data yang telah ada dari peneliti sebelumnya. Penelusuran literature disebut juga pengamatan tidak langsung.
3. **Penggunaan kuesioner (angket),** yaitu cara pengumpulan data dengan menggunkan daftar pertanyaan (angket) atau daftar isian terhadap subjek yang teliti.

4. **Wawancara (*interview*),** yaitu cara pengumpulan data dengan langsung mengadakan Tanya jawab kepada subjek yan diteliti. Data yang diperoleh disebut **data mentah.**

Berdasarkan banyaknya data yang diambil, cara pengumpulan data dibagi atas dua cara, yaitu sebagai berikut:

1. **Sensus**, yaitu cara pengumpulan data, di mana data diperoleh dari setiap anggota populasi.
2. **Sampling,** yaitu cara pengumpulan data, di mana hanya sebagian anggota populasi (sampel) saja yang diteliti. Akan tetapi, dari sebagian anggota populasi ini diharapkan dapat menggambarkan keadaan populasi yang sebenarnya.

Sebelum proses pengolahan data dilakukan uji validitas, Kegunaan uji validitas adalah untuk daya ketepatan mengukur, uji validitas terdiri dari uji logika dan empirik. Uji validitas logika digunakan untuk menguji apakah data sampel ini representatif atau tidak untuk penelitian tersebut. Uji validitas logika adalah uji yang diteliti dari segi susunan dan rekaan aspek: kognitif, afektif, dan psikomotor. Uji validitas empirik terdiri dari ramalan dan bandingan, ramalan adalah suatu kondisi yang menunjukkan seberapa jauh sebuah tes telah dapat dengan secara tepat menunjukkan kemampuannya untuk meramalkan apa yang bakal terjadi pada masa mendatang. Contoh : penerimaan mahasiswa baru. Bandingan adalah tes tersebut dalam kurun waktu yang sama dengan secara tepat telah mampu menunjukkan adanya hubungan searah antara tes yang pertama dan kedua (validitas sekarang/pengalaman).

Teknik yang digunakan untuk mengetahui kesejajaran adalah teknik Korelasi Product Moment yang dikemukakan oleh Pearson.

$$r_{xy} = \frac{n\sum XY - \sum X \sum Y}{\sqrt{(n\sum X^2 - (\sum X)^2)(n\sum Y^2 - (\sum Y)^2)}}$$

Koefisien Korelasi adalah sebagai berikut:

Antara 0,800 sampai dengan 1,00 = sangat tinggi

Antara 0,600 sampai dengan 0,800 = tinggi

Antara 0,400 sampai dengan 0,600 = cukup

Antara 0,200 sampai dengan 0,400 = rendah

Antara 0,00 sampai dengan 0,200 = sangat rendah

Uji validitas data adalah uji yang bertujuan untuk mengetahui kelayakan tiap-tiap pertanyaan dalam suatu kuisoner. Berikut adalah langkah-langkah uji validitas:

1. Jumlah uji validitas sesuai dengan jumlah butir pertanyaan dalam suatu kuisoner. Atau jumlah uji validitas sesuai dengan jumlah variabel yang digunakan.
2. Hitung nilai korelasi untuk tiap butir pertanyaan / variabel
$$r_{xy} = \frac{n\sum XY - \sum X \sum Y}{\sqrt{(n\sum X^2 - (\sum X)^2)(n\sum Y^2 - (\sum Y)^2)}}$$
dengan
X: poin tiap pertanyaan
Y: total poin pertanyaan untuk tiap responden
n: jumlah responden
3. Hitung nilai distribusi normal sampel (t)
$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$
4. Ambil kesimpulan dengan membandingkan nilai t hitung dengan nilai t standar (Lampiran 1)

Contoh. Diketahui suatu konversi poin pertanyaan dalam suatu kuisoner dengan 10 responden sebagai berikut. Lakukan uji validitas untuk pertanyaan pertama!

RESP	**PERTANYAAN**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
A	3	4	3	3	2	4	4	3	4	4
B	4	3	3	3	2	3	3	3	3	3
C	2	2	3	1	4	2	1	2	1	2
D	2	2	2	2	3	1	1	2	2	3
E	4	3	3	4	3	3	4	4	3	3
F	3	3	3	3	1	3	4	4	3	4
G	4	4	3	3	3	3	3	4	4	3
H	2	2	1	1	3	2	2	2	1	2
I	4	3	3	4	4	2	4	4	4	2
J	3	3	4	4	2	3	3	3	3	3

Jawab

X	Y	XY	X^2	Y^2
3	34	102	9	1156
4	30	120	16	900
2	20	40	4	400
2	20	40	4	400
4	34	136	16	1156
3	31	93	9	961
4	34	136	16	1156
2	18	36	4	324
4	34	136	16	1156
3	31	93	9	961
31	286	932	103	8570

Hitung nilai korelasi

$$r = \frac{10(932) - (31)(286)}{\sqrt{(10(103) - (31)^2)(10(8570) - (286)^2)}} = 0,874735$$

Hitung nilai distribusi normal sampel

$$t = \frac{0,874735\sqrt{10 - 2}}{\sqrt{1 - (0,874735)^2}} = 5,105476$$

Karena $t = 5,105476 > t_{0,05(8)} = 1,86$ maka terbukti pertanyaan satu valid dam data yang ada pada pertanyaan satu dapat digunakan.

C. RELIABILITAS DATA

Uji reliabilitas data adalah uji yang digunakan untuk mengetahui bahwa dalam suatu kuisoner, semua pertanyaan-pertanyaan dapat dipercaya dan konsisten. Teknik pengukuran reliabilitas Genap Ganjil, butir pertanyaan dikelompokkan jadi dua, genap dan ganjil.

RESP	**GANJIL (x)**					**GENAP (y)**				
	1	**3**	**5**	**7**	**9**	**2**	**4**	**6**	**8**	**10**
1										
2										
⋮										
⋮										
n										

Kelompok I dikorelasikan dengan II

$$r_{xy} = \frac{n\sum XY - \sum X \sum Y}{\sqrt{(n\sum X^2 - (\sum X)^2)(n\sum Y^2 - (\sum Y)^2)}}$$

$$r_{gg} = \frac{2r}{(1+r)}$$

Pertanyaan dikatakan handal jika $r_{gg} > r_{\alpha(n-2)}$

Contoh. Terdapat data-data berikut yang diperoleh dari 10 pertanyaan dan 10 responden.

RESP	**PERTANYAAN**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
A	3	4	3	3	2	4	4	3	4	4
B	4	3	3	3	2	3	3	3	3	3
C	2	2	3	1	4	2	1	2	1	2
D	2	2	2	2	3	1	1	2	2	3
E	4	3	3	4	3	3	4	4	3	3
F	3	3	3	3	1	3	4	4	3	4
G	4	4	3	3	3	3	3	4	4	3
H	2	2	1	1	3	2	2	2	1	2
I	4	3	3	4	4	2	4	4	4	2
J	3	3	4	4	2	3	3	3	3	3

Lakukan uji reliabilitas, apakah data yang diperoleh dapat digunakan dan pertanyaannya konsisten atau berkesinambungan

PERTANYAAN					
1	**3**	**5**	**7**	**9**	**x**
3	3	2	4	4	16
4	3	2	3	3	15
2	3	4	1	1	11
2	2	3	1	2	10
4	3	3	4	3	17
3	3	1	4	3	14
4	3	3	3	4	17
2	1	3	2	1	9
4	3	4	4	4	19
3	4	2	3	3	15
Jumlah					**143**

PERTANYAAN					
2	**4**	**6**	**8**	**10**	**Y**
4	3	4	3	4	18
3	3	3	3	3	15
2	1	2	2	2	9
2	2	1	2	3	10
3	4	3	4	3	17
3	3	3	4	4	17
4	3	3	4	3	17
2	1	2	2	2	9
3	4	2	4	2	15
3	4	3	3	3	16
Jumlah					**143**

XY	X2	Y2
288	256	324
225	225	225
99	121	81
100	100	100
289	289	289
238	196	289
289	289	289
81	81	81
285	361	225
240	225	256
2134	**2143**	**2159**

$$r_{xy} = \frac{n\sum XY - \sum X \sum Y}{\sqrt{(n\sum X^2 - (\sum X)^2)(n\sum Y^2 - (\sum Y)^2)}}$$

$$= \frac{10(2134) - (143)(143)}{\sqrt{(10(2143) - (143)^2)(10(2159) - (143)^2)}}$$

$$= 0{,}84$$

$$r_{gg} = \frac{2r}{(1+r)} = \frac{2 \times 0{,}84}{(1+0{,}84)} = 0{,}91$$

Karena $r_{gg} = 0{,}91 > r_{\alpha(n-2)} = r_{0{,}05(8)} = 0{,}63$ maka semua pertanyaan dalam suatu kuisoner konsisten atau berkesinambungan

D. SOAL

Jawablah pertanyaan-pertanyaan di bawah ini dengan benar!

1. Jelaskan alur pemeriksaan data!
2. Jelaskan perbedaan antara uji validitas dan uji reliabilitas!
3. Berdasarkan tabel berikut, tentukan validitas dan reliabilitasnya!

	PERTANYAAN									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
A	3	4	3	2	2	4	4	3	4	4
B	3	3	3	3	2	3	3	3	3	3
C	2	2	3	1	4	2	1	4	1	2
D	2	1	2	4	3	1	1	2	2	3
E	4	3	3	4	3	2	4	4	3	3
F	3	3	3	3	1	3	4	4	3	4
G	4	2	3	3	3	3	3	4	4	3
H	1	2	1	1	3	2	2	2	1	2
I	4	3	3	4	4	2	4	4	4	2
J	2	3	4	4	2	3	3	3	3	3

BAB III
TEKNIK PENYAJIAN DATA

Capaian Pembelajaran	menjelaskan cara penyajian data
Kemampuan Akhir Pembelajaran	mahasiswa dapat 1. menyebutkan jenis-jenis skala pengukuran dan manfaatnya 2. menyebutkan arti dan manfaat distribusi frekuensi 3. menyusun distribusi frekuensi data kualitatif dan data kuantitatif 4. menggambarkan grafik frekuensi, frekuensi relative, dan frekuensi kumulatif

A. PENDAHULUAN TEKNIK PENYAJIAN DATA

Pada bab sebelumnya telah disampaikan bagaimana cara menyiapkan data dan memeriksanya, data primer memiliki tingkat kesalahan atau eror yang lebih besar dibandingkan data sekunder. Hal ini terjadi karena data sekunder yang biasanya diperoleh dari instansi ataupun publikasi orang lain, cenderung menggunakan alat dan teknik yang lebih terpercaya dalam pengambilan data tersebut. Sedangkan data primer yang diperoleh secara langsung baik melalui kuisoner ataupun pengamatan langsung, memiliki resiko yang lebih besar.

Pada bab ini, akan disampaikan bagaimana cara menyajikan data yang merupakan penerapan dari bentuk statistika deskriptif. Materi yang akan disampaikan diantaranya penyusunan tabel distribusi frekuensi, histogram, polygon, dan diagram batang daun.

B. TABEL DISTRIBUSI FREKUENSI

Data statistik dapat disajikan dalam beberapa bentuk, sesuai dengan jenis data. Data statistik dapat berupa daftar bilangan yang mempunyai satuan yang sama atau disebut data tunggal. Data dapat dinyatakan dalam bentuk **daftar bilangan**. Data tunggal dapat dituliskan sebagai daftar bilangan sebagai contoh berikut. Data nilai Statistika 10 mahasiswa adalah : 60, 75, 65, 80, 95, 74, 88, 87, 76 dan 90.

Tabel distribusi frekuensi dapat dibedakan menjadi 2, yaitu tabel distribusi frekuensi data tunggal dan tabel distribusi frekuensi data berkelompok.

1. Tabel Distribusi Frekuensi Data Tunggal

   Contoh. Skor tes Statistika dari 50 mahasiswa, sajikan data di atas dalam daftar distribusi frekuensi tunggal !

29	25	28	22	24	25	28	26	26
24	23	25	26	21	23	26	27	23
28	30	27	27	24	26	25	25	24
21	25	22	25	25	27	24	23	27
25	26	23	26	23	27	25	24	26
25	24	22	24	26				

   Jawab:

Skor	Turus	Frekuensi
21	II	2
22	III	3
23	~~IIII~~ I	6
24	~~IIII~~ III	8
25	~~IIII~~ ~~IIII~~ I	11
26	~~IIII~~ ~~IIII~~	9
27	~~IIII~~ I	6
28	III	3
29	I	1
30	I	1

2. Tabel Distribusi Frekuensi Data Berkelompok

Jika sekumpulan data memiliki jumlah dan variasi data yang cukup banyak, maka data tersebut dapat disederhanakan dengan cara mengelompokkannya dalam kelas-kelas. Dengan demikian diperoleh tabel distribusi frekuensi data berkelompok.

Beberapa istilah yang penting dalam membuat tabel distribusi frekuensi berkelompok antara lain sebagai berikut

- **Interval Kelas**

  Interval kelas adalah kelas-kelas yang memuat beberapa data tertentu.

  $i = \frac{R}{k}$ i : interval Kelas
  R : jangkauan
  k : banyak kelas

- **Batas Kelas**

  Batas kelas adalah nilai terkecil dan nilai terbesar terdapat pada suatu kelas interval

- **Tepi kelas**

  Tepi kelas adalah setengah dari jumlah batas atas dan batas bawah dua kelas interval yang berurutan. Tepi atas kelas (t_a) adalah batas kelas ditambah setengah. Sedangkan tepi bawah kelas (t_b) adalah batas kelas dikurang setengah.

- **Panjang Kelas**

  Panjang kelas disebut juga lebar kelas atau interval kelas, yaitu selisih antara tepi atas dan tepi bawah dari tiap kelas dalam kelas interval yang sama

- **Titik Tengah Kelas**

  Nilai titik tengah kelas adalah setengah dari jumlah tepi bawah kelas dan tepi atas kelas.

3. Cara Menyusun Tabel Distribusi Kelompok

Beberapa langkah yang perlu diperhatikan dalam menyusun tabel distribusi frekuensi berkelompok adalah sebagai berikut. Menentukan nilai data terbesar (x_{maks}) dan nilai data terkecil (x_{min}) kemudian ditentukan jangkauannya (R) dengan rumus :

$$R = x_{maks} - x_{min}$$

Menentukan banyaknya kelas interval (k) dari n buah data adalah berdasarkan aturan Sturgess, yaitu :

$$k = 1 + 3{,}3 \log n$$

Menentukan panjang kelas (c) dengan rumus :

$$c = \frac{jangkauan}{banyak\ kelas}$$

Menentukan daftar distribusi frekuensi dengan menetapkan kelas – kelas sehingga nilai statistik minimum termuat dalam kelas interval terendah, tetapi tidak harus sebagai batas bawah kelas. Selanjutnya, menetapkan frekuensi tiap kelas yang dapat dilakukan dengan menggunakan turus atau dapat saja langsung dituliskan .

Contoh

Dari 48 kali pengukuran lembaran kain (ketelitian sampai cm terdekat), diperoleh data sebagai berikut.

54	50	53	54	60	56	62	54	58	65	71	58
58	65	56	58	52	70	74	62	52	62	58	60
70	73	45	60	56	54	52	53	67	54	59	64
57	49	48	56	58	58	60	64	63	68	57	59

Buatlah daftar distribusi frekuensi berkelompok dari data tersebut

Jawab:

Data pengukuran tersebut terdidi dari 48 data, sehingga n = 48

Nilai statistik minimum , x_{min} = 45 , dan nilai statistik maksimum, x_{maks} = 74

Jangkauan $(R) = x_{maks} - x_{min} = 74 - 45 = 29$

Banyaknya kelas (k) = 1 + 3,3 log n = 1 +3,3 log 48 = 6,548, dibulatkan ke atas menjadi k = 7

Panjang Kelas $c = \frac{R}{k} = \frac{29}{7} = 4{,}14$, dibulatkan ke atas menjadi tercakup dalam kelas interval.

Tabel distribusi frekuensi :

Hasil Pengukuran	Titik Tengah	Frekuensi
43 – 47	45	1
48 – 52	50	6
53 – 62	55	13
58 – 62	60	16
63 – 67	65	6
68 – 72	70	4
73 – 77	75	2

4. Tabel Distribusi Frekuensi Komulatif dan Frekuensi Relatif

Tabel distribusi frekuensi kumulatif dapat disusun dari tabel distribusi frekuensi berkelompok. Terdapat dua jenis frekuensi kumulatif, yaitu frekuensi kumulatif kurang dari tepi atas ($f_k \leq t_a$) dan frekuensi kumulatif lebih dari tapi bawah ($f_k \geq t_b$)

Setiap frekuensi (fi) dalam tabel distribusi frekuensi yang dinyatakan dalam persentase disebut frekuensi relatif. Frekuensi relatif (fr) dapat ditentukan denngan rumus :

$$f_r = \frac{f_i}{n} \times 100\%$$

Selanjutnya, daftar distribusi frekuensi kumulatif relative dapat disusun dari daftar distribusi frekuensi kumulatif.

Contoh

Dengan cara perhitungan yang sama, akan kita dapatkan tabel distribusi frekuensi kumulatif relatif berikut.

Hasil	Frek	Frekuensi Relatif (fr)	Frekuensi Kumulatif		Frekuensi Kumulatif Relatif	
			$f_k \leq t_a$	$f_k \geq t_b$	$f_{kr} \leq t_a$	$f_{kr} \geq t_b$
43 – 47	1	0,021	1	48	0,021	1
48 – 52	6	0,125	7	47	0,146	0,979
53 – 57	13	0,271	20	41	0,417	0,854
58 – 62	16	0,333	36	28	0,750	0,583
63 – 67	6	0,125	42	12	0,875	0,250
68 – 72	4	0,083	46	6	0,958	0,125
73 – 77	2	0,042	48	2	1	0,042

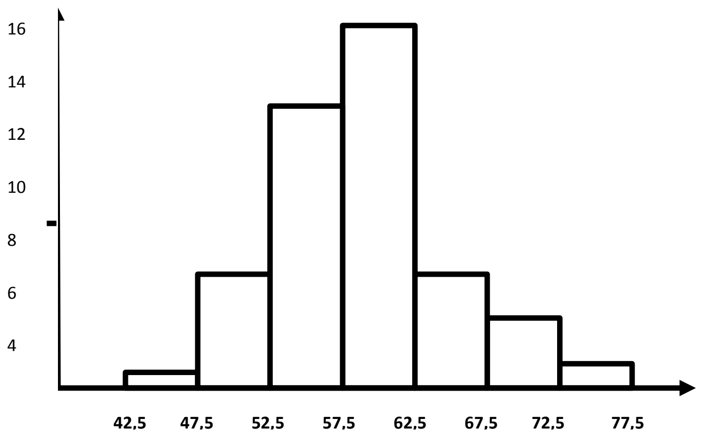
C. PENYAJIAN DATA

1. Histogram

Histogram adalah diagram garis yang daerah absisnya merupakan tepi kelas sehingga posisi antar batang berimpit. Sebagai contoh tabel distribusi frekuensi berikut, tentukan histogramnya.

Hasil Pengukuran	Frekuensi
43 – 47	1
48 – 52	6
53 – 62	13
58 – 62	16
63 – 67	6
68 – 72	4
73 – 77	2

Dengan mengikuti langkah-langkah membuat histogram suatu data berkelompok, histogram dari data tersebut sebagai berikut

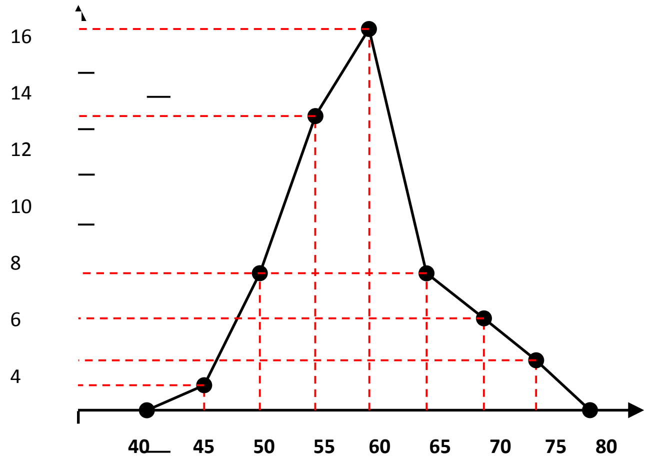


2. Poligon

Jika titik-titik tengah dari sisi atas tiap batang pada histogram dihubungkan, maka akan diperoleh grafik garis yang disebut poligon distribusi frekuesi. Selain dengan cara tersebut, poligon distribusi frekuensi dapat dibuat dengan langkah-langkah sebagai berikut :

- Menambahkan satu kelas interval sebelum kelas pertama dan satu kelas interval sesudah kelas terakhir.
- Menentukan titik tengah setiap kelas
- Menggambar sumbu horizontal dan sumbu vertical
- Menggambar titik-titik dengan titik tengah kelas interval sebagai absis dan frekuensi sebagai ordinat
- Menghubungkan titik – titik yang berdekatan dengan suatu aris lurus.

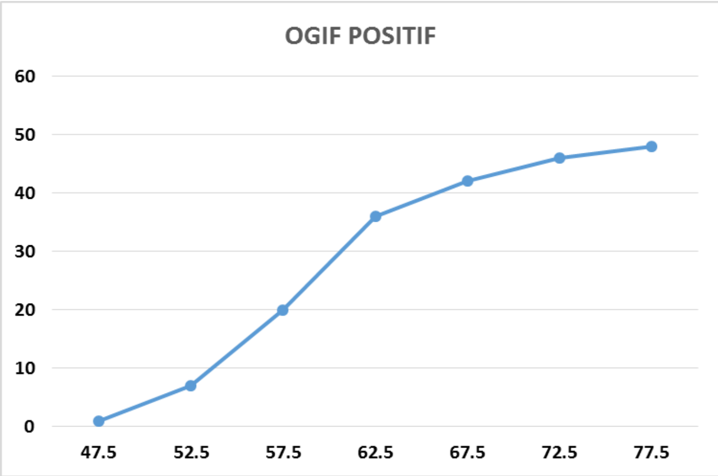
Poligon distribusi dari data tersebut diperlhatkan oleh gambar di bawah

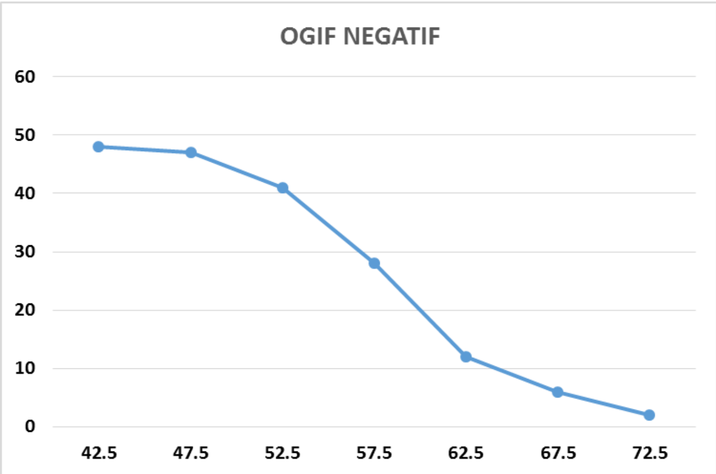


3. Ogif

Tabel distribusi frekuensi kumulatif yang disajikan dalam bentuk kurva, disebut polygon distribusi frekuensi kumulatif atau ogive. Ogive terdiri dari 2 macam yaitu ogive positif (ogive kurang dari) dan ogive negatif (ogive lebih dari). Ogive positif dibentuk dengan menghubungkan titik-titik , dengan tepi atas sebagai absis dan frekuensi kumulatif sebagai ordinat. Sementara itu, ogive negatif dapat dibentuk dengan cara menghubungkan titik-titik, dengan tepi bawah sebagai absis dan frekuensi kumulatif sebagai ordinat.

Hasil	Frek	Frekuensi Relatif (fr)	Frekuensi Kumulatif	
			$f_k \leq t_a$	$f_k \geq t_b$
43 – 47	1	0,021	1	48
48 – 52	6	0,125	7	47
53 – 57	13	0,271	20	41
58 – 62	16	0,333	36	28
63 – 67	6	0,125	42	12
68 – 72	4	0,083	46	6
73 – 77	2	0,042	48	2





D. SOAL

Jawablah pertanyaan-pertanyaan di bawah ini dengan benar!

1. Berdasarkan data-data di bawah ini, buatlah tabel distribusi frekuensinya!

42	41	32	38	35	26	21	17	42	44
51	24	21	23	20	33	27	27	47	48
52	25	22	20	31	34	28	29	31	51

2. Buatlah ogif negative, histogram, dan polygon dari data-data berikut!

Kelas	34 – 40	41 – 47	48 – 54	55 – 61	62 – 68	69 – 75
Frekuensi	1	5	13	16	6	4

BAB IV
UKURAN PEMUSATAN

Capaian Pembelajaran	menjelaskan ukuran pemusatan
Kemampuan Akhir Pembelajaran	mahasiswa dapat 1. Menyebutkan jenis ukuran pemusatan 2. Menjelaskan fungsi ukuran pemusatan 3. Menghitung ukuran pemusatan

A. PENDAHULUAN

Pada bab sebelumnya telah disampaikan salah satu statistika deskriptif, pada bab ini akan disampaikan statistika deskriptif lainnya yaitu ukuran pemusatan. Ukuran pemusatan adalah ukuran yang menunjukan pusat dari sekumpulan data, yang telah diurutkan dari yang terkecil sampai yang terbesar, atau sebaliknua dari yang terbesar sampai terkecil. Salah satu kegunaan dari ukuran pemusatan data adalah untuk membandingkan dua populasi atau sampel, karena sangat sulit membandingkan masing-masing anggota dari masing-masing anggota populasi atau sampel. Nilai ukuran pemusatan ini dibuat cukup mewakili seluruh nilai pada data yang bersangkutan. Ukuran pemusatan yang paling banyak digunakan adalah median, mean, dan modus. Tiap ukuran pemusatan data tersebut memiliki kekurangan. Nilai tengah atau median akan sangat dipengaruhi oleh nilai pencilan. Median terlalu bervariasi untuk dijadikan parameter populasi. Sedangkan modus hanya dapat diterapkan dalam data dengan ukuran yang besar.

B. UKURAN PEMUSATAN

1. Mean (Rataan Hitung)

Mean (rataan hitung) didefinisikan sebagai jumlah data kuantitatif dibagi banyaknya data. Atau dapat dinyatakan sebagai jumlah seluruh data dibagi banyaknya data.

Data tunggal : Data Kelompok :

$$\bar{x} = \frac{\sum x}{n} \qquad \bar{x} = \frac{\sum f_i x_i}{\sum f_i}$$

dengan : x_i = titik tengah kelas interval, f_i = frekuensi dari x_i

k = banyaknya kelas interval

Selain menggunakan rumus tersebut, menentukan mean atau rata-rata dari sekumpulan data dengan terlebih dahulu menentukan rataan sementaranya. Rataan sementara biasanya diambil dari nilai tengah data yang memiliki frekuensi terbesar. Untuk menghitung rata-rata dapat menggunakan rata-rata sementara. Kesulitan dalam menghitung rata-rata adalah apabila dijumpai bilangan besar atau tidak bulat. Untuk mengatasi hal ini, terlebih dahulu dilakukan penyederhanaan data, yaitu dengan cara memperkirakan nilai rata-rata yang disebut rata-rata sementara. Caranya adalah sebagai berikut:

a) Tetapkan rata-rata sementara $\overline{(x_0)}$, dipilih pada kelas yang mempunyai frekuensi tertinggi.
b) Tentukan simpangan (deviasi) terhadap rata – rata sementara, dengan rumus:
   $d_i = x_i - \overline{x_0}$
c) Tentukan rata – rata sesungguhnya, dengan rumus:
   $$\bar{x} = \overline{x_0} + \frac{\sum f_i d_i}{\sum f_i}$$

Contoh. Dua belas orang mengikuti pertandingan menembak pada jarak tertentu, setiap peserta menembak 10 kali. Hasil tembakan yang mengenai sasaran dari tiap-tiap peserta adalah 4, 8, 5, 8, 6, 4, 7, 7, 2, 3, 5, 7. Tentukan rataan tembakan yang menenai sasaran!

Jawab :

- **Data Tunggal**

Data di atas dipandang sebagai sampel, maka :

$$\sum x = 4+8+5+8+6+4+7+7+2+3+5+7 = 66 \; dan \; n = 12$$

$$\bar{x} = \frac{\sum x}{n} = \frac{66}{12} = 5{,}5$$

- **Data Kelompok**

Tentukan Rata – rata dari data berikut :

Nilai	Frekuensi (f_i)	Titik Tengah (x_i)	($f_i x_i$)
40 – 49	**4**	**44,5**	**178**
50 – 59	**6**	**54,5**	**327**
60 – 69	**10**	**64,5**	**645**
70 – 79	**4**	**74,5**	**298**
80 – 89	**4**	**84,5**	**338**
90 - 99	**2**	**94,5**	**189**
	$\sum f_i = 30$		$\sum f_i x_i = 1975$

$$\bar{x} = \frac{\sum f_i x_i}{f_i} = \frac{1975}{30} = 65{,}83$$

Jadi, rata – ratanya adalah 65,83

2. Modus (Nilai terbanyak)

Modus adalah nilai yang paling banyak muncul. Untuk data tunggal, modus sangat mudah ditentukan, yaitu data yang yang mempunyai frekuensi terbanyak. Modus mempunyai kelemahan, yaitu apabila kelompok data yang dimaksud memiliki dua nilai modus (bimodal) atau lebih, atau tidak memiliki modus, misal : Data 5, 7, 8, 10, 10,12,12 memiliki dua modus yaitu 10 dan 12.

Untuk data distribusi frekuensi dalam bentuk kelas – kelas interval, nilai modus tidak dapat ditentukan dengan tepat tetapi dengan pendekatan. Ada yang berpendapat nilai modus sama dengan nilai tengah kelas yang mempunyai frekuensi terbanyak. Cara lain yang dianggap lebih tepat, yaitu dengan memperhatikan frekuensi kelas sebelum dan sesudah kelas modus.

Rumus Modus :

$$Mod = L_0 + c\left\{\frac{(d_1)}{(d_1)+(d_2)}\right\}$$

dengan

L_0 : tepi kelas bawah pada kelas modus c : panjang kelas

d_1 : frekuensi kelas modus dikurangi frekuensi kelas sebelumnya

d_2 : frekuensi kelas modus dikurangi frekuensi kelas sesudahnya

Contoh. Suatu mesin yang memproduksi kaleng roti diperkirakan terdapat kesalahan. Dari penelitian terhadap 200 kaleng roti , dicatat berat kaleng roti, disajikan pada daftar di bawah ini:

Berat Kaleng	Frekuensi (f)
281 – 283	4
284 – 286	18
287 – 289	36
290 – 292	82
293 – 295	50
296 – 298	10

$$Mo = 289{,}5 + 3\left(\frac{46}{46+32}\right) = 291{,}26$$

3. Median

Median adalah nilai yan membagi data menjadi dua bagian yang sama banyaknya setelah data diurutkan dari yang terkecil hingga yang terbesar. Untuk mendapatkan nilai median dari daftar distribusi frekuensi kita dapat menggunakan rumus median, selain itu kita juga dapat mendapatkan nilai median menggunakan histogram, yang berarti median membagi histogram menjadi dua bagiab yang sama luasnya. Rumus Median :

$$Median = L_0 + c\left\{\frac{\frac{n}{2} - f_k}{f_m}\right\}$$

dengan

L_0 : tepi kelas bawah pada kelas median

c : panjang kelas median

n : ukuran sampel atau banyak data

F_k : jumlah semua fekuensi dengan tanda kelas lebih kecil dari tanda kelas median

f_m : frekuensi kelas median

Contoh. Suatu mesin yang memproduksi kaleng roti diperkirakan terdapat kesalahan. Dari penelitian terhadap 200 kaleng roti , dicatat berat kaleng roti, disajikan pada daftar di bawah ini:

Berat Kaleng	Frekuensi (f)
281 – 283	4
284 – 286	18
287 – 289	36
290 – 292	82
293 – 295	50
296 – 298	10

Langkah – langkah untuk mengerjakan median : $12n = 12 \times 200 = 100$, c = 3, $L_0 = 289{,}5$, $f_m = 82$, dan $F_k = 58$

$$Me = 289{,}5 + 3\left(\frac{100-58}{82}\right) = 291{,}03$$

4. Kuartil (Qi)

Kuartil adalah nilai yang membagi data menjadi 4 bagian yang sama banyak, setelah data diurutkan dari yang terkecil hingga yang terbesar. Terdapat 3 buah kuartil , yaitu kuartil bawah atau kuartil pertama dilambangkan Q_1, kuartil tengah atau kuartil kedua atau median dilambangkan q_2, dan kuartil atas atau kuartil ketiga dilambangkan Q_3. Sama halnya dengan median, maka nilai kuartil dapat dihitung dengan cara :

$$Q_i = L_o + C\left\{\frac{\frac{in}{4} - fk}{f_q}\right\}, \ i = 1,\ 2,\ 3$$

Contoh Data Tunggal

Tentukan Q_1, Q_2, dan Q_3 untuk data berikut!

1. 6, 8, 4, 2, 4, 7, 5, 4
2. 3, 5, 1, 5, 4, 7, 8, 4, 2

Jawab:

1. Banyak data, n = 8

   Data yang telah diurutkan :

   2, 4, 4, 4, 5, 6, 7, 8

   Q_1 Q_2 Q_3

   $Q1 = 124 + 4 = 4$; $Q2 = 124 + 5 = 4{,}5$; $Q3 = 126 + 7 = 6{,}5$

   Jadi, $Q_1 = 4$; $Q_2 = 4{,}5$; $Q_3 = 6{,}5$.

2. Banyak data, n = 9

   Data yang telah diurutkan :

   1, 2, 3, 4, 4, 5, 5, 7, 8

   Q_1 Q_2 Q_3

   $Q_1 = 122 + 3 = 2{,}5$; $Q3 = 4$; $Q3 = 125 + 7 = 6$

   Jadi, $Q_1 = 2{,}5$; $Q_2 = 4$; $Q_3 = 6$

Data Berkelompok

Suatu mesin yang memproduksi kaleng roti diperkirakan terdapat kesalahn. Dari penelitian terhadap 200 kaleng roti , dicatat berat kaleng roti, disajikan pada daftar di bawah ini:

Berat Kaleng	Frekuensi (f)	F kum
281 – 283	4	4
284 – 286	18	22
287 – 289	36	58
290 – 292	82	140
293 – 295	50	190
296 – 298	10	200

Jawab:

a) Dengan i = 3 dan n = 200
b) c = 3
c) $\frac{3}{4} \times 200 = 150$
d) L_0 = 292,5
e) fq = 50
f) F_k = 140

5. Desil (Di)

Desil adalah nilai yang membagi data menjadi 10 bagian yang sama banyak, setelah data diurutkan dari yang terkecil hingga yang terbesar. Untuk menentukan desil digunakan rumus sebagai berikut.

$$D_i = L_o + C\left\{\frac{\frac{in}{10} - fk}{f_d}\right\}, \mathrm{i} = 1,\ 2,\ \ 3, \ldots., 9$$

Contoh. Tentukan nilai desil ke-3 dari data berikut!

7   5   8   7   9   6   6   6   8   5   9   8   6   7   9

Jawab

Data yang telah diurutkan : 5   5   6   6   6   6   7   7   7   8   8   8   9   9   9

Bnayak data, n = 15.

Desil k-3 adalahnilai yan terletak pada urutan ke 3(15+1)10=4,8

$D_3 = x_4 + 0{,}8(x_5 – x_4) = 6 + 0{,}8\ (6 - 6) = 6$

Jadi, nilai D_3 adalah 6

6. Persentil (Pi)

Dalam hal ini kita juga dapat membagi sekelompok data menjadi seratus bbagian yang sama banyak, sehingga terdapat 99 nilai pembagi yang disebut persentil. Untuk menghitung nilai persentil digunakan rumus :

$$P_i = L_o + C\left\{\frac{\frac{in}{100} - fk}{f_p}\right\}, \text{ i} = 1,\ 2,\ \ 3,\ ....,\ \ 99$$

Contoh. Tentukan P_{76} dari tabel distribusi frekuensi di bawah ini :

Kelas	Fi
23 - 37	3
38 - 52	1
53 - 67	2
68 - 82	7
83 - 97	1
	14

$$P_{76} \text{ letak } \frac{in}{100} = \frac{76 * 14}{100} = 10.64$$

$$fk = 6\ ; \text{c} = 15\ ; \text{f}_\text{d} = 7 ; \text{L}_\text{o} = 67.5$$

$$P_{76} = 67.5 + 15\left\{\frac{10.64 - 6}{7}\right\} = 67.5 + (15 \text{ x } 0.7) = 78$$

C. SOAL

Jawablah pertanyaan-pertanyaan di bawah ini dengan benar!

1. Berdasarkan tabel distribusi frekuensi di bawah ini, tentukan rata-rata dengan rata-rata sementara kelas kedua!

Kelas	Frekuensi
34 – 40	1
41 – 47	5
48 – 54	10
55 – 61	14
62 – 68	6
69 – 75	4

2. Berdasarkan tabel distribusi frekuensi di bawah ini, tentukan modus dan median!

Kelas	Frekuensi
34 – 40	1
41 – 47	5
48 – 54	13
55 – 61	16
62 – 68	6
69 – 75	4

3. Berdasarkan tabel distribusi frekuensi di bawah ini, tentukan kuartil atas!

Kelas	Frekuensi
16 – 21	1
22 – 27	5
28 – 33	12
34 – 39	15
40 – 45	3
46 – 51	4

4. Berdasarkan tabel distribusi frekuensi di bawah ini, tentukan desil ke-6 dan persentil ke-59!

Kelas	Frekuensi
38 – 44	2
45 – 51	3
52 – 58	13
59 – 65	12
66 – 72	6
73 – 79	4

BAB V
UKURAN PENYEBARAN

Capaian Pembelajaran	menjelaskan ukuran penyebaran atau dispersi
Kemampuan Akhir Pembelajaran	mahasiswa dapat 1. Menyebutkan jenis ukuran penyebaran 2. Menjelaskan fungsi ukuran penyebaran 3. Menghitung ukuran penyebaran

A. PENDAHULUAN UKURAN PENYEBARAN

Dalam pengukuran statistika terdapat pula ukuran penyebaran. Ukuran penyebaran data adalah ukuran yang menunjukan seberapa jauh data menyebar dari rata-rata. Semakin kecil ukuran penyebaran maka semakin baik. Ukuran penyebaran yang disampaikan adalah *range*, hamparan, simpangan kuartil, simpangan rata-rata, ragam, dan simpangan baku. Tiap ukuran penyebaran memiliki kelebihan dan kelemahan. Kelebihan *range* sebagai salah satu ukuran penyebaran data adalah bahwa dengan menggunakan *range*, dalam waktu singkat dapat diperoleh gambaran umum mengenai luas penyebaran data yang dimiliki. Kelemahan *range* adalah *range* akan sangat tergantung pada nilai terendah dan nilai tertinggi yang terdapat di dalam distribusi sata yang dimiliki, dengan demikian *range* sifatnya kurang teliti.

B. UKURAN PENYEBARAN

Ukuran penyebaran data yang biasa digunakan untuk data tunggal antara lain rentang, hamparan simpangan kuartil, simpangan rata-rata, ragam dan simpangan baku.

1. **Rentang atau jangkauan (R)**

Jangkauan data atau rentang data adalah selisih antara data terbeasar (x_{maks}) dengan data terkecil (x_{min}). $R = X_{maks} - X_{min}$

2. **Hamparan (H)**

Jangkauan antarkuartil atau hamparan adalah selisih antara kuartil ketiga dengan kuartil pertama $H = Q_3 - Q_1$

3. Simpangan Kuartil (Qd)

Jangkauan semi antarkuartil atau simpangan kuartil adalah setengah kali panjang hamparan.

$$Q_d = \frac{1}{2}(Q_3 - Q_1)$$

Contoh. Diketahui data: 3, 4, 4, 5, 7, 8, 9, 9, 10. Tentukan jangkauan, jangkauan antarkuartil, dan simpangan kuartildari data tersebut

Jawab data

3, 4, 4, 5 7, 8, 9, 9, 10

Q_1 Q_2 Q_3

$$Q_1 = \frac{1}{2}(4+4) = 4\ ;\ \ Q_2 = 7\ ; Q_3 = \frac{1}{2}(9+9) = 9$$

Jangkauan = $x_{maks} - x_{min}$ = 10 – 3 = 7

4. Simpangan Rata-rata

Simpanagan rata-rata atau deviasi rata-rata merupakan rata-rata jarak suatu data terhadap rataan hitungannua. Nilai simpangan rata-rata (SR) untuk data tunggal dapat ditentukan dengan rumus:

$$SR = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$$

dengan

n = banyaknya data

xi = nilai data ke-i

$\bar{x}$ = rataan hitung

Data Kelompok

Nilai	Frek(f_i)	Titik Tengah	$\lvert x_I - \bar{x}\rvert$	$f\lvert x_i - \bar{x}\rvert$
40 – 49	4	44,5	21,17	84,68
50 – 59	6	54,5	11,17	67,02
60 – 69	10	64,5	1,17	11,70
70 – 79	4	74,5	8,83	35,.32
80 – 89	4	84,5	18,83	75,32
90 – 99	2	94,5	28,83	57,66

$$\mathrm{SR} = \frac{\sum \mathrm{f_i}|\mathrm{x_i} - \bar{\mathrm{x}}|}{\sum \mathrm{f_i}} = \frac{331{,}70}{30} = 11{,}06$$

Jadi, simpangan rata-ratanya adalah 11,06

5. Ragam dan Simpangan Baku

Misalnya data x_1 , x_2 , x_3 ,… x_n mempunyai rataan, maka **ragam** atau varians (S^2) dapat ditentukan dengan rumus:

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

Sementara itu, **simpanngan baku** atau deviasi baku (S) dapat ditentukan dengan rumus:

$$S = \sqrt{S^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$

dengan

n = banyaknya data

x_i = nilai data ke-i

$\bar{x}$ = rataan hitung

Contoh.

Hitunglah ragam dan simpangan bakudrai data: 1, 3, 4, 5, 8, 10, 12, 13

Jawab:

Data: 1, 3, 4,5, 8, 10, 12, 13

n = 8 dan $\bar{x}$ =7, maka:

$$\begin{aligned}\sum_{i=1}^{8} (x_i - \bar{x})^2 &= (1-7)^2 + (3-7)^2 + (4-7)^2 + (5-7)^2 + (8-7)^2 \\ &\quad + (10-7)^2 \ + (12-7)^2 \ + (13-7)^2 \\ &= 36 + 16 + 9 + 4 + 1 + 9 + 25 + 36 = 136\end{aligned}$$

$$S^2 = \frac{1}{8} \sum_{i=1}^{8} (x_i - \bar{x})^2 = \frac{1}{8}(136) = 17$$

$$S = \sqrt{S^2} = \sqrt{17} = 4{,}12$$

Jadi, data tersebut mempunyai ragam , S^2 = 17 dan simpangan baku , S = 4,12

C. SOAL

Jawablah pertanyaan-pertanyaan di bawah ini dengan benar!

1. Tentukan hamparan dari data berkelompok berikut!

Kelas	Frekuensi
16 – 21	2
22 – 27	5
28 – 33	13
34 – 39	14
40 – 45	6
46 – 51	10
52 – 57	5

2. Tentukan simpangan kuartil dari data berkelompok berikut!

Kelas	Frekuensi
26 – 31	1
32 – 37	5
38 – 43	12
44 – 49	15
50 – 55	3
56 – 61	4

3. Tentukan simpangan rata-rata dari tabel distribusi frekuensi!

Kelas	Frekuensi
6 – 11	1
12 – 17	5
18 – 23	12
24 – 29	15
30 – 35	3
36 – 41	4

4. Tentukan variansi dari data-data berikut ini!

32	33	23	24	28	29	31	33
24	23	20	21	35	37	22	23
12	17	18	21	25	27	28	28
22	23	24	27	28	29	31	33

5. Tentukan simpangan baku dari data-data berikut!

Kelas	Frekuensi
24 – 28	1
29 – 33	5
34 – 38	12
39 – 43	15
44 – 48	3

Kelas	Frekuensi
49 – 53	4

BAB VII
ANALISIS REGRESI

Capaian Pembelajaran	menyajikan, mengolah, dan menganalisis data
Kemampuan Akhir Pembelajaran	mahasiswa dapat 1. Menjelaskan pentingnya analisis hubungan 2. Menghitung koefisien korelasi dan regresi sederhana 3. Menerapkan regresi dan korelasi 4. Menggunakan teknik ramalan dan melakukan analisis regresi

A. PENDAHULUAN

Analisa regresi adalah analisa ketergantungan antara variabel *independent* terhadap variabel *dependent*. Variabel *independent*/ variabel bebas/indikator adalah variabel yang nilainya berdiri sendiri, tidak dipengaruhi oleh variabel yang lain. Variabel *dependent*/variabel tak bebas/respon adalah variabel yang nilainya mempengaruhi variabel yang lain.

A. ESTIMASI PARAMETER

Untuk menduga nilai parameter β_0 dan β_1 terdapat bermacam-macam metode, misalnya metode kuadrat terkecil (*least square method*), metode kemungkinan maksimum (*maximum likelihood method*), metode kuadrat terkecil terboboti (*weighted least square method)*. Metode yang digunakan adalah metode kuadrat terkecil, karena mudah dikerjakan secara manual. Prinsip dasar metode kuadrat terkecil adalah meminimumkan jumlah kuadrat simpangan atau Jumlah Kuadrat Galat

$$JKG = \sum_{i=1} e_i^2 = \sum_{i=1} \left(Y_i - \widehat{Y_\iota}\right)^2$$

Dengan menggunakan bantuan pelajaran kalkulus, diperoleh nilai dugaan parameter regresi sebagai berikut:

$$b_1 = \frac{n\sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2} \qquad b_0 = \frac{\sum_{i=1}^{n} Y_i \sum_{i=1}^{n} X_i^2 - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} X_i Y_i}{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}$$

Dengan demikian dapat diperoleh hubungan;

$$b_0 = \frac{1}{n}\left(\sum Y_i - b_1 \sum X_i\right) = \bar{Y} - b_1 \bar{X}$$

Contoh Seorang manajer pemasaran akan meneliti apakah terdapat pengaruh iklan terhadap penjualan perusahaan di Kabupaten Pemalang, untuk kepentingan penelitian tersebut diambil 8 perusahaan sejenis yang telah melakukan promosi.

Penjualan (Y)	64	61	84	70	88	92	72	77
Promosi (X)	20	16	34	23	27	32	18	22

Tentukan estimasi parameter model regresi berikut

Y	X	XY	X^2	Y^2
64	20	1280	400	4096
61	16	976	256	3721
84	34	2856	1156	7056
70	23	1610	529	4900
88	27	2376	729	7744
92	32	2944	1024	8464
72	18	1296	324	5184
77	22	1694	484	5929
608	192	15032	4902	47094

$\sum_{i=1}^{8} X_i = 192$ $\sum_{i=1}^{8} Y_i = 608$ $\sum_{i=1}^{8} X_i Y_i = 15032$

$\sum_{i=1}^{8} X_i^2 = 4902$ $\bar{X} = 24$ $\sum_{i=1}^{8} Y_i^2 = 47094$ $\bar{Y} = 76$

$$b_1 = \frac{8(15032) - (192)(608)}{8(4902) - (192)^2} = 1{,}497$$

$$a = \frac{(608) - 1{,}497(192)}{8} = 40{,}072$$

Jadi persamaan regresi dugaan $\hat{Y}$ = 40,072 + 1,497X

B. KOEFISIEN DETERMINASI

Koefisien determinasi adalah nilai yang digunakan untuk mengetahui seberapa besar pengaruh variabel independen terhadap variabel dependen. Dalam analisis regresi, koefisien korelasi yang dihitung tidak untuk diartikan sebagai ukuran keeratan hubungan variabel bebas (X) dan variabel tidak bebas (Y), sebab dalam analisis regresi asumsi normal bivariat tidak terpenuhi.

Indeks determinasi yang diperoleh tersebut digunakan untuk menjelaskan persentase variasi dalam variabel tidak bebas (Y) yang disebabkan oleh bervariasinya variabel bebas (X). Hal ini untuk menunjukkan bahwa variasi dalam variabel tak bebas (Y) tidak semata-mata disebabkan oleh bervariasinya variabel bebas (X), bisa saja variasi dalam variabel tak bebas tersebut juga disebabkan oleh bervariasinya variabel bebas lainnya yang mempengaruhi variabel tak bebas tetapi tidak dimasukkan dalam model persamaan regresinya.

$$R^2 = 1 - \frac{\sum (Y - \hat{Y})^2}{\sum (Y - \bar{Y})^2}$$

dengan $\hat{Y}$ adalah nilai prediksi model regresi

Untuk kasus sebelumnya diperoleh nilai berikut

Y	**X**	**Y_{pred}**	**(Y-Ypred)^2**	**(Y-Yrata)^2**
64	*20*	*70.012*	*36.144144*	*144*
61	*16*	*64.024*	*9.144576*	*225*
84	*34*	*90.97*	*48.5809*	*64*
70	*23*	*74.503*	*20.277009*	*36*
88	*27*	*80.491*	*56.385081*	*144*
92	*32*	*87.976*	*16.192576*	*256*
72	*18*	*67.018*	*24.820324*	*16*
77	*22*	*73.006*	*15.952036*	*1*
608	*192*	*608*	*227.496646*	*886*

$$R^2 = 1 - \frac{\sum (Y-\hat{Y})^2}{\sum (Y-\bar{Y})^2} = 1 - \frac{227,496646}{886} = 1 - 0.2568 = 0,7432$$

Untuk kasus tersebut dapat diambil kesimpulan biaya iklan yang dikeluarkan memiliki pengaruh 74,32% terhadap penjualan perusahaan tersebut.

C. ESTIMASI BAKU

Estimasi baku adalah nilai yang digunakan untuk mengetahui besar kesalahan model regresi

$$Se = \sqrt{\frac{\sum (Y-\hat{Y})^2}{n-k}}$$

dengan k adalah banyaknya variabel yang terlibat

untuk kasus pengaruh biaya iklan terhadap penjualan perusahaan sebelumnya diperoleh nilai estimasi baru berikut

$$Se = \sqrt{\frac{(227,467)}{8-2}} = 6,1576$$

D. ESTIMASI EROR STANDAR

Estimasi eror standar adalah nilai yang digunakan untuk mengetahui besar kesalahaan koefisien model regresi

$$Sb = \frac{Se}{\sqrt{\sum X^2 - \frac{(\sum X)^2}{n}}}$$

untuk kasus pengaruh biaya iklan terhadap penjualan perusahaan sebelumnya diperoleh nilai estimasi eror standar berikut

$$Sb_1 = \frac{6,1576}{\sqrt{(4902) - \frac{(192)^2}{8}}} = 0,359$$

E. UJI KECOCOKAN MODEL/GOODNESS OF FIT TEST

Goodness of fit test atau sering disebut dengan uji Anova adalah untuk uji dalam analisa regresi yang digunakan untuk mengetahui kecocokan model terdapat data atau kasus yang terjadi. Uji F digunakan untuk menentukan pengaruh secara bersama-sama variabel bebas terhadap variabel tergantung:

Ho: ditolak jika $F_{hitung} > F_{tabel} (k - 1\ ;\ n - k\ ;\ \alpha)$

$$F = \frac{R^2/(k-1)}{1-R^2/(n-k)}$$

untuk kasus pengaruh biaya iklan terhadap penjualan perusahaan sebelumnya diperoleh nilai statistik hitung F berikut

$$F = \frac{0,743/(2-1)}{1-0,743/(8-2)} = 17,367$$

Karena F hitung (17,367) > dari F tabel (5,99) maka secara bersama variabel bebas berpengaruh terhadap variabel tergantung. Atau dengan kata lain secara bersama-sama biaya iklan berpengaruh terhadap hasil penjualan.

F. SOURCE CODE MATLAB

```
%% Linear Regression (Simplest Implementation)
%% Generating Data
%%
% Taking any random data
X=randn(100,1);

% Adding noise to equation y = 3x + 7
Y=3*(X + 0.3*randn(100,1)) + (7 + 0.3*randn(100,1));

% If data is not normalized, Zscore Normalization (Advised for
SGD)
% data=[zscore(X),Y];
data=[X,Y];

% Define a symbolic varible for function plot
syms x
%% CV Partition
% * 70% to training set + 30% to testing set
```

```
[train_set,test_set ] = holdout(data,70 );
X_train=train_set(:,1:end-1); Y_train=train_set(:,end);
X_test=test_set(:,1:end-1); Y_test=test_set(:,end);

% Number of training instances
N=length(X_train)

% Number of testing instances
M=length(X_test)
%% 1.  Using Direct Method
% * Append a vectors of one to _X_train _for calculating *bias*.
%%
W=pinv([ones(N,1) X_train])*Y_train
%% Mean Square Error

predicted_values=[ones(M,1) X_test]*W;
mse1=sqrt(mean((predicted_values-Y_test).^2))
%% Plot

figure
hold on
scatter(X_test,Y_test)
fplot(W(1)+W(2)*x)
xlabel({'X_1'})
ylabel({'Y'})
title({'Regression Using Direct Method'})
xlim([-3 3])
hold off
%% 2. Using Inbuilt MATLAB Function
%%
```

```
Test_mdl = fitlm(X_train,Y_train);
W=Test_mdl.Coefficients{:,1}
%% Mean Square Error

predicted_values=predict(Test_mdl,X_test);
mse2=sqrt(mean((predicted_values-Y_test).^2))
%% Plot

figure
hold on
scatter(X_test,Y_test)
fplot(W(1)+W(2)*x)
xlabel({'X_1'});
ylabel({'Y'});
title({'Regression Using Inbuilt MATLAB Function'});
xlim([-3 3])
hold off
%% 3. Using Stochastic Gradient Descent
% * Learning Parameter, alpha = 0.1
% * Append a vectors of one to _X_train _for calculating *bias*.
% * Tolerence = 10^-5
%%
X_train=[ones(N,1), X_train];

W=zeros(size(X_train,2),1);
W_old=ones(size(X_train,2),1);

while(norm(W_old-W) > 10^-5)
    W_old=W;
  W = W - 0.1/N*X_train'*(X_train*W - Y_train);
```

```
end
W
%% Mean Square Error
predicted_values=[ones(length(X_test),1),X_test]*W;
mse3=sqrt(mean((predicted_values-Y_test).^2))
%% Plot
figure
hold on
scatter(X_test,Y_test)
fplot(W(1)+W(2)*x)
xlabel({'X_1'});
ylabel({'Y'});
title({'Regression using Stochastic Gradient Descent'});
xlim([-3 3])
hold off
%% Comparing Mean Square
%%
figure
mse=[mse1,mse2,mse3];
plot(mse,'--^m')
title('Comparison: Mean Square Error')
ylabel({'MSE \rightarrow'});
xlabel({'Different methods   [--Direct_{Method} --
Inbuilt_{Function} --SGD_{Method}]'});
xlim([0, 4]); ylim([min(mse)-1e-6, max(mse)+1e-6]);
%%
```

G. SOAL

1. Tabel di bawah ini menyajikan skore motivasi belajar Matematika (X) dan prestasi belajar Matematika (Y) dari 20 siswa yang dipilih secara acak dari suatu sekolah menengah pertama.

X	78 60 57 40 59 70 65 66 68 58 44 38 70 60 65 68 50 74 46 54
Y	85 70 65 45 78 89 50 60 75 50 50 40 87 75 78 80 45 90 50 58

   a. Gambarkan diagram pencar untuk data di atas!
   b. Tentukan nilai dugaan koefisien regresi b_1 dan b_0!
   c. Tuliskan persamaan regresi dugaannya!
   d. Gambarkan garis regresi dugaan tersebut pada gambar diagram pencar pada soal a!

2. Suatu penelitian telah dilakukan untuk menentukan hubungan antara peubah bebas X dan peubah tak bebas Y. Data hasil penelitian telah dihitung dan didapatkan hasil sebagai berikut: $\sum X = 63{,}6$; $\sum X^2 = 339{,}18$; $\sum Y = 62$; $\sum Y^2 = 390$; $\sum XY = 339{,}1$; n=12.
   a. Tentukan nilai dugaan b_1 dan b_0!
   b. Tuliskan persamaan regresi dugaannya!

3. Tabel berikut adalah hasil observasi terhadap sampel acak yang terdiri dari 8 desa di kota "Alfabet" mengenai pendapatan dan pengeluaran kesehatan penduduk desa bersangkutan selama tahun 2010.

Desa	Pendapatan (juta rupiah)	Peng Kesehatan (juta rupiah)
A	21	4
B	15	3
C	15	3.5
D	9	2

Desa	Pendapatan (juta rupiah)	Peng Kesehatan (juta rupiah)
E	12	3
F	18	3.5
G	6	2.5
H	12	2.5

a. Dengan menggunakan least square error methods, tentukan persamaan regresi linear sederhana pengeluaran kesehatan terhadap pendapatan. Kemudian jelaskan arti koefisien yang terdapat dalam persamaan tersebut.
b. Berapakah rata-rata pengeluaran kesehatan penduduk suatu desa yang memiliki rata-rata pendapatan penduduknya sebesar Rp 25 juta per tahun.
c. Hitung indeks determinasinya, kemudian jelaskan artinya.
d. Lakukan uji t dan uji F dengan menggunakan $\alpha = 5\%$, bagaimana kesimpulan dari kedua pengujian koefisien regresi tersebut.

BAB III. ANALISA REGRESI BERGANDA

A. PENDAHULUAN

Analisis regresi berganda merupakan perluasan dari analisis regresi linier sederhana. Dalam regresi linier sederhana, dibuat analisis hubungan dua variabel (satu variabel *independent* dengan satu variabel *dependent*) yang dinyatakan dengan persamaan linier

$\hat{Y}$ = a + bX, dengan tujuan membuat prediksi tentang besarnya nilai

Y (variabel dependent) berdasarkan nilai X (variabel independent) tertentu. Prediksi perubahan variabel dependent (Y) akan menjadi lebih baik apabila dimasukkan lebih dari satu variabel independent dalam persamaan liniernya (X_1, X_2,……..X_n).

Hubungan antara lebih dari satu variabel *independent* dengan satu variabel *dependent* inilah yang dibicarakan dalam analisis regresi linier berganda. Hubungan antara banyak variabel inilah yang sesungguhnya terjadi dalam dunia nyata, karena sebenarnya kebanyakan hubungan antar variabel dalam ilmu soisal merupakan hubungan statistikal, artinya bahwa perubahan nilai Y tidak mutlak hanya dipengaruhi oleh satu nilai X tertentu tetapi dipengaruhi oleh banyak nilai X. Model regresi berganda dengan 1 variabel dependent (Y) dengan n variabel independent (X) adalah :

$$\hat{Y} = a + b_1.X_1 + b_2.X_2 + \ldots\ldots + b_n.X_n + e$$

Misalnya untuk n = 2, model regresinya adalah :

$$\hat{Y} = a + b_1.X_1 + b_2.X_2 + e$$

dengan

$\hat{Y}$ = nilai Y prediksi

X_1 = Variabel bebas 1

X_2 = Variabel bebas 2

b_1 = Koefisien regresi variabel bebas 1, adalah perubahan pada Y untuk setiap perubahan X_1 sebesar 1 unit dengan asumsi X_2 konstan

b_2 = Koefisien regresi variabel bebas 2, adalah perubahan pada Y untuk setiap perubahan X_2 sebesar 1 unit dengan asumsi X_1 konstan

e = Kesalahan Prediksi (error)

B. ESTIMASI PARAMETER

Analisis regresi linier berganda, berdasarkan penelitian sampel dinyatakan dengan persamaan linier :

$$\hat{Y} = a + b_1.X_1 + b_2.X_2 + \ldots\ldots + b_n.X_n + e$$

Untuk kasus dua variabel independent, persamaan liniernya dinyatakan sebagai :

$$\hat{Y} = a + b_1.X_1 + b_2.X_2$$

Untuk mendapatkan nilai a, b_1 dan b_2 digunakan rumus-rumus sebagai berikut :

$$a = \overline{Y} - b_1\overline{X}_1 - b_2\overline{X}_2$$

$$b_1 = \frac{\left(\sum X_2^2\right)\left(\sum X_1 Y\right) - \left(\sum X_1 X_2\right)\left(\sum X_2 Y\right)}{\left(\sum X_1^2\right)\left(\sum X_2^2\right) - \left(\sum X_1 X_2\right)^2}$$

$$b_2 = \frac{\left(\sum X_1^2\right)\left(\sum X_2 Y\right) - \left(\sum X_1 X_2\right)\left(\sum X_1 Y\right)}{\left(\sum X_1^2\right)\left(\sum X_2^2\right) - \left(\sum X_1 X_2\right)^2}$$

Untuk memberikan gambaran bagaimana membuat analisis regresi linier berganda, diberikan contoh sebagai berikut : Misalnya kita hendak memprediksi besarnya pengeluaran untuk bahan makanan per bulan (variabel Y) berdasarkan penghasilan keluarga per bulan (variabel X_1) dan banyaknya/besar keluarga (variabel X_2). Berdasarkan sampel random 15 keluarga diperoleh informasi sebagai berikut :

Penghasilan Kelurga X_1	Besar Keluarga X_2	Pengeluaran Bahan Makanan (Y)
5,5	1	0,8
8,9	1	1,0
21,8	1	1,7
6,8	2	1,4
7,5	2	1,2
17,2	2	1,8
22,1	2	1,9
19,0	3	2,3
12,0	3	1,7

Penghasilan Kelurga X_1	Besar Keluarga X_2	Pengeluaran Bahan Makanan (Y)
14,0	4	1,5
10,9	4	1,8
7,5	5	2,0
14,0	5	2,2
13,7	6	2,8
6,0	7	2,1

Untuk memperoleh persamaan garis regresi linier tentang hubungan antara variabel penghasilan keluarga (X_1) dan besar keluarga (X_2) dengan variabel pengeluaran untuk bahan makanan (Y) periksa tabel berikut :

x1	x2	y	x1 y	x2 y	x1 x2	x1^2	x2^2	y^2
5.5	1	0.8	4.4	0.8	5.5	30.25	1	0.64
8.9	1	1	8.9	1	8.9	79.21	1	1
21.8	1	1.7	37.06	1.7	21.8	475.24	1	2.89
6.8	2	1.4	9.52	2.8	13.6	46.24	4	1.96
7.5	2	1.2	9	2.4	15	56.25	4	1.44
17.2	2	1.8	30.96	3.6	34.4	295.84	4	3.24
22.1	2	1.9	41.99	3.8	44.2	488.41	4	3.61
19	3	2.3	43.7	6.9	57	361	9	5.29
12	3	1.7	20.4	5.1	36	144	9	2.89
14	4	1.5	21	6	56	196	16	2.25
10.9	4	1.8	19.62	7.2	43.6	118.81	16	3.24
7.5	5	2	15	10	37.5	56.25	25	4
14	5	2.2	30.8	11	70	196	25	4.84
13.7	6	2.8	38.36	16.8	82.2	187.69	36	7.84
6	7	2.1	12.6	14.7	42	36	49	4.41
186.9	48	26.2	343.31	93.8	567.7	2767.19	204	49.54

$\sum_{i=1}^{15} X1_i = 186,9$ $\sum_{i=1}^{15} X2_i = 48$ $\sum_{i=1}^{15} X1_i Y_i = 343,31$ $\sum_{i=1}^{15} X1_i X2_i = 567,7$

$\sum_{i=1}^{15} Y_i = 26,2$ $\sum_{i=1}^{15} Y_i^2 = 49,54$ $\sum_{i=1}^{15} X2_i Y_i = 93,8$ $\sum_{i=1}^{15} X1_i^2 = 2767,19$

Koefisien regresinya adalah :

$$b_1 = \frac{(\sum X_2^2)(\sum X_1 Y) - (\sum X_1 X_2)(\sum X_2 Y)}{(\sum X_1^2)(\sum X_2^2) - (\sum X_1 X_2)^2} = \frac{(204)(343,31) - (567,7)(93,8)}{(2767,19)(204) - (567,7)^2} = 0,069295$$

$$b_2 = \frac{(\sum X_1^2)(\sum X_2 Y) - (\sum X_1 X_2)(\sum X_1 Y)}{(\sum X_1^2)(\sum X_2^2) - (\sum X_1 X_2)^2} = \frac{(2767,19)(93,8) - (567,7)(343,31)}{(2767,19)(204) - (567,7)^2} = 0,266966$$

$$a = \overline{Y} - b_1 \overline{X}_1 - b_2 \overline{X}_2 = 1,746667 - 0,069295(12,6) - 0,266966(3,2) = 0.028956$$

Persamaan regresi linier bergandanya adalah :

$$\hat{Y} = 0,028956 + 0,069295.X_1 + 0,266966.X_2$$

Pengertian persamaan tersebut adalah : Pertama, apabila X_2 konstan, pertambahan satu unit pada X_1 akan mempunyai pengaruh menaikkan 0,069295 unit pada Y. Kedua, apabila X_1 konstan, pertambahan satu unit pada X_2, akan mempunyai pengaruh menaikkan 0,266966 unit pada Y. Ketiga, apabila X_1 dan X_2 sama dengan nol, besarnya Y adalah 0,028956 satuan.

Berdasarkan persamaan tersebut dapat dibuat prediksi/ramalan nilai-nilai Y berdasarkan kombinasi nilai X_1 dan X_2 tertentu misalnya nilai $X_1 = 5,5$ dan $X_2 = 1$, maka nilai Y adalah

$\hat{Y}$ = 0, 028956 + 0, 069295 (5,5) + 0, 266966 (1) = 0,6770445

C. STANDAR EROR

Standard error of estimates dinyatakan dengan rumus :

$$S_{y12} = \sqrt{\frac{\sum Y^2 - b_1 \sum X_1 Y - b_2 \sum X_2 Y}{n-k}}$$

dengan

n = jumlah observasi

k = banyak koefisien

Berdasarkan contoh tersebut dimuka, besarnya standard error of estimate adalah :

$$S_{y12} = \sqrt{\frac{\sum Y^2 - b_1 \sum X_1 Y - b_2 \sum X_2 Y}{n-k}}$$

$$S_{y12} = \sqrt{\frac{49{,}54 - 0{,}069295(343{,}31) - 0{,}266966(93{,}9)}{15-3}}$$

$$S_{y12} = 0{,}243038$$

D. ANALISA KORELASI BERGANDA

Analisis korelasi berganda merupakan perluasan dari analisis korelasi sederhana. Dalam analisis korelasi berganda bertujuan untuk mengetahui bagaimana derajat hubungan antara beberapa variabel independent (Variabel X_1, X_2, ……., X_k) dengan variabel dependent (Variabel Y) secara bersama-sama.

Asumsi-asumsi sehubungan dengan analisis regresi berganda tersebut adalah :

1. Variabel-Variabel independent dan variabel dependent mempunyai hubungan linier
2. Semua variabel, baik variabel-variabel independent maupun variabel dependent, merupakan variabel-variabel random kontinyu.
3. Distribusi kondisional nilai masing-masing variabel berdistribusi normal (multivariate normal distribution)
4. Untuk berbagai kombinasi nilai variabel yang satu dengan yang lain tertentu, varaince dari distribusi kondisional masing-masing variabel adalah homogen (asumsu homoscedasticity berlaku untuk semua variabel)
5. Untuk masing-masing variabel, nilai observasi yang satu dengan yang lain, tidak berkaitan.

Berdasarkan korelasi berganda, yang diberi notasi $R_{Y.12.....n}$ dihitung melalui jalur terjadinya hubungan antara beberapa variabel independent (X_1, X_2, ……., X_n) dengan satu variabel dependent (Y), yakni yang berupa regresi linier berganda

$$\hat{Y} = a + b_1.X_1 + b_2.X_2 + \ldots\ldots + b_n.X_n.$$

Berdasarkan adanya regresi berganda tersebut, koefisien korelasi linier berganda tersebut dihitung dengan menggunakan rumus sebagai berikut :

$$R_{Y.12} = \sqrt{\frac{b_1 \sum X_1 Y + b_2 \sum X_2 Y + ...+ b_n \sum X_n Y}{\sum Y^2}}$$

Berdasarkan contoh tentang hubungan antara penghasilan keluarga (X_1) dan besar keluarga (X_2) dengan pengeluaran untuk bahan makanan (Y), koefisien korelasi linier bergandanya dinyatakan dengan :

$$R_{Y.12} = \sqrt{\frac{b_1 \sum X_1 Y + b_2 \sum X_2 Y}{\sum Y^2}}$$

Jadi koefisien korelasi berganda dari contoh tersebut adalah :

$$R_{Y.12} = \sqrt{\frac{(0{,}069295)(343{,}31) + (0{,}266966)(93{,}8)}{49{,}54}} = 0{,}99282$$

Sedang koefisien determinasi berganda (R^2)dari contoh tersebut adalah :

$$R^2_{Y.12} = \frac{b_1 \sum X_1 Y + b_2 \sum X_2 Y}{\sum Y^2} \quad atau \quad \frac{SSR}{SST}$$

$$R^2_{Y.12} = \frac{(0{,}069295)(343{,}31) + (0{,}266966)(93{,}8)}{49{,}54} = 0{,}985692$$

Angka tersebut menunjukkan bahwa sekitar 98,5692% dari variasi pengeluaran untuk bahan makanan (Y) dijelaskan oleh kombinasi dari penghasilan keluarga (X_1) dan besar keluarga (X_2). Sisanya yakni 1,44% dijelaskan oleh variabel independent lainnya yang tidak teramati.

Dari kasus korelasi linier berganda, peneliti dapat menghitung koefisien korelasi parsialnya. Korelasi parsial (partial correlation) adalah korelasi antara sebuah variabel dependent (Y) dengan sebuah variabel independent (X), sementara sejumlah variabel independent lainnya konstan.

Apabila variabel independentnya ada dua buah yaitu X_1 dan X_2, maka koefisien parsial yang ada ialah r_{Y12} dan r_{Y21}, yang masing-masing menunjukkan koefisien korelasi antara Y dengan X_1 apabila X_2 konstan dan koefisien korelasi antara Y dengan X_2 apabila X_1 konstan. Seperti dalam contoh tersebut dimuka, r_{Y12} menunjukkan korelasi antara penghasilan keluarga (X_1) dengan pengeluaran untuk bahan makanan (Y) apabila besar keluarga (X_2) konstan. Dan r_{Y21} menunjukkan korelasi antara besar keluarga (X_2) dengan pengeluaran untuk bahan makanan (Y) apabila penghasilan keluarga (X_1) konstan.

Rumus-rumusnya adalah :

$$r_{Y.1(2)} = \frac{rX_1Y - (rX_2Y)(rX_1X_2)}{\sqrt{(1-r^2X_2Y)(1-r^2X_1X_2)}}$$

$$r_{Y.2(1)} = \frac{rX_2Y - (rX_1Y)(rX_1X_2)}{\sqrt{(1-r^2X_1Y)(1-r^2X_1X_2)}}$$

Untuk menghitung koefisien korelasi parsialnya terlebih dahulu harus dihitung koefisien korelasi sederhana antara X_1 dengan Y, X_2 dengan Y dan antara X_1 dengan X_2.

$$r_{X_1Y} = \frac{\sum X_1Y}{\sqrt{(\sum X_1^2)(\sum Y^2)}} = \frac{343{,}31}{370{,}2521} = 0{,}927233$$

$$r_{X_2Y} = \frac{\sum X_2Y}{\sqrt{(\sum X_2^2)(\sum Y^2)}} = \frac{93{,}8}{100{,}5294} = 0{,}93306037$$

$$r_{X_1X_2} = \frac{\sum X_2X_2}{\sqrt{(\sum X_1^2)(\sum X_2^2)}} = \frac{567{,}7}{751{,}3366} = 0{,}75558677$$

Koefisien korelasi parsial :

$$r_{Y.1(2)} = \frac{0{,}927 - (0{,}933)(0{,}76)}{\sqrt{(1-(0{,}933)^2)(1-(0{,}76)^2)}} = \frac{0{,}218}{(0{,}3605)(0{,}6499)} = 0{,}9304$$

$$r_{Y.2(1)} = \frac{0{,}933 - (0{,}927)(0{,}76)}{\sqrt{(1-(0{,}927)^2)(1-(0{,}76)^2)}} = \frac{0{,}2285}{(0{,}3751)(0{,}6499)} = 0{,}9372$$

Koefisien determinasi dan pengertiannya :

1. $r^2{}_{X_1Y}$ = $(0{,}927233)^2$ = 0,8598

   Sekitar 85,98% dari variasi pengeluaran untuk bahan makanan (Y) dapat dijelaskan oleh variasi dalam penghasilan keluarga (X_1) dimana faktor lain tidak dipertimbangkan.

2. $r^2{}_{X_2Y}$ = $(0{,}9331)^2$ = 0,8707

Sekitar 87,07% dari variasi pengeluaran untuk bahan makanan (Y) dapat dijelaskan oleh variasi dalam besar keluarga (X_2) dimana faktor lain tidak dipertimbangkan.

3. $r^2_{Y_{1(2)}} = (0{,}9304)^2 = 0{,}8656$

   Sekitar 86,56% dari variasi pengeluaran untuk bahan makanan (Y) dapat dijelaskan oleh variasi dalam penghasilan keluarga (X_1), apabila pengaruh dari besar keluarga (X_2) dianggap konstan.

4. $r^2_{Y_{2(1)}} = (0{,}9372)^2 = 0{,}8783$

   Sekitar 87,83% dari variasi pengeluaran untuk bahan makanan (Y) dapat dijelaskan oleh variasi dalam besar keluarga (X_2), apabila pengaruh dari penghasilan keluarga (X_1) dianggap konstan.

B. SOAL

Jawablah pertanyaan-pertanyaan di bawah ini dengan benar!

1. Berikut adalah data kompetensi (X) dan kinerja pegawai (Y) kita ambil sampel acak 15 orang pegawai sebagai berikut :

X	*Y*
40	4
55	16
32	12
55	24
50	15
52	24
61	22
44	17
30	4
22	14
40	24
64	26
58	20
48	9
44	14

   a. Apa terdapat hubungan antara kompetensi dengan kinerja pegawai?
   b. Berapa besar hubungan kompetensi mempengaruhi kinerja pegawai?
   c. Tentukan persamaan regresi antara kompetensi dengan kinerja pegawai!

2. Suatu sampel acak terdiri atas 20 keluarga di suatu daerah, memberikan data:

X	15	20	25	20	25	30	16	15	25	20
Y	10	15	20	16	22	25	15	14	10	18

X	16	18	20	25	30	25	19	10	20	20
Y	12	15	15	20	25	23	16	8	15	17

X = pendapatan keluarga perbulan dalam ratusan ribu rupiah

Y = pengeluaran keluarga perbulan dalam ratusan ribu rupiah

a. Jika diduga bahwa hubungan antara pendapatan keluarga dan pengeluaran keluarga linear, tentukan persamaan regresi dugaannya

b. Bila dianggap asumsi-asumsi dalam analisis regresi linear terpenuhi, ujilah apakah ada hubungan antara pendapatan keluarga perbulan dan pengeluaran keluarga perbulan. Gunakan α = 0,05.

3. Suatu penelitian dilakukan terhadap 20 mahasiswa semester satu yang diambil secara acak untuk menentukan apakah nilai mutu rata-rata (NMR) pada akhir tahun pertama (Y) dapat diprediksi dari nilai ujian masuk (X). Data yang diperoleh sbb.

X	5,5	4,8	4,7	3,9	4,5	6,2	6,0	5,2	4,7	4,3
Y	3,1	2,3	3,0	1,9	2,5	3,7	3,4	2,6	2,8	1,6

X	4,9	5,4	5,0	6,3	4,6	4,3	5,0	5,9	4,1	4,7
Y	2,0	2,9	2,3	3,2	1,8	1,4	2,0	3,8	2,2	1,5

a. Jika hubungan antar NMR dan nilai ujian masuk dapat dinyatakan dengan garis linear, tentukan persamaan regresi linear dugaannya.

b. Bila dianggap asumsi-asumsi dalam analisis regresi linear terpenuhi, ujilah apakah ada hubungan antara nilai ujian masuk dan nilai mutu rata-rata (NMR) pada akhir tahun pertama. Gunakan α = 0,05.

c. Tentukan nilai dugaan untuk NMR jika nilai ujian masuk 6,0

4. Bagian kepegawaian suatu perusahaan menggunakan 12 orang dalam suatu penelitian untuk menentukan hubungan antara nilai prestasi kerja (Y) dan nilai empat tes, yaitu tes kemampuan di bidang IT (X_1), kemampuan berbahasa

Inggris (X_2), kemampuan bekerja sama (X_3), dan kemampuan berkomunikasi (X_4). Datanya adalah sebagai berikut

Y	X_1	X_2	X_3	X_4
11,2	56,5	71,0	38,5	43,0
14,5	59,5	72,5	38,2	44,8
17,2	69,2	76,0	42,5	49,0
17,8	74,5	79,5	43,5	56,3
19,3	81,2	84,0	47,5	60,2
24,5	88,0	86,2	47,4	62,0
21,2	78,2	80,0	44,5	58,1
16,9	69,0	72,0	41,8	48,1
14,8	58,1	68,0	42,1	46,0
20,0	80,5	85,0	48,1	60,3
13,2	58,3	71,0	37,5	47,1
22,5	84,0	87,2	51,0	65,2

a. Ujilah apakah ada hubungan linear antara nilai prestasi kerja (y) dan nilai empat tes, yaitu tes kemampuan di bidang IT, kemampuan berbahasa Inggris, dan kemampuan bekerja sama, kemampuan berkomunikasi. Gunakan α = 0,05.
b. Manakah diantara empat variable yang secara signifikan berpengaruh terhadap prestasi kerja?
c. Berdasarkan hasil b) Tentukan persamaan regresi linear dugaannya.
d. Lakukan uji asumsi dalam analisis regresi linear dan simpulkan hasilnya.

5. Daya rentang produk fiber sintetis diperkirakan berhubungan dengan persentase bahan katun dalam fiber, waktu pengeringan fiber. Hasil percobaan terhadap 10 potong fiber yang diproduksi dalam beberapa kondisi yang berbeda diberikan pada Tabel berikut

Y	X_1	X_2
213	13	2,1

220	15	2,3
216	14	2,2
225	18	2,5
235	19	3,2
218	20	2,4
239	22	3,4
243	17	4,0
233	16	4,
240	18	4.3

a. Lakukan analisis regresi untuk menguji apakah ada hubungan linear antara persentase bahan katun dalam fiber dan waktu pengeringan dengan daya rentang fiber sintetis.

b. Tentukan persaman regresi dugaannya.

c. Lakukan uji asumsi dalam analisis regresi linear dan simpulkan hasilnya.

6. Membahas hubungan antara kompetensi lulusan (X_1) dan kompetensi waktu kerja (X_2) dengan kinerja pegawai (Y). Untuk tujuan itu maka kita ambil sampel acak sebagai berikut :

No Subyek	$\boldsymbol{Y}$	$\boldsymbol{X_1}$	$\boldsymbol{X_2}$
1	6	12	10
2	7	14	11
3	8	10	14
4	8	16	13
5	9	18	15
6	10	24	20
7	5	12	8
8	12	30	16
9	6	10	12
10	7	6	9

Tentukan persamaan regresi ganda $\boldsymbol{Y}$ atas X_1 dan X_2

BAB VIII
ANALISA RUNTUN WAKTU

Capaian Pembelajaran	menyajikan, mengolah, dan menganalisis data
Kemampuan Akhir Pembelajaran	mahasiswa dapat menjelaskan aplikasi statistika di bidang teknik informatika

A. PENDAHULUAN

Runtun waktu adalah susunan observasi berurut menurut waktu. Suatu runtun waktu z_t dapat dipandang sebagai suatu realisasi dari proses stokastik (statistik). Jika fkp gabungan dari runtun waktu $z_1, z_2, ..., z_t, ..., z_n$ tidak dipengaruhi oleh perubahan waktu maka runtun waktu tersebut disebut stasioner. Jika tidak demikian maka disebut runtun waktu nonstasioner.

B. MODEL STASIONER

1. Metode Pemulusan Eksponensial Tunggal (*Single Exponential Smoothing*)

Digunakan untuk data runtut waktu yang mengikuti pola stasioner.

Bentuk umum yang digunakan untuk menghitung ramalan adalah:

$$\hat{Y}_{t+1} = \alpha Y_t + (1-\alpha)\hat{Y}_t$$

dengan

$\hat{Y}_{t+1}$: nilai ramalan untuk periode berikutnya

α : konstanta pemulusan

Y_t : data baru atau nilai Y yg sebenarnya pada periode t

$\hat{Y}_t$: nilai pemulusan yang lama atau rata-rata pemulusan hingga periode t-1

Contoh. Peramalan pengiriman alat pembuka kaleng listrik dengan menggunakan pemulusan eksponensial. Agar dapat memulai sistem peramalan SES kita memerlukan $\hat{Y}_1$, karena $\hat{Y}_2 = \alpha Y_1 + (1-\alpha)\hat{Y}_1$. Karena nilai $\hat{Y}_1$ tidak diketahui, maka kita dapat menggunakan nilai observasi pertama (Y_1) sebagai ramalan pertama.

Bulan	waktu	Data	Nilai pemulusan eksponensial			Nilai Kesalahan			Kuadrat nilai kesalahan		
			α = 0,1	α = 0,5	α = 0,9	α = 0,1	α = 0,5	α = 0,9	□ = 0,1	□ = 0,5	□ = 0,9
Jan	1	200.0	200.0	200.0	200.0	-	-	-	-	-	-
Peb	2	135.0	200.0	200.0	200.0	-65.0	-65.0	-65.0	4225.0	4225.0	4225.0
Maret	3	195.0	193.5	167.5	141.5	1.5	27.5	53.5	2.3	756.3	2862.3
April	4	197.5	193.7	181.3	189.7	3.8	16.3	7.8	14.8	264.1	61.6
Mei	5	310.0	194.0	189.4	196.7	116.0	120.6	113.3	13447.9	14550.4	12833.5
Juni	6	175.0	205.6	249.7	298.7	-30.6	-74.7	-123.7	938.3	5578.2	15294.6
Juli	7	155.0	202.6	212.3	187.4	-47.6	-57.3	-32.4	2262.7	3288.3	1047.6
Agustus	8	130.0	197.8	183.7	158.2	-67.8	-53.7	-28.2	4598.4	2880.7	797.3
September	9	220.0	191.0	156.8	132.8	29.0	63.2	87.2	839.2	3989.7	7599.7
Oktober	10	277.0	193.9	188.4	211.3	83.1	88.6	65.7	6901.1	7846.8	4318.8
Nopember	11	235.0	202.2	232.7	270.4	32.8	2.3	-35.4	1073.6	5.2	1255.2
Desember	12	-	205.5	233.9	238.5	-	-	-	-	-	-
							Jumlah		34303.3	43384.6	50295.6
							Periode pengujian		10	10	10
							MSE		3430.3273	4338.4625	5029.5628

2. **Pemulusan Eksponensial Ganda: Metode Satu Parameter dari Brown**

Digunakan dalam peramalan data runtut waktu yang mengikuti suatu trend linier. Bentuk umum yang digunakan untuk menghitung ramalan adalah

$$A_t = \alpha Y_t + (1-\alpha)A_{t-1}$$

$$A'_t = \alpha A_t + (1-\alpha)A'_{t-1}$$

$$a_t = 2A_t - A'_t$$

$$b_t = \frac{\alpha}{1-\alpha}(A_t - A'_t)$$

Persamaan yang digunakan untuk membuat peramalan pada periode p yang akan datang adalah:

$$\hat{Y}_{t+p} = a_t + b_t p$$

dengan

A_t : nilai pemulusan eksponensial

A'_t : nilai pemulusan eksponensial ganda

α : konstanta pemulusan

a_t : perbedaan antara nilai-nilai pemulusan eksponensial

b_t : faktor penyesuai tambahan = pengukuran slope suatu kurva

Y_t : nilai aktual pada periode t

p : jumlah periode ke depan yang akan diramalkan

Agar dapat memulai sistem peramalan metode Brown kita memerlukan A_1 dan A'_1, karena $A_2 = \alpha Y_2 + (1-\alpha)A_1$ dan $A'_2 = \alpha A_2 + (1-\alpha)A'_1$

Karena pada saat t = 1, nilai A_1 dan A'_1 tidak diketahui, maka kita dapat menggunakan nilai observasi pertama (Y_1).

Contoh. Pemulusan eksponensial dari Brown pada data permintaan suatu produk. Perhitungan pada contoh di bawah ini menggunakan nilai α = 0,2.

Periode	Permintaan suatu produk	A_t	A'_t	a_t	b_t	Nilai ramalan	
1	143	143	143	143.0	0.0		
2	152	144.8	143.4	146.2	0.4	143.0	
3	161	148.0	144.3	151.8	0.9	146.6	
4	139	146.2	144.7	147.8	0.4	152.7	
5	137	144.4	144.6	144.1	-0.1	148.2	
6	174	150.3	145.8	154.9	1.1	144.1	
7	142	148.6	146.3	151.0	0.6	156.0	
8	141	147.1	146.5	147.7	0.2	151.5	
9	162	150.1	147.2	153.0	0.7	147.9	
10	180	156.1	149.0	163.2	1.8	153.7	
11	164	157.7	150.7	164.6	1.7	164.9	
12	171	160.3	152.6	168.0	1.9	166.3	
13	-	-	-	-	-	169.9	(p =1)
14	-	-	-	-	-	171.9	(p=2)
15	-	-	-	-	-	173.8	(p=3)
16	-	-	-	-	-	175.7	(p=4)

Periode	Permintaan suatu produk	A_t	A'_t	a_t	b_t	Nilai ramalan
17	-	-	-	-	-	177.6

(p=5)

3. Pemulusan Eksponensial Ganda: Metode Dua Parameter dari Holt

Digunakan dalam peramalan data runtut waktu yang mengikuti suatu trend linier.

Bentuk umum yang digunakan untuk menghitung ramalan adalah:

$$A_t = \alpha Y_t + (1-\alpha)(A_{t-1} + T_{t-1})$$

$$T_t = \beta(A_t - A_{t-1}) + (1-\beta)T_{t-1}$$

Persamaan yang digunakan untuk membuat peramalan pada periode p yang akan datang adalah:

$$\hat{Y}_{t+p} = A_t + T_t p$$

dengan

A_t : nilai pemulusan eksponensial

α: konstanta pemulusan untuk data ($0 \leq \alpha \leq 1$)

β : konstanta pemulusan untuk estimasi trend ($0 \leq \beta \leq 1$)

Y_t : nilai aktual pada periode t

T_t : estimasi trend

p : jumlah periode ke depan yang akan diramalkan

Agar dapat memulai sistem peramalan metode Brown kita memerlukan A_1, karena $A_2 = \alpha Y_2 + (1-\alpha)(A_1 + T_1)$, karena pada saat t = 1, nilai A_1 tidak diketahui, maka kita dapat menggunakan nilai observasi pertama (Y_1). Untuk estimasi trend pada saat t = 1, nilai T_1 tidak diketahui, maka kita dapat menggunakan selisih nilai observasi kedua (Y_2) dengan nilai observasi pertama (Y_1), yaitu $T_1 = Y_2 - Y_1$.

Contoh. Pemulusan eksponensial dari Holt pada data permintaan suatu produk. Perhitungan pada contoh di bawah ini menggunakan nilai α = 0,2 dan β = 0,3.

Periode	Permintaan suatu produk	A_t	T_t	Nilai ramalan	
1	143	143	9		
2	152	152.0	9.0	152.0	
3	161	161.0	9.0	161.0	
4	139	163.8	7.1	170.0	
5	137	164.2	5.1	170.9	
6	174	170.2	5.4	169.3	
7	142	168.9	3.4	175.6	
8	141	166.0	1.5	172.2	
9	162	166.4	1.2	167.5	
10	180	170.1	1.9	167.6	
11	164	170.4	1.4	172.0	
12	171	171.6	1.4	171.8	
13	-	-	-	173.0	(p =1)
14	-	-	-	174.4	(p=2)
15	-	-	-	175.8	(p=3)
16	-	-	-	177.2	(p=4)
17	-	-	-	178.6	(p=5)

4. **Pemulusan Eksponensial Ganda: Metode Tiga Parameter dari Winter**

Digunakan dalam peramalan data runtut waktu yang mengikuti suatu pola musiman.

Didasarkan pada 3 persamaan pemulusan, yaitu: untuk unsur stasioner, untuk trend, dan untuk musiman.

Bentuk umum yang digunakan untuk menghitung ramalan adalah:

Pemulusan eksponensial $A_t = \alpha \frac{Y_t}{S_{t-L}} + (1-\alpha)(A_{t-1} + T_{t-1})$

Estimasi trend $T_t = \beta(A_t - A_{t-1}) + (1-\beta)T_{t-1}$

Estimasi musiman $S_t = \mu \frac{Y_t}{A_t} + (1-\mu)S_{t-L}$

Persamaan yang digunakan untuk membuat peramalan pada periode p yang akan datang adalah:

$$\hat{Y}_{t+p} = (A_t + T_t p) S_{t-L+p}$$

dengan

A_t : nilai pemulusan eksponensial

α : konstanta pemulusan untuk data ($0 \leq \alpha \leq 1$)

β : konstanta pemulusan untuk estimasi trend ($0 \leq \beta \leq 1$)

μ : konstanta pemulusan untuk estimasi musiman ($0 \leq \mu \leq 1$)

Y_t : nilai aktual pada periode t

T_t : estimasi trend

S_t : estimasi musiman

L : panjangnya musim

p : jumlah periode ke depan yang akan diramalkan

Dalam metode ini diperlukan estimasi nilai awal yg akan digunakan untuk mendapatkan nilai pemulusan awal, estimasi trend awal, dan keempat estimasi musiman. Nilai pemulusan awal dapat diestimasi dengan menggunakan nilai aktual awal. Nilai trend awal dapat diestimasi dengan menggunakan nilai 0 (slope persamaan trend yang diperoleh dari data masa masa lalu tidak ada). Nilai estimasi pengaruh musiman awal dengan menggunakan nilai 1 (untuk menghilangkan penaruh musiman dalam data asli $Y_1 \rightarrow Y_1/S_1 = Y_1/1 = Y_1$

Contoh. Pemulusan eksponensial dari Winter pada data musiman di bawah ini. Perhitungan pada contoh di bawah ini menggunakan nilai α = 0,4 ; β = 0,1 ; dan μ = 0,3. Panjang musim L = 4

Periode	Aktual	A_t	T_t	S_t	Ramalan
1	500	500	0	1	
2	350	440.0	-6.0	0.94	
3	250	360.4	-13.4	0.91	
4	400	368.2	-11.2	1.03	
5	450	394.2	-7.5	1.04	357.0
6	350	381.2	-8.1	0.93	362.9

Periode	Aktual	A_t	T_t	S_t	Ramalan
7	200	311.9	-14.2	0.83	338.8
8	300	295.6	-14.4	1.02	305.5
9	350	303.0	-12.2	1.08	293.2
10	200	…	…	…	…
11	150	…	…	…	…
12	400	…	…	…	…
13	550	…	…	…	…
14	350	…	…	…	…
15	250	…	…	…	…
16	550	…	…	…	…
17	550	…	…	…	…
18	400	…	…	…	…
19	350	…	…	…	…
20	600	…	…	…	…
21	750	…	…	…	…
22	500	…	…	…	…
23	400	…	…	…	…
24	650	…	…	…	…
25					…
26					…
27					…
28					…

D. MODEL NON STASIONER

Floros (2005) menjelaskan bahwa ARMA merupakan bentuk model runtun waktu linear yang berusaha untuk mengidentifikasikan persamaan dengan hanya menggunakan nilai masa lalunya atau kombinasi nilai masa lalu dan eror masa lalunya. Model ARMA mengandung dua komponen yaitu model AR dan MA dengan order dari AR adalah p dan order dari MA adalah q.

Berikut adalah model stasioner menurut Cryer (2010)

1. *Autoregressive* (AR)

Autoregressive (AR) adalah model rata-rata yang menggambarkan suatu pengamatan pada waktu t dipengaruhi pada nilai-nilai pengamatan sepanjang p periode sebelumnya. Bentuk umum model *autoregressive* orde p adalah

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t \qquad (2.1)$$

Model AR(1)

Model AR(1) adalah besarnya nilai-nilai pengamatan pada waktu t dipengaruhi oleh nilai-nilai pengamatan sepanjang 1 periode sebelumnya, didefinisikan berikut

$$Y_t = \phi_1 Y_{t-1} + e_t$$

Model AR(1) dengan $e_t \sim WN(\mu_e, \sigma_e^2)$. Model AR(1) merupakan model stasioner. Suatu proses dikatakan stasioner jika tidak dipengaruhi pada nilai t.

$$Y_t = \phi_1 Y_{t-1} + e_t$$

$$Y_{t-1} = \phi_1 Y_{t-2} + e_{t-1}$$

$$Y_{t-2} = \phi_1 Y_{t-3} + e_{t-2}$$

$$Y_{t-3} = \phi_1 Y_{t-4} + e_{t-3}$$

$$Y_t = e_t + \phi_1 e_{t-1} + \phi_1^2 e_{t-2} + \cdots + \phi_1^k e_{t-k} + \cdots + \phi_1^{t-1} e_1$$

Mean model autoregresi orde 1 diperoleh sebagai berikut

$$E(Y_t) = \frac{\mu_e(1 - \phi_1^t)}{(1 - \phi_1)}$$

Untuk $t \to \infty$ dan $|\phi_1| < 1$ maka

$$E(Y_t) = \frac{\mu_e}{(1 - \phi_1)}$$

Variansi model autoregresi orde 1 diperoleh sebagai berikut

$$Var(Y_t) = \frac{\sigma_e^2(1 - \phi_1^{2t})}{(1 - \phi_1^2)}$$

Untuk $t \to \infty$ dan $|\phi_1| < 1$ maka

$$Var(Y_t) = \frac{\sigma_e^2}{(1 - \phi_1^2)}$$

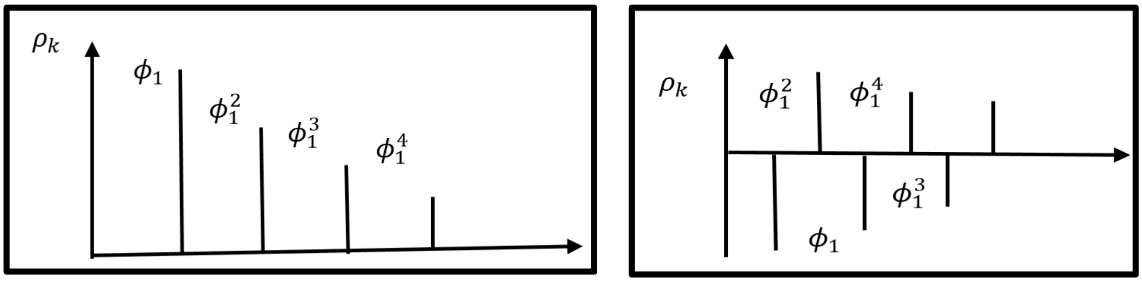
Kovariansi model AR(1) untuk $k = 1$ diperoleh sebagai berikut

$$Cov(Y_tY_{t-1}) = \frac{\phi_1\sigma_e^2(1-\phi_1^{2t})}{(1-\phi_1^2)}$$

Untuk $t \to \infty$ dan $|\phi_1| < 1$ maka

$$\gamma_1 = \frac{\phi_1\sigma_e^2}{(1-\phi_1^2)}$$

kemudian secara umum untuk $k = p$ diperoleh $\gamma_p = \frac{\phi_1^p\sigma_e^2}{(1-\phi_1^2)}$. Autokorelasi (ACF) diperoleh $\rho_p = \phi_1^p$ sehingga diperoleh grafik ACF sebagai berikut



(a) $0 < \phi_1 < 1$ (b) $-1 < \phi_1 < 0$

Gambar 2.1 Grafik ACF Model AR

Berdasarkan Gambar 2.1 terlihat grafik ACF untuk $0 < \phi_1 < 1$ turun cepat secara eksponensial dan untuk $-1 < \phi_1 < 0$ turun cepat secara sinusoidal.

Model AR(p)

Model AR(p) adalah besarnya nilai-nilai pengamatan pada waktu t dipengaruhi oleh nilai-nilai pengamatan sepanjang p periode sebelumnya, didefinisikan seperti pada Persamaan (2.1). Model AR(p) dengan $e_t \sim WN(\mu_e, \sigma_e^2)$. Autokovariansi model AR(p) diperoleh sebagai berikut

$$\gamma_k = cov(Y_tY_{t-k})$$

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \cdots + \phi_p\gamma_{k-p}$$

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \cdots + \phi_p\rho_{k-p}$$

untuk $k = 1$ diperoleh $\rho_1 = \phi_1 + \phi_2\rho_1 + \cdots + \phi_p\rho_{p-1}$

untuk $k = 2$ diperoleh $\rho_2 = \phi_1\rho_1 + \phi_2 + \cdots + \phi_p\rho_{2-p}$

untuk $k = p$ diperoleh $\rho_p = \phi_1\rho_{p-1} + \phi_{p-2} + \cdots + \phi_p$

sehingga diperoleh variansi model AR(p) sebagai berikut

$$\gamma_0 = \phi_1\gamma_1 + \phi_2\gamma_2 + \cdots + \phi_p\gamma_p + \sigma_e^2$$

2. *Moving Average* (MA)

Moving average (MA) adalah model stasioner yang nilai-nilai pengamatan pada waktu t dipengaruhi oleh penyimpangan pengamatan sepanjang satu periode sebelumnya. Bentuk umum model *moving average* orde q adalah

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \tag{2.2}$$

Model MA(1)

Model MA(1) adalah model stasioner yang nilai-nilai pengamatan pada waktu t dipengaruhi oleh penyimpangan pengamatan sepanjang satu periode sebelumnya, didefinisikan berikut

$$Y_t = e_t - \theta_1 e_{t-1}$$

dengan $e_t \sim WN(\mu_e, \sigma_e^2)$. Mean model MA(1) diperoleh sebagai berikut

$$E(Y_t) = E(e_t) - \theta_1 E(e_{t-1}) = (1 - \theta_1)\mu_e$$

Variansi model MA(1) diperoleh sebagai berikut

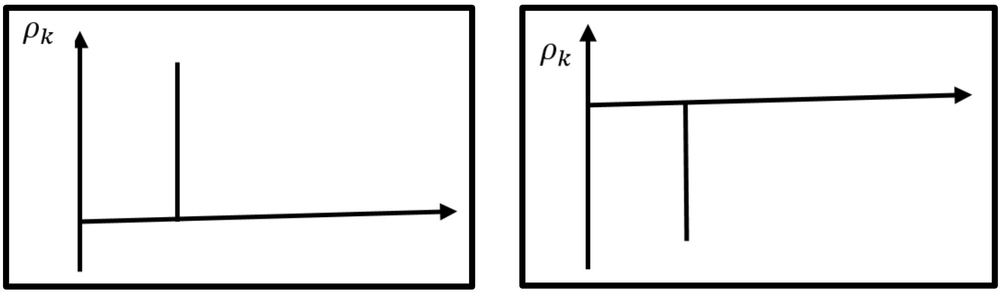
$$var(Y_t) = \gamma_0 = (1 + \theta_1^2)\sigma_e^2$$

Autokovariansi model MA(1) untuk $k = 1$ diperoleh $cov(Y_t Y_{t-1}) = \gamma_1 = -\theta_1\sigma_e^2$. Untuk $k = 2,3,4, \ldots$ diperoleh $cov(Y_t Y_{t-2}) = \gamma_2 = 0$

Fungsi autokorelasi model MA(1)

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} \dfrac{-\theta_1}{1 + \theta_1^2} & ; k = 1 \\ 0 & ; k \geq 2 \end{cases}$$

sehingga diperoleh grafik ACF sebagai berikut



(a) $-1 < \theta_1 < 0$ (b) $0 < \theta_1 < 1$

Gambar 2.2 Grafik ACF Model MA(1)

Berdasarkan Gambar 2.2 grafik ACF model MA(1) terpotong setelah lag pertama. Hal ini dapat digunakan untuk mengidentifikasi suatu data mengikuti model MA(1).

Model MA(2)

Model MA(2) adalah model stasioner yang nilai-nilai pengamatan pada waktu t dipengaruhi oleh penyimpangan pengamatan sepanjang dua periode sebelumnya, didefinisikan berikut

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$$

dengan $e_t \sim WN(\mu_e, \sigma_e^2)$. Mean model MA(2) diperoleh sebagai berikut

$$E(Y_t) = (1 - \theta_1 - \theta_2)\mu_e$$

Variansi model MA(2) diperoleh sebagai berikut

$$var(Y_t) = \gamma_0 = (1 + \theta_1^2 + \theta_2^2)\sigma_e^2$$

Autokovariansi model MA(2)

untuk $k = 1$ diperoleh $cov(Y_t Y_{t-1}) = \gamma_1 = -\theta_1(1 - \theta_2)\sigma_e^2$

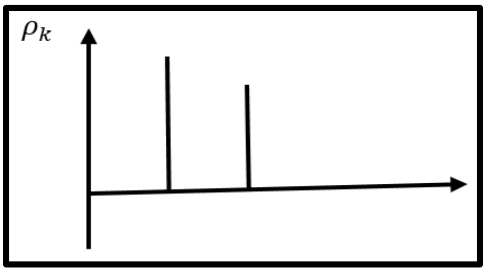
untuk $k = 2$ diperoleh $cov(Y_t Y_{t-2}) = \gamma_2 = -\theta_2\sigma_e^2$

untuk $k = 3, 4, \ldots$ diperoleh $cov(Y_t Y_{t-3}) = \gamma_3 = 0$

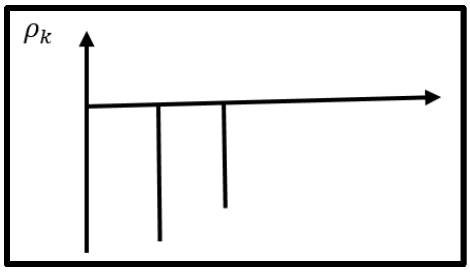
Fungsi autokorelasi model MA(2)

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} \dfrac{-\theta_1(1 - \theta_2)}{(1 + \theta_1^2 + \theta_2^2)} & ; k = 1 \\ \dfrac{-\theta_2}{(1 + \theta_1^2 + \theta_2^2)} & ; k = 2 \\ 0 & ; k \geq 3 \end{cases}$$

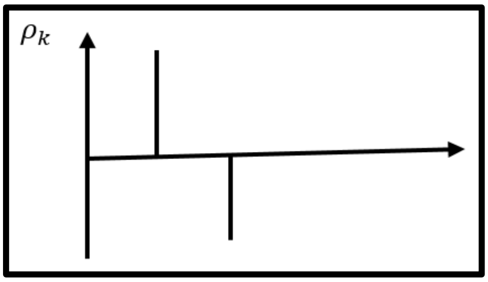
sehingga diperoleh grafik ACF sebagai berikut



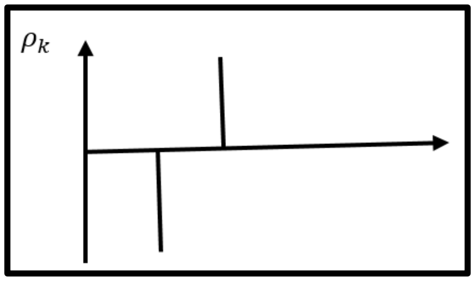
(a) $-1 < \theta_1 < 0; -1 < \theta_2 < 0$



(b) $0 < \theta_1 < 1; 0 < \theta_2 < 1$



(c) $-1 < \theta_1 < 0; 0 < \theta_2 < 1$



(d)$0 < \theta_1 < 1; -1 < \theta_2 < 0$

Gambar 2.3 Grafik ACF Model MA(2)

Berdasarkan Gambar 2.3 grafik ACF model MA(2) terpotong setelah lag kedua. Hal ini dapat digunakan untuk mengidentifikasi suatu data mengikuti model MA(2).

Model MA(q)

Model MA(q) adalah model stasioner yang nilai-nilai pengamatan pada waktu t dipengaruhi oleh penyimpangan pengamatan sepanjang q periode sebelumnya, didefinisikan seperti pada Persamaan (2.2). dengan $e_t \sim WN(\mu_e, \sigma_e^2)$. Mean model MA($q$) diperoleh $E(Y_t) = \left(1 - \theta_1 - \theta_2 - \cdots - \theta_q E\left(e_{t-q}\right)\right)\mu_e$. Variansi model MA($q$) diperoleh $var(Y_t) = \gamma_0 = \left(1 + \theta_1^2 \ldots + \theta_q^2\right)\sigma_e^2$. Autokovariansi model MA(q) untuk $k = 1, 2, \ldots, q$ diperoleh

$$cov(Y_t Y_{t-k}) = \gamma_k = \left(-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_k\right)\sigma_q^2$$

Untuk $k = q + 1$ diperoleh $cov\left(Y_t Y_{t-q-1}\right) = \gamma_{q+1} = 0$.

Fungsi autokorelasi model MA(q)

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} \frac{\left(-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_k\right)}{\left(1 + \theta_1^2 \ldots + \theta_q^2\right)} & ; k = 1,2,\ldots,q \\ 0 & ; k > q \end{cases}$$

3. *Autoregressive Moving Average* (ARMA)

Autoregressive Moving Average (ARMA) adalah gabungan antara AR dan MA, berikut adalah model umum ARMA (*p,q*).

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \quad (2.3)$$

ARMA(1,1)

ARMA(1,1) adalah proses autoregresif orde 1 dan proses *moving average* orde 1 sebagai berikut

$$Y_t = \phi Y_{t-1} + e_t - \theta e_{t-1}$$

sehingga diperoleh fungsi autokovariansi sebagai berikut

untuk $k = 0$ diperoleh

$$E(Y_t Y_t) = \gamma_0 = \phi\gamma_1 + \sigma_e^2 - \theta(\phi - \theta)\sigma_e^2 \quad (2.4)$$

untuk $k = 1$ diperoleh

$$E(Y_t Y_{t-1}) = \gamma_1 = \phi\gamma_0 - \theta\sigma_e^2 \quad (2.5)$$

dengan substitusi Persamaan (2.5) ke Persamaan (2.4) diperoleh

$$\gamma_0 = \frac{(1-2\theta\phi+\theta^2)}{(1-\phi^2)}\sigma_e^2 \text{ dan } \gamma_1 = \frac{(1-\theta\phi)(\phi-\theta)}{(1-\phi^2)}\sigma_e^2$$

untuk $k = 2$, diperoleh fungsi autokovariansi

$$\gamma_2 = \frac{(1-\theta\phi)(\phi-\theta)}{(1-\phi^2)}\phi\sigma_e^2$$

untuk $k = k$ diperoleh

$$\gamma_k = \frac{(1-\theta\phi)(\phi-\theta)}{(1-\phi^2)}\phi^{k-1}\ \sigma_e^2$$

Sedangkan fungsi autokorelasi, untuk $k = 1$

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{(1-\theta\phi)(\phi-\theta)}{1-2\theta\phi+\theta^2}$$

untuk $k = 2$

$$\rho_2 = \frac{\gamma_2}{\gamma_0} = \frac{(1-\theta\phi)(\phi-\theta)\phi}{1-2\theta\phi+\theta^2}$$

untuk $k = k$

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{(1-\theta\phi)(\phi-\theta)\phi^{k-1}}{1-2\theta\phi+\theta^2}$$

ARMA$(\boldsymbol{p}, \boldsymbol{q})$

ARMA(p, q) adalah proses autoregresif orde p dan proses *moving average* orde q seperti pada Persamaan (2.3). Kemudian dengan mengalikan kedua ruas dengan Y_{t-k} sehingga diperoleh bentuk

$$Y_t Y_{t-k} = \phi_1 Y_{t-1} Y_{t-k} + \cdots + \phi_p Y_{t-p} Y_{t-k} + e_t Y_{t-k} - \theta_1 e_{t-1} Y_{t-k} - \cdots \\ - \theta_q e_{t-q} Y_{t-k}$$

sehingga fungsi autokovariansi

$$\gamma_k = \phi_1 \gamma_{k-1} + \cdots + \phi_p \gamma_{k-p} + E(e_t Y_{t-k}) - \theta_1 E(e_{t-1} Y_{t-k}) - \cdots \\ - \theta_q E\left(e_{t-q} Y_{t-k}\right)$$

karena $E(e_{t-i} Y_{t-k}) = 0$ untuk $k > i$ maka $\gamma_k = \phi_1 \gamma_{k-1} + \cdots + \phi_p \gamma_{k-p}$ untuk $k \geq q + 1$ dan fungsi autokorelasi diperoleh sebagai berikut

$$\rho_k = \phi_1 \rho_{k-1} + \cdots + \phi_p \rho_{k-p} \text{ untuk } k \geq q + 1$$

2.1 Estimasi Parameter Model Stasioner

2.8.1 Estimasi Parameter Model Autoregressive (AR)

Model AR(1)

Berikut adalah model autoregresi orde satu dengan mean μ

$$Y_t - \mu = \phi_1 (Y_{t-1} - \mu) + e_t$$

Persamaaan tersebut dapat dilihat sebagai model regresi dengan variabel prediktor Y_{t-1} dan variabel respon Y_t. Estimasi kuadrat terkecil dihasilkan dengan meminimalkan jumlah kuadrat erornya.

$$(Y_t-\mu)-\phi_1(Y_{t-1}-\mu)$$

Karena hanya $Y_1, Y_2, \dots, Y_n$ yang diamati, maka diperoleh jumlahan dari $t=2$ sampai $t=n$ sebagai berikut

$$S_c(\phi_1,\mu)=\sum_{t=2}^{n}[(Y_t-\mu)-\phi_1(Y_{t-1}-\mu)]^2$$

kemudian estimasi dari mean $(\hat{\mu})$ diperoleh melalui

$$\frac{\partial S_c}{\partial \mu}=\sum_{t=2}^{n}2[(Y_t-\mu)-\phi_1(Y_{t-1}-\mu)]\,(-1+\phi_1)=0$$

$$\mu=\frac{1}{(n-1)(1-\phi_1)}\left[\sum_{t=2}^{n}Y_t-\phi_1\sum_{t=2}^{n}Y_{t-1}\right]$$

untuk n besar dapat diperoleh

$$\frac{1}{(n-1)}\sum_{t=2}^{n}Y_t\approx\frac{1}{(n-1)}\sum_{t=2}^{n}Y_{t-1}\approx\bar{Y}$$

$$\hat{\mu}\approx\frac{1}{(1-\phi_1)}[\bar{Y}-\phi_1\bar{Y}]=\bar{Y}$$

kemudian estimasi dari $\phi_1$diperoleh

$$\frac{\partial S_c(\phi_1,\bar{Y})}{\partial \phi_1}=\sum_{t=2}^{n}2[(Y_t-\bar{Y})-\phi_1(Y_{t-1}-\bar{Y})]\,(Y_{t-1}-\bar{Y})=0$$

$$\hat{\phi}_1=\frac{\sum_{t=2}^{n}(Y_t-\bar{Y})(Y_{t-1}-\bar{Y})}{\sum_{t=2}^{n}(Y_{t-1}-\bar{Y})^2}$$

Model AR(2)

Berikut adalah model autoregresi orde dua dengan mean μ

$$Y_t-\mu=\phi_1(Y_{t-1}-\mu)+\phi_2(Y_{t-2}-\mu)+e_t$$

Persamaaan tersebut dapat dilihat sebagai model regresi dengan variabel prediktor Y_{t-1} dan variabel respon Y_t. Estimasi kuadrat terkecil dihasilkan dengan meminimalkan jumlah kuadrat erornya.

$$(Y_t-\mu)-\phi_1(Y_{t-1}-\mu)-\phi_2(Y_{t-2}-\mu)$$

$$S_c(\phi_1,\phi_2,\mu)=\sum_{t=3}^{n}[(Y_t-\mu)-\phi_1(Y_{t-1}-\mu)-\phi_2(Y_{t-2}-\mu)]^2$$

kemudian estimasi dari mean ($\hat{\mu}$) diperoleh melalui

$$\frac{\partial S_c}{\partial \mu} = \sum_{t=3}^{n} 2[(Y_t - \mu) - \phi_1(Y_{t-1} - \mu) - \phi_2(Y_{t-2} - \mu)]\,(-1 + \phi_1 + \phi_2) = 0$$

$$\mu = \frac{1}{(n-1)(1-\phi_1-\phi_2)}\left[\sum_{t=3}^{n} Y_t - \phi_1 \sum_{t=3}^{n} Y_{t-1} - \phi_2 \sum_{t=3}^{n} Y_{t-2}\right]$$

untuk n besar dapat diperoleh

$$\frac{1}{(n-1)}\sum_{t=3}^{n} Y_t \approx \frac{1}{(n-1)}\sum_{t=3}^{n} Y_{t-1} \approx \frac{1}{(n-1)}\sum_{t=3}^{n} Y_{t-2} \approx \bar{Y}$$

$$\hat{\mu} \approx \frac{1}{(1-\phi_1-\phi_2)}[\bar{Y} - \phi_1\bar{Y} - \phi_2\bar{Y}] = \bar{Y}$$

kemudian diperoleh estimasi dari $\phi_1$dan ϕ_2 berikut

$$\hat{\phi}_1 = \frac{r_1(1-r_2)}{1-r_1^2} \text{ dan } \hat{\phi}_2 = \frac{r_1^2 - r_2}{r_1^2 - 1}$$

dengan

$$r_1 = \frac{\sum_{t=3}^{n}(Y_t - \bar{Y})(Y_{t-1} - \bar{Y})}{\sum_{t=3}^{n}(Y_{t-1} - \bar{Y})^2} \text{ dan } r_2 = \frac{\sum_{t=3}^{n}(Y_t - \bar{Y})(Y_{t-2} - \bar{Y})}{\sum_{t=3}^{n}(Y_{t-2} - \bar{Y})^2}$$

Model AR(p)

Berikut adalah model autoregresi orde p dengan mean μ

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \cdots + \phi_p(Y_{t-p} - \mu) + e_t$$

Persamaaan tersebut dapat dilihat sebagai model regresi dengan variabel prediktor Y_{t-1} dan variabel respon Y_t. Estimasi kuadrat terkecil dihasilkan dengan meminimalkan jumlah kuadrat erornya.

$$S_c(\phi_1, \phi_2, \dots, \phi_p, \mu) = \sum_{t=p+1}^{n} [(Y_t - \mu) - \phi_1(Y_{t-1} - \mu) - \phi_2(Y_{t-2} - \mu) - \cdots - \phi_p(Y_{t-p} - \mu)]^2$$

kemudian estimasi dari mean ($\hat{\mu}$) diperoleh

$$\hat{\mu} \approx \frac{1}{(1-\phi_1-\phi_2-\cdots--\phi_p)}[\bar{Y} - \phi_1\bar{Y} - \phi_2\bar{Y} - \cdots - \phi_2\bar{Y}] = \bar{Y}$$

Estimasi dari $\phi_1$diperoleh

$$\frac{\partial S_c(\phi_1, \phi_2, \dots, \phi_p, \bar{Y})}{\partial \phi_1} = 0$$

$\sum_{t=3}^{n} 2\left[(Y_t - \bar{Y}) - \phi_1(Y_{t-1} - \bar{Y}) - \phi_2(Y_{t-2} - \bar{Y}) - \cdots - \phi_p\left(Y_{t-p} - \bar{Y}\right)\right](Y_{t-1} - \bar{Y}) = 0$

$$\hat{\gamma}_1 = \hat{\gamma}_0\, \phi_1 + \hat{\gamma}_1\phi_2 + \hat{\gamma}_2\phi_3 + \cdots + \hat{\gamma}_{p-1}\phi_p$$

$$\hat{\rho}_1 = \phi_1 + \hat{\rho}_1\phi_2 + \hat{\rho}_2\phi_3 + \cdots + \hat{\rho}_{p-1}\phi_p$$

$$r_1 = \phi_1 + r_1\phi_2 + r_2\phi_3 + \cdots + r_{p-1}\phi_p$$

Estimasi dari ϕ_2 diperoleh

$$\frac{\partial S_c(\phi_1, \phi_2, \dots, \phi_p \bar{Y})}{\partial \phi_2} = 0$$

$\sum_{t=3}^{n} 2\left[(Y_t - \bar{Y}) - \phi_1(Y_{t-1} - \bar{Y}) - \phi_2(Y_{t-2} - \bar{Y}) - \cdots - \phi_p\left(Y_{t-p} - \bar{Y}\right)\right](Y_{t-2} - \bar{Y}) = 0$

$$\hat{\gamma}_2 = \hat{\gamma}_1\, \phi_1 + \hat{\gamma}_0\phi_2 + \hat{\gamma}_1\phi_3 + \cdots + \hat{\gamma}_{p-2}\phi_p$$

$$\hat{\rho}_2 = \hat{\rho}_1\phi_1 + \phi_2 + \hat{\rho}_1\phi_3 + \cdots + \hat{\rho}_{p-2}\phi_p$$

$$r_2 = r_1\phi_1 + \phi_2 + r_1\phi_3 + \cdots + r_{p-2}\phi_p$$

Kemudian estimasi dari ϕ_p diperoleh

$$\frac{\partial S_c(\phi_1, \phi_2, \dots, \phi_p \bar{Y})}{\partial \phi_p} = 0$$

$\sum_{t=3}^{n} 2\left[(Y_t - \bar{Y}) - \phi_1(Y_{t-1} - \bar{Y}) - \phi_2(Y_{t-2} - \bar{Y}) - \cdots - \phi_p\left(Y_{t-p} - \bar{Y}\right)\right]\left(Y_{t-p} - \bar{Y}\right) = 0$

$$\hat{\gamma}_p = \hat{\gamma}_{p-1}\, \phi_1 + \hat{\gamma}_{p-2}\phi_2 + \hat{\gamma}_{p-3}\phi_3 + \cdots + \phi_p$$

$$\hat{\rho}_p = \hat{\rho}_{p-1}\phi_1 + \hat{\rho}_{p-2}\phi_2 + \hat{\rho}_{p-3}\phi_3 + \cdots + \phi_p$$

$$r_p = r_{p-1}\phi_1 + r_{p-2}\phi_2 + r_{p-3}\phi_3 + \cdots + \phi_p$$

sehingga diperoleh p persamaan linear sebagai berikut

$$r_1 = \phi_1 + r_1\phi_2 + r_2\phi_3 + \cdots + r_{p-1}\phi_p$$

$$r_2 = r_1\phi_1 + \phi_2 + r_1\phi_3 + \cdots + r_{p-2}\phi_p$$

$$\vdots \qquad\qquad\qquad\qquad \vdots$$

$$r_p = r_{p-1}\phi_1 + r_{p-2}\phi_2 + r_{p-3}\phi_3 + \cdots + \phi_p$$

persamaan linear tersebut dapat diubah ke bentuk matriks sebagai berikut

$$\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_p \end{bmatrix} = \begin{bmatrix} 1 & r_1 & r_2 & \cdots & r_{p-1} \\ r_1 & r_2 & r_3 & \cdots & r_{p-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{p-1} & r_{p-2} & r_{p-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} \qquad (2.6)$$

misal untuk $k = 2$ diperoleh

$$\begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} 1 & r_1 \\ r_1 & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}$$

maka diperoleh estimasi parameter AR(2)

$$\hat{\phi}_1 = \frac{r_1(1 - r_2)}{1 - r_1^2} \quad \text{dan} \quad \hat{\phi}_2 = \frac{r_2 - r_1^2}{1 - r_1^2}$$

untuk $k = 3$ diperoleh

$$\begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} = \begin{bmatrix} 1 & r_1 & r_2 \\ r_1 & r_2 & r_3 \\ r_2 & r_3 & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{bmatrix}$$

maka diperoleh estimasi parameter AR(3)

$$\hat{\phi}_1 = \frac{\begin{vmatrix} r_1 & r_1 & r_2 \\ r_2 & r_2 & r_3 \\ r_3 & r_3 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & r_1 & r_2 \\ r_1 & r_2 & r_3 \\ r_2 & r_3 & 1 \end{vmatrix}} \quad \hat{\phi}_2 = \frac{\begin{vmatrix} 1 & r_1 & r_2 \\ r_1 & r_2 & r_3 \\ r_2 & r_3 & 1 \end{vmatrix}}{\begin{vmatrix} 1 & r_1 & r_2 \\ r_1 & r_2 & r_3 \\ r_2 & r_3 & 1 \end{vmatrix}} \quad \text{dan} \quad \hat{\phi}_3 = \frac{\begin{vmatrix} 1 & r_1 & r_1 \\ r_1 & r_2 & r_2 \\ r_2 & r_3 & r_3 \end{vmatrix}}{\begin{vmatrix} 1 & r_1 & r_2 \\ r_1 & r_2 & r_3 \\ r_2 & r_3 & 1 \end{vmatrix}}$$

Untuk $k = p$ diperoleh

$$\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_p \end{bmatrix} = \begin{bmatrix} 1 & r_1 & r_2 & \cdots & r_{p-1} \\ r_1 & r_2 & r_3 & \cdots & r_{p-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{p-1} & r_{p-2} & r_{p-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix}$$

misal

$$\mathbf{R} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_p \end{bmatrix}, \mathbf{P} = \begin{bmatrix} 1 & r_1 & r_2 & \cdots & r_{p-1} \\ r_1 & r_2 & r_3 & \cdots & r_{p-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{p-1} & r_{p-2} & r_{p-3} & \cdots & 1 \end{bmatrix}, \mathbf{\Phi} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix}$$

maka Persamaan (2.6) dapat ditulis menjadi

$$\mathbf{R} = \mathbf{P\Phi}$$

$$\mathbf{\Phi} = \mathbf{P}^{-1}\mathbf{R}$$

maka diperoleh estimasi parameter AR(p) adalah $\widehat{\mathbf{\Phi}} = \widehat{\mathbf{P}}^{-1}\mathbf{R}$.

2.8.2 Estimasi Parameter Model ARMA

Model ARMA(1,1)

Model ARMA(1,1) adalah model stasioner yang nilai pengamatan ke-t dipengaruhi oleh nilai pengamatan dan penyimpangan pada periode $t-1$. Untuk membentuk fungsi *likelihood* dari n pengamatan, dilakukan pengamatan mulai dari pengamatan pertama sampai pengamatan ke-n. Pengamatan pertama proses ARMA(1,1) misal adalah vektor $\mathbf{Y}_{11} = \begin{bmatrix} y_1 \\ e_1 \end{bmatrix}$ dengan $\boldsymbol{\mu}_{11} = \begin{bmatrix} \mu \\ 0 \end{bmatrix}$ diperoleh matriks variansi-kovariansi sebagai berikut

$$\mathbf{\Omega} = \sigma_e^2 \boldsymbol{V}_{11} = \begin{bmatrix} \gamma_0 & \sigma_e^2 \\ \sigma_e^2 & \sigma_e^2 \end{bmatrix}$$

maka *pdf* untuk pengamatan pertama proses ARMA(1,1) sebagai berikut

$$f(y_1;\phi_1;\theta_1)$$

$$= (2\pi\sigma_e^2)^{\frac{1}{2}} \left(\frac{1}{\sigma_e^{-2}\gamma_0 - 1}\right)^{\frac{1}{2}} \exp\left(-\frac{1}{2\sigma_e^2} [y_1 - \mu \quad e_1] \begin{bmatrix} \dfrac{1}{\sigma_e^{-2}\gamma_0 - 1} & \dfrac{-1}{\sigma_e^{-2}\gamma_0 - 1} \\ \dfrac{-1}{\sigma_e^{-2}\gamma_0 - 1} & \dfrac{\sigma_a^{-2}\gamma_0}{\sigma_e^{-2}\gamma_0 - 1} \end{bmatrix} \begin{bmatrix} y_1 - \mu \\ e_1 \end{bmatrix} \right)$$

Pengamatan kedua dinyatakan sebagai berikut

$$y_2 - \mu = \phi_1(y_1 - \mu) + e_2 - \theta_1 e_1 \qquad (2.7)$$

Persamaan (2.7) memperlihatkan bahwa y_2 adalah jumlahan dari nilai konstan $\phi_1(y_1 - \mu) - \theta_1 e_1$ dengan $e_1 \sim N(0, \sigma_e^2)$ dan $e_2 \sim N(0, \sigma_e^2)$. Pengamatan kedua y_2 dengan syarat y_1, $y_2|y_1 \sim N(\phi_1(y_1 - \mu) - \theta_1 e_1, \sigma_e^2)$, sehingga *pdf* pengamatan kedua dengan syarat y_1 diperoleh sebagai berikut

$$f(y_2|y_1,\phi_1,\theta_1)=(2\pi\sigma_e^2)^{1/2}\exp\left(-\frac{1}{2\sigma_e^2}\big((y_2-\mu)-\phi_1(y_1-\mu)+\theta_1 e_1\big)^2\right)$$

kemudian pdf bersama untuk pengamatan kedua dan pengamatan pertama diperoleh

$$f(y_2,y_1,\phi_1,\theta_1)$$

$$=(2\pi\sigma_e^2)^{-1}\left(\frac{1}{\sigma_e^{-2}\gamma_0-1}\right)^{\frac{1}{2}}\exp\left(-\frac{1}{2\sigma_e^2}\left(\frac{1}{\sigma_e^{-2}\gamma_0-1}\right)((y_1-\mu)^2\right.$$
$$\left.-2(y_1-\mu)e_1+e_1^2\sigma_e^{-2}\gamma_0)+\big((y_2-\mu)-\phi_1(y_1-\mu)+\theta_1 e_1\big)^2\right)$$

sehingga pdf dari n pengamatan ARMA(1,1) dinyatakan dengan

$$f(y_n,\dots,y_1,\phi_1,\theta_1)=f(y_1,\phi_1,\theta_1)\prod_{t=2}^{n}f(y_t|y_{t-1},\phi_1,\theta_1)$$

$$f(y_n,\dots,y_1,\phi_1,\theta_1)=(2\pi\sigma_e^2)^{-n/2}\left(\frac{1}{\sigma_{ee}^{-2}\gamma_0-1}\right)^{1/2}\exp\left(-\frac{1}{2\sigma_e^2}\left(\frac{1}{\sigma_e^{-2}\gamma_0-1}\right)((y_1-\mu)^2-2(y_1-\mu)e_1+e_1^2\sigma_e^{-2}\gamma_0)+\sum_{t=2}^{n}\big((y_t-\mu)-\phi_1(y_{t-1}-\mu)+\theta_1 e_{t-1}\big)^2\right)$$

Fungsi *likelihood* diperoleh sebagai berikut

$$L(y_n,\dots,y_1,\phi_1,\theta_1)=(2\pi\sigma_e^2)^{-n/2}\left(\frac{1}{\sigma_e^{-2}\gamma_0-1}\right)^{1/2}\exp\left(-\frac{1}{2\sigma_e^2}\left(\frac{1}{\sigma_e^{-2}\gamma_0-1}\right)((y_1-\mu)^2-2(y_1-\mu)e_1+e_1^2\sigma_e^{-2}\gamma_0)+\sum_{t=2}^{n}\big((y_t-\mu)-\phi_1(y_{t-1}-\mu)+\theta_1 e_{t-1}\big)^2\right)$$

Kemudian estimasi untuk tiap parameter diperoleh dari turunan pertama ln fungsi *likelihood* terhadap tiap parameter sama dengan nol, diperoleh sebagai berikut

$$\frac{\partial \ln L(\phi_1,\theta_1)}{\partial\phi_1}=-2\left(\sum_{t=2}^{n}\big((y_t-\mu)-\phi_1(y_{t-1}-\mu)+\theta_1 e_{t-1}\big)\right)(y_{t-1}-\mu)$$
$$=0$$

$$\hat{\phi}_1=\frac{\sum_{t=2}^{n}(y_t-\mu)(y_{t-1}-\mu)+\theta_1\sum_{t=2}^{n}e_{t-1}(y_{t-1}-\mu)}{\sum_{t=2}^{n}(y_{t-1}-\mu)(y_{t-1}-\mu)}$$

$$\frac{\partial \ln L(\phi_1,\theta_1)}{\partial\theta_1}=2\left(\sum_{t=2}^{n}\big((y_t-\mu)-\phi_1(y_{t-1}-\mu)+\theta_1 e_{t-1}\big)\right)(e_{t-1})=0$$

$$\hat{\theta}_1=\frac{-\sum_{t=2}^{n}(y_t-\mu)(e_{t-1})+\phi_1\sum_{t=2}^{n}e_{t-1}(y_{t-1}-\mu)}{\sum_{t=2}^{n}(e_{t-1})(e_{t-1})}$$

Model ARMA$(\boldsymbol{p}, \boldsymbol{q})$

Model ARMA(p, q) adalah model stasioner yang nilai pengamatan ke-t dipengaruhi oleh nilai pengamatan sepanjang p periode sebelumnya dan penyimpangan sepanjang q periode sebelumnya. Bentuk umum model ARMA(p, q) adalah

$$y_t - \mu = \phi_1(y_{t-1} - \mu) + \cdots + \phi_p\left(y_{t-p} - \mu\right) + e_t - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q} \qquad (2.10)$$

dengan mengalikan kedua ruas Persamaan (2.10) dengan $y_t - \mu$ maka diperoleh variansi sebagai berikut

$$\gamma_0 = \phi_1\gamma_1 + \phi_2\gamma_2 + \cdots + \phi_p\gamma_p + \left(1 - \theta_1 - \theta_2 - \cdots - \theta_q\right)\sigma_e^2$$

Untuk membentuk fungsi *likelihood* dari n pengamatan, dilakukan pengamatan mulai dari pengamatan pertama sampai pengamatan ke- n. Pandang p pengamatan pertama sebagai vector $\mathbf{Y}_{p,q}$ berukuran $[(p+q) \times 1]$ yang merupakan realisasi dari variable berdimensi $p + q$ dan berdistribusi Normal. Rata-rata dari $\mathbf{Y}_{p,q}$ dinotasikan $\boldsymbol{\mu}_{p,q}$ berukuran $[(p+q) \times 1]$ dengan elemen μ dan $e_t \sim N(0, \sigma_e^2)$ sedangkan matriks varian kovarian dinotasikan $\boldsymbol{\Omega}$

$$\mathbf{Y}_{p,q} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \\ e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \qquad \boldsymbol{\mu}_{p,q} = \begin{bmatrix} \mu \\ \mu \\ \vdots \\ \mu \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\boldsymbol{\Omega} = \sigma_e^2 \boldsymbol{V}_{p,q} = E\left[\mathbf{Y}_{p,q}\mathbf{Y}_{p,q}{}'\right]$$

$\boldsymbol{\Omega} =$

$$\begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{p-1} & \sigma_e^2 & 0 & \cdots & 0 \\ \gamma_1 & \gamma_2 & \cdots & \gamma_{p-2} & (\phi_1-\theta_1)\sigma_e^2 & \sigma_e^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0 & (\phi_{p-1}-\theta_{q-1})\sigma_e^2 & \cdots & \cdots & \sigma_e^2 \\ \sigma_e^2 & (\phi_1-\theta_1)\sigma_e^2 & \cdots & (\phi_{p-1}-\theta_{q-1})\sigma_e^2 & \sigma_e^2 & 0 & \cdots & 0 \\ 0 & \sigma_e^2 & \cdots & (\phi_{p-1}-\theta_{q-1})\sigma_e^2 & 0 & \sigma_e^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_e^2 & 0 & 0 & \cdots & \sigma_e^2 \end{bmatrix}$$

Distribusi untuk p pengamatan pertama adalah $(y_1, y_2, \dots, y_p) \sim WN(\boldsymbol{\mu}_{p,q}, \sigma_e^2 \boldsymbol{V}_{p,q})$ pdf bersama untuk pengamatan pertama proses ARMA(p,q) sebagai berikut

$$f(y_p, \dots, y_1; \varphi) = (2\pi\sigma_e^2)^{-p/2} \left|\boldsymbol{V}_{p,q}^{-1}\right|^{1/2} \exp\left(-\frac{1}{2\sigma_e^2}(\mathbf{Y}_{p,q} - \boldsymbol{\mu}_{p,q})' \boldsymbol{V}_{p,q}^{-1} (\mathbf{Y}_{p,q} - \boldsymbol{\mu}_{p,q})\right)$$

dengan $\varphi = (\phi_1, \phi_2, \dots, \phi_p; \theta_1, \theta_2, \dots, \theta_q)$. Sedangkan (y_{p+1}) bersyarat dengan $y_1, y_2, \dots, y_p$ mempunyai distribusi

$$y_{p+1} | y_1, y_2, \dots, y_p \sim N\left(\phi_1(y_p - \mu) + \cdots + \phi_p(y_1 - \mu) - \theta_1 e_q - \cdots - \theta_q e_1, \sigma_e^2\right)$$

Sehingga pdf bersama dari *n* pengamatan pada model ARMA(p,q) diperoleh

$$f(y_n, y_{n-1}, \dots, y_1; \varphi) = f(y_1, y_2, \dots, y_p; \varphi) \prod_{t=p+1}^{n} f(y_t | y_1, y_2, \dots, y_{t-1}; \varphi)$$

$$f(y_n, y_{n-1}, \dots, y_1; \varphi) = (2\pi\sigma_e^2)^{-p/2} \left|\boldsymbol{V}_{p,q}^{-1}\right|^{1/2} \exp\left(-\frac{S(\varphi)}{2\sigma_e^2}\right)$$

Fungsi *likelihood* diperoleh sebagai berikut

$$L(y_n, y_{n-1}, \dots, y_1; \varphi) = (2\pi\sigma_e^2)^{-p/2} \left|\boldsymbol{V}_{p,q}^{-1}\right|^{1/2} \exp\left(-\frac{S(\varphi)}{2\sigma_e^2}\right)$$

dengan

$$S(\varphi) = (\mathbf{Y}_{p,q} - \boldsymbol{\mu}_{p,q})' \boldsymbol{V}_{p,q}^{-1} (\mathbf{Y}_{p,q} - \boldsymbol{\mu}_{p,q}) + \sum_{t=p+1}^{n} \left((y_t - \mu) + \cdots + \phi_p(y_1 - \mu) - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q}\right)^2$$

Kemudian estimasi untuk tiap parameter diperoleh dari turunan pertama ln fungsi *likelihood* terhadap tiap parameter sama dengan nol, diperoleh sebagai berikut

$$\hat{\phi}_1 = \frac{\sum_{t=p+1}^{n}(y_t-\mu)(y_{t-1}-\mu)-\phi_2\sum_{t=p+1}^{n}(y_{t-2}-\mu)(y_{t-1}-\mu)}{\sum_{t=p}^{n}(y_{t-1}-\mu)(y_{t-1}-\mu)}$$

$$\frac{+\theta_1\sum_{t=p+1}^{n}e_{t-1}(y_{t-1}-\mu)+\cdots+\theta_q\sum_{t=p+1}^{n}e_{t-q}(y_{t-1}-\mu)}{\sum_{t=p}^{n}(y_{t-1}-\mu)(y_{t-1}-\mu)}$$

$$\hat{\phi}_2 = \frac{\sum_{t=p+1}^{n}(y_t-\mu)(y_{t-2}-\mu)-\phi_1\sum_{t=p+1}^{n}(y_{t-1}-\mu)(y_{t-2}-\mu)}{\sum_{t=p}^{n}(y_{t-2}-\mu)(y_{t-2}-\mu)}$$

$$\frac{+\theta_1\sum_{t=p+1}^{n}e_{t-1}(y_{t-2}-\mu)+\cdots+\theta_q\sum_{t=p+1}^{n}e_{t-q}(y_{t-2}-\mu)}{\sum_{t=p}^{n}(y_{t-2}-\mu)(y_{t-2}-\mu)}$$

$$\hat{\phi}_p = \frac{\sum_{t=p+1}^{n}(y_t-\mu)(y_1-\mu)-\cdots-\phi_{p-1}\sum_{t=p+1}^{n}(y_2-\mu)(y_1-\mu)}{\sum_{t=p}^{n}(y_1-\mu)(y_1-\mu)}$$

$$\frac{+\theta_1\sum_{t=p+1}^{n}e_{t-1}(y_1-\mu)+\cdots+\theta_q\sum_{t=p+1}^{n}e_{t-q}(y_1-\mu)}{\sum_{t=p}^{n}(y_1-\mu)(y_1-\mu)}$$

$$\hat{\theta}_1 = \frac{-\sum_{t=p+1}^{n}(y_t-\mu)(e_{t-1})+\cdots+\phi_p\sum_{t=p+1}^{n}(y_1-\mu)(e_{t-1})}{\sum_{t=p}^{n}(e_{t-1})(e_{t-1})}$$

$$\frac{-\theta_2\sum_{t=p+1}^{n}e_{t-1}(e_{t-2})-\cdots-\theta_q\sum_{t=p+1}^{n}e_{t-q}(e_{t-1})}{\sum_{t=p}^{n}(e_{t-1})(e_{t-1})}$$

$$\hat{\theta}_q = \frac{-\sum_{t=p+1}^{n}(y_t-\mu)(e_{t-q})+\cdots+\phi_p\sum_{t=p+1}^{n}(y_1-\mu)(e_{t-q})}{\sum_{t=p}^{n}(e_{t-q})(e_{t-q})}$$

$$\frac{-\theta_1\sum_{t=p+1}^{n}e_{t-1}(e_{t-q})-\cdots-\theta_{q-1}\sum_{t=p+1}^{n}e_{t-q-1}(e_{t-q})}{\sum_{t=p}^{n}(e_{t-q})(e_{t-q})}$$

BAB IX
APLIKASI STATISTIKA

Capaian Pembelajaran	menyajikan, mengolah, dan menganalisis data
Kemampuan Akhir Pembelajaran	mahasiswa dapat menjelaskan aplikasi statistika di bidang teknik informatika

A. LANDASAN TEORI

Para mahasiswa tingkat awal yang mempelajari Statistika seringkali bingung dengan dua istilah ini, analisis korelasi dan analisis regresi. Kebingungannya berkisar kepada fungsi kedua alat analisis, apakah fungsinya sama atau beda ? kalau beda, kapan kita menggunakan analisis korelasi dan kapan dan pada situasi bagaimana kita menggunakan analisis regresi. Berikut ini akan dijelaskan perbedaan kedua alat analisis tersebut.

Tujuan analsisi korelasi adalah ingin mengetahui apakah ada hubungan antara dua variabel atau lebi. Sedangkan tujuan analisis regresi adalah untuk memprediksi seberapa jauh pengaruh yang ada tersebut (yang telah dianalisis melalui analisis korelasi). Misal dengan analisis korelasi ingin diketahui apakah ada hubungan antara penggunaan jejaring social dengan interaksi sosial remaja. Bila melalui analisis korelasi terbukti bahwa ada hubungan diantara kedua variabel tersebut, maka analisis regresi akan memperkirakan jika jumlah pengguna jejaring social bertambah maka akan berpengaruh dengan interaksi social remaja.

Mungkin ada juga yang bertanya dan kebetulan sudah pernah mengenal alat analisis statistika lainnya, yaitu uji-t dan uji-F (ANOVA). Perbedaan antara uji-t dengan uji-F adalah bila uji-t dan ANOVA menguji ada tidaknya perbedaan dua sampel atau lebih, maka analisis regresi menguji ada tidaknya hubungan dua variabel atau lebih.

Contoh. Mulyaningsih (2015), dalam penelitiannya membahas besarnya pengaruh pengguna jejaring social dengan besarnya interaksi social. Pengguna jejaring social facebook, twitter, dan instagram dengan indikator intensitas dan frekuensi pengguna juga interaksi sesama pengguna. Sedangkan variabel dependen interaksi social dengan indikator kontak social dan komunikasi. Berikut adalah tabel tabulasi dari penelitian tersebut.

	JEJARING SOSIAL FACEBOOK				JEJARING SOSIAL TWITTER				JEJARING SOSIAL INSTAGRAM				INTERAKSI SOSIAL			
Resp	P1	P2	P3	X1	P4	P5	P6	X2	P7	P8	P9	X3	P10	P11	P12	Y
1	3	5	4	12	3	4	3	10	4	5	2	11	5	5	4	14
2	3	3	4	10	4	4	3	11	3	4	2	9	3	3	3	9
3	4	4	4	12	5	5	4	14	3	4	3	10	5	5	5	15
4	2	5	3	10	5	5	3	13	4	4	3	11	4	4	3	11
5	3	2	2	7	5	5	3	13	4	4	3	11	5	4	3	12
6	5	3	3	11	4	5	3	12	3	4	3	10	4	3	4	11
7	4	2	3	9	3	4	4	11	3	4	4	11	3	4	3	10
8	5	5	4	14	3	4	3	10	5	4	3	12	4	3	5	12
9	4	4	4	12	3	4	4	11	4	4	3	11	4	3	4	11
10	3	4	4	11	3	4	4	11	4	4	3	11	4	4	3	11
11	3	5	4	12	5	3	3	11	4	4	4	12	4	3	4	11
12	3	3	3	9	5	3	4	12	3	4	2	9	3	3	5	11
13	3	4	2	9	4	3	5	12	5	4	3	12	3	4	3	10
14	3	5	3	11	4	4	5	13	4	4	3	11	4	3	4	11

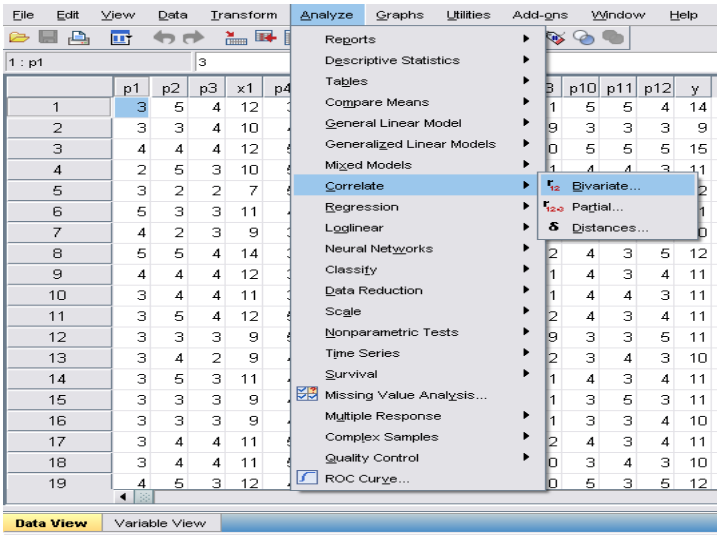
	JEJARING SOSIAL FACEBOOK				JEJARING SOSIAL TWITTER				JEJARING SOSIAL INSTAGRAM				INTERAKSI SOSIAL			
Resp	P1	P2	P3	X1	P4	P5	P6	X2	P7	P8	P9	X3	P10	P11	P12	Y
15	3	3	3	9	4	4	5	13	4	4	3	11	3	5	3	11
16	3	3	3	9	4	4	4	12	4	3	4	11	3	3	4	10
17	3	4	4	11	5	4	3	12	5	3	4	12	4	3	4	11
18	3	4	4	11	5	4	4	13	3	3	4	10	3	4	3	10
19	4	5	3	12	4	5	4	13	4	3	3	10	5	3	5	12
20	2	4	3	9	3	4	3	10	4	2	3	9	2	3	4	9
21	3	3	3	9	5	5	3	13	4	3	3	10	4	3	4	11
22	4	3	2	9	5	5	4	14	4	4	3	11	4	4	4	12
23	3	2	3	8	5	4	4	13	3	4	4	11	3	4	4	11
24	3	3	3	9	4	4	3	11	3	4	4	11	3	2	5	10
25	4	3	3	10	4	4	3	11	4	5	4	13	3	5	5	13
26	2	3	3	8	5	4	5	14	4	5	4	13	5	4	3	12
27	3	2	2	7	5	4	3	12	5	3	4	12	3	4	3	10
28	5	3	3	11	4	4	4	12	5	5	5	15	5	5	5	15
29	4	3	3	10	3	5	4	12	5	5	5	15	5	4	5	14
30	5	3	4	12	4	4	5	13	4	5	5	14	5	3	4	12
31	4	4	4	12	4	4	5	13	4	4	3	11	4	3	4	11
32	3	3	4	10	4	5	4	13	4	4	4	12	4	4	3	11
33	3	5	4	12	5	5	3	13	4	4	3	11	3	5	4	12
34	3	3	4	10	5	5	4	14	5	4	4	13	4	3	5	12
35	3	4	4	11	5	5	4	14	5	3	3	11	3	5	4	12
36	3	4	3	10	4	4	4	12	3	3	4	10	4	3	4	11
37	3	4	3	10	3	4	3	10	3	3	4	10	3	4	4	11
38	2	3	3	8	5	4	2	11	4	3	4	11	4	3	3	10
39	4	3	3	10	5	3	3	11	4	3	3	10	5	4	3	12
40	3	3	4	10	5	3	5	13	4	4	4	12	5	5	3	13
41	5	2	4	11	4	4	3	11	5	4	3	12	4	3	5	12
42	5	4	4	13	4	5	5	14	4	4	3	11	4	5	4	13
43	4	4	4	12	5	3	5	13	3	4	3	10	4	4	4	12
44	4	4	3	11	5	4	5	14	4	4	3	11	4	5	3	12
45	3	4	3	10	4	3	3	10	4	4	4	12	4	4	3	11
46	3	5	4	12	3	3	5	11	4	3	4	11	3	4	5	12
47	4	5	4	13	4	4	3	11	4	3	4	11	4	5	3	12
48	4	5	4	13	5	3	3	11	3	3	4	10	3	5	3	11
49	3	2	4	9	4	5	5	14	4	3	3	10	4	5	5	14
50	4	2	3	9	4	3	3	10	3	4	3	10	2	5	3	10
51	3	4	3	10	4	4	4	12	4	4	3	11	3	4	4	11
52	4	5	4	13	4	3	3	10	4	4	5	13	5	3	4	12
53	3	4	3	10	5	3	3	11	4	4	5	13	4	5	3	12
54	3	3	4	10	5	3	3	11	4	4	4	12	3	4	4	11
55	4	3	3	10	4	3	3	10	3	5	4	12	3	5	3	11
56	2	3	4	9	5	4	4	13	3	5	4	12	4	3	4	11
57	3	4	3	10	4	3	3	10	3	3	3	9	3	3	4	10
58	5	4	4	13	4	4	5	13	4	3	5	12	4	5	4	13
59	4	3	4	11	5	4	5	14	3	4	4	11	5	3	4	12
60	5	3	3	11	5	5	4	14	3	4	5	12	4	5	4	13

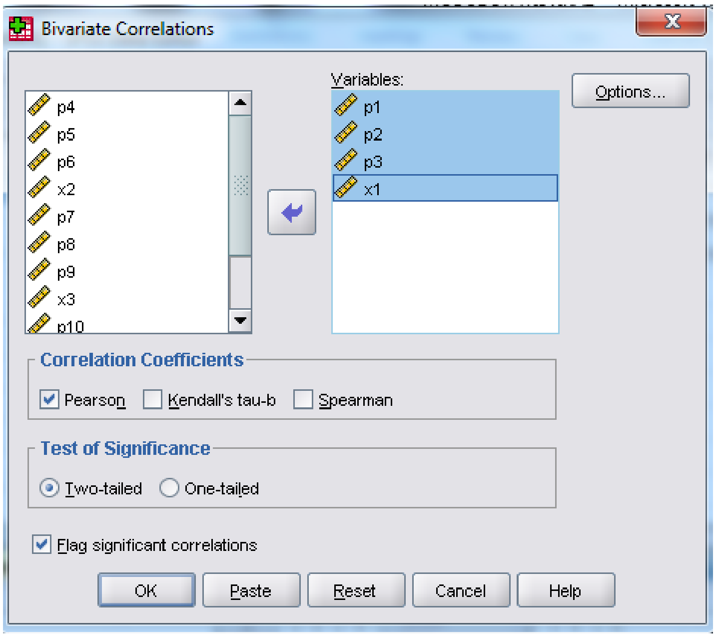
1. Uji Validitas

Uji Validitas (uji kesahihan) digunakan untuk mengetahui apakah kuesioner yang disusun tersebut itu valid atau sahih, maka perlu diuji dengan uji korelasi antara skor (nilai) tiap-tiap item pertanyaan dengan skor total kuesioner tersebut. Untuk item-item pertanyaan yang tidak valid harus dibuang atau tidak dipakai sebagai instrumen pertanyaan. Sedangkan Uji Reliabilitas (uji keandalan) merupakan suatu ukuran yang menunjukkan sejauhmana suatu alat ukur dapat dipercaya (dapat diandalkan) atau dengan kata lain menunjukkan sejauhmana hasil pengukuran tersebut tetap konsisten jika dilakukan pengukuran dua kali atau lebih terhadap gejala yang sama. Uji reliabilitas harus dilakukan hanya pada pertanyaan-pertanyaan yang sudah memenuhi uji validitas dan yang tidak memenuhi maka tidak perlu diteruskan untuk uji reliabilitas. Kedua uji tersebut di atas biasanya dilakukan pada penelitian yang nilai masing-masing variabelnya merupakan penjumlahan dari item-item pertanyaan per variabel yang diajukan (model skor).

Uji validitas dengan SPSS

Untuk melakukan Uji Validitas item pertanyaan Variabel X1 adalah dengan langkah: klik **Analiyze**, pilih **Correlate**, dan klik **Bivariate**.





Uji Validitas tiap Item Pertanyaan untuk Variabel Jejaring sosial facebook

Correlations

		P1	P2	P3	jejaring sosial facebook
P1	Pearson Correlation	1	-.041	.159	.574**
	Sig. (2-tailed)		.756	.226	.000
	N	60	60	60	60
P2	Pearson Correlation	-.041	1	.338**	.708**
	Sig. (2-tailed)	.756		.008	.000
	N	60	60	60	60
P3	Pearson Correlation	.159	.338**	1	.683**
	Sig. (2-tailed)	.226	.008		.000
	N	60	60	60	60
jejaring sosial facebook	Pearson Correlation	.574**	.708**	.683**	1
	Sig. (2-tailed)	.000	.000	.000	
	N	60	60	60	60

**. Correlation is significant at the 0.01 level (2-tailed).

Berdasarkan data tersebut pada Tabel di atas, dengan melihat nilai Pearson Correlation antara pertanyaan (p1) dengan jejaring sosial facebook (X1), pertanyaan 2 (p2) dengan jejaring sosial facebook (X1), dan pertanyaan 3 (p3) dengan jejaring sosial facebook (X1) berada pada taraf signifikansi korelasi sebesar 0,01 (lihat tanda bintang) dan nilai sig. (2-tailed) = 0,000 < 0,01 maka dapat dinyatakan bahwa item-item pertanyaan untuk Variabel jejaring sosial facebook (X1) dinyatakan valid. Semua data pada pertanyaan tersebut dapat digunakan untuk olah data berikutnya.

Uji Validitas tiap Item Pertanyaan untuk Variabel Jejaring sosial twitter

Correlations

		P4	P5	P6	jejaring sosial twitter
P4	Pearson Correlation	1	.033	-.029	.531**
	Sig. (2-tailed)		.801	.823	.000
	N	60	60	60	60
P5	Pearson Correlation	.033	1	.086	.601**
	Sig. (2-tailed)	.801		.514	.000
	N	60	60	60	60
P6	Pearson Correlation	-.029	.086	1	.647**
	Sig. (2-tailed)	.823	.514		.000
	N	60	60	60	60
jejaring sosial twitter	Pearson Correlation	.531**	.601**	.647**	1
	Sig. (2-tailed)	.000	.000	.000	
	N	60	60	60	60

**. Correlation is significant at the 0.01 level (2-tailed).

Berdasarkan data tersebut pada Tabel di atas, dengan melihat nilai Pearson Correlation antara pertanyaan 4 (p4) dengan jejaring sosial twitter (X2), pertanyaan 5 (p5) dengan jejaring sosial twitter (X2), dan pertanyaan 6 (p6) dengan jejaring sosial twitter (X2) berada pada taraf signifikansi korelasi sebesar 0,01 (lihat tanda bintang) dan nilai sig. (2-tailed) = 0,000 < 0,01 maka dapat dinyatakan bahwa item-item pertanyaan untuk Variabel jejaring sosial twitter (X2) dinyatakan valid. Semua data pada pertanyaan tersebut dapat digunakan untuk olah data berikutnya.

Uji Validitas tiap Item Pertanyaan untuk Variabel Jejaring sosial instagram

Correlations

		P7	P8	P9	jejaring sosial instagram
P7	Pearson Correlation	1	.013	.076	.560**
	Sig. (2-tailed)		.920	.566	.000
	N	60	60	60	60
P8	Pearson Correlation	.013	1	.112	.594**
	Sig. (2-tailed)	.920		.395	.000
	N	60	60	60	60
P9	Pearson Correlation	.076	.112	1	.687**
	Sig. (2-tailed)	.566	.395		.000
	N	60	60	60	60
jejaring sosial instagram	Pearson Correlation	.560**	.594**	.687**	1
	Sig. (2-tailed)	.000	.000	.000	
	N	60	60	60	60

**. Correlation is significant at the 0.01 level (2-tailed).

Berdasarkan data tersebut pada Tabel di atas, dengan melihat nilai Pearson Correlation antara pertanyaan 7 (p7) dengan jejaring sosial instagram (X3), pertanyaan 8 (p8) dengan jejaring sosial instagram (X3), dan pertanyaan 9 (p9) dengan jejaring sosial instagram (X3) berada pada taraf signifikansi korelasi sebesar 0,01 (lihat tanda bintang) dan nilai sig. (2-tailed) = 0,000 < 0,01 maka dapat dinyatakan bahwa item-item pertanyaan untuk Variabel jejaring sosial instagram (X3) dinyatakan valid. Semua data pada pertanyaan tersebut dapat digunakan untuk olah data berikutnya.

Uji Validitas tiap Item Pertanyaan untuk Variabel Interaksi sosial (Y)

Correlations

		P10	P11	P12	Interaksi sosial
P10	Pearson Correlation	1	-.030	.122	.653**
	Sig. (2-tailed)		.822	.354	.000
	N	60	60	60	60
P11	Pearson Correlation	-.030	1	-.267*	.512**
	Sig. (2-tailed)	.822		.039	.000
	N	60	60	60	60
P12	Pearson Correlation	.122	-.267*	1	.448**
	Sig. (2-tailed)	.354	.039		.000
	N	60	60	60	60
Interaksi sosial	Pearson Correlation	.653**	.512**	.448**	1
	Sig. (2-tailed)	.000	.000	.000	
	N	60	60	60	60

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Berdasarkan data tersebut pada Tabel di atas, dengan melihat nilai Pearson Correlation antara pertanyaan 10 (p10) dengan Interaksi sosial (Y), pertanyaan 11 (p11) dengan Interaksi sosial (Y), dan pertanyaan 12 (p12) dengan Interaksi sosial (Y) berada pada taraf signifikansi korelasi sebesar 0,01 (lihat tanda bintang) dan nilai sig. (2-tailed) = 0,000 < 0,01 maka dapat dinyatakan bahwa item-item pertanyaan untuk Variabel Interaksi sosial (Y) dinyatakan valid. Semua data pada pertanyaan tersebut dapat digunakan untuk olah data berikutnya.

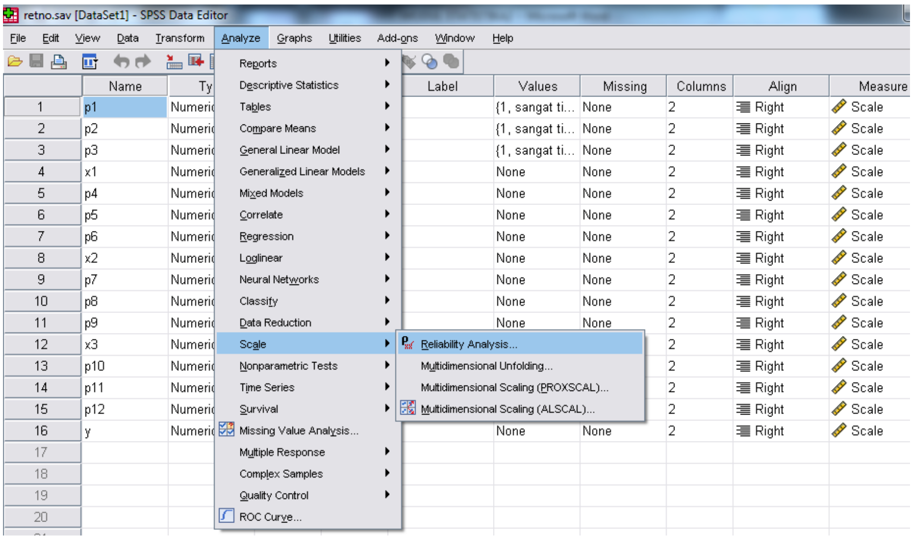
2. Uji Reliabilitas

Reliabilitas berasal dari kata reliability berarti sejauh mana hasil suatu pengukuran dapat dipercaya. Suatu hasil pengukuran dapat dipercaya apabila dalam beberapa kali pelaksanaan pengukuran terhadap kelompok subyek yang sama, diperoleh hasil pengukuran yang relatif sama, selama aspek yang diukur dalam diri subyek memang belum berubah.

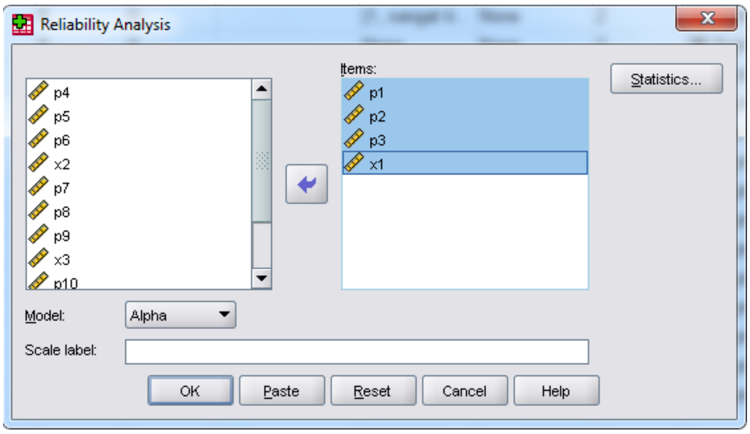
Untuk mengetahui apakah tanggapan terhadap tes atau instrumen itu mantap, konsisten atau tidak plin-plan, dapat dilakukan dengan cara memberikan tes yang sama secara berulang kali (dua kali) kepada obyek ukur atau responden yang sama. Pengetesan dua kali merupakan syarat minimal untuk mengetahui apakah tanggapan obyek ukur terhadap tes tersebut konsisten atau tidak.

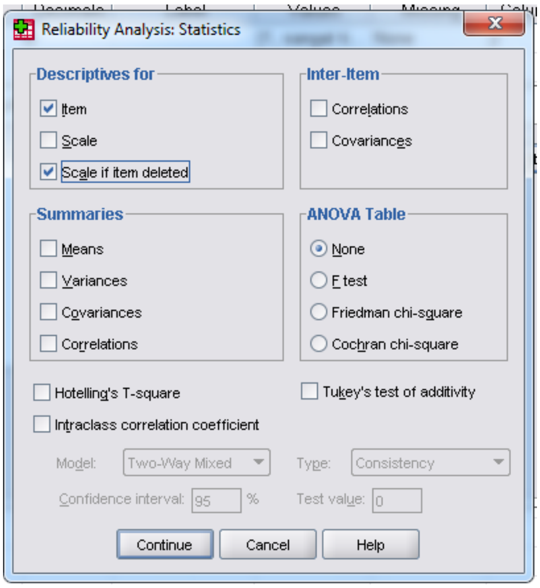
Uji Reliabilitas dengan SPSS

Langkah untuk melakukan Uji Reliabilitas, pertama kali harus kembali pada kotak SPSS Data Editor. Misalnya menguji Reliabilitas untuk Variabel X: klik **Analiyze**, pilih **Scale**, klik **Reliability Analysis….**



Setelah tampil kotak **Reliability Analysis**, pindahkan item-item pertanyaan dan Variabel X_1 ke kotak Items dengan jalan blok (p1), (p2), (p3), dan (X_1), kemudian klik tanda panah dan klik **Statistics** maka akan muncul kotak **Reliability Analysis: Statistics**.





Uji Reliabilitas Variabel Jejaring sosial facebook (X1)

Reliability Statistics

Cronbach's Alpha	N of Items
.734	4

Berdasarkan Tabel di atas, dapat dilihat bahwa nilai Cronbach's Alpha (lihat kotak Reliability Statistics) sebesar 0,734 sehingga item pertanyaan untuk mendapatkan nilai Variabel X1 dapat dikatakan reliable atau andal. Dari beberapa literatur disebutkan bahwa kriteria indeks reliabilitas adalah sebagai berikut:

Tabel Kriteria Indeks Reliabilitas

No	Interval	Kriteria
1	< 0,200	Sangat rendah

No	Interval	Kriteria
2	0,200 – 0,399	Rendah
3	0,400 – 0,599	Cukup
4	0,600 – 0,799	Tinggi
5	0,800 – 1,000	Sangat Tinggi

Uji Reliabilitas Variabel Jejaring sosial twitter (X2)

Reliability Statistics

Cronbach's Alpha	N of Items
.685	4

Berdasarkan Tabel di atas, dapat dilihat bahwa nilai Cronbach's Alpha (lihat kotak Reliability Statistics) sebesar 0,685 sehingga item pertanyaan untuk mendapatkan nilai Variabel X2 dapat dikatakan reliable atau andal.

Uji Reliabilitas Variabel Jejaring sosial instagram (X3)

Reliability Statistics

Cronbach's Alpha	N of Items
.707	4

Berdasarkan Tabel di atas, dapat dilihat bahwa nilai Cronbach's Alpha (lihat kotak Reliability Statistics) sebesar 0,707 sehingga item pertanyaan untuk mendapatkan nilai Variabel X3 dapat dikatakan reliable atau andal.

Uji Reliabilitas Variabel Interaksi sosial (Y)

Reliability Statistics

Cronbach's Alpha	N of Items
.753	4

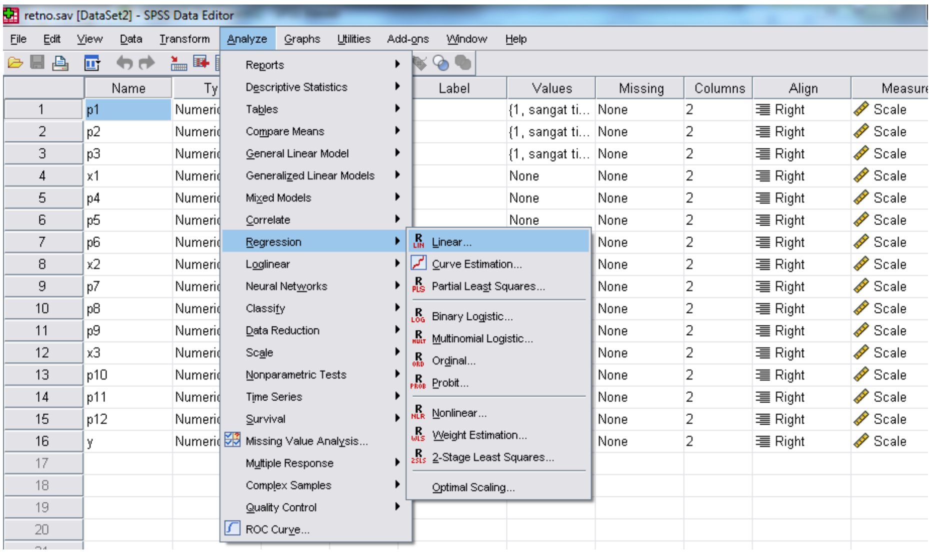
Berdasarkan Tabel di atas, dapat dilihat bahwa nilai Cronbach's Alpha (lihat kotak Reliability Statistics) sebesar 0,753 sehingga item pertanyaan untuk mendapatkan nilai Variabel X3 dapat dikatakan reliable atau andal.

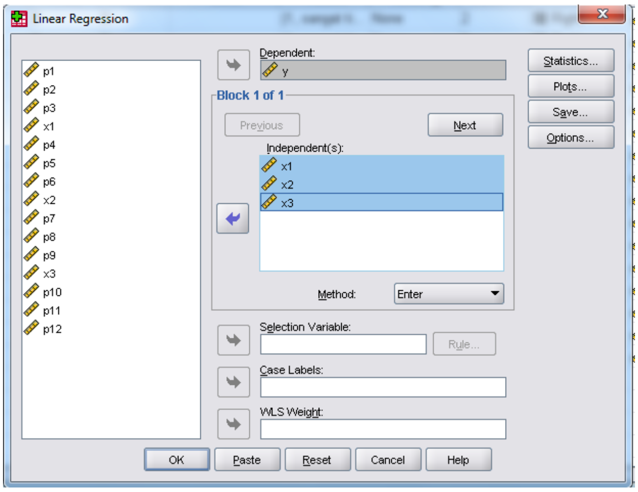
3. Persamaan Regresi Ganda

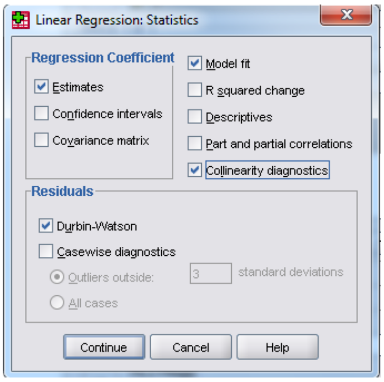
Suatu analisis yang bertujuan untuk mengetahui pengaruh suatu variabel terhadap variabel lain. Dalam analisis regresi, variabel yang mempengaruhi disebut *Independent Variable* (variabel bebas) dan variabel yang dipengaruhi disebut *Dependent Variable* (variabel terikat). Dalam persamaan regresi hanya terdapat satu variabel bebas dan satu variabel terikat, maka disebut sebagai persamaan regresi sederhana, Sedangkan jika variabel bebasnya lebih dari satu, maka disebut sebagai persamaan regresi berganda.

Analisa Regresi dengan SPSS

untuk mendapatkan nilai yang dibutuhkan untuk Uji T, Uji F, dan R^2 dengan langkah: setelah tabulasi data peneli tian sudah masuk pada kotak **Data View**, pilih menu **Analyze**, pilih sub menu **Regression**, dan klik **Linier**. Ketika tampi l kotak **Linier Regression: Statistic**, beri tanda centang pada **Covariance matrix**, **Collinearity diagnostics** (digunakan untuk Uji Multikolinieritas), dan **Durbin-Watson** (digunakan untuk Uji Autokorelasi), kemudian klik **Continue**, maka akan kembali ke kotak **Linier Regression** dan klik OK.







Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-.040	1.780		-.023	.982		
	jejaring sosial facebook	.306	.083	.371	3.683	.001	.990	1.010
	jejaring sosial twitter	.339	.096	.355	3.523	.001	.990	1.010
	jejaring sosial instagram	.382	.100	.387	3.834	.000	.989	1.011

a. Dependent Variable: interaksi sosial

Berdasarkan hasil pengolahan data analisis regresi linier pada Tabel 7.8. di atas, persamaan regresi yang dibentuk adalah : Y = -0,040 + 0,306 X1 + 0,339 X2 + 0,382 X3. Persamaan regresi linier tersebut memberikan gambaran bahwa:

a. Variabel Jejaring sosial facebook (X1) mempunyai nilai koefisien regresi sebesar 0,306 yang berarti Jejaring sosial facebook mempunyai pengaruh positif terhadap Interaksi sosial (Y), yaitu jika terjadi kenaikan jejaring sosial facebook maka akan menurunkan interaksi sosial;

b. Variabel Jejaring sosial twitter (X2) mempunyai nilai koefisien regresi sebesar 0,339 yang berarti Jejaring sosial twitter mempunyai pengaruh positif terhadap Interaksi sosial (Y), yaitu jika terjadi kenaikan jejaring sosial twitter maka akan menurunkan interaksi sosial;

c. Variabel Jejaring sosial instagram (X3) mempunyai nilai koefisien regresi sebesar 0,382 yang berarti Jejaring sosial instagram mempunyai pengaruh positif terhadap Interaksi sosial (Y), yaitu jika terjadi kenaikan jejaring sosial instagram maka akan menurunkan interaksi sosial;

d. Konstanta mempunyai nilai -0,040 yang artinya jika variabel X1, X2, dan X3 dalam mempunyai nilai nol atau tidak ada maka interaksi sosial sebesar -0,040 dan nilai tersebut merupakan pengaruh dari variabel lain yang tidak dimasukkan dalam model regresi linier atau tergabung dalam Variabel Penggangu (e).

Nilai koefisien regresi dan model regresi linier tersebut belum dapat digunakan, baik sebagai alat pengambilan keputusan maupun alat peramalan, sebelum dilakukan uji hipotesis.

Uji Hipotesis Parsial (Uji T)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-.040	1.780		-.023	.982		
	jejaring sosial facebook	.306	.083	.371	3.683	.001	.990	1.010
	jejaring sosial twitter	.339	.096	.355	3.523	.001	.990	1.010
	jejaring sosial instagram	.382	.100	.387	3.834	.000	.989	1.011

a. Dependent Variable: interaksi sosial

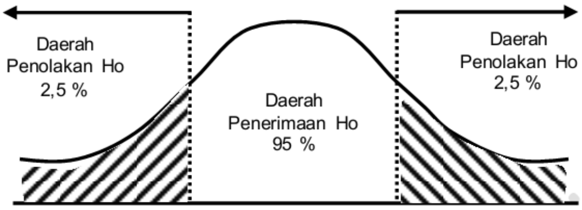
Hipotesis penelitian yang diajukan adalah Interaksi sosial (Y) dipengaruhi oleh Variabel Jejaring sosial facebook (X1), Jejaring sosial twitter (X2), dan Jejaring sosial instagram (X3). Berdasarkan hipotesis penelitian tersebut dapat dibuat hipotesis statistik (uji parsial), yaitu:

Ho : b1 = 0 Ha : b1 ≠ 0

Ho : b2 = 0 Ha : b2 ≠ 0

Ho : b3 = 0 Ha : b3 ≠ 0

Nilai T tabel untuk data sebanyak 60 responden, jumlah variabel sebanyak 4 variabel, dan tingkat signifikan yang digunakan 5% (uji dua arah), adalah sebesar ± 2,001717 (lihat Tabel T students pada df: 58 dan α: 2,5%). Daerah penolakannya adalah jika nilai t < -2,001717 atau t > 2,001717. Daerah penerimaan dan penolakan Ho dapat digambarkan sebagai berikut:



Hasil pengolahan data dapat dilihat bagian Coefficiens, diketahui bahwa nilai T hitung untuk Variabel jejaring sosial facebook sebesar 3,683 (Sig. 0,001), Variabel jejaring sosial twitter sebesar 3,523 (Sig. 0,001), dan Variabel jejaring sosial instagram sebesar 3,834 (Sig. 0,000). Nilai T hitung untuk masing-masing variabel independen tersebut berada pada daerah penolakan Ho atau mempunyai nilai Sig. di bawah 0,05 (5%).

Uji Hipotesis Serampak (Uji F)

ANOVA[b]

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	42.271	3	14.090	14.435	.000[a]
	Residual	54.663	56	.976		
	Total	96.933	59			

a. Predictors: (Constant), jejaring sosial instagram, jejaring sosial twitter, jejaring sosial facebook

b. Dependent Variable: interaksi sosial

Untuk melakukan uji hipotesis secara serempak (Uji F) hipotesis statistik yang diajukan adalah:

Ho : b1 = b2 = b3 = 0 dan Ha : b1 ≠ b2 ≠ b3 ≠ 0

Nilai F tabel dengan df: 3 ; 58 dan tingkat signifikan (α) 5% adalah sebesar 2,180727. Sedangkan untuk nilai F hitung hasil pengolahan data adalah sebesar 14,435 (lihat Tabel bagian ANOVA). Dengan membandingkan nilai F hitung dengan F tabel, diketahui bahwa F hitung lebih besar dari F tabel atau 14,435 > 2,180727 (nilai Sig. di bawah 0,05), maka dapat disimpulkan bahwa Variabel Independen secara keseluruhan mempunyai pengaruh yang signifikan terhadap Variabel Dependen atau Variabel jejaring sosial facebook, jejaring sosial twitter, dan jejaring sosial instagram secara keseluruhan mempunyai pengaruh yang signifikan terhadap Variabel penurunan interaksi sosial .

Koefisien Determinasi dan Korelasi

Analisis determinasi dalam regresi linear berganda digunakan untuk mengetahui prosentase sumbangan pengaruh variabel independen (X1, X2,……Xn) secara serentak terhadap variabel dependen (Y). Koefisien ini menunjukkan seberapa besar prosentase variasi variabel independen yang digunakan dalam model mampu menjelaskan variasi variabel dependen. R^2 sama dengan 0, maka tidak ada sedikitpun prosentase sumbangan pengaruh yang diberikan variabel independen terhadap variabel dependen, atau variasi variabel independen yang digunakan dalam model tidak menjelaskan sedikitpun variasi variabel dependen. Sebaliknya R^2 sama dengan 1, maka prosentase sumbangan pengaruh yang diberikan variabel independen terhadap variabel dependen adalah sempurna, atau variasi variabel independen yang digunakan dalam model menjelaskan 100% variasi variabel dependen.

Dari hasil analisis regresi, lihat pada output model summary dan disajikan sebagai berikut:

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.873[a]	.762	.762	.988	2.037

a. Predictors: (Constant), jejaring sosial instagram, jejaring sosial twitter, jejaring sosial facebook

Nilai Koefisien Determinasi atau R Square (R^2) dari hasil pengolahan data adalah sebesar 0,762 atau 76,2% (lihat tabel Model Summary). Nilai tersebut memberikan gambaran bahwa sumbangan Variabel Independen (Variabel jejaring sosial facebook, jejaring sosial twitter, dan jejaring sosial instagram) dalam pengaruhnya terhadap naik turunnya Variabel Dependen (Variabel penurunan interaksi sosial) adalah sebesar 76,2% dan sisanya sebesar 23,8% merupakan sumbangan variabel-variabel lain yang tidak dimasukkan dalam model (tidak diteliti) dan tergabung dalam variabel pengganggu (e) dalam model regresi linier.

Nilai korelasi berganda (R) dari hasil pengolahan data adalah sebesar 87,3%. Nilai korelasi tersebut menggambarkan bahwa hubungan antara Variabel

Independen dengan Variabel Dependen adalah mempunyai hubungan yang erat atau hubungan antara Variabel jejaring sosial facebook, jejaring sosial twitter, dan jejaring sosial instagram dengan Variabel penurunan interaksi sosial adalah mempunyai hubungan yang erat.

Adjusted R Square adalah nilai R Square yang telah disesuaikan, nilai ini selalu lebih kecil atau sama dengan R Square dan angka ini dapat memiliki harga negatif. Menurut Santoso (2001) bahwa untuk regresi dengan lebih dari dua variabel bebas digunakan Adjusted R^2 sebagai koefisien determinasi.

Standard Error of the Estimate adalah suatu ukuran banyaknya kesalahan model regresi dalam memprediksikan nilai Y. Dari hasil regresi di dapat nilai 0,988, hal ini berarti banyaknya kesalahan dalam prediksi variable penurunan interaksi sosial sebesar 0,988. Sebagai pedoman jika Standard error of the estimate kurang dari standar deviasi Y, maka model regresi semakin baik dalam memprediksi nilai Y.

Asumsi Klasik Analisis Regresi

Uji asumsi klasik dilakukan untuk mengetahui apakah model estimasi telah memenuhi kriteria ekonometrika, dalam arti tidak terjadi penyimpangan yang cukup serius dari asumsi-asumsi yang harus dipenuhi dalam metode Ordinary Least Square (OLS). Terdapat tiga asumsi yang diperlukan dalam penaksiran OLS, yaitu:

1. Tidak adanya Autokorelasi antar gangguan (e);
2. Tidak adanya Multikolinearitas;
3. Variansi kesalahan pengganggu tetap atau homoskedastisitas (tidak terjadi Heteroskedastisitas); dan
4. Normalitas antar gangguan (e).

Uji Autokorelasi digunakan untuk menguji apakah dalam sebuah model regresi linier terdapat korelasi antara kesalahan pengganggu pada periode t dengan kesalahan pada periode t-1 (sebelumnya). Autokorelasi jarang dijumpai pada data cross section dan biasanya terjadi pada data time series (serial waktu). Multikolinieritas adalah suatu kondisi dimana terjadi korelasi yang kuat diantara Variabel Independen yang diikutsertakan dalam pembentukan model regresi linier berganda. Sedangkan Uji Heteroskedastisitas digunakan untuk menguji apakah dalam model regresi liner kesalahan pengganggu mempunyai varians yang sama atau tidak dari satu pengamatan ke pengamatan yang lain.

Uji Autokorelasi

Uji Autokorelasi digunakan untuk menguji apakah dalam sebuah model regresi linier terdapat korelasi antara kesalahan pengganggu pada periode t dengan kesalahan pada periode t-1 (sebelumnya). Untuk menguji Autokorelasi dapat dilihat dari nilai Durbin Waston (DW), yaitu

- jika nilai DW terletak antara du dan (4 – dU) atau du ≤ DW ≤ (4 – dU) berarti bebas dari Autokorelasi, sebaliknya
- jika nilai DW < dL atau DW > (4 – dL) berarti terdapat Autokorelasi.
- Nilai dL dan dU dapat dilihat pada tabel Durbin Waston, yaitu nilai dL ; dU ; α ; n ; (k – 1). Keterangan: n adalah jumlah sampel, k adalah jumlah variabel, dan α adalah taraf signifikan.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.873[a]	.762	.762	.988	2.037

a. Predictors: (Constant), jejaring sosial instagram, jejaring sosial twitter, jejaring sosial facebook

Nilai tabel Durbin Watson pada α = 5%; n = 60; k – 1 = 3 adalah dL = 1,480 dan dU = 1,689. Hasil pengolahan data pada Tabel di atas,

- menunjukkan nilai Durbin Watson sebesar 2,037 dan nilai tersebut berada di antara dU dan (4 – dU) atau 1,689 < 2,037 < 2,311 maka dapat disimpulkan bahwa dalam regresi linier tersebut tidak terdapat Autokorelasi atau tidak terjadi korelasi diantara kesalahan pengganggu.

Uji Multikolinearitas

Uji Multikolinearitas digunakan untuk mengetahui apakah terjadi korelasi yang kuat di antara variabel-variabel independen yang diikutsertakan dalam pembentukan model. Untuk mendeteksi apakah model regresi linier mengalami multikolinearitas dapat diperiksa menggunakan Variance Inflation Factor (VIF) untuk masing-masing Variabel Independen, yaitu jika suatu Variabel Independen mempunyai nilai VIF > 10 berarti telah terjadi multikolinearitas.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-.040	1.780		-.023	.982		
	jejaring sosial facebook	.306	.083	.371	3.683	.001	.990	1.010
	jejaring sosial twitter	.339	.096	.355	3.523	.001	.990	1.010
	jejaring sosial instagram	.382	.100	.387	3.834	.000	.989	1.011

a. Dependent Variable: interaksi sosial

Berdasarkan table Coefficients di atas maka diperoleh nilai VIF variable jejaring sosial facebook (X1) adalah 1,010 kurang dari 10 artinya tidak terjadi multikolinearitas pada variable tersebut. Nilai VIF variable jejaring sosial twitter (X2) adalah 1,010 kurang dari 10 artinya tidak terjadi multikolinearitas pada variable tersebut. Nilai VIF variable jejaring sosial instagram (X3) adalah 1,011 kurang dari 10 artinya tidak terjadi multikolinearitas pada variable tersebut.

Uji Heteroskedastisitas

Uji Heteroskedastisitas digunakan untuk menguji apakah dalam model regresi liner kesalahan pengganggu (e) mempunyai varians yang sama atau tidak dari satu pengamatan ke pengamatan yang lain. Untuk menguji Heteroskedastisitas dapat diketahui dari nilai signifikan korelasi Rank Spearman antara masing-masing variabel independen dengan residualnya. Jika nilai signifikan lebih besar dari α (5%) maka tidak terdapat Heteroskedastisitas, dan sebaliknya jika lebih kecil dari α (5%) maka terdapat Heteroskedastisitas.

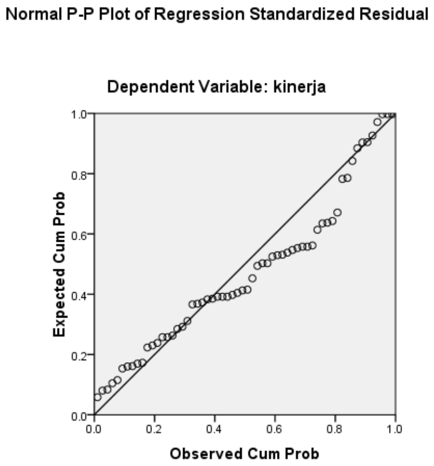
Correlations

			jejaring sosial facebook	jejaring sosial twitter	jejaring sosial instagram	residual
Spearman's rho	jejaring sosial facebook	Correlation Coefficient	1.000	-.028	.067	.040
		Sig. (2-tailed)	.	.832	.609	.759
		N	60	60	60	60
	jejaring sosial twitter	Correlation Coefficient	-.028	1.000	.042	-.031
		Sig. (2-tailed)	.832	.	.747	.815
		N	60	60	60	60
	jejaring sosial instagram	Correlation Coefficient	.067	.042	1.000	-.113
		Sig. (2-tailed)	.609	.747	.	.389
		N	60	60	60	60
	Residual	Correlation Coefficient	.040	-.031	-.113	1.000
		Sig. (2-tailed)	.759	.815	.389	.
		N	60	60	60	60

Berdasarkan tabel 8.5 tersebut di atas, pada kolom Residual dapat dilihat bahwa nilai Correlation Coefficient adalah rendah atau nilai signifikan (Sig. (2-tailed) masing - masing Variabel Independen di atas 5%, artinya masing-masing Variabel Independen (Variabel jejaring sosial facebook, variabel jejaring sosial twitter, dan variabel jejaring sosial instagram) tidak mempunyai hubungan dengan Residualnya. Dengan demikian, dapat disimpulkan bahwa tidak terdapat Heteroskedastisitas.

Asumsi Normalitas Residual

Uji normalitas bertujuan untuk menguji apakah dalam model regresi, variabel pengganggu atau residual memiliki distribusi tidak normal. Kalau asumsi ini dilanggar maka uji statistik menjadi tidak valid untuk jumlah sampel kecil (Ghozali, 2006).



Berdasarkan pola grafik di samping maka asumsi normalitas residual terpenuhi. Persamaan regresi berganda

Y = - 0,040 + 0,306 X1 + 0,339 X2 + 0,382 X3

Dapat digunakan untuk peramalan variabel penurunan interaksi sosial jika diketahui variabel jejaring sosial facebook, jejaring sosial twitter, dan jejaring sosial instagram

DAFTAR PUSTAKA

Hasan. (1999), *Pokok-pokok Materi Statistika 1 (Statistik Deskriptif)*. Bumi Aksara. Jakarta.

Mulyaningsih. (2015), *Analisa Pengaruh Penggunaan Jejaring Sosial terhadap Interaksi Sosial Remaja*. Teknik Informatika. Sekolah Tinggi Manajemen Infomatika dan Komputer Sinar Nusantara Surakarta.

Sugiarto. (1997), *Statistika Himpunan Soal dan Penyelesaian*. Andi. Yogyakarta.

Supranto, J. (2000), *Statistik Teori dan Aplikasi*, Erlangga, Jakarta.

Winarno, W. W. (2007), *Analisis Ekonometrika dan Statistika dengan Eviews*, Sekolah Tinggi Ilmu Manajemen YKPN, Yogyakarta.

BIODATA PENULIS

Penulis dilahirkan di Surakarta, Jawa Tengah pada tanggal 13 Maret 1988 dan merupakan anak ketiga dari tiga bersaudara. Pendidikan formal didapatkan penulis mulai dari TK Siwi Peni 11, SD N Tegalsari 12, SMP N 1 Surakarta, SMA N 4 Surakarta, sampai ke Universitas Negeri Sebelas Maret pada tahun 2006 dan lulus pada tahun 2010. Kemudian melanjutkan S2 di jurusan matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Institut Teknologi Sepuluh Nopember Surabaya pada tahun 2013 dengan beasiswa BPP DN Calon Dosen. Penulis menekuni ilmu matematika optimasi khususnya pemodelan dan simulasi. Saran dan kritik silakan menghubungi retno.tv@gmail.com.