PENERAPAN STATISTIKA DESKRIPTIF DAN STATISTIKA INFERENSIAL

Sanksi Pelanggaran Pasal 113
Undang-undang No. 28 Tahun 2014 Tentang Hak Cipta

1. **Setiap Orang** yang dengan tanpa hak melakukan pelanggaran hak ekonomi sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf i untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 1 (satu) tahun dan/atau pidana denda paling banyak Rp100.000.000 (seratus juta rupiah).
2. Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf c, huruf d, huruf f, dan/atau huruf h untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 3 (tiga) tahun dan/atau pidana denda paling banyak Rp500.000.000,00 (lima ratus juta rupiah).
3. Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf a, huruf b, huruf e, dan/atau huruf g untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 4 (empat) tahun dan/atau pidana denda paling banyak Rp1.000.000.000,00 (satu miliar rupiah).
4. Setiap Orang yang memenuhi unsur sebagaimana dimaksud pada ayat (3) yang dilakukan dalam bentuk pembajakan, dipidana dengan pidana penjara paling lama 10 (sepuluh) tahun dan/atau pidana denda paling banyak Rp4.000.000.000,00 (empat miliar rupiah).

PENERAPAN STATISTIKA DESKRIPTIF DAN STATISTIKA INFERENSIAL

Dr. Abd. Basir A., M.Si

AMERTA
MEDIA

PENERAPAN STATISTIKA DESKRIPTIF DAN STATISTIKA INFERENSIAL

Anggota IKAPI
No 192JTE/2020
Cetakan Pertama:
15,5 cm x 23 cm
ISBN

Penulis:

Editor:

Desain Cover:
Tim Penerbit Amerta Media

Tata Letak:
Tim Penerbit Amerta Media

Diterbitkan Oleh:
Penerbit Amerta Media

NIB. 0220002381476

Jl. Raya Sidakangen, RT 001 RW 003, Kel, Kebanggan, Kec. Sumbang, Purwokerto, Banyumas 53183, Jawa Tengah. Telp. 081-356-3333-24
Email: mediaamerta@gmail.com
Website: amertamedia.co.id
Whatsapp : 081-356-3333-24

KATA PENGANTAR

DAFTAR ISI

B A B I

PENYAJIAN DATA DENGAN TABEL DAN PERAGAAN

Bab ini berisikan pelajaran tentang penyajian data dengan tabel dan diagram. Pelajaran dimulai dari bagaimana data mentah, berupa himpunan skor hasil pengukuran, dikelompokan dalam subhimpun yang disebut kelas. Selanjutnya dilakukan *tolly* atau *taurus* untuk memperoleh frekuensi masing-masing kelas dan kemudian disajikan dalam tabel frekuensi. Penyajikan data dengan tabel, grafik atau diagram, meliputi penyajian data dengan tabel frekuensi, histogram atau diagram batang, polygon frekuensi, ogive atau grafik garis, dan diagram dahan daun.

Sesungguhnya masih terdapat penyajian data dengan diagram yang juga penting diketahui dan disajikan dalam buku ini, yaitu penyajian data dengan diagram kotak garis (*box plot*). Oleh karena cara penyajian data dengan diagram kotak garis memerlukan pengetahuan tentang ukuran letak dan ukuran penyebaran, yaitu cara menentukan median dan kuartil, maka pembahasan tentang diagram kotak garis pada buku ini disajikan pada Bab 2, setelah bagian bab yang membahas tentang ukuran letak dan ukuran penyebaran.

Capaian Pembelajaran MK :

Menggunakan ilmu statistika dasar, baik **statistika deskriptip** maupun statistika inferensial dalam penelitian dan pemecahan masalah kehidupan sehari-hari.

1.1. Penyajian Data Dengan Tabel Frekuensi

Deskripsi data dengan tabel sering menggunakan tabel frekuensi. Umumnya tabel frekuensi disajikan melalu pengelompokan data. Himpunan data terlebih dahulu dipartisi atas subhimpunan, disebut kelas. Lebar kelas disebut interval kelas. Umumnya lebar kelas dibuat sama, meskipun untuk kepentingan atau kondisi tertentu dari data, lebar kelas boleh tidak sama.

Penentuan interval kelas dan jumlah kelas bagi suatu himpunan data dilakukan dengan terlebih dahulu meninjau jangkauan data, nilai pengamatan terkecil (skor minimum) dan nilai pengamatan terbesar (skor maksimum). Jangkauan data adalah beda antara skor minimum dengan skor maksimum. Semakin banyak kelas maka deskripsi data akan semakin tampak sederhana, tetapi informasi yang hilang juga semakin banyak.

Tabel 1.1 Data mentah : Massa Bagasi (kg) Suatu Pengangkutan Barang

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	
[1,]	8	51	20	37	19	40	11	43	55	
[2,]	30	23	25	7	20	9	21	28	-	tdk lurus
[3,]	55	55	40	34	49	33	7	7	-	
[4,]	17	45	50	23	49	20	26	47	-	
[5,]	27	20	46	42	39	12	24	34	-	
[6,]	50	42	26	10	51	49	48	14	-	
[7,]	21	25	27	32	20	12	35	27	-	
[8,]	20	47	18	55	15	12	17	31	-	

Tabel 1.2 Data Massa Bagasi Setelah Diurutkan

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	
[1,]	7	7	7	8	9	10	11	12	
[2,]	12	12	14	15	17	17	18	19	
[3,]	20	20	20	20	20	20	21	21	
[4,]	23	23	24	25	25	26	26	27	
[5,]	27	27	28	28	30	31	32	33	
[6,]	34	34	35	37	39	40	40	42	
[7,]	42	43	45	46	47	47	48	49	
[8,]	49	49	50	50	51	51	55	55	55

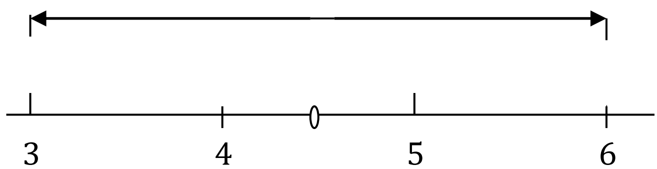
Setelah dilakukan penentuan kelas dan interval kelas, dilakukan *taurus/tolly* terhadap setiap amatan dari data. Jumlah amatan dalam suatu kelas merupakan frekuensi dari kelas tersebut. Untuk keperluan tertentu, pada tabel frekuensi sering ditambahkan kolom frekuensi kumulatif, frekuensi relatif, dan titik tengah dari setiap interval kelas.

Tabel 1.3 adalah contoh penyajian data dengan tabel frekuensi. Cara membuat tabel frekuensi menggunakan suatu himpunan data tentang berat bagasi suatu angkutan penerbangan pada Tabel 1.1 Skor minimum 7 dan skor maksimum 55, sehingga jangkauan data adalah 48. Diambil keputusan untuk jumlah kelas adalah 6, dengan lebar kelas 10. Penentuan jumlah dan lebar kelas tidak perlu menggunakan rumus tertentu, seperti dituliskan di beberapa buku. Syarat penting adalah kelas yang dibuat dapat mencakup seluruh data, dan mempertimbangkan seberapa sederhana penyajian yang diinginkan. Hal utama yang juga perlu diperhatikan adalah cara mendefinisikan interval kelas, yaitu menentukan batas bawah dan batas atas interval kelas.

Tabel 1.3 Sebaran Frekuensi

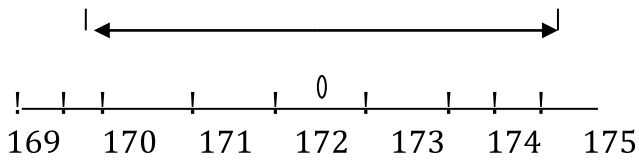
Kelas (kg)	Frekuensi (f_i)	Titik tengah (x_i)	Frekuensi kumulatif	Frekuensi Relatif (%)
0 - 9	5	5	5	(5/65) x 100 = 7,7
10 - 19	11	15	16	(11/65)x100 = 16,9
20 - 29	20	25	36	30,8
30 - 39	9	35	45	13,8
40 - 49	13	45	58	20,0
50 - 60	7	55	65	10,8
	N = 65	180	--	100

Terdapat dua cara yang sering digunakan untuk mendefinisikan interval kelas. Pemilihan ditentukan oleh sifat alami data yang digunakan. Sebagai contoh variabel umur dalam satuan tahun. Misalkan interval kelas ditulis “3 - 5 “. Anak yang termasuk dalam kelas ini adalah anak yang telah berulang tahun ke tiga sampai anak yang belum berulang tahun ke enam. Sehingga batas bawah aktual adalah 3 dan batas atas aktual adalah 6. Titik tengah interval kelas adalah 4,5.



Gambar 1.1 Interval Kelas Variabel Umur

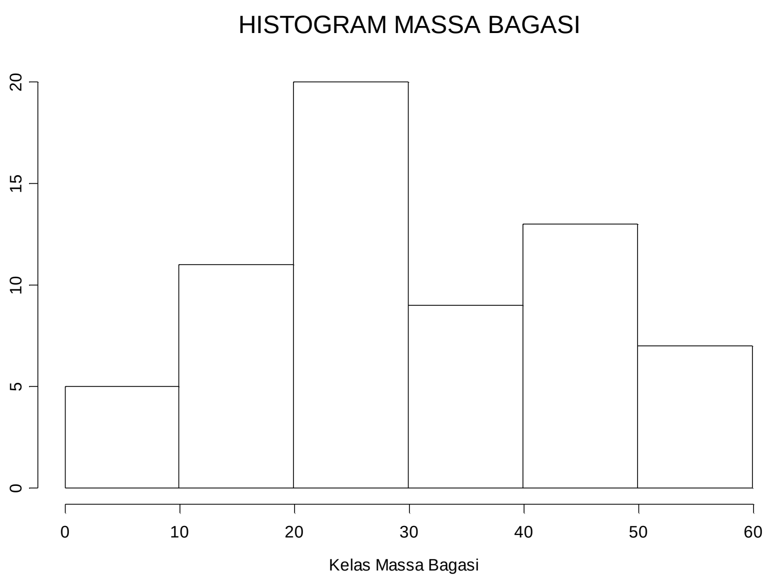
Berbeda dengan variabel tinggi badan dengan menggunakan satuan tertentu, misalkan centimeter (cm). Kesalahan relatif adalah setengah dari satuan cm, yaitu 0,5 cm. Misalkan interval kelas ditulis 170 - 174 ”, maka batas-batas kelas aktual adalah 169,5 dan 174,5. Titik tengah dari interval kelas adalah 172. Dua cara mendefinisikan interval kelas ini ditunjukkan berturt-turut pada Gambar 1.1 dan Gambar 1.2.



Gambar 1.2 Interval Kelas Variabel Tinggi Badan

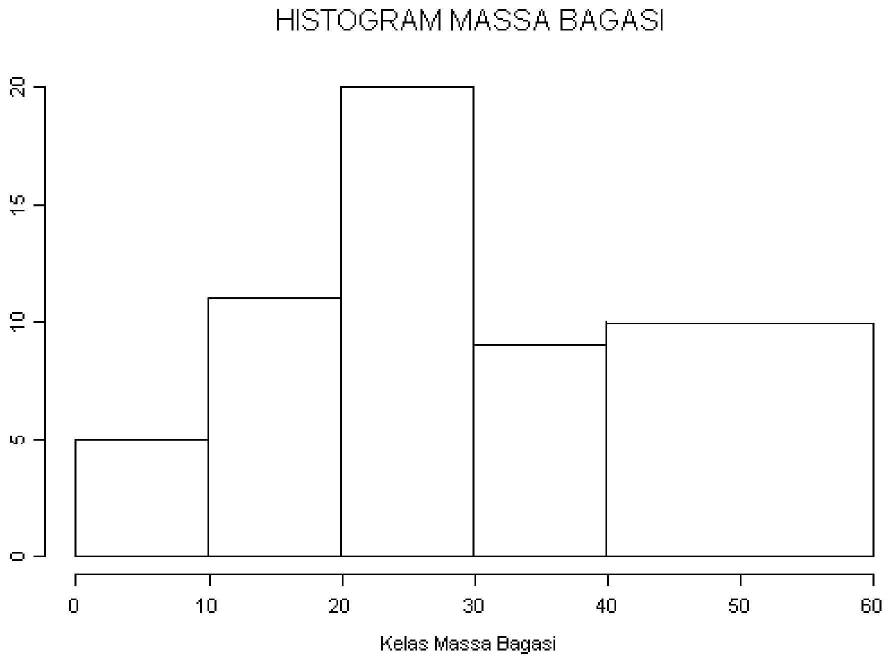
1.2 Histogram Dan Polygon Frekuensi

Batas-batas kelas aktual dan titik tengah kelas menjadi penting ketika kita ingin memperagakan sebaran data dengan histogram maupun polygon frekuensi. Histogram bagi data kelompok adalah suatu grafik lajur (boleh juga baris) dengan luas setiap lajur persegipanjang proporsional dengan masing-masing frekuensinya. Histogram bagi data kelompok pada Tabel 1.3, diperlihatkan pada Gambar 1.3. Lebar setiap lajur persegipanjang adalah 10 kg, mengikuti interval kelas. Sisi kiri dan sisi kanan lajur berimpit dengan batas-batas kelas aktual pada sumbu horizontal. Demikian pula tengah setiap lajur terletak pada titik tengah interval kelas.



Gambar 1.3 Histogram bagi Data Massa Bagasi Suatu Pengankutan

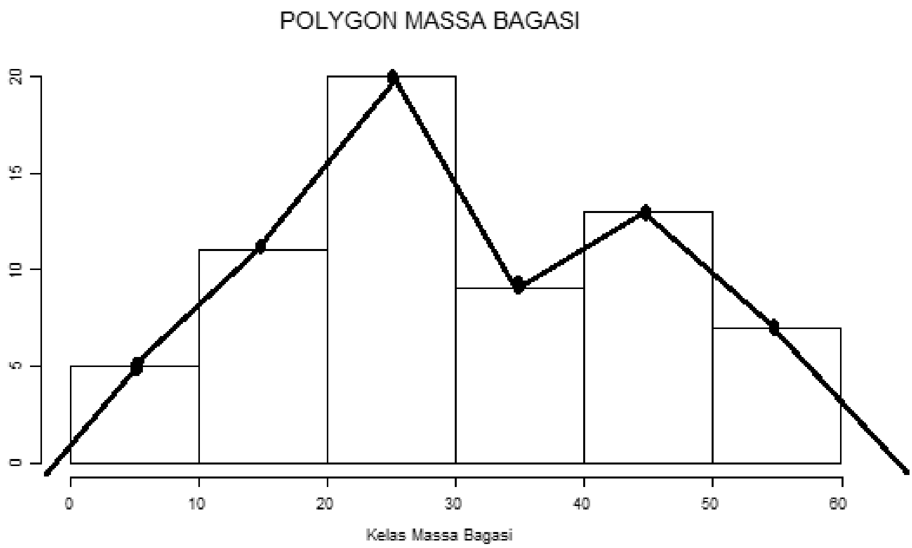
Jika pada data kelompok dibuat interval kelas yang tidak sama, misalkan pada contoh ini dua kelas terakhir digabungkan, lebar atau interval kelas menjadi 20, maka tinggi lajur persegipanjang akan berkurang dari setengahnya, yaitu berkurang 0,5 dari (13 + 7). Tinggi lajur persegipanjang terakhir adalah 10. Cara ini menjadikan luas lajur proporsional terhadap frekuensinya. Hasil pembuatan histogram ini diperlihatkan pada Gambar 1.4.



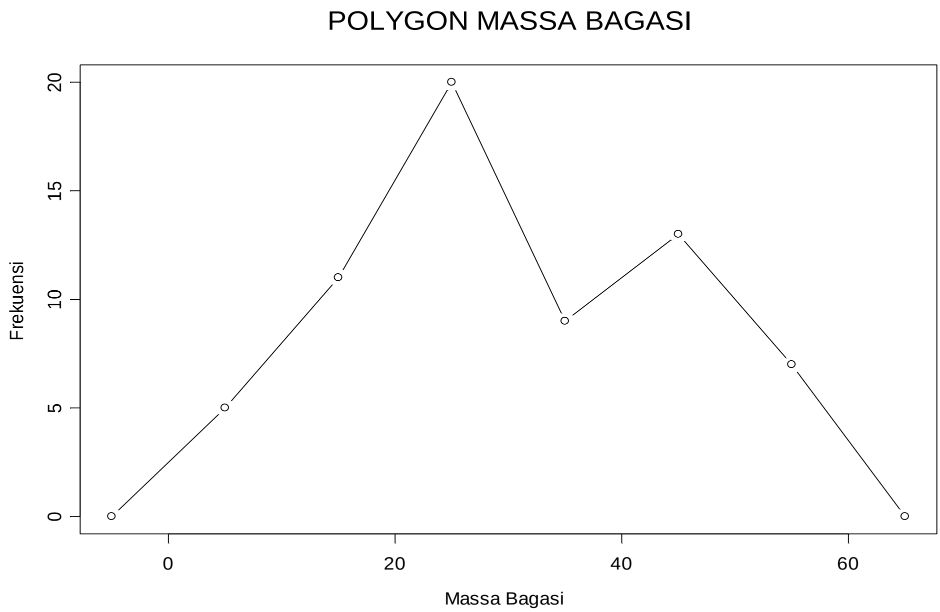
Gambar 1.4 Histogram dengan Interval Kelas yang Tidak Sama

Polygon frekuensi adalah suatu plot antara frekuensi setiap kelas dengan masing-masing titik tengah interval kelas, dengan membubuhkan garis yang menghubungkan setiap titik plot. Selanjutnya kedua ujung garis diperpanjang ke bawah sampai memotong sumbu horizontal. Garis memotong sumbu pada titik yang berjarak sama dengan lebar kelas, untuk jarak yang dihitung dari titik tengah dua kelas pinggir. Kelas yang dimaksud adalah dua kelas pinggir kiri dan kanan.

Contoh cara menggambarkan polygon frekuensi bagi histogram di atas, diperlihatkan pada Gambar 1.5. Titik-titik tengah interval kelas, yaitu 5, 15, 25, 35, 45, 55 diplot dengan berturut-turut frekuensi masing-masing kelas, yaitu 5, 11, 20, 9, 13, 7. Selanjutnya titik-titik hasil plot dihubungkan dengan garis. Kedua ujung garishubung diperpanjang melalui atau memotong ½ tinggi sisi masing-masing lajur persegipanjang kiri dan kanan, sampai mencapai sumbu horizontal. Kedua ujung garishubung memotong sumbu horizontal berturut-turut di titik –5 dan 65. Perhatikan bahwa luas daerah di bawah kurva polygon, atau luas daerah yang dibatasi oleh kurva dan sumbu horizontal



Gambar 1.5 Cara Membuat Polygon Frekuensi



Gambar 1.6 Polygon Massa Bagasi

adalah sama dengan jumlah luas persegipanjang pada histogram. Polygon frekuensi menjadi alternatif pengganti histogram, khususnya ketika bertujuan membandingkan dua himpunan data di atas satu grafik. Hasil lukisan polygon bagi data massa bagasi diperlihatkan pada Gambar 1.6.

1.3 Diagram Dahan Dan Daun

Alternatif lain sebagai pengganti histogram, untuk tujuan memperlihatkan sebaran data adalah diagram dahan dan daun (*Stem and leaf diagrams*). Diagram dahan dan daun dibuat langsung dari data mentah, atau tanpa melalui pengelompokan data. Pada diagram dahan daun, informasi yang terbuang oleh hanya proses *tolly* pada kelas dapat dihindari. Nilai pengamatan dapat langsung terlihat pada diagram dahan daun, sekaligus menampilkan bentuk sebaran data.

Contoh suatu diagram dahan dan daun untuk data massa bagasi disajikan pada Gambar 1.7. Diagram terdiri dari dua bagian, yaitu bagian dahan dan bagian daun. Kemudian dilengkapi dengan keterangan frekuensi kumulatif pada lajur pertama. Angka pada dahan menyatakan bilangan puluhan (juga dapat menyatakan bilangan ratusan dan seterusnya). Angka pada daun menyatakan bilangan satuan. Setiap amatan dalam himpunan data dipilah menjadi dua bagian, yaitu untuk suatu dahan dan suatu daun. Misalkan nilai amatan 48, dipilah menjadi angka 4 diletakkan pada bagian dahan yang menyatakan bilangan empatpuluhan, dan angka 8 diletakkan pada bagian daun yang menyatakan bilangan delapan satuan.

5	0 : 77789
16	1 : 01222457789
(20)	2 : 00000011334556677788
29	3 : 012344579
20	4 : 0022356778999
7	5 : 0011555

Gambar 1.7 Diagram Dahan dan Daun

Misalkan pandangan kita pusatkan pada dahan tigapuluhan, untuk contoh diagram dahan daun pada Gambar 7, yaitu pandangan pada baris keempat :

"29 3 : 012344579 ". Peragan ini mewakili 9 amatan, yaitu

30 31 32 33 34 34 35 37 39

Bilangan tigapuluhan diletakkan pada dahan dengan menuliskan angka 3, selanjutnya setiap nilai satuan dituliskan pada bagian daun dengan menuliskan " 012344579 ". Sedangkan bilangan 29 yang tertulis pada lajur pertama menyatakan frekuensi kumulatif dihitung mulai dari baris paling bawah.

Frekuensi kumulatif dilampirkan pada lajur depan suatu diagram dahan daun sebagai tambahan keterangan. Pertama-tama, frekuensi kumulatif dihitung mulai dari baris paling atas bergerak ke baris dibawahnya atau baris kedua, dan berhenti setelah mencapai baris atau dahan yang memuat median. Pada dahan atau baris yang memuat median, dituliskan berapa frekuensi dahan tersebut, yang ditulis bukan frekuensi kumulatif. Penulisan frekuensi dahan memuat median dilengkapi dengan tanda kurun. Selanjutnya frekuensi kumulatif dihitung mulai dari baris paling bawah sampai baris di bawah dahan yang memuat median.

5	0 : 77789
11	1 : 012224
16	1 : 57789
27	2 : 00000011334
(9)	2 : 556677788
29	3 : 012344
23	3 : 579
20	4 : 00223
15	4 : 56778999
7	5 : 0011
3	5 : 555

Gambar 1.8 Diagram Dahan dan Daun yang Diperoleh dengan Pemilahan Dahan Atas Dua Bagian

Jika pada suatu diagram dahan daun, terdapat suatu atau beberapa dahan dengan frekuensi yang mencolok sangat besar, maka penulisan barisan angka-angka pada bagian daun akan sangat panjang kearah pinggir kanan dari lembar halaman. Penyajian diagram dapat diperbaiki dengn memilah setiap dahan menjadi dua bagian. Bagian pertama untuk memuat daun 0 1 2 3 4, dan bagian kedua untuk memuat daun 5 6 7 8 9. Kita dapat memutuskan untuk memilah setiap dahan menjadi 2 bagian, 5 bagian, dan 10 bagian. Contoh diagram dahan daun jika dahan dipilah menjadi 2 bagian, bagi data massa bagasi diperlihatkan pada Gambar 8.

DAFTAR PUSTAKA

Anderson, David R., Sweeney, Dennis J and Thomas W.A, (1993) *Statistics for*
Business and Economics, 5th edn, St Paul, MN: West Publishing.
Fleming, M. C, and Nellis, J.G., (1994), *Principles of Applied Statistics*, First edition, Routledge, New York
Lane D., et al (2003), Introduction to Statistics, Online Edition
https://onlinestatbook.com/Online_Statistics_Education.pdf
Nolan, B. (1994), *Data Analysis: An Introduction* , Cambridge, Polity Press.
Walpole, R. E., (1982), *Introduction to Statistics*, 3 rd edition, Macmillan Publishing Co. Inc., New York.

LATIHAN UNTUK BAB 1

Soal 1 :

Pada Tabel 1. 4, terdapat himpunan data yang mencatat tinggi badan bagi 100 mahasiswa laki-laki. Data telah terurut dari amatan paling rendah sampai amatan paling tinggi. Data dikelompokkan menurut 8 kelas, seperti kelas yang dibentuk pada lajur pertama dari Tabel 1.5.

Tabel 1.4 Himpunan Data Tinggi Badan bagi 100 Mahasiswa Laki-laki

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
[1,]	156	157	159	162	162	163	164	165	165	166
[2,]	166	166	166	167	167	168	168	168	169	169
[3,]	169	169	169	170	170	170	170	170	171	171
[4,]	171	171	171	172	172	172	172	172	173	173
[5,]	173	173	173	173	173	173	174	174	174	174
[6,]	174	174	174	174	175	175	175	175	175	175
[7,]	175	175	175	175	176	176	176	176	177	177
[8,]	177	177	177	178	178	179	179	179	179	179
[9,]	179	180	180	181	181	181	181	181	182	182
[10,]	182	183	183	184	186	186	188	188	193	194

a. Periksa berapa jangkaun bagi data tinggi badan di atas. Apakah semua amatan dapat termuat pada kelas-kelas yang ada, seperti pengelompokan pada lajur pertama Tabel 1.5.
b. Lakukan tolly untuk melengkapi lajur dua atau frekuensi setiap kelas pada Tabel 1.5. Selanjutnya lengkapi setiap lajur pada tabel tersebut
c. Pada kelas yang memuat median, jelaskan batas-batas kelas aktual dan titik tengahnya, melalui suatu penyajian gambar.
d. Gambar histogram dan polygon bagi himpunan data tinggi badan yang telah dikelompokkan.

e. Buat diagram dahandaun bagi himpunan data tinggi badan. Jelaskan apa yang terjadi pada dahan seratus tujuhpuluhan, yaitu pada bagian daunnya !

f. Buatkan diagram dahandaun bagi himpunan data tinggi badan, dengan cara memilah setiap dahan atas 2 bagian.

Pemecahan :

a. Jangkauan : Jangkauan ; $Rg = X_{maks} - X_{mim}$

Rg = ……. - ………. = ……….

Komentar : ……………………………………………………………

……………………………………………………………

……………………………………………………………

b. Tabel 1.5 Sebaran Frekuensi

Kelas (cm)	**Frekuensi (f_i)**	**Titik tengah (x_i)**	**Frekuensi kumulatif**	**Frekuensi Relatif (%)**
155 - 159	3	157	3	
160 - 164	4	162	7	
165 - 169	….	….	….	
170 - 174	….	….	….	
175 - 179	….	….	….	
180 - 184				
185 – 189				
190 – 194				
	N = 100	…..	--	100

c. Median terletak pada data ke n/2

data ke … / 2

data ke …..

Sehingga median terletak pada kelas ke …., yaitu interval kelas “ …… - ……. “

Soal 2:

Hasil tes untuk 50 mahasiswa pada matakuliah satistika adalah

Tabel 1.6 Hasil Tes Atas 50 Mahasiswa

26	42	8	28	39	78	32	54	93	11
27	33	22	78	75	83	62	76	77	67
77	80	7	26	18	10	34	30	36	43
79	41	24	91	90	63	87	55	60	48
35	35	51	67	10	76	34	47	51	33

Tidak lurus kolomnya

a. Periksa berapa jangkaun bagi data di atas.
b. Buat Tabel frekuensi, lengkap dengan lajur untuk titik tengah, dan lajur untuk frekuensi kumulatif.
c. Pada kelas yang memuat median, jelaskan batas-batas kelas aktual dan titik tengahnya, melalui suatu penyajian gambar.
d. Gambar histogram dan polygon bagi himpunan data hasil tes statistika.
e. Buat diagram dahan daun bagi himpunan data hasil tes statistika
f. Buatkan diagram dahan daun bagi himpunan data hasil tes statistika, dengan cara memilah setiap dahan atas 2 bagian.

BAB II

PENYAJIAN DATA SECARA NUMERIK

Bab ini berisikan pelajaran tentang penyajian data secara numerik, yaitu menyajikan ukuran pemusatan (letak) dan ukuran penyebaran suatu himpunan data. Ukuran pemusatan meliputi rata-rata aritmatika, median dan modus, sedangkan ukuran penyebaran meliputi jangkauan, kuartil, desil, persentil dan jangkauan termodifikasi.

Perlu diketahui bahwa meskipun kuartil, desil dan persentil di atas disebutkan sebagai ukuran-ukuran penyebaran, namun ketiga statistik tersebut sesungguhnya juga dapat berfungsi sebagai ukuran letak atau pemusatan.

Adapun kata termodifikasi pada istilah jangkauan termodifikasi dimaksudkan bahwa jangkauan dihitung berdasarkan selisih nilai maksimun dengan nilai minimum masing-masing statistik quartil, desil dan persentil. Berbeda dengan menghitung jangkauan biasa, yaitu dihitung langsung dari selisih skor maksimun dengan skor minimum, jangkauan termodifikasi hanya dapat ditentukan setelah kita menghitung statistik-statististik yang bersesuaian, baik itu kuartile, desil, maupun persentil.

Pada bagian akhir bab ini, juga dibahas tentang diagram kotak garis yang merupakan materi pelajaran dari penyajian data dengan diagram. Penyajian materi dengan diagram sesungguhnya merupakan bagian materi yang telah dibahas pada bab sebelumnya, namun khusus pembahasan tentang diagram kotak garis yang memerlukan pengetahuan tentang penentuan kuartil, ditempatkan pada akhir bab ini.

Capaian Pembelajaran MK :

Menggunakan ilmu statistika dasar, baik **statistika deskriptip** maupun statistika inferensial dalam penelitian dan pemecahan masalah kehidupan sehari-hari.

Capaian pembelajaran pada bab ini adalah melakukan penyarian data numerik dengan ukuran pemusatan dan letak dengan persisten, teliti dan jujur serta bertanggung jawab.

2.1 Ukuran Pemusatan

Terdapat empat ukuran pemusatan yang sering digunakan, yaitu rata-rata aritmatika, median, modus, dan rata-rata geometri. Pemilihan yang mana dari salah satu ukuran pemusatan yang digunakan, tergantung dari keadaan alami data dan tujuan pengukuran.

2.1.1 Ukuran Pemusatan Untuk Data Tunggal

a. Rata-rata Aritmatika

Bagi himpunan data $X_1, X_2, \ldots, X_n$. Rata rata aritmatika dinyatakan sebagai $\overline{X}$ (dibaca : X bar), dirumuskan sebagai

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Keterangan :

$\overline{X}$ = nilai rata-rata

n = banyaknya pengamatan

Contoh pengerjaan 2.1 :

Data pada pengambilan sampel adalah 2, 6, 6, 7, 3, 1, 3

<u>Pemecahan</u> :

Rata-rata aritmatika adalah

$$\overline{X} = \frac{2+6+6+7+3+1+3}{7} = \frac{28}{7} = 4$$

b. Median

Median dari suatu himpunan data adalah nilai tengah data setelah data diurutkan (*ordered*) dari nilai amatan paling kecil ke nilai amatan paling besar. Setelah data diurutkan, median data dirumuskan sebagai

Med = nilai data pada pengamatan ke (n + 1) / 2

Med = X (n + 1) / 2 $X_{(n+1)/2}$

Keterangan :

Med = median

n = banyaknya pengamatan

Median untuk data berukuran ganjil dapat langsung diidentifikasi, karena median tercantum pada data. Jika data berukuran genap maka median adalah rata-rata dua pengamatan tengah dalam data terurut.

Contoh pengerjaan 2.2:

Tentukan median dari data berikut

a. 8, 4, 4, 3, 5, 9, 10, 8, 8, 7, 9
b. 5, 5, 4, 7, 8, 9, 8, 10, 10, 3, 4, 10

Pemecahan :

a. Data terurut : 3, 4, 4, 5, 7, 8, 8, 8, 9, 9, 10

Med = nilai data pada pengamatan ke (11+1) / 2
= nilai data pada pengamatan ke 6

Med = X_6
= 8

b. Data terurut : 3, 4, 4, 5, 5, 7, 8, 8, 9, 10, 10, 10

Med = nilai data pada pengamatan ke (12+1) / 2
= nilai data pada pengamatan ke 6,5

Med = $X_6 + 0,5 (X_7 - X_6)$
= 7 + 0,5 (8-7)
= 7,5

c. Modus

Modus adalah kejadian yang paling sering terjadi pada suatu peristiwa. Dengan demikian modus suatu himpunan data adalah nilai data yang memiliki frekuensi terbesar atau paling sering muncul.

Modus suatu data dapat tidak ada dan dapat tidak unik. Jika terdapat dua nilai yang menjadi modus pada suatu himpunan data, maka modus demikian dinamakan bimodus. Jika terdapat tiga modus dinamakan trimodus.

Modus sering lebih cocok digunakan sebagai ukuran pemusatan pada hasil pengukuran yang berupa angka-angka yang bulat saja.

Contoh pengerjaan 2.3:

Tentukan modus dari data berikut

Data : a. 6, 4, 3, 5, 5, 6, 6, 6, 7, 9, 8
b. 5, 6, 3, 4, 7, 7, 8, 8, 9, 8, 7

<u>Pemecahan :</u>

a. Nilai data yang paling sering muncul adalah nilai 6, dengan frekuensi 4 sehingga modus dari data ini adalah 6 , ditulis Mod = 6.
b. Nilai data yang paling sering muncul atau dengan frekuensi terbesar ada dua, yaitu nilai 7 dan nilai 8, keduanya memiliki frekuensi 3. Bimodus dari data ini adalah 7 dan 8.

Ukuran pemusatan lain, tetapi lebih sering digunakan sebagai ukuran penyebaran adalah kuartil 1, kuartil 3, nilai minimum, dan nilai maksimum. Kuartil bersama-sama dengan nilai minimun dan nilai maksimum sering diistilahkan statistik lima serangkai atau statistik tegar. Bersifat tegar karena disamping sebagai ukuran pemusatan, juga sebagai ukuran penyebaran.

2.1.2 Ukuran Pemusatan Untuk Data Kelompok

Data kelompok adalah data yang diringkas menurut kelas-kelas. Himpunan data terlebih dahulu dipartisi atas subhimpunan, disebut kelas. Lebar kelas disebut interval kelas. Umumnya lebar kelas dibuat sama, meskipun untuk kepentingan atau kondisi tertentu dari data, lebar kelas boleh tidak sama. Penentuan lebar kelas dilakukan dengan terlebih dahulu melakukan tinjauan terhadap sejauhmana penyederhanaan data, jangkauan data dan berapa lebar kelas yang kita inginkan. Jangkauan data adalah lebar antara nilai maksimum dengan nilai minimum atau rentangan dari nilai paling rendah sampai nilai paling tinggi. Sekelompok pengamatan dalam suatu kelas tertentu dihitung jumlahnya, dan dinyatakan sebagai frekuensi kelas tersebut.

Rumus menentukan ukuran pemusatan adalah :

Rata-rata :

$$\overline{X} = \frac{\sum_{i=1}^{p} f_i \cdot x_i}{n}$$

Median :

$$Md = L_{med} + \frac{\frac{n}{2} - F_{p-1}}{f_{med}} \cdot W_{med}$$

Modus :

$$Mod = L + \frac{d_1}{d_1 + d_2} \cdot W$$

Contoh data kelompok :

Pindah hal 20

Tabel 2.1 Data Kelompok yang Disajikan Melalui Tabel Frekuensi

Kelas	Frekuensi Kelas (f_i)	Titik Tengah (x_i)	$f_i . x_i$	Frekuensi Kumulatif (F_i)
0 - < 5	2	2,5	5	2
5 - < 15	6	10	60	8
15 - < 25	20	20	400	28
25 - < 35	15	30	450	43
35 - < 45	8	40	320	51
45 - < 100	4	72,5	290	55
Total	n = 55		1.525	

Contoh pengerjaan 2.4:
Hitunglah rata-rata, median, dan modus bagi data kelompok pada Tabel ~~6~~. 2.1

Pemecahan :

Rata-rata : $\overline{X} = \frac{5+60+...+290}{55} = \frac{1.525}{55} = 27,7273$

Kelas median adalah kelas yang memuat nilai yang terletak di tengah data, yaitu data ke n / 2 , dengan n / 2 = 55 / 2 = 27,5. Pada lajur frekuensi kumulatif Tabel 2.1, dari kelas pertama sampai pada kelas ke 2, terdapat 8 pengamatan, dan sampai pada kelas ke 3 telah terdapat 28 pengamatan. Ini berarti data ke 27,5 berada dalam kelas ke 3, yaitu interval kelas “ 15 - < 25”. Kelas ke 3 merupakan kelas median, dan memiliki batas bawah aktual 15 dan frekuensi 20 .

Median : $Md = L_{med} + \dfrac{\frac{n}{2} - F_{p-1}}{f_{med}} \cdot W_{med}$

Median : $Md = 15 + \left(\dfrac{27{,}5 - 8}{20}\right).10 = 15 + \left(\dfrac{19{,}5}{20}\right) 10 = 24{,}75$

Modus : $Mod = 15 + \dfrac{(20-6)}{(20-6)+(20-15)} \cdot 10$

$= 15 + \dfrac{14}{14+5} \cdot 10 = 22{,}07$

2.2 Ukuran Penyebaran (Variabilitas)

Dikenal beberapa ukuran penyebaran data, seperti :

a. Jangkauan dan jangkauan termodifikasi
b. Deviasi rata-rata
c. Standar deviasi atau Simpangan baku
d. Ragam atau variansi
e. Koefisien variasi

Tidak semua atau hanya beberapa dari ukuran penyebaran di atas yang dibahas pada tulisan ini.

2.2.1 Ukuran Penyebaran Untuk Data Tunggal

a. Jangkauan

Jangkauan suatu himpunan data menyatakan lebar dari data tersebut. Untuk suatu data hasil pengambilan sampel (*sampling data*), jangkauan dihitung dari selisih (perbedaan) antara nilai maksimum (X_{maks}) dengan nilai minimum (X_{mim}). Pada himpunan data terurut $X_{mim}, X_2, X_3, \ldots, X_{maks}$.

Jangkauan : $Rg = X_{maks} - X_{mim}$

b. Jangkauan Termodifikasi

Jangkauan termodifikasi didasarkan atas ukuran bervariasi dari pembagian (partisi) suatu himpunan data terurut, menurut beberapa bagian yang sama. Pembagian data atas beberapa bagian, dapat berupa pembagian Kuartil (empat bagian), pembagian Desil (sepuluh bagian), dan pembagian persentil (seratus bagian).

Untuk pembagian kuartil, ukuran kuartil didefenisikan sebagai berikut

1. Kuartil satu (Q_1) adalah pengamatan yang terletak pada lokasi ke (n+1)/4 dalam data terurut. Mempartisi data atas 25 % ke bawah dan 75 % ke atas.
2. Kuartil dua (Q_2) adalah pengamatan yang terletak pada lokasi ke 2 (n+1)/4 dalam data terurut. Mempartisi data atas 50 % ke bawah dan 50 % ke atas. Kuartil dua suatu himpunan data sama dengan median data tersebut.
3. Kuartil tiga (Q_3) adalah pengamatan yang terletak pada lokasi ke 3 (n+1) /4 dalam data terurut. Mempartisi data atas 75 % ke bawah dan 25 % ke atas.

c. Ragam (Variansi)

Bagi himpunan data di populasi $X_1, X_2, \ldots, X_N.$, ragam bagi populasi dapat dihitung dengan rumus berikut.

Ragam : $$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \overline{X})^2}{N}$$

Bagi himpunan data sampel (*sampling data*) $X_1, X_2, \ldots, X_n.$, ragam bagi sampel dapat dihitung dengan rumus berikut.

Ragam : $$S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n - 1}$$ atau

$$S^2 = \frac{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}}{n - 1} \quad \text{atau}$$

$$S^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\,\overline{X}^2}{n - 1}$$

Tiga rumus menghitung ragam di atas hanya berbeda pada pembilang, namun tetap menghasilkan nilai akhir yang sama atau nilai ragam yang sama. Ketiga rumus ini dapat diterapkan untuk menghitung ragam sampel maupun ragam populasi.

d. Simpangan Baku (Standar Deviasi)

Simpangan baku adalah akar kuadrat dari ragam, yaitu bagi himpunan data di populasi $X_1, X_2, \ldots, X_N$. Simpangan baku populasi dapat dihitung dengan rumus berikut.

Simpangan baku : $$\sigma = \sqrt{\frac{\sum X_i^2 - N\,\overline{X}^2}{N}}$$

Bagi himpunan data sampel $X_1, X_2, \ldots, X_n$. Simpangan baku sampel dapat dihitung dengan rumus berikut.

Simpangan baku : $$S = \sqrt{\frac{\sum X_i^2 - n\,\overline{X}^2}{n - 1}}$$

Contoh pengerjaan 2.5:

Data : 4 7 8 9 22 26 29 37 40 48

Hitunglah :

a. Jangkauan
b. Jangkauan antar kuartil
c. Ragam
d. Simpangan baku

Pemecahan :

a. Jangkauan : $Rg = X_{maks} - X_{mim} = 48 - 4$
$= 44$

b. Jangkauan antarkuartil :
Q_1 terletak pada pengamatan ke (n+1)/4 . dengan (n+1)/4 = (10 +1)/ 4 = 2,75

Sehingga $Q_1 = X_2 + 0{,}75\ (X_3 - X_2)$
$Q_1 = 7 + 0{,}75\ (8-7) = 7{,}75$

Q_3 terletak pada pengamatan ke 3(n+1)/4
Dengan 3(n+1)/4 = 3 (10 +1)/ 4 = 8,25

Sehingga $Q_3 = X_8 + 0{,}25\ (X_9 - X_8)$
$Q_3 = 37 + 0{,}25\ (40-37) = 37{,}75$
Jangkauan antarkuartil $Q_3 - Q_1 = 37{,}75 - 7{,}75 = 30$
Jangkauan data setelah terpangkas 25 % ke bawah dan 25% ke atas adalah 30.

c. Ragam:

Untuk menghitung ragam, terlebih dahulu dilakukan hitungan pendahuluan berikut

$\sum X_i = 4 + 7 + \ldots + 48 = 230$

$\sum X_i^2 = 16 + 49 + 64 + 81 + 484 + 676 + 841 + 1369 + 1600 + 2304 = 7484$

Ragam S^2

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

$$= \frac{7484 - ((230)^2 / 10)}{10 - 1}$$

$= 243.78$

d. Simpangan baku :

Simpangan baku adalah akar kuadrat dari ragam. Jika telah diperoleh ragam S^2 = 243,78 maka simpangan baku adalah S = $\sqrt{243,78}$ = 15,61

2.2.2 Ukuran Penyebaran Untuk Data Kelompok

Cara menghitung kuartil satu (Q_1) maupun kuartil tiga (Q_3) bagi data kelompok sama dengan cara menghitung median bagi data kelompok.

Kuartil ke i : $$Q_i = L_{q_i} + \frac{i\frac{n}{4} - F_{p-1}}{f_{q_i}} . W_{q_i}$$

Keterangan :

Q_i = Kuartil ke i (i = 1, 2 , 3)
Lq_i = Batas bawah dari interval kelas kuartil
F_{p-1} = frekuensi kumulatif di kelas sebelum kelas kuartil i
fq_i = Frekuensi kelas kuartil i
Wq_i = Lebar kelas kuartil i

Rumus untuk menghitung ragam untuk data kelompok adalah

$$S^2 = \frac{\sum_{i=1}^{p} f_i (X_i)^2 - \left(\left(\sum_{i=1}^{p} f_i X_i \right)^2 / n \right)}{n - 1}$$

Simpangan baku bagi data kelompok adalah akarkuadrat dari ragam bagi data kelompok, yaitu $S = \sqrt{S^2}$

Contoh pengerjaan 2.6:
Hitunglah jangkauan antarkuartil, ragam, dan simpangan baku bagi data kelompok pada Tabel 2.2.

Tabel 2.2 Tabel Frekuensi

Kelas	Frekuensi Kelas (f_i)	Titik Tengah (x_i)	$f_i . x_i$	$f_i . (x_i)^2$	Frekuensi Kumulatif (F_i)
0 - < 5	2	2,5	5	12,50	2
5 - < 10	5	7,5	37,5	281,25	7
10 - < 15	8	12,5	100,0	1250,00	15
15 - < 20	6	17,5	105,0	1837,50	21
20 - < 25	4	22,5	90,0	2025,00	25
25 - < 45	1	35,0	35,0	1225,00	26
Total	n = 26		372,5	6631,25	

Pemecahan pengerjaan 2.6:
Q_1 terletak pada pengamatan n /4, dimana n /4 = 26/4 = 6,5. Jika dilihat dari frekuensi kumulatif, pengamatan ke 6,5 berada pada kelas ‘ 5 - < 10’. Sehingga kelas dari Q_1 adalah ‘ 5 - < 10’. Kelas kuartil Q_1 ini memiliki batas bawah, frekuensi dan lebar berturut-turut L = 5, f = 5, dan W = 5.

$$Q_1 = L_{q1} + \frac{1\frac{n}{4} - F_{p-1}}{f_{q_1}} . W_{q_1}$$

$$Q_1 = 5 + \left(\frac{6{,}5 - 2}{5}\right).5 = 5 + 4{,}5 = 9{,}5$$

Q_3 terletak pada pengamatan 3n /4, dimana 3n /4 = 3(26)/4 = 19,5. Jika dilihat dari frekuensi kumulatif, pengamatan ke 19,5 berada pada kelas ' 15 - < 20'. Sehingga kelas dari Q_3 adalah ' 15 - < 20'. Kelas kuartil Q_3 ini memiliki batas bawah, frekuensi dan lebar berturut-turut L = 15, f = 6, dan W = 5.

$$Q_3 = L_{q3} + \frac{3\frac{n}{4} - F_{p-1}}{f_{q_3}} . W_{q_3}$$

$$Q_3 = 15 + \left(\frac{19{,}5 - 15}{6}\right).5 = 15 + \left(\frac{4{,}5}{6}\right) 5 = 18{,}75$$

Jangkauan antarkuartil = $Q_3 - Q_1$ = 18,75 – 9,5 = 9,25

Ragam :

$$S^2 = \frac{\sum_{i=1}^{p} f_i (X_i)^2 - \left(\left(\sum_{i=1}^{p} f_i X_i\right)^2 / n\right)}{n - 1}$$

$$= \frac{6631{,}25 - ((372{,}5)^2 / 26)}{26 - 1}$$

= 51,7789

Simpangan baku :

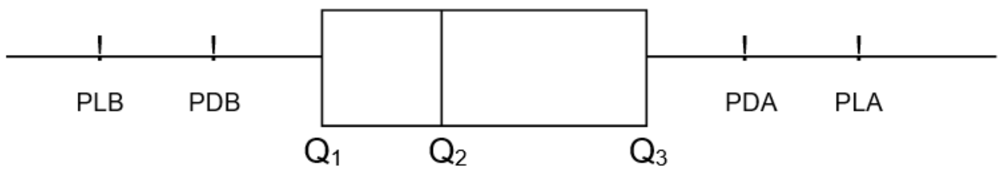
Simpangan baku adalah akar kuadrat dari ragam, sehingga simpangan baku

$S = \sqrt{51{,}7789} = 7{,}1958$

2.3 Diagram Kotakgaris (*Boxplot Diagram*)

Selain diagram dahan dan daun, penyebaran data juga dapat diperagakan dengan diagram kotakgaris. Peragaan jangkauan antarkuartil tampak lebih jelas melalui peragaan diagram kotakgaris, karena dilengkapi dengan peragaan bagi penyebaran amatan-amatan diluar jangkauan antarkuartil. Disamping itu, juga disajikan hasil identifikasi amatan yang merupakan pencilan (*outlier*).

Jangkauan antarkuartil digambarkan sebagai lebar kotak. Di sekitar kotak terdapat pagar dalam dan pagar luar untuk membantu mengidentifikasi pencilan. Bentuk diagram kotakgaris diperlihatkan pada Gambar 9.



Gambar 2.1 Diagram Kotakgaris

Keterangan gambar :

Q_1 = Kuartil satu
Q_2 = Kuartil dua
Q_3 = Kuartil tiga

PLB = Pagar luar bawah
= $Q_1 - 3\ (Q_3 - Q_1)$

PDB = Pagar dalam bawah
= $Q_1 - 3/2\ (Q_3 - Q_1)$

PDA = Pagar dalam atas
= $Q_3 + 3/2\ (Q_3 - Q_1)$

PLA = Pagar luar atas
= $Q_3 + 3\ (Q_3 - Q_1)$

Contoh pengerjaan 2.7:
Gambar diagram kotakgaris dari data berikut

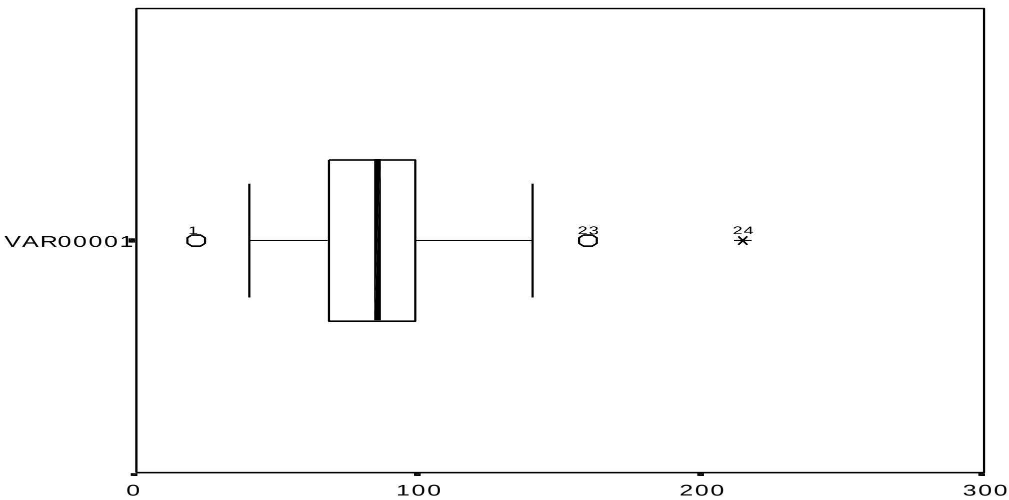
21	40	42	50	60	64	72	78	80	84	85	86	86	90	90	90
92	98	100	114	120	140	160	215								

<u>Pemecahan</u> :
Q_1 Terletak pada pengamatan ke 6,25 sehingga
$Q_1 = 64 + 0{,}25\ (72 - 64) = 66$

Q_2 Terletak pada pengamatan ke 12,5 sehingga
$Q_2 = 86 + 0{,}5\ (86 - 86) = 86$

Q_3 Terletak pada pengamatan ke 18,75 sehingga
$Q_3 = 98 + 0{,}75\ (100 - 98) = 99{,}5$

Jangkauan antarkuartil = 99,5 – 66 = 33,5



Gambar 2.2 Diagram Kotakgaris bagi Data Contoh Pengerjaan 2.7.

$$PDB = Q_1 - 3/2\ (Q_3 - Q_1) = 66 - 3/2\ (33{,}5) = 15{,}75$$

$$PLB = Q_1 - 3\ (Q_3 - Q_1) = 66 - 3\ (33{,}5) = -34{,}5$$

$$PDA = Q_3 + 3/2\ (Q_3 - Q_1) = 99{,}5 + 3/2\ (33{,}5) = 149{,}75$$

$$PLA = Q_3 + 3\ (Q_3 - Q_1) = 99{,}5 + 3\ (33{,}5) = 200$$

Diagram kotak garis yang dihasilkan diperlihatkan pada Gambar 2.2.

DAFTAR PUSTAKA

Fleming, M. C, and Nellis, J.G., (1994), *Principles of Applied Statistics*, First edition, Routledge, New York

Lane D., et al (2003), Introduction to Statistics, Online Edition https://onlinestatbook.com/Online_Statistics_Education.pdf

Nolan, B. (1994), *Data Analysis: An Introduction* , Cambridge, Polity Press.

Walpole, R. E., (1982), *Introduction to Statistics*, 3 rd edition, Macmillan Publishing Co. Inc., New York.

LATIHAN UNTUK BAB 2

Soal 1 :
Seorang inspektor pengendalian mutu menemukan jumlah bagian cacat hasil produksi, selama 15 periode produksi sebagai berikut

3, 10, 9, 4, 6, 10, 12, 6, 10, 7, 11, 9, 1, 13, 3

a. Hitung rata-rata aritmatika bagi data jumlah bagian cacat
b. Tentukan median dan modus bagi data jumlah bagian cacat

Pemecahan :

a. $\overline{X} = \dfrac{X_1 + X_2 + \ldots + X_n}{n} = \ldots.$

b. Data terurut : ……………………………………

Med = nilai data pada pengamatan ke (15+1) / 2
= nilai data pada pengamatan ke ….
Med = …………………

Soal 2 :

Data tentang massa bagasi pada Tabel 3, disajikan lagi pada Tabel 2.3 di bawah ini. Lengkapi lajur yang kosong, selajutnya tentukan ukuran-ukuran pemusatan rata-rata, median dan modus dari data tersebut.

Tabel 2.3 Sebaran Frekuensi

Kelas (kg)	Frekuensi (f_i)	Titik tengah (x_i)	$f_i . x_i$	Frekuensi kumulatif
0 - 9	5	5		5
10 - 19	11	15		16
20 - 29	20	25		36
30 - 39	9	35		45
40 - 49	13	45		58
50 - 60	7	55		65
	N = 65	180		--

Pemecahan :

a. Rata-rata :

$$\overline{X} = \frac{\sum_{i=1}^{p} f_i . x_i}{n}$$

$$\overline{X} = \frac{(5 \text{ x } 5) + (11\text{x}15) + . \; . \; . + (7\text{x}55)}{65} = \frac{\ldots\ldots\ldots}{65} = \ldots\ldots\ldots\ldots\ldots$$

b. Kelas median adalah kelas yang memuat nilai yang terletak di tengah data, yaitu data ke n / 2 , dengan n / 2 = … / 2 = …….
Pada lajur frekuensi kumulatif Tabel 2.3, sampai pada kelas ke … telah terdapat …. pengamatan. Ini berarti kelas ke … adalah kelas median, yaitu
"………… " . Kelas median memiliki batas bawah ….. dan frekuensi ….. .

Median :

$$\text{Md} = L_m + \frac{\frac{n}{2} - F_{p\text{-}1}}{f_m} . W_m$$

Median : $Md = \ldots\ldots\ldots + \left(\frac{\ldots\ldots\ldots - \ldots\ldots}{\ldots\ldots\ldots}\right)\ldots\ldots\ldots = \ldots\ldots\ldots$

c. Kelas yang memiliki frekuensi terbesar pada Tabel 2.3 adalah ……………., dengan frekuensi ……..
Kelas sebelum kelas modus memiliki frekuensi ………,
sehingga d_1 = …. - …..
Kelas setelah kelas modus memiliki frekuensi ……….. ,
sehingga d_2 = …. - ……

Modus : $Mod = L + \frac{d_1}{d_1 + d_2} . W$

Mod = ………………..

Soal 3 :
Seorang pemain basket menyumbangkan poin untuk timnya pada 12 kali pertandingan, sebagai berikut.

18, 3, 21, 15, 9, 34, 27, 10, 42, 6, 54, 62

Hitunglah :
a. Jangkauan
b. Jangkauan antarkuartil
c. Ragam
d. Simpangan baku

Pemecahan :

a. Jangkauan : $Rg = X_{maks} - X_{mim}$ = ……. – ……..
= ………

b. Jangkauan antarkuartil :
Q_1 terletak pada pengamatan ke (n+1)/4 .
dengan (n+1)/4 = (…. +1)/ 4 = …………

Sehingga $Q_1 = X_{\ldots} + \ldots\ldots (X_{\ldots} - X_{\ldots})$
Q_1 = …… + …… (…..-…….) = ………..

Q_3 terletak pada pengamatan ke 3(n+1)/4
dengan 3(n+1)/4 = 3 (…… +1)/ 4 = ……….

Sehingga Q_3 = $X_{....}$ + …… ($X_{...}$ - $X_{...}$)
Q_3 = …… + ……. (…….- ……..) = ……….

Jangkauan antarkuartil $Q_3 - Q_1$ = ……. – …… = ………

c. Ragam :

Untuk menghitung ragam, terlebih dahulu dilakukan hitungan pendahuluan berikut

$\sum X_i$ = 18 + 3 + . . . + 62 = ……..

$\sum X_i^2$ = $18^2 + 3^2$ + …. + 62^2 = ……

$$\text{Ragam } S^2 = \frac{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}}{n-1}$$

$$= \frac{\ldots\ldots\ldots - ((\ldots\ldots)^2/12)}{12-1}$$

= ………

d. Simpangan baku :
Simpangan baku adalah akar kuadrat dari ragam. Jika telah diperoleh ragam S^2 = …….. maka simpangan baku adalah S = $\sqrt{\ldots\ldots\ldots}$ = …….

Soal 4 :

Data tinggi 100 mahasiswi dicatat pada Tabel 2.4. Urutkan amatan dari yang terendah ke amatan tertinggi. Lakukan *tolly* untuk dapat melengkapi isi Tabel 2.5. Selanjutnya hitunglah :

a. Jangkauan
b. Jangkauan antarkuartil
c. Ragam
d. Simpangan baku

Tabel 2.4 Data Tinggi 100 Mahasiswi

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
[1,]	155	169	162	166	176	162	160	144	159	159
[2,]	162	149	167	162	161	163	153	163	153	143
[3,]	164	156	159	166	162	158	150	166	160	169
[4,]	168	148	170	168	172	156	160	161	157	158
[5,]	158	165	170	163	165	162	156	155	157	164
[6,]	167	164	154	155	161	154	157	160	162	169
[7,]	163	165	158	164	169	161	163	166	153	170
[8,]	156	163	161	156	150	171	160	151	163	167
[9,]	165	174	170	159	167	165	168	150	172	157
[10,]	165	160	154	164	158	160	170	160	164	159

Tabel 2.5 Sebaran Frekuensi

Kelas (cm)	**Frekuensi (f_i)**	**Titik tengah (x_i)**	**Frekuensi kumulatif**	**Frekuensi Relatif (%)**
140 - 144	2	142	2	
145 - 149	2	147	4	
150 - 154	….	….	….	
155 - 159	….	….	….	
160 - 164	….	….	….	
165 - 169				
170 - 174				
175 - 179				
	N = 100	…..	--	100

Pemecahan :

Q_1 terletak pada pengamatan n /4, dimana n /4 = …./4 = ….. Jika dilihat dari frekuensi kumulatif, pengamatan ke ….. berada pada kelas ‘ … - … ’. Sehingga,

$$Q_1 = L_{q1} + \frac{1\frac{n}{4} - F_{p-1}}{f_{q_1}} \cdot W_{q_1}$$

$$Q_1 = \ldots\ldots + \left(\frac{\ldots\ldots - \ldots\ldots}{\ldots\ldots}\right)\ldots\ldots = \ldots\ldots + \ldots\ldots = \ldots\ldots$$

Q_3 terletak pada pengamatan 3n /4, dimana 3n /4 = 3(……)/4 = …….. Jika dilihat dari frekuensi kumulatif, pengamatan ke …… berada pada kelas ‘ …. - …..’. Sehingga,

$$Q_3 = L_{q3} + \frac{3\frac{n}{4} - F_{p-1}}{f_{q_3}} \cdot W_{q_3}$$

$$Q_3 = \ldots\ldots + \left(\frac{\ldots\ldots - \ldots\ldots}{\ldots\ldots}\right)\ldots\ldots = \ldots\ldots + \ldots\ldots = \ldots\ldots$$

Jangkauan antarkuartil = $Q_3 - Q_1$ = ……. – ……. = …….

Ragam :

$$S^2 = \frac{\sum_{i=1}^{p} f_i (X_i)^2 - \left(\left(\sum_{i=1}^{p} f_i X_i\right)^2 / n\right)}{n-1}$$

$$= \frac{\ldots\ldots\ldots - ((\ldots\ldots)^2 / \ldots\ldots)}{\ldots\ldots - 1}$$

$$= \ldots\ldots$$

Simpangan baku : Simpangan baku adalah akar kuadrat dari ragam. Jika telah diperoleh ragam S^2 = …….. maka simpangan baku adalah S = $\sqrt{\ldots\ldots\ldots}$ = …….

BAB III

SEBARAN NORMAL

Pada bab ini akan dikemukakan tentang sebaran normal, yang meliputi ciri-ciri suatu sebaran yang menyebar normal, cara mentransformasi (membakukan) suatu variabel acak yang menyebar normal menjadi variabel acak yang menyebar normal baku, menentukan besar peluang suatu interval yang menyebar normal, dan menerapkan sebaran normal dalam pemecahan masalah. Khususnya penentuan besar peluang suatu interval, kita mempelajari bagaimana menentukan besar peluang suatu interval melalui penggunaan tabel sebaran normal baku. Demikian juga dengan menggunakan tabel sebaran normal baku, kita akan mempelajari cara menentukan nilai variabel acak yang diketahui besaran peluang kumulatifnya, baik besaran peluang kumulatif dari atas maupun besaran peluang kumulatif dari bawah. Penggunaan tabel sebaran normal baku dengan cara terakhir ini sering disebut penggunaan tabel invers sebaran normal baku.

.

Capaian Pembelajaran MK :

Menggunakan ilmu statistika dasar, baik **statistika deskriptip** maupun statistika inferensial dalam penelitian dan pemecahan masalah kehidupan sehari-hari.

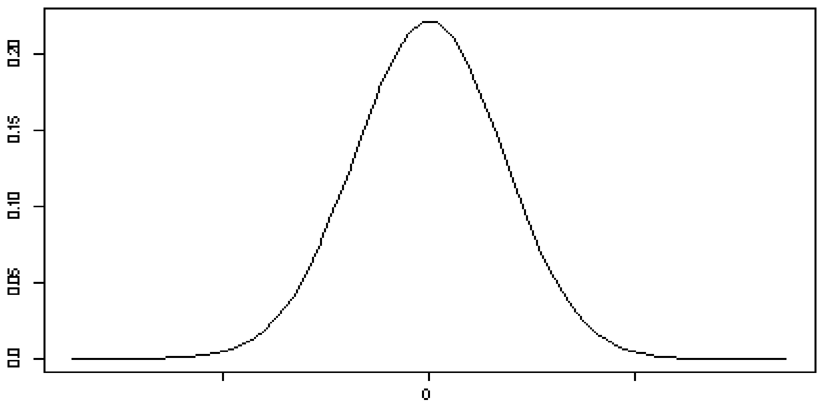
Capaian Pembelajaran BAB III :

Memahami sebaran peluang Normal; menjelaskan ciri-ciri sebaran peluang normal, menentukan besar peluang suatu inteval sebaran normal dengan bantuan tabel, dan menerapkan sebaran normal dalam pemecahan masalah

3.1 Kurva Normal

Sebaran peluang kontinu yang paling sering digunakan pada statistika inferensial adalah **sebaran normal.** Grafik sebaran normal sering disebut **kurva normal**, yaitu suatu gambar kurva yang berbentuk genta seperti diperlihatkan pada Gambar 3.1. Banyak sekali gugusan data yang dijumpai sehari-hari dapat didekati sebagai sebaran yang berbentuk sebaran normal atau atau sebaran peluangnya menyerupai sebaran normal. Akibatnya, sebaran normal dan grafik kurva normalnya sangat sering digunakan untuk gugusan data yang berasal dari berbagai bidang, seperti industri, ekonomi, pertanian, maupun untuk penelitian.

Sebaran normal sering disebut **sebaran Gauss**, untuk menghormati Gauss (1777 – 1855), yang berhasil menemukan persamaan sebaran normal melalui studi tentang galat dalam pengukuran yang berulang-ulang terhadap benda yang sama. Sebelumnya, DeMoivre telah berhasil menurunkan persamaan matematik bagi kurva normal ini pada tahun 1733.



Gambar 3.1 Kurva Sebaran Normal

Suatu variabel acak kontinu X yang memiliki sebaran berbentuk genta seperti dalam Gambar 3.1 disebut **variabel acak normal**. Persamaan matematik bagi sebaran peluang variabel acak normal ditentukan oleh dua parameter μ dan σ, yaitu nilai tengah dan simpangan bakunya. Oleh karena itu kita lambangkan nilai-nilai fungsi kepekatan bagi X ini dengan $n(x;\ \mu,\ \sigma)$

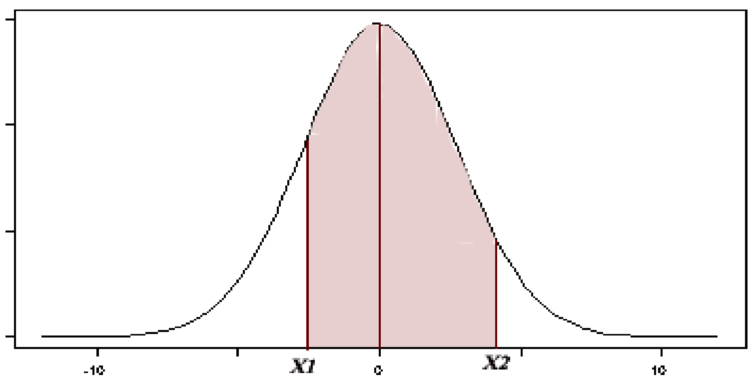
Bila X adalah suatu variabel acak normal dengan nilai tengah μ dan ragam σ^2, maka persamaan *kurva normal*nya adalah

$$n(x;\ \mu,\ \sigma) = \frac{1}{\sqrt{2\pi}\ \sigma} e^{-\frac{1}{2}\left(\frac{\chi-\mu}{\sigma}\right)^2},\ \text{untuk}\ -\infty < x < \infty\ ,\ \textbf{(Z1)} \longrightarrow (z1)$$

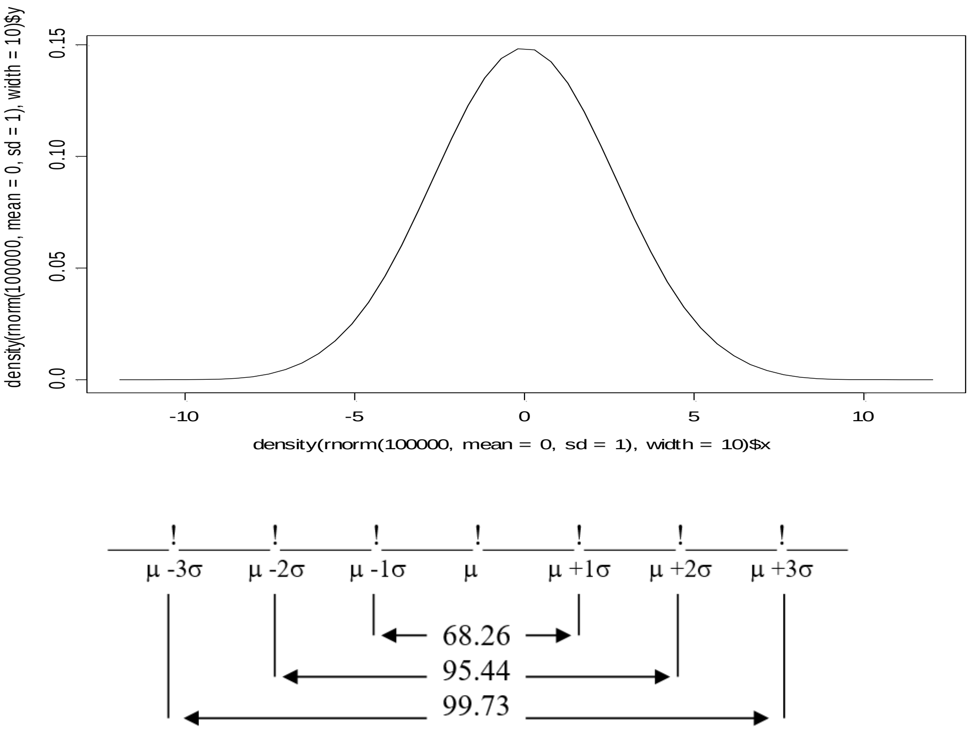
sedangkan dalam hal ini π = 3.14159... dan e = 2.71828.

3.2 Luas Daerah Dibawah Kurva Normal

Kurva sembarang sebaran peluang kontinu atau fungsi kepekatan dibuat sedemikian rupa sehingga luas daerah di bawah kurva itu dibatasi oleh $x = x_1$ dan $x = x_2$ sama dengan peluang bahwa variabel acak X mengambil nilai antara $x = x_1$ dan $x = x_2$. Jadi, bagi kurva normal dalam Gambar 3.2. Gambar $P(x_1 < X < x_2)$ dinyatakan oleh luas daerah yang diarsir.



Gambar 3.2 $P(x_1 < X < x_2)$ = luas daerah gelap



Gambar 3.3 Sifat-sifat Sebaran Normal

Sifaf-sifat sebaran normal sedemikian hingga, peluang semua pengamatan dari suatu variabel acak X yang menyebar normal, berada dalam rentangan *n* standar deviasi atas kedua sisi dari nilai rata-rata adalah :

a. 68.26 persen dari semua nilai X berada dalam rentangan satu standar deviasi atas kedua sisi nilai rata-rata, yaitu μ - 1σ sampai μ + 1σ.
b. 95.44 persen dari semua nilai X berada dalam rentangan dua standar deviasi atas kedua sisi nilai rata-rata, yaitu μ - 2σ sampai μ + 2σ.
c. 99.73 persen dari semua nilai X berada dalam rentangan tiga standar deviasi atas kedua sisi nilai rata-rata, yaitu μ - 3σ sampai μ + 3σ.

Sifat-sifat a, b, dan c di atas secara visual diperlihatkan pada Gambar 3.3.

3.3 Sebaran Normal Baku Z

Tidak efisien jika kita selalu berusaha menyusun tabel yang terpisah untuk setiap kurva normal bagi setiap pasangan nilai μ dan σ yang mungkin. Tetapi kita harus menggunakan tabel bila kita ingin menghindar dari keharusan menggunakan kalkulus integral. Untunglah kita dapat mentransformasikan setiap pengamatan yang berasal dari sembarang variabel acak normal X menjadi suatu nilai variabel acak normal baku Z dengan nilai tengah nol ($\mu = 0$) dan ragam satu ($\sigma^2 = 1$).

Gb 3.4. Pindah ke sini agar hal tak terpotong

Cara ini dapat dilakukan melalui transformasi $Z = \dfrac{X - \mu}{\sigma}$ **(Z2)** → ( )

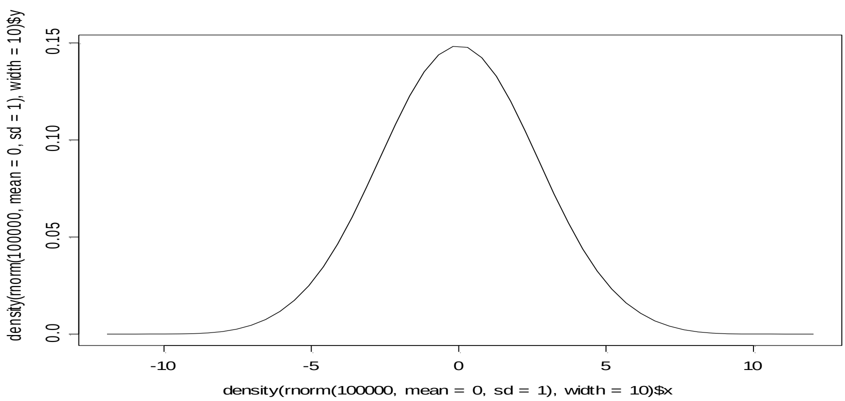
Nilai tengah Z adalah nol, karena $E(Z) = \dfrac{1}{\alpha} E(X - \mu) = \dfrac{1}{\alpha}(\mu - \mu) = 0$

Sedangkan ragamnya adalah $\alpha_z^2 = \alpha_{(X-\mu)/\sigma}^2 = \alpha_{X/\sigma}^2 = \dfrac{1}{\sigma^2}\sigma_x^2 = \dfrac{\sigma^2}{\sigma^2} = 1.$

Definisi : *Sebaran Normal Baku*. Sebaran variabel acak normal dengan nilai tengah nol dan simpangan baku 1 disebut *sebaran normal baku.*
Bila X berada diantara $x = x_1$ dan $x = x_2$, maka variabel acak Z akan berada diantara nilai-nilai padanannya

$$z_1 = \frac{x_1 - \mu}{\sigma} \quad \text{dan} \quad z_2 = \frac{x_2 - \mu}{\sigma} \qquad \textbf{(Z3)}$$

( )



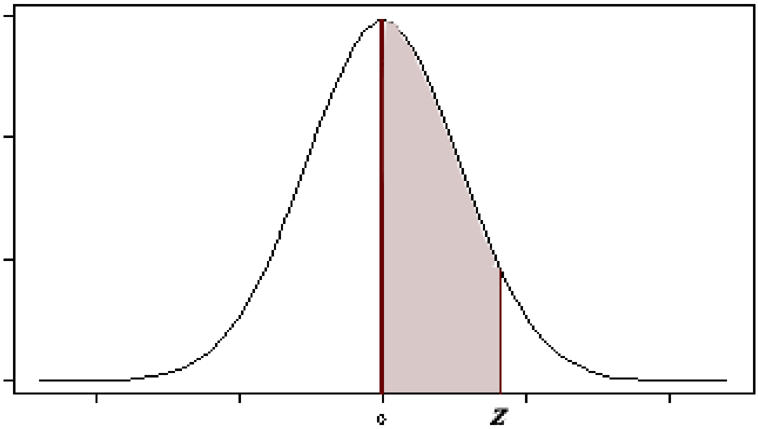
```
____!_______!_______!_______!_______!_______!_______!____  X
  μ -3σ    μ -2σ    μ -1σ     μ      μ +1σ   μ +2σ   μ +3σ
    !       !       !       !       !       !       !
____!_______!_______!_______!_______!_______!_______!____  Z
   -3      -2      -1       0       1       2       3
```

Gambar 3.4 Nilai variabel acak normal X dan Hasil Transformasi ke Normal Baku Z

Sebaran asal dan hasil transformasi ke sebaran normal baku diilustrasikan dalam Gambar 3.4. Karena semua nilai *X* yang jatuh antara x_1 dan x_2 mempunyai nilai-nilai *z* padanannya antara z_1 dan z_2, maka luas daerah dibawah kurva normal X antara $x = x_1$ dan $x = x_2$ dalam Gambar 3.1 sama dengan luas daerah dibawah kurva normal baku Z antara nilai hasil transformasi $z = z_1$ dan $z = z_2$. Dengan demikian $P(x_1 < X < x_2) = P(z_1 < Z < z_2)$

3.4 Penggunaan Tabel Sebaran Normal Baku

Dalam menggunakan tabel normal baku, kita menyederhanakan atau mengurangi banyaknya tabel luas kurva-normal menjadi hanya satu, yaitu yang berasal dari sebaran normal baku. Tabel A.1 mencantumkan luas daerah dibawah kurva normal baku yang merupakan nilai $P(0 < Z < z)$ untuk berbagai nilai *z* dari 0 sampai 5.49. Untuk ilustrasi penggunaan tabel ini, marilah kita hitung peluang *Z* yang mengambil nilai-nilai antara 0 sampai 1.74. Pertama-tama carilah nilai *z* yang sama dengan 1.7 pada kolom paling kiri, dan kemudian susurilah sepanjang garis tersebut sampai kolom dibawah 0.04, disitu kita membaca 0.4591. Jadi $P(0 < Z < 1.74) = 0.4591$.



Gambar 3.5 Peluang Z antara 0 sampai z
P (0 < Z < z)

Adakalanya, kita diminta untuk mencari nilai z untuk suatu nilai peluang tertentu, dan nilai z itu berada diantara nilai-nilai yang tercantum dalam Tabel A.1. Untuk mudahnya, kita akan mengambil nilai z yang ada dalam tabel itu yang nilai peluangnya yang paling mendekati nilai peluang yang diketahui. Tetapi, bila nilai peluang yang diketahui itu jatuh tepat ditengah-tengah antara nilai kedua tabel, maka nilai z kita dapatkan dari rata-rata dua nilai z padanan kedua nilai tabel tersebut. Misalnya, untuk mencari nilai z yang menghasilkan peluang sebesar 0.2975, yang terletak antara 0.2967 dan 0.2995 dalam tabel A.1, kita akan mengambil z = 0.83, karena 0.2975 lebih dekat ke 0.2967. Tetapi untuk peluang sebesar 0.2981, yang jatuh tepat ditengah-tengah 0.2967 dan 0.2995, kita akan mengambil z = 0.835.

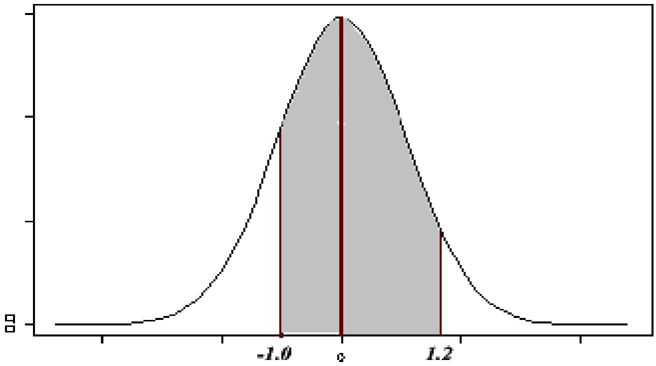
Contoh pengerjaan 3.1. Untuk sebaran normal dengan μ = 50 dan σ = 10, hitunglah peluang bahwa X mengambil sebuah nilai antara 40 dan 62.

Pemecahan : Nilai-nilai padanan x_1 = 40 dan x_2 = 62 adalah

$$z_1 = \frac{40-50}{10} = -1.0$$

$$z_2 = \frac{62-50}{10} = 1.2.$$

Dengan demikian, $P(40 < X < 62) = P(-1.0 < Z < 1.2)$



Gambar 3.6 P (-1.0 < Z < 1.2)

Nilai = $P(-1.0 < Z < 1.2)$ diberikan oleh daerah gelap dalam Gambar 3.6. Luas ini dapat diperoleh dengan menambahkan luas daerah dari -1 sampai 0 dengan luas daerah dari 0 sampai 1.2.. Dengan menggunakan Tabel A.1, kita memperoleh hitungan sebagai berikut

$$\begin{aligned} P(45 < X < 62) &= P(-1.0 < Z < 1.2) \\ &= P(-1.0 < Z < 0) + P(0 < Z < 1.2) \\ &= P(0 < Z < 1) + P(0 < Z < 1.2) \quad \text{(Normal Z simetri)} \\ &= 0.3413 + 0.3849 \\ &= 0.7262 \end{aligned}$$

Contoh pengerjaan 3.2.
Untuk sebaran normal dengan μ = 20 dan σ = 5, hitunglah peluang bahwa variabel acak *X* mengambil suatu nilai :

a. Antara 8 sampai 16
b. Lebih kecil sama dengan 15.

Pemecahan :

a. P (8 < X < 16)

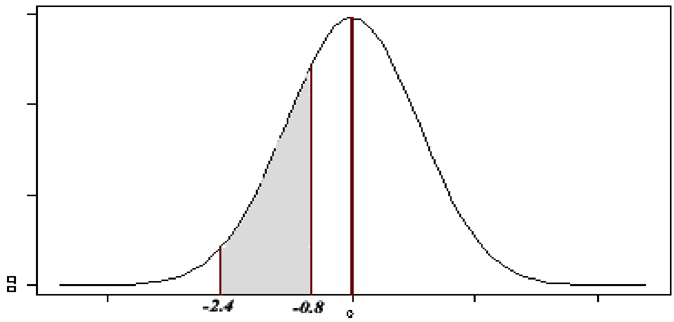
$$= P\left(\frac{8-20}{5} < Z < \frac{16-20}{5}\right)$$

= P (-2.4 < Z < -0.8)

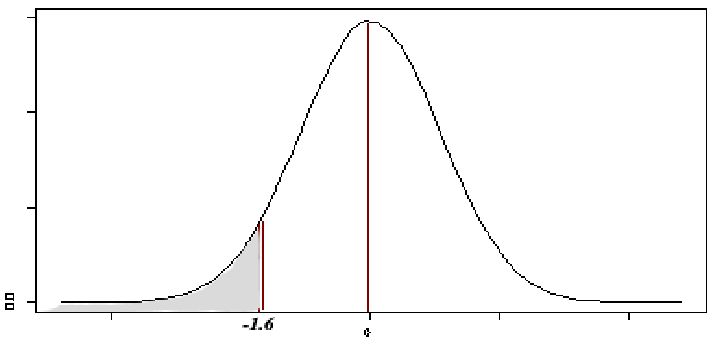
= P (0.8 < Z < 2.4)

= P (0 < Z < 2.4) - P (0 < Z < 0.8)

= 0.4918 – 0.2881 = 0.2037



Gambar 3.7 P (-2.4 < Z < -0.8)



Gambar 3.8 P (Z < -1.6)

b. P (X < 12)

$$= P\left(Z < \frac{12 - 20}{5} \right)$$

$$= P (Z < -1.6)$$

$$= P (Z > 1.6)$$

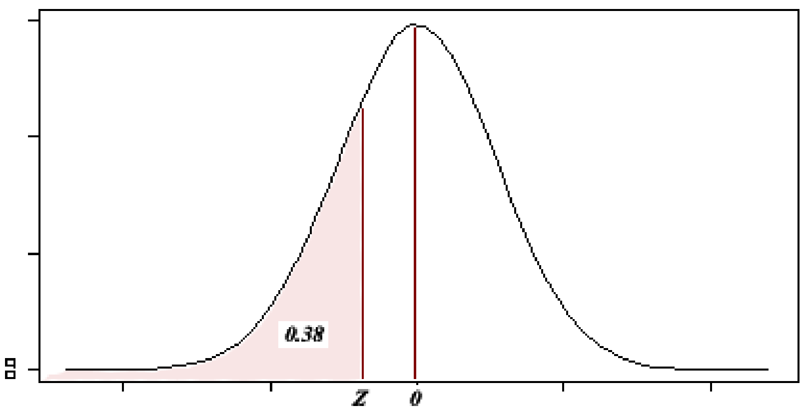
$$= 0.5 - P (0 < Z < 1.6)$$

$$= 0.5000 – 0.4452 = 0.0548$$

Contoh pengerjaan 3.3.
Diberikan sebuah sebaran normal dengan μ = 40 dan σ = 6. Hitunglah nilai *x* yang luas daerah dibawahnya ada 38 % .

Pemecahan : Dua teladan sebelumnya diselesaikan dengan bekerja dari nilai *x* ke nilai *z* dan kemudian dilanjutkan dengan menghitung luas yang diinginkan. Dalam teladan ini kita balik prosesnya; mulai dengan peluang atau luas daerah yang diketahui, kemudian menghitung nilai *z*-nya dan terakhir menentukan nilai *x* dengan cara mengubah rumus

$$z = \frac{x - \mu}{\sigma} \quad \text{menjadi} \quad x = \sigma z + \mu$$



Gambar 3.9 P (Z < z) = 0.38

Daerah seluas 0,38 berada disebelah kiri suatu nilai z, P(Z < z) = 0.38 , diperlihatkan pada Gambar 3.9. Karena z terletak disebelah kiri 0 maka z bernilai negatip. P(Z < z) = 0.5 - P (z < Z < 0), selanjutnya P (z < Z < 0) = 0.12. Jika diterapkan sifat simetri sebaran normal, P (z < Z < 0) = P (0 < Z < -z). Dari Tabel A.1 kita mendapatkan P (0 < Z < 0.31) = 0.12, selanjutnya diperoleh - z = 0.31 atau z = -0.31. Dengan demikian

$$x = (6)(-0.31) + 40$$
$$= 38.14.$$

Contoh pengerjaan 3.4.

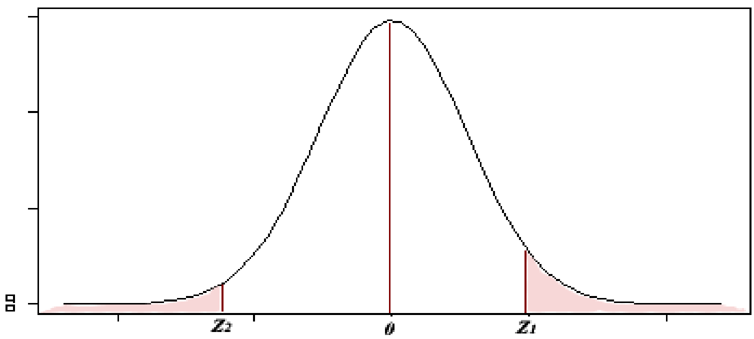
Diberikan sebuah sebaran Z-normal baku.

a. Hitunglah nilai z_1 yang luas daerah diatasnya ada 5 % .

b. Hitunglah nilai z_2 yang luas daerah dibawahnya 0.025.

Pemecahan :

a. Nilai z_1 yang dicari adalah nilai yang memberikan luas daerah P (Z > z_1) = 0.05. menyatu . Tidak Pisah



Gambar 3.10 P (Z > z_1) dan P (Z < z_2)

P (Z > z_1) = 0.5 - P (0 < Z < z_1)

$\Rightarrow$ 0.5 - P (0 < Z < z_1) = 0.05

$\Rightarrow$ P (0 < Z < z_1) = 0.5 - 0.05

$\Rightarrow$ P (0 < Z < z_1) = 0.45

Dari Tabel A1 : P (0 < Z < z_1) = 0.45, maka z_1 = 1.64

b. Nilai z_2 yang dicari adalah nilai yang memberikan luas daerah $P(Z < z_2) = 0.025$. Pada Gambar 3.10 nilai z_2 terletak disebelah kiri 0, maka z_2 bernilai negatip .

$P(Z < z_2) = P(Z > -z_2)$ (sifat simetri sebaran normal)

$\Rightarrow \quad P(Z > -z_2) = 0.5 - P(0 < Z < -z_2)$

$\Rightarrow \quad 0.5 - P(0 < Z < -z_2) = 0.025$

Dari Tabel A1: $P(0 < Z < -z_2) = 0.475$, maka $-z_2 = 1.96$, selanjutnya $z_2 = -1.96$

DAFTAR PUSTAKA

Baarda, D.B., de Goede, M.P.M. and van Dijkum, C.J. (2019) Introduction to Statistics with SPSS, Introduction to Statistics with SPSS. Available at: https://doi.org/10.4324/9781003021704.

Fleming, M. C, and Nellis, J.G., (1994), *Principles of Applied Statistics*, First edition, Routledge, New York

Lane D., et al (2003), *Introduction to Statistics*, Online Edition https://onlinestatbook.com/Online_Statistics_Education.pdf

Lumen Learning, (2023), *Introduction to Statistics*, Simple book production, Online Course Provided by: Lumen Learning. License: CC BY: Attribution https://courses.lumenlearning.com/introstats1/

Walpole, R. E., (1982), *Introduction to Statistics*, 3 rd edition, Macmillan Publishing Co. Inc.

Soal latihan

1. Jika variabel acak Z menyebar normal baku, dengan menggunakan tabel Z-normal
   baku hitunglah peluang :

   a. P (-2 < Z < 1,5)
   b. P (Z< 1,97)
   c. P (Z > 1,96)

2. Jika variabel acak X menyebar normal dengan nilai tengah 43 dan simpangan baku 5 dengan bantuan tabel Z-normal baku hitunglah peluang :

   a. P (40 < X < 49)
   b. P (X < 41)
   c. P (X > 50)

3. Dengan bantuan tabel, tentukan nilai z (Z_{hitung}) jika luas daerah disebelah kanannya adalah α= 0.05 ($Z_{0.05}$).

BAB IV

PENGUJIAN HIPOTESIS

Aspek utama dalam penerapan statistika inferensial, setelah pendugaan adalah pengujian hipotesis. Dalam statistika, pengujian hipotesis merupakan bagian terpenting untuk mengambil keputusan. Dengan melakukan pengujian hipotesis seorang peneliti akan dapat menjawab pertanyaan-pertanyaan yang diajukan, dengan menyatakan penolakan atau penerimaan hipotesis.

Pada bab ini akan dipelajari bagaimana langkah langkah dalam melakukan pengujian suatu hipotesis. Berkenaan dengan pengujian hipotesis, maka isi dari bab ini akan mengemukakan bagaimana merumuskan suatu hipotesis uji berdasarkan tujuan pengujiannya, menentukan kriteria penolakan hipotesis nol dengan menunjukan daerah kritis untuk suatu pemilihan taraf signifikan pengujian. Demikian juga bagaimana menentukan dan menghitung nilai dari statistik yang digunakan, meliputi statistik Z, statistik T, dan statistik F, serta menyimpulkan hasil pengujian hipotesis.

Hipotesis adalah jawaban sementara sebelum percobaan dilaksanakan, yang didasarkan hasil studi literatur. Hipotesis biasanya memuat pernyataan yang bersifat netral atau yang umum terjadi. Kebenaran hipotesis secara pasti tidak pernah diketahui kecuali jika dilakukan pengamatan terhadap seluruh anggota populasi. Untuk melakukan hal ini sangatlah tidak efisien apalagi bila ukuran populasi sangat besar.

Penarikan sejumlah sampel acak dari suatu populasi, diamati karakteristiknya dan kemudian dibandingkan dengan hipotesis yang diajukan merupakan suatu langkah melakukan uji hipotesis. Apabila sampel acak ini memberikan indikasi yang mendukung hipotesis yang diajukan, maka hipotesis tersebut diterima. Sebaliknya bila sampel acak memberikan indikasi yang bertentangan dengan hipotesis yang diajukan, maka hipotesis tersebut ditolak.

Pengertian diterima atau ditolak suatu hipotesis tidak bersifat mutlak. Suatu hipotesis ditolak tidak berarti bahwa hipotesis tersebut salah, melainkan data telah memberi petunjuk bahwa telah ada perubahan pada karakteristik populasi yang dihipotesiskan. Penerimaan hipotesis berarti bahwa belum cukup bukti untuk menerima hipotesis tandingan.

Capaian Pembelajaran MK :

Menggunakan ilmu statistika dasar, baik statistika deskriptip maupun **statistika inferensial** dalam penelitian dan pemecahan masalah kehidupan sehari-hari.

Capaian Pembelajaran Bab 4 :

Mampu melakukan pengujian hipotesis kesamaan rata-rata dari suatu populasi dan kesamaan rata-rata dua populasi yang independen.

4.1 Pengujian Hipotesis

Hipotesis statistik dibedakan menjadi dua, yaitu hipotesis nol (H_0) dan hipotesis tandingan (H_1). Pernyatan yang ingin ditolak kebenarannya ditetapkan sebagai hipotesis nol, sedangkan hipotesis lawannya ditetapkan sebagai hipotesis tandingan.

Dalam pengujian hipotesis dikenal dua macam kesalahan, yaitu kesalahan jenis I, dan kesalahan jenis II. Kesalahan jenis I adalah kesalahan menolak H_0 padahal sesungguhnya H_0 benar. Kesalahan jenis II adalah kesalahan menerima H_0 padahal sesungguhnya H_1 yang benar.

Peluang terjadinya kesalahan jenis I dilambangkan dengan α (alpha), yang sering juga disebut taraf nyata atau taraf signifikan pengujian. Sedangkan peluang terjadinya kesalahan jenis II dilambangkan dengan β (beta). Dalam pengujian hipotesis statistik, besar peluang kedua jenis kesalahan diharapkan berimbang dan disesuaikan dengan permasalahan yang dihadapi. Peluang $1-\alpha$ disebut interval kepercayaan (confidence interval), menyatakan peluang menerima H_0 dan H_0 memang benar. Peluang $1-\beta$ disebut kuasa pengujian (power of test), menyatakan peluang menolak H_0 jika H_0 memang salah.

Tahapan yang dilakukan dalam menguji suatu hipotesis adalah :

1. Merumuskan hipotesis uji atau hipotesis statistiknya
2. Memilih taraf signifikan pengujian yang ingin digunakan, α (alpha)
3. Menentukan dan menghitung statistik yang digunakan. Statistik yang digunakan disesuaikan dengan hipotesis yang diuji, jumlah sampel, dan pengetahuan tentang asumsi sebaran yang digunakan.
4. Memutuskan hasil pengujian apakah menolak atau masih menerima hipotesis nol (H_0) berdasarkan kriteria pengujian.

4.2 Pengujian Nilai Rata-Rata Untuk Satu Populasi

Data suatu sampel yang diambil dari suatu populasi dapat digunakan untuk mengkaji karakteristik dari populasi asal. Salah satu karakteristik populasi yang paling menarik untuk dikaji adalah nilai rata-rata atau pemusatan data. Ada tiga bentuk perumusan hipotesis uji untuk nilai rata-rata satu populasi, yaitu :

1. $H_0 : \mu = \mu_o$
   $H_1 : \mu \neq \mu_o$

2. $H_0 : \mu \geq \mu_o$
   $H_1 : \mu < \mu_o$ **(A1)** (A1)

3. $H_0 : \mu \leq \mu_o$
   $H_1 : \mu > \mu_o$

Sesuai dengan tahapan pengujian hipotesis yang telah dikemukakan pada bagian sebelumnya, maka langkah-langkah penting untuk melakukan uji hipotesis di atas adalah

a. Menetapkan hipotesis uji, yaitu hipotesis nol dan hipotesis tandingan. Hipotesis uji ini adalah salah satu dari rumusan hipotesis di atas.

b. Menghitung statistik sampel, yaitu nilai rata-rata sampel $\overline{X}$

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \qquad \textbf{(A2)}$$

( )

c. Menentukan dan menghitung statistik uji.

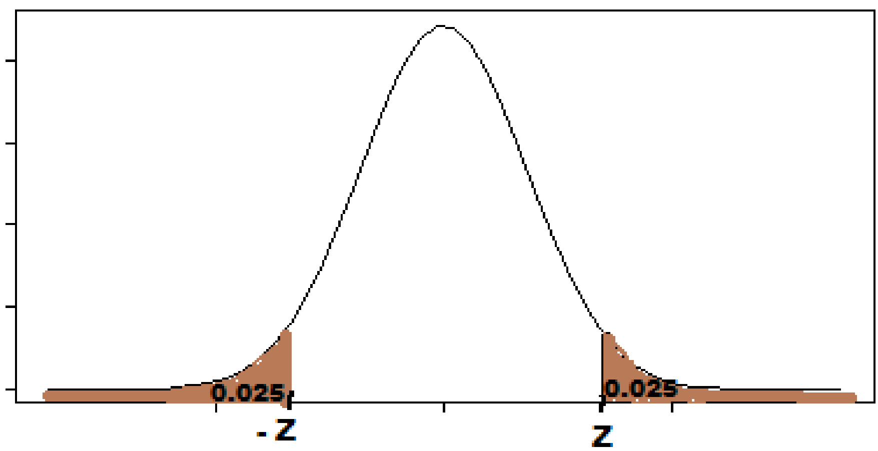
Jika ragam populasi (σ^2) diketahui atau ragam sampel (S^2) telah stabil digunakan untuk menduga ragam populasi (σ^2), statistik uji yang digunakan adalah normal baku (z). Ragam sampel S^2 telah stabil digunakan untuk menduga ragam populasi σ^2 apabila ukuran sampel $n \geq 30$. Selanjutnya ditulis $S^2 = \hat{\sigma}^2$. Statistik uji yang digunakan adalah Z-normal baku :

$$z_{\text{hitung}} = \frac{\overline{X} - \mu_0}{\sigma_{\overline{x}}} = \frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \quad \text{atau} \quad z_{\text{hitung}} = \frac{\overline{X} - \mu_0}{\hat{\sigma} / \sqrt{n}} \quad \textbf{(A3)}$$

Jika ragam populasi tidak diketahui dan ukuran sampel n < 30 ,maka statistik uji yang digunakan adalah t-student, sebagai berikut :

$$t_{\text{hitung}} = \frac{\overline{X} - \mu_0}{S_{\overline{x}}} = \frac{\overline{X} - \mu_0}{S / \sqrt{n}} \quad \textbf{(A4)}$$ dengan derajat bebas n-1.

d. Menetapkan daerah kritis untuk menolak hipotesis nol (*critical region to reject H_0*). Penetapan daerah kritis sangat tergantung pada tiga hal yaitu bentuk hipotesis tandingan, statistik uji yang digunakan dan besarnya taraf signifikan pengujian. Arah penolakan hipotesis nol searah dengan hipotesis tandingan.

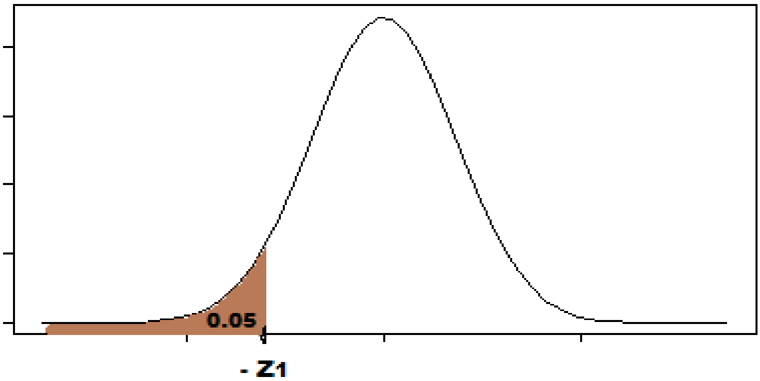


Gambar 4.1 Daerah penolakan H_0 jika H_1: $\mu \neq \mu_0$

Jika H_1: $\mu \neq \mu_0$ maka daerah kritisnya

$|z_{hitung}| > z_{\alpha/2}$ atau $|t_{hitung/2}| > t_{\alpha/2,db=n-1}$ **(A5)**

Daerah penolakan ini diperlihatkan pada Gambar 4.1



Gambar 4.2 Daerah penolakan Jika H_1: $\mu < \mu_0$

Jika H_1: $\mu < \mu_0$ maka daerah kritisnya,

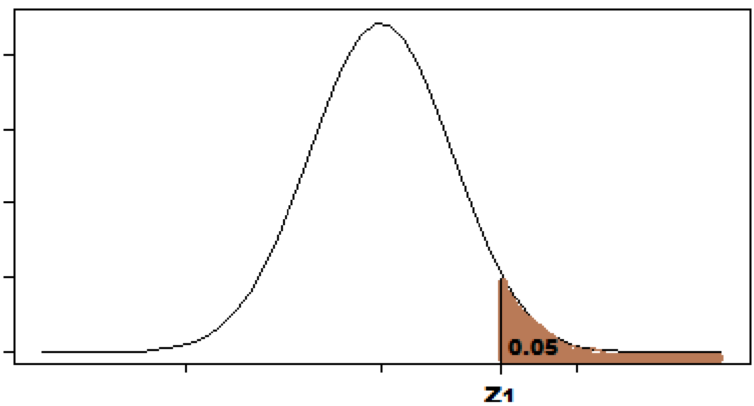
$z_{hitung} < -z_{\alpha}$ atau $t_{hitung} < -t_{\alpha,db=n-1}$ **(A6)** ( )

Daerah atau nilai penolakan ini diperlihatkan pada Gambar 4.2

Jika H_1: $\mu > \mu_0$ maka daerah kritisnya

enter

$z_{hitung} > z_{\alpha}$ atau $t_{hitung} > t_{\alpha,db=n-1}$ **(A7)** Daerah penolakan ini diperlihatkan pada Gambar 4.3



Gambar 4.3 Daerah penolakan Jika H_1: $\mu > \mu_0$

Contoh Pengerjaan 4.1:

Pada suatu pengumpulan data dengan menarik sampel dari suatu populasi didapat data setelah diurutkan sebagai berikut :

Tabel 4.1 Data Penarikan Sampel

7.8910	8.7619	8.8901	9.1033	9.8530
9.8728	10.0135	10.3151	10.3221	10.8333
11.0250	11.5518	11.9744	12.3591	12.7180

Apabila kita mengetahui bahwa data tersebut dari populasi yang menyebar normal dengan ragam 2 dan ingin diketahui apakah populasi tersebut masih memiliki nilai tengah 10, maka kita akan melakukan uji hipotesis berikut, jika dimisalkan ingin digunakan taraf signifikan α = 0,05.

Pemecahan pengerjaan 4.1:

Bentuk hipotesis :

$H_0 : \mu = 10$

$H_1 : \mu \neq 10$

Taraf signifikan α = 0,05.

Karena ragam populasi diketahui maka statistik uji yang akan digunakan adalah statistik Z, yaitu:

$Z = \dfrac{10.366 - 10}{\sqrt{2}/\sqrt{15}} = 1.002$ dan dari tabel : $|Z_{\alpha/2}| = |Z_{0.05/2}| = 1.96$.

Hasil Pengujian: Karena $|Z_{hitung}| < |Z_{\alpha/2}|$, yaitu 1.002 < 1.96 maka kita menerima H_0, artinya belum cukup data (bukti) untuk menolak hipotesis H_0.

Contoh pengerjaan 4.2:

Suatu pabrik aki (battery) mengklaim bahwa rata-rata hidup aki mencapai 55 jam. Pada hasil tes yang dilakukan terhadap suatu hasil produksi *batch* yang terdiri dari 40 aki, diperoleh rata-rata hidup 50 jam, dengan simpangan baku 11.734 jam. Lakukan pengujian hipotesis dengan taraf signifikan 1 persen bahwa

a. rata-rata hidup aki memang 55 jam
b. rata-rata hidup aki kurang dari 55 jam

Pemecahan pengerjaan 4.2:

Data : $\mu_0 = 55$; $\bar{x} = 50$; s = 11.734; n = 40

a. Hipotesis : $H_0 : \mu_0 = 55$
$H_1 : \mu_0 \neq 55$

Taraf Signifikan : $\alpha = 0.01$

Simpangan baku dari $\bar{X}$: $s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{11.734}{\sqrt{(40)}} = 1.86$

Statistik Z-hitung : $z_{hitung} = \frac{\bar{X} - \mu_0}{s_{\bar{x}}} = \frac{50 - 55}{1.86}$
$= -2.69$

Nilai kritis : Z-tabel = $z_{\alpha/2} = 2.58$ (dari table normal baku)

Hasil pengujian : $|z_{hitung}| > z_{\alpha/2}$, yaitu 2.69 > 2.58 . Selanjutnya H_0 ditolak dengan taraf signifikan 1 persen. Disimpulkan bahwa rata-rata hidup aki tidak sama dengan 55 jam.

b. Hipotesis : H_0 : $\mu_0 = 55$;

H_1 : $\mu_0 < 55$

Taraf Signifikan : $\alpha = 0.01$

Z-tabel = z_α = 2.33 (dari table normal baku)

Hasil pengujian : $z_{hitung} < -z_\alpha$, yaitu -2.69 > -2.33 . Selanjutnya H_0 ditolak dengan taraf signifikan 1 persen. Disimpulkan bahwa rata-rata hidup aki kurang dari 55 jam.

Contoh Pengerjaan 4.3:

Hasil survey yang lalu terhadap angkatan kerja menyatakan bahwa dalam satu tahun, rata-rata banyaknya hari setiap pekerja absen karena sakit adalah 15 hari. Seorang peneliti menggunakan sampel acak 25 pekerja, dan mencatat ketidakhadiran masing-masing pekerja dalam setahun sebagai berikut.

Tabel 4.2 Data Absen Pekerja

5	25	10	0	3	50	12	14	40
12	32	8	4	47	20	14	16	10
1	22	58	5	23	18	9		

Lakukan pengujian hipotesis dengan taraf signifikan 5 persen bahwa

a. rata-rata pekerja absen dalam setahun adalah 15 hari
b. rata-rata pekerja absen lebih besar dari 15 hari.

Pemecahan (a) :

Data : $\mu_0 = 15$; n = 25

Dengan perhitungan $\bar{x} = 18.32$; s = 15.845

Hipotesis : H_0 : $\mu_0 = 15$;

H_1: $\mu_0 \neq 15$

Taraf signifikan : α = 0.05
Nilai kritis : t = $t_{\alpha/2}$ = 2.064 (dari tabel dengan (25-1) derajat bebas)

Simpangan baku $\overline{X}$: $s_{\overline{x}} = \frac{s}{\sqrt{n}} = \frac{15.845}{\sqrt{(25)}} = 3.169$

$$t_{hitung} = \frac{\overline{X} - \mu_0}{s_{\overline{x}}} = \frac{18.32 - 15}{3.169}$$
$$= 1.05$$

Hasil pengujian : $|t_{hitung/2}| < t_{\alpha/2, db=n-1}$, yaitu 1.05 < 2.064, selanjutnya H_0 diterima atau belum cukup bukti untuk menolak H_0. Disimpulkan bahwa rata-rata pekerja absen dalam satu tahun memang sama dengan 15 hari.

b. Hipotesis : $H_0 : \mu_0 = 15$
$H_1 : \mu_0 > 15$

Taraf signifikan : α = 0.05
Nilai kritis : t = t_{α} = …….. (dari tabel-t dengan 25-1 derajat bebas)
Hasil pengujian : ………………..,yaitu……………., selanjutnya H_0 ……………..
Disimpulkan bahwa ……………………….

4.3 Uji Perbedaan Nilai Rata-Rata Dua Populasi

Suatu penelitian sering bertujuan membandingkan nilai rata-rata (aritmatika) dua populasi. Dua populasi ini adalah populasi-populasi dimana masing-masing dua sampel berasal. Metoda statistika menyediakan beberapa cara menguji perbedaan nilai rata-rata dua populasi. Pada metoda statistika parametrik, uji perbedaan dapat dilakukan dengan pendekatan sebaran normal (uji normal baku-z), dan pendekatan sebaran t-student (uji t-student). Kedua pendekatan sebaran ini dapat dilakukan apabila terhadap data hasil pengukuran layak diterapkan asumsi sebaran normal. Disebabkan sebaran normal adalah sebaran dari variabel acak kontinu, maka data yang layak disesuaikan dengan pendekatan asumsi sebaran normal adalah data pengukuran kuantitatif dan kontinu. Pendekatan asumsi sebaran normal layak diterapkan pada data hasil pengukuran skala interval ataupun skala rasio.

4.3.1 Uji Perbedaan Rata-rata Dengan Uji Normal Baku

Pendekatan sebaran normal, dengan uji normal baku-Z digunakan jika ragam (variansi) dua populasi diketahui atau ragam sampel (S^2) telah stabil digunakan untuk menduga ragam populasi (σ^2). Ragam sampel S^2 telah stabil digunakan untuk menduga ragam populasi σ^2 apabila ukuran sampel $n \geq 30$. Demikian pula untuk ukuran sampel $n \geq 30$, sebaran bagi $\overline{X}$ dari peubah acak X (tidak diketahui sebarannya) akan membentuk sebaran normal

Hipotesis uji untuk menguji kesamaan rata-rata dua populasi yang saling bebas (uji dua arah) dirumuskan sebagai berikut :

H_0 : $\mu_1 = \mu_2$ (N1) (N1)

H_1 : $\mu_1 \neq \mu_2$

Keterangan :

μ_1 = rata-rata populasi pertama

μ_2 = rata-rata populasi kedua

Untuk menguji hipotesis di atas dengan pendekatan sebaran normal, digunakan rumus untuk nilai Z-hitung sebagai berikut :

a. Jika kedua ragam populasi (σ_1^2 dan σ_2^2) diketahui, maka

$$\text{Z-hitung} = \frac{\overline{X}_1 - \overline{X}_2}{\sigma_{\overline{X}_1 - \overline{X}_2}} \quad (\text{N2}) \qquad (\ \ )$$

Dimana

$$\sigma_{\overline{X}_1-\overline{X}_2} = \sqrt{\frac{\sigma_1^{\,2}}{n_1}+\frac{\sigma_2^{\,2}}{n_2}} \quad (\text{N3})$$

keterangan :

$\sigma_{\overline{X}_1-\overline{X}_2}$ = simpangan baku populasi untuk selisih dua nilai rata rata.
σ_1^2 = ragam dari populasi pertama
σ_2^2 = ragam dari populasi kedua
n_1 = jumlah pengamatan pada sampel pertama
n_2 = jumlah pengamatan pada sampel kedua

b. Jika kedua ragam populasi tidak diketahui tetapi $n_1 \geq 30$ dan $n_2 \geq 30$, maka

$$\text{Z-hitung} = \frac{\overline{X}_1-\overline{X}_2}{\hat{\sigma}_{\overline{X}_1-\overline{X}_2}} \quad (\text{N4})$$

Dimana

$$\hat{\sigma}_{\overline{X}_1-\overline{X}_2} = \sqrt{\frac{\hat{\sigma}_1^{\,2}}{n_1}+\frac{\hat{\sigma}_2^{\,2}}{n_2}} \quad (\text{N5})$$

Keterangan :

$\hat{\sigma}_{\overline{X}_1-\overline{X}_2}$ = dugaan simpangan baku populasi untuk selisih dua nilai rata-rata
$\hat{\sigma}_1^{\,2}$ = dugaan ragam populasi pertama, diduga dari ragam sampel S_1^2
$\hat{\sigma}_2^{\,2}$ = dugaan ragam populasi kedua, diduga dari ragam sampel kedua S_2^2
n_1 = jumlah pengamatan pada sampel pertama
n_2 = jumlah pengamatan pada sampel kedua

Kriteria pengujian:

Jika pada pengujian ini digunakan taraf signifikan sebesar α dan Z-hitung diperoleh dari (N2) atau (N4), maka

H_0 diterima jika $|$ Z-hitung $| \leq Z_{\alpha/2}$-tabel , atau
$Pr = [\, P\,(\, Z \leq -|z_{hitung}|\,) + P\,(\, Z \geq |z_{hitung}|\,)\,] \geq \alpha$, sebaliknya

H_0 ditolak jika $|$ Z-hitung $| > Z_{\alpha/2}$-tabel , atau
$Pr = [\, P\,(\, Z \leq -|z_{hitung}|\,) + P\,(\, Z \geq |z_{hitung}|\,)\,] < \alpha$.

Untuk menguji apakah rata-rata populasi pertama lebih kecil dari rata-rata populasi kedua, digunakan uji satu arah dengan rumusan hipotesis uji sebagai berikut

H_0 : $\mu_1 = \mu_2$ (N6) →
H_1 : $\mu_1 < \mu_2$ ( )

Pengujian hipotesis (N6) dengan pendekatan sebaran normal, menggunakan rumus untuk nilai Z-hitung yang sama dengan nilai Z-hitung pada pengujian hipotesis (N1), yaitu :

a. Jika kedua ragam populasi (σ_1^2 dan σ_2^2) diketahui, maka

$$\text{Z-hitung} = \frac{\overline{X}_1 - \overline{X}_2}{\sigma_{\overline{X}_1 - \overline{X}_2}} \qquad (\text{ N2 }) \rightarrow \qquad (\ )$$

b. Jika kedua ragam tidak diketahui tetapi $n_1 \geq 30$ dan $n_2 \geq 30$, maka

$$\text{Z-hitung} = \frac{\overline{X}_1 - \overline{X}_2}{\hat{\sigma}_{\overline{X}_1 - \overline{X}_2}} \qquad (\text{ N4 }) \rightarrow \qquad (\ )$$

Kriteria pengujian:

Jika pada pengujian ini digunakan taraf signifikan sebesar α , maka

H_0 diterima jika Z-hitung $\geq$ - Z_α-tabel atau Pr ($Z \leq z_{\text{hitung}}$) $\geq \alpha$, sebaliknya

H_0 ditolak jika Z-hitung < - Z_α-tabel atau Pr ($Z \leq z_{\text{hitung}}$) < α .

Contoh Pengerjaan 4.4:

Suatu pabrik bola golf memperkenalkan produksi terbaru dari bola golfnya (tipe baru), dan menyatakan lebih bagus dari bola golf yang lama (tipe lama). Jarak tempuh kedua tipe bola golf masing masing memiliki simpangan baku : tipe lama σ_1 = 9.8 meter dan tipe baru σ_2 = 7.1 meter. Seorang golfer ingin menguji pernyataan di atas dengan mengambil secara acak 35 bola tipe lama dan 40 bola tipe baru. Jarak tempuh bola diukur satu persatu, dan hasilnya dicatat bahwa rata-rata jarak tempuh bola tipe lama $\overline{X}_1$ = 20.1 meter, dan bola tipe baru $\overline{X}_2$ = 23.6 meter. Dengan menggunakan taraf signifikan pengujian α = 0.05 ,ujilah hipotesis bahwa

a. rata-rata jarak tempuh kedua tipe bola golf sama
b. rata-rata jarak tempuh bola golf tipe lama lebih pendek dari rata-rata jarak tempuh bola golf tipe baru.

Pemecahan Pengerjaan 4.4:

σ_1	= 9.8 meter	σ_2	= 7.1 meter
$\overline{X}_1$	= 20.1 meter	$\overline{X}_2$	= 23.6 meter
n_1	= 35	n_2	= 40

Hipotesis (a) H_0 : $\mu_1 = \mu_2$

H_1 : $\mu_1 \neq \mu_2$

(b) H_0 : $\mu_1 = \mu_2$

H_1 : $\mu_1 < \mu_2$

Taraf signifikan pengujian : $\alpha = 0.05$

Simpangan baku : $\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

$$= \sqrt{\frac{(9.8)^2}{35} + \frac{(7.1)^2}{40}} = 2.0$$

Z-hitung $= \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\bar{X}_1 - \bar{X}_2}} = \frac{20.1 - 23.6}{2.0} = -1.75$

Hasil pengujian :

a). Pada tabel normal baku ; $\pm Z_{\alpha/2}$-tabel = $\pm$ 1.96
Kesimpulan : $|$ Z-hitung $|$ > $Z_{\alpha/2}$-tabel, yaitu 1.75 < 1.96, sehingga belum cukup bukti untuk menolak H_0. Kita simpulkan bahwa tidak terdapat perbedaan yang signifikan dari rata-rata jarak tempuh kedua tipe bola golf.

b). Pada tabel normal baku ; - Z_{α}-tabel = - 1.64
Kesimpulan : Z-hitung < - Z_{α}-tabel, yaitu - 1.75 < -1.64, selanjutnya H_0 ditolak dengan taraf signifikan 5 persen. Kita simpulkan bahwa rata-rata jarak tempuh bola golf tipe lama signifikan lebih pendek dari rata-rata jarak tempuh bola golf tipe baru.

Contoh Pengerjaan 4.5:

Data tentang skor tes dari hasil dua metode pembelajaran, yaitu metode mengajar dengan lembar kerja peserta didik (LKPD) dan metode mengajar drill dengan soal-soal (DRILL) dicatat berturut-turut pada kolom 2 dan kolom 3 dari Tabel 4.3. Ujilah hipotesis bahwa rata-rata skor tes antara kedua hasil metoda pengajaran (a) tidak berbeda, (b) metode Drill lebih bagus dari metode Lkpd.

Pemecahan (Hitung Manual):

Rata-rata : $\overline{X}_1 = \dfrac{\sum_{i=1}^{30} X_i}{30} = \dfrac{1976}{30} = 65.867$ $\quad \overline{X}_2 = \dfrac{\sum_{i=1}^{30} X_i}{30} = \dfrac{2218}{30} = 73.933$

Tabel 4.3 Skor Hasil Pembelajaran Metode Lkpd dan Metode Drill

N0.	LKS	DRILL		
	X_1	X_2	X_1^2	X_2^2
[1,]	72	71	5184	5041
[2,]	67	78	4489	6084
[3,]	63	71	3969	5041
[4,]	70	66	4900	4356
[5,]	63	79	3969	6241
[6,]	69	75	4761	5625
[7,]	72	75	5184	5625
[8,]	64	72	4096	5184
[9,]	63	77	3969	5929
[10,]	62	77	3844	5929
[11,]	56	71	3136	5041
[12,]	68	79	4624	6241
[13,]	65	74	4225	5476
[14,]	68	71	4624	5041
[15,]	59	66	3481	4356
[16,]	69	78	4761	6084
[17,]	66	72	4356	5184
[18,]	64	71	4096	5041
[19,]	60	76	3600	5776
[20,]	61	77	3721	5929
[21,]	68	76	4624	5776
[22,]	65	77	4225	5929
[23,]	67	77	4489	5929
[24,]	71	70	5041	4900
[25,]	74	78	5476	6084
[26,]	64	73	4096	5329
[27,]	69	81	4761	6561
[28,]	65	66	4225	4356

[29,]	64	72	4096	5184
[30,]	68	72	4624	5184
	$\sum X_1 = 1976$	$\sum X_2 = 2218$	$\sum X_1^2 = 130646$	$\sum X_2^2 = 164456$

Ragam : $$S^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\,\overline{X}^2}{n-1}$$

$$S_1^{\,2} = \frac{(130646) - 30\,(65.867)^2}{30-1} = 17.016 \qquad S_2^{\,2} = \frac{(164456) - 30\,(73.933)^2}{30-1} = 16.271$$

Hipotesis (a) $H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

(b) $H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 < \mu_2$

Taraf signifikan pengujian : $\alpha = 0.01$

Simpangan baku : $$\hat{\sigma}_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{\hat{\sigma}_1^{\,2}}{n_1} + \frac{\hat{\sigma}_2^{\,2}}{n_2}}$$

$$= \sqrt{\frac{(17.016)}{30} + \frac{(16.271)}{30}} = 1.053$$

Z-hitung $$= \frac{\overline{X}_1 - \overline{X}_2}{\hat{\sigma}_{\overline{X}_1 - \overline{X}_2}}$$

$$= \frac{65.867 - 73.933}{1.053} = -7.66$$

Hasil pengujian :

a).Pada tabel normal baku ; $\pm Z_{\alpha/2}$-tabel = $\pm$ 2.57
Kesimpulan : $|$ Z-hitung $|$ > $Z_{\alpha/2}$-tabel, yaitu 7.66 > 2.57, selanjutnya H_0 ditolak dengan taraf signifikan 1 persen. Kita simpulkan bahwa terdapat perbedaan yang signifikan antara rata-rata skor tes dari metode Lkpd dengan rata-rata skor tes dari metode Drill.

b).Pada tabel normal baku ; $- Z_{\alpha}$-tabel = - 2.33
Kesimpulan : Z-hitung < $- Z_{\alpha}$-tabel, yaitu -7.66 < -2.33, selanjutnya H_0 ditolak dengan taraf signifikan 1 persen. Kita simpulkan bahwa rata-rata skor tes dari metode Lkpd secara signifikan lebih kecil dari rata-rata skor tes dari metode Drill.

<u>Pemecahan (Dengan Komputerisasi)</u> :

Pada Tabel 4.4 disajikan *output* hasil komputerisasi dengan SPSS untuk data pada Tabel 4.3. Untuk pengujian hipotesis dari hasil komputer, keputusan menolak atau menerima hipotesis nol H_0 dilakukan dengan membandingkan taraf signifikan statistik yang digunakan dengan taraf signifikan pengujian. Jika taraf signifikan statistik (sig) kurang dari taraf signifikan pengujian (α), maka H_0 ditolak, jika sebaliknya maka H_0 diterima.

Hasil komputerisasi pada Tabel 4.4, menunjukan bahwa nilai statistik t = -7,685 dengan taraf sigifikan 0,000. Oleh karena taraf signifikan statistik T kurang dari taraf signifikan pengujian, yaitu 0,000 < 0,05 maka H_0 ditolak. Disimpulkan bahwa terdapat perbedaan yang signifikan antara rata-rata skor tes dari metode Lkpd dengan rata-rata skor tes dari metode Drill.

Tabel 4.4 Hasil Komputerisasi dengan SPSS untuk Uji T

	Levene's Test for Equality of Variances		t-test for Equality of Means		
	F	Sig.	t	df	Sig. (2-tailed)
Equal variances assumed	.015	.901	-7.658	58	.000
Equal variances not assumed			-7.658	57.971	.000

4.3.2 Uji Perbedaan Rata-Rata Dua Populasi Dengan Uji t-Student

Pendekatan sebaran normal, dengan uji t-student digunakan jika ragam (variansi) dari kedua populai tidak diketahui dan ragam sampel (S^2) belum stabil digunakan untuk menduga ragam populai (σ^2). Ragam sampel S^2 tidak stabil digunakan menduga ragam populasi σ^2 apabila ukuran sampel $n < 30$. Dengan demikian uji t-student untuk perbedaan rata-rata dua populasi digunakan apabila kedua ragam tidak diketahui , dan $n_1 < 30$ atau $n_2 < 30$.

1. Hipotesis uji untuk menguji kesamaan rata-rata dua populasi yang saling bebas (uji dua arah) dirumuskan sebagai berikut :

   H_0 : $\mu_1 = \mu_2$ (T1) → ( )

   H_1 : $\mu_1 \neq \mu_2$

   Dimana :

   μ_1 = rata-rata populasi pertama

   μ_2 = rata-rata populasi kedua

2. Hipotesis uji untuk menguji rata-rata populasi pertama lebih kecil dari rata-rata populasi kedua (Uji satu arah untuk dua populasi saling bebas) dirumuskan sebagai berikut :

H_0 : $\mu_1 = \mu_2$ (T2)

H_1 : $\mu_1 < \mu_2$

Untuk menguji hipotesis di atas (baik hipotetesis (T1) maupun (T2)) dengan pende-katan sebaran t-student, digunakan rumus untuk nilai t-hitung sebagai berikut :

a. Jika ada indikasi bahwa kedua ragam populasi homogen ($\sigma_1^2 = \sigma_2^2 = \sigma^2$) , maka

$$\text{t-hitung} = \frac{\overline{X}_1 - \overline{X}_2}{S_{\overline{X}_1 - \overline{X}_2}} \quad (\text{T3})$$

Dimana

$$S_{\overline{X}_1 - \overline{X}_2} = S_{gab}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad (\text{T4})$$

$$S_{gab} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \quad (\text{T5})$$

keterangan :

$S_{\overline{X}_1 - \overline{X}_2}$ = simpangan baku dari selisih dua nilai rata-rata sampel.

S_{gab} = simpangan baku gabungan sampel pertama dengan sampel kedua

S_1^2 = ragam sampel pertama

S_2^2 = ragam sampel kedua

n_1 = jumlah pengamatan pada sampel pertama

n_2 = jumlah pengamatan pada sampel kedua

Derajat bebas untuk t-hitung pada (T3) dengan simpangan baku yang diperoleh dari (T4), adalah Db = $n_1 + n_2 - 2$

b. Jika kedua ragam populasi tidak homogen ($\sigma_1^2 \neq \sigma_2^2$), maka

$$\text{t-hitung} = \frac{\overline{X}_1 - \overline{X}_2}{S_{\overline{X}_1 - \overline{X}_2}} \quad (\text{T6})$$

Dimana

$$S_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{S_1^{\ 2}}{n_1} + \frac{S_2^{\ 2}}{n_2}} \quad (\text{T7})$$

Keterangan :

$S_{\overline{X}_1 - \overline{X}_2}$ =simpangan baku dari selisih dua nilai rata-rata sampel.
S_1^2 = ragam sampel pertama
S_2^2 = ragam sampel kedua
n_1 = jumlah pengamatan pada sampel pertama
n_2 = jumlah pengamatan pada sampel kedua

Derajat bebas efektif untuk t-hitung pada (T6) dengan simpangan baku yang diperoleh dari (T7) adalah

$$Db_{ef} = \frac{[(S_1^2/n_1) + (S_2^2/n_2)]^2}{\dfrac{(S_1^2/n_1)^2}{n_1 - 1} + \dfrac{(S_2^2/n_2)^2}{n_1 - 1}} \quad (\text{T8})$$

Kriteria pengujian:
Jika pada pengujian ini digunakan taraf signifikan sebesar α , maka
a. Kriteria pengujian dua arah untuk hipotesis uji pada (T1) adalah H_0 diterima jika $|\text{t-hitung}| \leq t_{\alpha/2;Db}$-tabel , atau Pr = [P ($t \leq -|t_{\text{hitung}}|$) + P ($t \geq |t_{\text{hitung}}|$)] $\geq \alpha$, sebaliknya H_0 ditolak jika $|\text{t-hitung}| > t_{\alpha/2;Db}$-tabel , atau Pr = [P ($t \leq -|t_{\text{hitung}}|$) + P ($t \geq |t_{\text{hitung}}|$) $< \alpha$.

b. Kriteria pengujian satu arah untuk hipotesis uji pada (T2) adalah

H_0 diterima jika t-hitung $\geq$ - $t_{\alpha;Db}$-tabel , atau Pr (T $\leq$ t_{hitung}) $\geq$ α , sebaliknya

H_0 ditolak jika t-hitung < - $t_{\alpha;Db}$-tabel , atau Pr (T $\leq$ t_{hitung}) < α .

Contoh Pengerjaan 4.6 :

Data tentang skor tes dari hasil belajar menggunakan pembelajaran *Discovery Learning* (DL) dan *Problem-based Learning* (PBL) dicatat berturut-turut pada kolom 2 dan 3 dari Tabel 4.5. Ujilah hipotesis bahwa rata-rata skor tes antara kedua hasil metoda pembelajaran: (a) tidak berbeda, (b) metode *Problem-based Learning* (PBL) lebih bagus dari metode *Discovery Learning* (DL).

Tabel 4.5 Hasil Belajar dari Metode *Discovery Learning* (DL) dan Metode *Problem-based Learning* (PBL)

N0.	DL X_1	PBL X_2	X_1^2	X_2^2
[1,]	61	76	3721	5776
[2,]	65	78	4225	6084
[3,]	69	66	4761	4356
[4,]	72	70	5184	4900
[5,]	68	86	4624	7396
[6,]	64	77	4096	5929
[7,]	60	84	3600	7056
[8,]	67	78	4489	6084
[9,]	61	83	3721	6889
[10,]	61	74	3721	5476
[11,]	65	75	4225	5625
[12,]	68	85	4624	7225
[13,]	65	79	4225	6241
[14,]	64	69	4096	4761
[15,]	62	75	3844	5625
[16,]	60	78	3600	6084
[17,]	56	75	3136	5625
[18,]	68	71	4624	5041

[19,]	57	78	3249	6084
[20,]	68	82	4624	6724
[21,]	69	79	4761	6241
[22,]	61	73	3721	5329
[23,]	61	60	3721	3600
[24,]	61	84	3721	7056
[25,]	67	83	4489	6889
$\sum X_1 = 1600$		$\sum X_2 = 1918$	$\sum X_1^2 = 102802$	$\sum X_2^2 = 148096$

Pemecahan (hitung manual) :

Rata-rata : $\overline{X}_1 = \frac{\sum_{i=1}^{25} X_i}{25} = \frac{1600}{25} = 64$

$$\overline{X}_2 = \frac{\sum_{i=1}^{25} X_i}{25} = \frac{1918}{25} = 76.72$$

Ragam : $S^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\,\overline{X}^2}{n-1}$

$$S_1{}^2 = \frac{(102802) - 25\,(64)^2}{25-1} = 16.75 \qquad S_2{}^2 = \frac{(148096) - 25\,(76.72)^2}{25-1} = 39.46$$

Hipotesis (a) H_0 : $\mu_1 = \mu_2$

H_1 : $\mu_1 \neq \mu_2$

(b) H_0 : $\mu_1 = \mu_2$

H_1 : $\mu_1 < \mu_2$

Taraf signifikan pengujian : $\alpha = 0.05$

Jika ada indikasi ragam tidak sama, maka :

$$S_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{S_1^{\,2}}{n_1}+\frac{S_2^{\,2}}{n_2}} = \sqrt{\frac{16.75}{25}+\frac{39.46}{25}} = 1.4994$$

$$\text{t-hitung} = \frac{\bar{X}_1-\bar{X}_2}{S_{\bar{X}_1-\bar{X}_2}} = \frac{64-76.72}{1.4994} = -8.48$$

Derajat bebas efektif : $Db_{ef} = \dfrac{[(S_1^2/n_1)+(S_2^2/n_2)]^2}{\dfrac{(S_1^2/n_1)^2}{n_1-1}+\dfrac{(S_2^2/n_2)^2}{n_1-1}}$

$$= \frac{((16.75/25)+(39.46/25))^2}{((16.75/25)^2/(25-1))+((39.46/25)^2/(25-1))}$$

Db_{ef} = 41.264

a. Pada Tabel t-student : $t_{\alpha/2;Db}$-tabel = $t_{0.025;41}$-tabel = 2.021
Oleh karena | t-hitung | > $t_{\alpha/2;Db}$-tabel , yaitu 8.48 > 2.021 maka H_0 ditolak pada taraf signifikan 5 persen. Disimpulkan bahwa terdapat perbedaan yang signifikan antara rata-rata skor tes dari metode *Discovery Learning* (DL) dengan metode *Problem-based Learning* (PBL).

b. Pada Tabel t-student : - $t_{\alpha;Db}$-tabel = - $t_{0.05;41}$-tabel = - 1.684
Oleh karena t-hitung < - $t_{0.05;41}$-tabel, yaitu - 8.48 < -1.684, maka H_0 ditolak pada taraf signifikan 5 persen. Disimpulkan bahwa rata-rata skor tes dari metode *Discovery Learning* signifikan lebih rendah dari rata-rata skor tes dari metode *Problem-based Learning* (PBL).

Pemecahan (Dengan Komputerisasi) :

Tabel 4.6 Hasil Komputerisasi dengan SPSS

	Levene's Test for Equality of Variances		t-test for Equality of Means		
	F	Sig.	t	df	Sig. (2-tailed)
Equal variances assumed	2.296	.136	-8.483	48	.000
Equal variances not assumed			-8.483	41.264	.000

Pada Table 4.6, terlebih dahulu diperlihatkan hasil pengujian kesamaan dua ragam populasi dengan menggunakan statistik F. Pengujian kesamaan dua ragam dimaksudkan untuk memilih salah satu dari dua statistik T yang akan digunakan pada uji perbedaan rata-rata, yaitu uji statistik t dengan asumsi ragam sama (homogen) dan uji statistit T dengan asumsi ragam berbeda. Hasil komputerisasi pada Tabel 4.6, menunjukan bahwa nilai statistik F = 2,296 dengan taraf sigifikan 0,136. Oleh karena taraf signifikan statistik F lebih dari taraf signifikan pengujian, yaitu 0,136 > 0,05 maka H_0 diterima atau belum cukup bukti untuk menolak H_0 . Disimpulkan bahwa kedua ragam populasi tidak berbeda. Dengan demikian, untuk pengujian kesamaan rata-rata digunakan uji t dengan asumsi ragam homogen.

Hasil komputerisasi pada Tabel 4.6, menunjukan bahwa nilai statistik T = -8,483 dengan taraf sigifikan 0,000. Oleh karena taraf signifikan statistik T kurang dari taraf signifikan pengujian, yaitu 0,000 < 0,05 maka H_0 ditolak. Disimpulkan bahwa terdapat perbedaan yang signifikan antara rata-rata skor tes dari metode *Discovery Learning* (DL) dengan metode *Problem-based Learning* (PBL).

DAFTAR PUSTAKA

Fleming, M. C, and Nellis, J.G., (1994), *Principles of Applied Statistics*, First edition, Routledge, New York

Lane D., et al (2003), *Introduction to Statistics*, Online Edition https://onlinestatbook.com/Online_Statistics_Education.pdf

Lumen Learning, (2023), *Introduction to Statistics*, Simple book production, Online Course Provided by: Lumen Learning. License: CC BY: Attribution https://courses.lumenlearning.com/introstats1/

Walpole, R. E., (1982), *Introduction to Statistics*, 3 rd edition, Macmillan Publishing Co. Inc.

LATIHAN UNTUK BAB 4

a. Uji Perbedaan Rata-rata Dua Populasi Dengan Uji Normal Baku

Soal 1 :

Suatu pabrik bola golf memperkenalkan produksi terbaru dari bola golfnya (tipe baru), dan menyatakan lebih bagus dari bola golf yang lama (tipe lama). Jarak tempuh kedua tipe bola golf masing masing memiliki simpangan baku : tipe lama σ_1 = 8.5 meter dan tipe baru σ_2 = 9.2 meter. Seorang golfer ingin menguji pernyataan di atas dengan mengambil secara acak 30 bola tipe lama dan 35 bola tipe baru. Jarak tempuh bola diukur satu persatu, dan hasilnya dicatat bahwa rata-rata jarak tempuh bola tipe lama $\overline{X}_1$ = 21.3 meter, dan bola tipe baru $\overline{X}_2$ = 26.5 meter. Dengan taraf signifikan pengujian $\alpha = 0.05$,ujilah hipotesis bahwa

a. rata-rata jarak tempuh kedua tipe bola golf sama
b. rata-rata jarak tempuh bola golf tipe lama lebih pendek dari rata-rata jarak tempuh bola golf tipe baru.

<u>Pemecahan :</u>

σ_1	= ……. meter	σ_2	= …… meter
$\overline{X}_1$	= …….. meter	$\overline{X}_2$	= …….. meter
n_1	= 30	n_2	= 35

Hipotesis (a) $H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

(b) $H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 < \mu_2$

Taraf signifikan pengujian : $\alpha = 0.05$

Simpangan baku : $\sigma_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

$$= \sqrt{\frac{(\ldots\ldots)^2}{30} + \frac{(\ldots\ldots)^2}{\ldots\ldots}} = \ldots\ldots$$

$$\text{Z-hitung} = \frac{\overline{X}_1 - \overline{X}_2}{\sigma_{\overline{X}_1 - \overline{X}_2}}$$

$$= \frac{\ldots\ldots\ldots - \ldots\ldots\ldots}{\ldots\ldots\ldots} = \ldots\ldots\ldots$$

Hasil pengujian :

a). Pada tabel normal baku ; $\pm Z_{\alpha/2}$-tabel = $\pm$ 1.96
Kesimpulan : $|$ Z-hitung $|$ …… $Z_{\alpha/2}$-tabel, yaitu …………… 1.96, selanjutnya H_0 ………. dengan taraf signifikan 5 persen. Kita simpulkan bahwa …………

b). Pada tabel normal baku ; - Z_α-tabel = - 1.64
Kesimpulan : Z-hitung …… - Z_α-tabel, yaitu -……… -1.64, selanjutnya H_0 …….. dengan taraf signifikan 5 persen. Kita simpulkan bahwa ……………..

b. Uji Perbedaan Rata-Rata Dua Populasi Dengan Uji t-Student

Soal 2 :

Data tentang skor tes dari hasil metode pembelajaran ekspositori dan pembelajaran kelompok dicatat berturut-turut pada kolom 2 dan 3 dari Tabel 4.7. Ujilah hipotesis bahwa rata-rata skor tes antara kedua hasil metoda pembelajaran (a) tidak berbeda, (b) metode kelompok lebih bagus dari metode ekspositori.

Tabel 4.7 Hasil Belajar dari Metode pembelajaran Ekspositori dan Metode Pembelajaran Kelompok

N0.	Ekspositori X_1	Kelompok X_2	X_1^2	X_2^2
[1,]	59	68	3481	4624
[2,]	65	71	4225	5041
[3,]	59	72	3481	5184
[4,]	60	72	3600	5184
[5,]	60	72	3600	5184
[6,]	65	69	4225	4761
[7,]	61	79	3721	6241
[8,]	60	74	3600	5476
[9,]	66	68	4356	4624
[10,]	62	70	3844	4900
[11,]	66	68	4356	4624
[12,]	65	67	4225	4489
[13,]	59	71	3481	5041
[14,]	66	66	4356	4356
[15,]	64	76	4096	5776
[16,]	57	73	3249	5329
[17,]	61	68	3721	4624
[18,]	61	72	3721	5184
[19,]	60	75	3600	5625
[20,]	68	64	4624	4096
[21,]	69	70	4761	4900
[22,]	58	71	3364	5041
[23,]	62	82	3844	6724
[24,]	63	75	3969	5625
[25,]	60	74	3600	5476
[26,]	63	71	3969	5041
	$\sum X_1 = 1619$	$\sum X_2 = 1858$	$\sum X_1^2 = 101069$	$\sum X_2^2 = 133170$

Pemecahan (hitung manual) :

Rata-rata : $\overline{X}_1 = \dfrac{\sum_{i=1}^{26} X_i}{26} = \dfrac{\ldots\ldots\ldots}{26} = \ldots.$ $\overline{X}_2 = \dfrac{\sum_{i=1}^{26} X_i}{26} = \dfrac{\ldots\ldots\ldots}{26} = \ldots.$

Ragam : $S^2 = \dfrac{\sum_{i=1}^{n} X_i^2 - n\,\overline{X}^2}{n-1}$

$$S_1{}^2 = \frac{(\ldots\ldots\ldots..) - 26\,(\ldots\ldots..)^2}{26-1} = \ldots\ldots\ldots..$$

$$S_2{}^2 = \frac{(\ldots\ldots\ldots..) - 26\,(\ldots\ldots..)^2}{26-1} = \ldots\ldots\ldots..$$

Hipotesis (a) H_0 : $\mu_1 = \mu_2$

H_1 : $\mu_1 \neq \mu_2$

(b) H_0 : $\mu_1 = \mu_2$

H_1 : $\mu_1 < \mu_2$

Taraf signifikan pengujian : $\alpha = 0.05$

Jika ada indikasi ragam sama, maka :

$$S_{gab} = \sqrt{\frac{(26-1)\ldots\ldots. + (26-1)\ldots\ldots\ldots}{26+26-2}}$$

$= \ldots\ldots..$

$$S_{\bar{X}_1-\bar{X}_2} = S_{gab}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = \text{……..}\sqrt{\frac{1}{26}+\frac{1}{26}}$$

$$= \text{………..}$$

$$\text{t-hitung} = \frac{\bar{X}_1-\bar{X}_2}{S_{\bar{X}_1-\bar{X}_2}} = \frac{\text{………}-\text{……….}}{\text{……….}} = \text{………}$$

Derajat bebas untuk ragam gabungan : $n_1 + n_2 - 2$ = ……….

a. Pada Tabel t-student : $t_{\alpha/2;Db}$-tabel = $t_{0.025;41}$-tabel = 2.021
Oleh karena $|$ t-hitung $|$ … $t_{\alpha/2;Db}$-tabel , yaitu ……… 2.021 maka H_0 …….
pada taraf signifikan 5 persen. Disimpulkan bahwa ………..

b. Pada Tabel t-student : - $t_{\alpha;Db}$-tabel = - $t_{0.05;41}$-tabel = - 1.684
Oleh karena t-hitung …. - $t_{0.05;41}$-tabel, yaitu ……… -1.684 maka H_0 ………. pada taraf signifikan 5 persen. Disimpulkan bahwa ……………………..

3. Suatu survey terhadap 210 rumah tangga di Kab. Kutai kartanegara mencatat bahwa rata-rata pendapatan rumahtangga perbulan adalah 1.2 juta dengan simpangan baku 0.12 juta. Survey yang sama dilakukan terhadap 210 rumahtangga di Kab. Kutai timur, mencatat bahwa rata-rata pendapatan rumahtangga perbulan adalah 0.97 juta dengan simpangan baku 0.13 juta. Ujilah hipotesis dengan taraf signifikan 5 persen bahwa
   a. Rata-rata pendapatan rumahtangga Di Kab. Kutai timur lebih rendah dari rata-rata pendapatan rumahtangga di Kab. Kutai kartanegara.
   b. Rata-rata pendapatan rumahtangga kedua Kabupaten tidak berbeda.

4. Sebuah pabrik bohlam memproduksi dua jenis bohlam, dan mengklaim bahwa bohlam jenis kedua lebih bagus dari bohlam jenis pertama. Dari sampel acak berukuran 24 pada kedua jenis bohlam,

dicatat rata-rata dan simpangan baku atas lama menyala bohlam (jam), yaitu bohlam jenis I : $\overline{X}_1$ = 5800 jam s = 31 jam, dan bohlan jenis II : $\overline{X}_2$ = 6000 jam , s = 30 jam. Ujilah hipotesis dengan taraf signifikan 1 persen bahwa

a. rata-rata dari lama menyala kedua bohlam sama.
b. rata-rata dari lama menyala bohlam jenis II lebih rendah dari Jenis I.

5. Hasil belajar dari dua model pembelajaran, yaitu Inquiry learning dan Discovey learning disajikan pada Tabel 4.8. Lakukan pengujian dengan taraf signifikan α = 0,05, untuk menjawab pertanyaan apakah kedua hasil pembelajaran tersebut tidaksignifikan berbeda.

Tabel 4.8 Hasil Belajar dari model pembelajaran Inquiry learning dan Discovey learning

No.	Inquiry	Discovery	X1^2	X2^2	X1.X2
	X1	X2			
1	85	76	7225	5776	6460
2	75	76	5625	5776	5700
3	77	72	5929	5184	5544
4	73	76	5329	5776	5548
5	77	74	5929	5476	5698
6	75	71	5625	5041	5325
7	82	59	6724	3481	4838
8	80	65	6400	4225	5200
9	82	74	6724	5476	6068
10	83	75	6889	5625	6225
11	83	60	6889	3600	4980
12	80	63	6400	3969	5040
13	79	68	6241	4624	5372
14	81	74	6561	5476	5994
15	80	69	6400	4761	5520
16	80	78	6400	6084	6240
17	80	77	6400	5929	6160
18	76	68	5776	4624	5168
19	78	64	6084	4096	4992
20	73	67	5329	4489	4891

No.	Inquiry	Discovery	X1^2	X2^2	X1.X2
	X1	X2			
21	77	65	5929	4225	5005
22	73	65	5329	4225	4745
23	82	68	6724	4624	5576
24	81	69	6561	4761	5589
25	80	63	6400	3969	5040
26	84	77	7056	5929	6468
	2056	1813	162878	127221	143386

BAB V

REGRESI LINIER

Bab ini berisikan pelajaran tentang analisis regresi linier. Analisis regresi adalah salah satu teknik analisis statistika untuk keperluan generalisasi dari sampel ke populasi (inferensia). Analisis regresi biasanya diterapkan pada data hasil penelitian yang diperoleh dari metode pengumpulan data dengan survey. Metode pengumpulan data dengan survey ditujukan untuk data yang telah tersedia di lapangan atau tidak diperlukan eksperimen untuk membangkitkan data yang belum ada. Hal ini berakibat bahwa pengumpulan data dengan survey tidak melakukan pengendalian terhadap penyelidikan tentang bagaimana variabel bebas berpengaruh terhadap variabel terikat. Data dari hasil pengumpulan data dengan eksperimen juga dapat dianalisis menggunakan regresi linier, dimana faktor perlakukan dinyatakan sebagai variabel kategori dalam persamaan.

Analisis regresi meliputi analisis regresi linier dan analisis regresi nonlinier. Pada bab ini, pembahasan kita hanya pada analisis regresi linier, yang meliputi analisis regresi linier sederhana dan analisis regresi linier ganda, termasuk analisis regresi menggunakan variabel kategori. Pada bagian materi analisis regresi ganda, disajikan tahapan perhitungan dalam pengolahan data yang menggunakan aljabar matriks.

Demikian pula untuk memperluas dan memperdalam jangkauan analisis maka pembahasan tentang analisis regresi ganda dalam bab ini akan dilengkapi dengan penerapan komputerisasi, menggunakan perangkat lunak *Statistical Process for the Social Sciences* (SPSS). Pada materi analisis regresi ganda, pengujian hipotesis tidak terbatas hanya pada pengujian keberartian model regresi dan pengujian koefisien secara partial, namun juga menyajikan pengujian hipotesis tentang uji koefisien sequential dan uji hipotesis linier gabungan.

Capaian Pembelajaran MK :

Menggunakan ilmu statistika dasar, baik statistika deskriptip maupun **statistika inferensial** (pada bab ini khususnya analisis regresi) dalam penelitian dan pemecahan masalah kehidupan sehari-hari.

5.1 Pengertian Regresi Linier

Regresi linier (Univariat) merupakan analisis statistika yang memodelkan hubungan beberapa variabel menurut bentuk hubungan persamaan linier eksplisit. Persamaan linier bentuk eksplisit adalah persamaan linier yang menempatkan suatu variabel secara tunggal pada salah satu ruas persamaan.

Satu variabel eksplisit pada model tersebut merupakan variabel acak, dan kemungkinan besar memiliki perilaku yang tergantung oleh variabel-variabel yang lain. Variabel yang menjadi perhatian utama ini dinyatakan sebagai variabel terikat (respon), dengan simbol Y. Sebagai contoh variabel ini, dapat berupa kematian yang disebabkan oleh suatu penyakit, tingkat harga menurut kondisi pasar, dan prestasi belajar dari suatu metode mengajar.

Variabel lainnya dalam model persamaan linier ini adalah variabel yang mungkin dapat memberikan informasi tentang perilaku variabel terikat Y. Variabel ini ditempatkan sebagai variabel prediktor atau variabel penjelas di dalam model persamaan linier. Variabel-variabel ini merupakan variabel tetap yang telah diketahui (tidak acak), selanjutnya dinyatakan sebagai variabel bebas, dengan simbol X.

Secara umum pemodelan regresi linier ini bertujuan menyajikan bagaimana nilai rata-rata variabel terikat " E (Y) " berubah menurut perubahan masing-masing variabel bebas. Diasumsikan bahwa ragam Y tidak dipengaruhi oleh perubahan masing-masing variabel bebas. Selanjutnya persamaan **regresi linier dinyatakan sebagai tempat kedudukan nilai harapan Y pada setiap nilai-nilai X tetap**. Nilai-nilai harapan ini memiliki sebaran yang identik dengan ragam yang sama.

5.2 Regresi Linier Sederhana

5.2.1 Model dan Pendugaan Koefisien

Model regresi linier sederhana melibatkan hanya satu variabel bebas X. Model ini menyatakan perubahan secara konstan nilai rata-rata variabel respon Y menurut perubahan (peningkatan atau penurunan) nilai variabel penjelas X. Hubungan ini dinyatakan dalam bentuk persamaan linier

$E(Y_j) = \beta_0 + \beta_1 X_j$ **(R.1)** (R1)

Dimana β_0 adalah intersep atau nilai $E(Y_j)$ ketika X= 0, dan β_1 adalah slope atau tingkat perubahan $E(Y_j)$ menurut perubahan satu unit X

Pengamatan ke-j dari variabel respon Y, ditulis Y_j, diasumsikan sebagai pengamatan acak dari populasi variabel acak, dengan nilai rata-rata setiap populasi adalah $E(Y_j)$. Deviasi setiap pengamatan Y_j dari rataan populasi, yaitu $E(Y_j)$ dinyatakan sebagai error acak ε_j. Selanjutnya model regresi linier yang digunakan dimodelkan sebagai berikut.

$Y_j = \beta_0 + \beta_1 X_j + \varepsilon_j$ **(R.2)** ( )

Dimana :

J = pengamatan ke 1, 2, . . . n
Yj = nilai pengamatan ke-j bagi variabel respon
Xj = nilai pengamatan ke-j bagi variabel penjelas.
β0 = parameter koefisien konstanta regresi
β1 = parameter koefisien variabel penjelas
εj = error ke- j dari model

Nilai rata-rata dari error acak ε_j adalah 0, dan diasumsikan mempunyai ragam bersama σ^2, serta setiap error ke j saling independen. Oleh karena elemen acak pada model $Y_j = \beta_0 + \beta_1 X_j + \varepsilon_j$ hanya ε_j, maka asumsi ini mengakibatkan Y_j juga mempunyai ragam bersama σ^2, dan setiap pengamatan Y_j saling independen. Untuk keperluan pendugaan yang signifikan, diasumsikan ε_j menyebar normal, yang menyebabkan Y_j juga menyebar normal. Keseluruhan asumsi ini dinyatakan secara ringkas sebagai berikut.

$\varepsilon_j \sim NID\ (0, \sigma^2)$, dan mengakibatkan $Y_j \sim NID\ (\beta_0 + \beta_1 X_j, \sigma^2)$

Pendugaan terhadap koefisien regresi menggunakan penduga jumlah kuadrat terkecil, dan untuk keperluan pengujian yang signifikan atau pendugaan interval, diasumsikan bahwa ε_j menyebar normal independen dan identik dengan nilai tengah 0 dan ragam bersama σ^2, atau $\varepsilon_j \sim NID\ (0, \sigma^2)$. Prosedur pendugaan dengan metode jumlah kuadrat terkecil adalah upaya memperkecil (terkecil) jumlah kuadrat dari deviasi antara Y_j dengan nilai dugaan $\hat{Y}_j$, yaitu jumlah kuadrat bagi $e_j = Y_j - \hat{Y}_j$ minimal. Deviasi ini disebut sisaan, dan akan diperoleh setelah memperoleh koefisien dugaan $\hat{\beta}_0$ dan $\hat{\beta}_1$.

$$JKS = \sum (Y_j - \hat{Y}_j)^2$$

$$JKS = \sum (Y_j - (\hat{\beta}_0 + \hat{\beta}_1 X_j))^2$$

Dengan menggunakan kalkulus, derivatif JKS terhadap masing-masing $\hat{\beta}_0$ dan $\hat{\beta}_1$ disamakan dengan 0. Selanjutnya diperoleh sistem persamaan yang disebut persamaan normal, yaitu

$$(n)\ \hat{\beta}_0 + (\sum X_j)\ \hat{\beta}_1 = \sum Y_j \quad \textbf{(R.3)}$$

$$(\sum X_j)\ \hat{\beta}_0 + (\sum X_j^2)\ \hat{\beta}_1 = \sum X_j Y_j$$

Solusi dari sistem persamaan akan memperoleh nilai dugaan terhadap masing-masing β_0 dan β_1, yaitu

$$\hat{\beta}_1 = \sum (X_j - \bar{\mathbf{X}})(Y_j - \bar{\mathbf{Y}}) / \sum (X_j - \bar{\mathbf{X}})^2 = [\sum X_i Y_i - (\sum X_i)(\sum Y_i) / n\ ] / [\sum X_i^2 - (\sum X_i)^2 / n\ ] \quad \textbf{(R.4)}$$

$$\hat{\beta}_0 = \bar{\mathbf{Y}} - \hat{\beta}_1 \bar{\mathbf{X}}$$

Sebagai hasil dari pendugaan kedua koefisien parameter, diperoleh persamaan regresi dugaan, yaitu

$$\hat{Y}_j = \hat{\beta}_0 + \hat{\beta}_1 X_j \quad \textbf{(R.5)}$$

5.2.2 Analisis Ragam Dari Regresi Linier

Analisis ragam pada regresi linier menyajiakan peninjauan partisi masing-masing suku jumlah kuadrat pada persamaan yang menyatakan : *deviasi nilai dugaan terhadap pengamatan sebenarnya*, yaitu $e_j = Y_j - \hat{Y}_j$ atau dituliskan sebagai $Y_j = \hat{Y}_j + e_j$.

$$\sum Y_j^2 = \sum(\hat{Y}_j + e_j)^2$$

$$= \sum \hat{Y}_j^2 + \sum e_j^2 \qquad \text{“(karena perkalian } \sum \hat{Y}_j e_j = 0 \text{)”}$$

$$= \sum \hat{Y}_j^2 + \sum (Y_j - \hat{Y}_j)^2$$

JK(Total)Tkk JK(Model) + JK(Sisa) **(R.6)**

Jumlah kuadrat total dipartisi atas jumlah kuadrat yang terjelaskan (JK(Model)) dan jumlah kuadrat tak terjelaskan (JK(Sisa)). Jumlah kuadrat pada kedua ruas partisi ini adalah jumlah kuadrat yang belum terkoreksi. Partisi ini selanjutnya dapat dibuat menjadi partisi jumlah kuadrat yang terkoreksi. Koreksi dilakukan pada kedua ruas persamaan oleh faktor koreksi $n\bar{\mathbf{Y}}^2$.

$$\sum Y_j^2 - n\bar{\mathbf{Y}}^2 = (\sum \hat{Y}_j^2 - n\bar{\mathbf{Y}}^2) + \sum (Y_j - \hat{Y}_j)^2$$

JK(Total) = JK(Reg) + JK(Sisa) **(R.7)**

Jika digunakan nilai dugaan $\hat{\beta}_1$, jumlah kuadrat regresi pada partisi ini dapat dituliskan sebagai berikut.

$$\text{JK(Total)} = \text{JK(Reg)} + \text{JK(Sisa)}$$

$$\sum Y_j^2 - n\bar{\mathbf{Y}}^2 = (\sum \hat{Y}_j^2 - n\bar{\mathbf{Y}}^2) + \sum (Y_j - \hat{Y}_j)^2$$

$$\sum Y_j^2 - n\bar{\mathbf{Y}}^2 = (\hat{\beta}_1^2 \sum (X_j - \bar{\mathbf{X}})^2) + \sum (Y_j - \hat{Y}_j)^2 \quad \textbf{(R.8)}$$

Derajat bebas masing-masing jumlah kuadrat ditentukan oleh ukuran sampel dan banyaknya parameter di dalam model (*p*). Setiap derajat bebas dari jumlah kuadrat terkoreksi mengalami pengurangan dengan satu sebagai akibat dilakukan koreksi faktor. Derajat bebas JK(total) adalah *n* – 1. Derajat bebas JK(Reg) adalah derajat bebas dari JK(model) dikurang satu. Derajat bebas JK(Model) sama dengan banyaknya parameter dalam model regresi, yaitu terdapat p = 2 parameter koefisien. Dengan demikian derajat bebas JK(Reg) adalah p-1 = 2-1. Derajat bebas JK(Sisa) adalah n-p, yaitu n-2.

Rata-rata dari jumlah kuadrat (RJK) adalah jumlah kuadrat dibagi masing-masing derajat bebasnya. Perhitungan masing-masing jumlah kuadrat (JK) dengan rata-ratanya (RJK) dirumuskan sebagai berikut

$$\begin{aligned}
\text{JK(Total)} &= \sum Y_j^2 - n\overline{\mathbf{Y}}^2 \\
&= \sum Y_j^2 - \left(\sum Y_j\right)^2 / n \\
\text{RJK(Total)} &= \text{JK(Total)}/\,n\text{-}1 \\
\text{JK(Reg)} &= \hat{\beta}_1{}^2 \sum (X_j - \overline{\mathbf{X}})^2 \\
&= \hat{\beta}_1{}^2 \left[\sum X_i^2 - \left(\sum X_i\right)^2 / n\right] \quad \textbf{(R.9)} \\
\text{RJK(Reg)} &= \text{JK(Reg)}/\,p\text{-}1 \\
\text{JK(Sisa)} &= \text{JK(Total)} - \text{JK(Reg)} \\
\text{RJK(Sisa)} &= \text{JK(Sisa)}/n\text{-}p
\end{aligned}$$

Hasil analisis ragam ini disajikan dalam tabel analisis ragam. Pada tabel analisis ragam ini dicantumkan nilai F-hitung untuk keperluan pengujian menggunakan pendekatan sebaran F.

Tabel 5.1 Tabel Analisi Ragam

Sumber Variasi	**Jumlah Kuadrat (JK)**	**Derajat Bebas (Db)**	**Rata-rata Jum. Kuad. (RJK)**	**F-hitung**
Regresi	JK(reg)	p-1	RJK(Reg)	$F = \dfrac{\text{RJK(Reg)}}{\text{RJK(Sisa)}}$
Sisaan	JK(Sisa)	n-p	RJK(Sisa)	
Total	JK(Total)	n-1		

Pada partisi jumlah kuadrat total ini, JK(Reg) mewakili jumlah kuadrat yang terjelaskan atau dapat dikendalikan, dan JK(Sisa) mewakili jumlah kuadrat yang tak terjelaskan. RJK(Reg) menduga $\sigma^2 + \beta_1^2 \sum (X_j - \overline{\mathbf{X}})^2$, dan RJK(Sisa) menduga σ^2. Pada kejadian hipotesis $\beta_1 = 0$ benar, kedua rata-rata jumlah kuadrat, yaitu RJK(Reg) dan RJK(Sisa) menduga σ^2. Jika β_1 menjauh dari 0 maka RJK(Reg) meningkat lebih besar dari RJK(Sisa). Suatu hasil analisis ragam yang baik apabila jumlah kuadrat yang terjelaskan jauh lebih besar dari pada jumlah kuadrat tak terjelaskan. Rasio antara RJK(Reg) dengan RJK(Sisa) yang besar mengindikasikan bahwa β_1 tidak sama dengan nol. Jika asumsi bahwa sisaan menyebar normal adalah sahih, dan hipotesis $\beta_1 = 0$ benar, maka ratio antara RJK(Reg) dengan RJK(Sisa) mengikuti sebaran F.

Dengan menggunakan pendekatan sebaran F dalam menguji keberartian suatu persamaan regresi dugaan $\hat{Y}_j = \hat{\beta}_0 + \hat{\beta}_1 X_j$, maka RJK(Reg) menyatakan porsi komponen terjelaskan yang telah terkoreksi, dan JK(Sisa) menyatakan porsi komponen tak terjelaskan. Metode ini dapat pula digunakan untuk menguji hipotesis,

H_0 : $\beta_1 = 0$. **(R.10)**

Selanjutnya rumus untuk menghitung statistik F-hitung adalah

F_{hit} = RJK(Reg) / RJK(Sisa) **(R.11)**

Keterangan :

RJK(Reg) = Rata-rata jumlah kuadrat regresi

RJK(Sisa) = Rata-rata jumlah kuadrat sisaan

Derajat bebas pembilang dan penyebut statistik F-hitung berturut-turut adalah 1 dan *n*-2.

Kriteria pengujian:

Jika pada pengujian ini digunakan taraf signifikan α, maka Terima H_0 jika F-hitung $\leq F_{(\alpha;1;n-2)}$-tabel atau $Pr = P(F > F_{-hitung}) \geq \alpha$

Tolak H_0 jika F-hitung $> F_{(\alpha;1;n-2)}$-tabel atau Pr = P (F > $F_{-hitung}$) $< \alpha$. **(R.12)**

( )

5.2.3 Koefisien Determinasi, R^2

spasi

Suatu ukuran seberapa tepat model regresi linier dugaan dapat menjelaskan hubungan antara variabel bebas dengan variabel terikat dapat dilihat dari besarnya nilai kofisien determinasi (R^2). Pada partisi jumlah kuadrat total, indikasi model semakin tepat apabila jumlah kuadrat terjelaskan semakin besar mendekati jumlah kuadrat total. Koefisien determinasi R^2

dirumuskan sebagai proporsi JK(Reg) terhadap JK(Total). Dengan demikian koefisien determinasi memiliki rentangan nilai antara 0 sampai 1. Nilai R^2 yang mendekati 1 berarti variasi variabel terikat semakin terjelaskan oleh hubungan liniernya dengan variabel bebas.

R^2 = JKReg / JKTotal

$$= \hat{\beta}_1^{\,2}\ [\sum X_i^2 - (\sum X_i)^2 / n]\ / \sum Y_j^2 - n\bar{\mathbf{Y}}^2 \quad \textbf{(R.13)}$$

5.2.4 Pengujian Koefisien Variabel Penjelas, β_1

Penyelidikan secara langsung apakah ada perubahan yang berarti pada variabel Y oleh perubahan variabel X, dilakukan melalui pengujian koefisien β_1. Besaran koefisien β_1 diartikan sebagai tingkat perubahan pada nilai rata-rata Y oleh perubahan satu unit X. Untuk menguji apakah β_1 sama dengan nol atau tidak, dirumuskan hipotesis uji

H0 : $\beta 1 = 0$

H1 : $\beta_1 \neq 0$ **(R.14)**

Pengujian hipotesis ini menggunakan pendekatan sebaran t-student. Rumus t-hitung yang digunakan adalah

$$\text{t-hitung} = (\hat{\beta}_1 - 0) / S_{\hat{\beta}1} \qquad \textbf{(R.15)}$$

dengan

$$S_{\hat{\beta}1} = \sqrt{\text{RJKS} / \sum (Xj - \bar{X})^2}$$

$$= \sqrt{\text{RJKS} / (\sum X_j^2 - (\sum X_j)^2 / n)}$$

Keterangan :

t–hitung = nilai statistik t

$\hat{\beta}_1$ = nilai dugaan bagi koefisien variabel penjelas

$S_{\hat{\beta}_1}$ = simpangan baku bagi dugaan koefisien variabel penjelas

Statistik t-hitung di atas menyebar t-student dengan derajat bebas n-2.

Kriteria pengujian :

Jika pada pengujian ini digunakan taraf signifikan sebesar α , maka

H_0 diterima jika $|$t-hitung$| \leq t_{\alpha/2;Db}$-tabel , atau

$Pr = [\ P\ (\ t \leq -|t_{hitung}|\ ) + P\ (\ t \geq |t_{hitung}|\ )\ ] \geq \alpha$, sebaliknya

R.16) (

H_0 ditolak jika $|$t-hitung$| > t_{\alpha/2;Db}$-tabel , atau

$Pr = [P\ (\ t \leq -|t_{hitung}|\ ) + P\ (\ t \geq |t_{hitung}|\ ) < \alpha$.

Contoh Pengerjaan 5.1:

Tabel 5.2 Konsumsi Bensin dan Penjualan Mobil Selama Setahun

Bln Ke	Bensin Y	Mobil X	Y^2	X^2	X Y
1	5.33	28.3	28.4089	800.89	150.839
2	6.00	29.0	36.0000	841.00	174.000
3	5.72	28.7	32.7184	823.69	164.164
4	5.55	28.2	30.8025	795.24	156.510
5	5.51	28.0	30.3601	784.00	154.280
6	5.12	27.8	26.2144	772.84	142.336
7	5.33	28.5	28.4089	812.25	151.905
8	5.80	28.9	33.6400	835.21	167.620
9	5.75	28.9	33.0625	835.21	166.175
10	5.91	29.2	34.9281	852.64	172.572
11	6.21	32.0	38.5641	1024.00	198.720
12	6.14	31.3	37.6996	979.69	192.182
$\Sigma Y = 68.37$		$\Sigma X = 348.8$	$\Sigma Y^2 = 390.8075$	$\Sigma X^2 = 10156.66$	$\Sigma YX = 1991.303$

Suatu penelitian dilakukan untuk mengetahui hubungan antara banyaknya pemakaian bensin (Y) dengan tingkat penjualan mobil (X) di Kota Samarinda. Data dicatat setiap bulan selama satu tahun.

Pemakaian bensin (satuan 1000 ltr), dan tingkat penjualan mobil berturut-turut ditunjukkan pada kolom dua dan tiga dalam Tabel 5.2.
Jika hubungan kedua variabel di atas ingin diselidiki menurut model regresi linier,

a. Dugalah persamaan regresi liniernya
b. Lakukan pengujian terhadap keberartian model dugaan melalui uji F.
c. Hitung koefisien determinasi R^2
d. Lakukan pengujian terhadap koefisien $\beta_1 = 0$

<u>Pemecahan pengerjaan 5.1 (Hitung Manual):</u>

Rata-rata : $\overline{X} = \frac{\sum_{j=1}^{12} Xj}{12} = \frac{348.8}{12} = 29.07$

$$\overline{Y} = \frac{\sum_{j=1}^{12} Y_j}{12} = \frac{68.37}{12} = 5.6975$$

a. Perhitungan koefisien dugaan $\hat{\beta}_1$ dan $\hat{\beta}_0$:

$\hat{\beta}_1 = \sum(X_j - \overline{\mathbf{X}})(Y_j - \overline{\mathbf{Y}}) / \sum (X_j - \overline{\mathbf{X}})^2$

$\hat{\beta}_1 = [\sum X_j Y_j - (\sum X_j)(\sum Y_j) / n] / [\sum X_j^2 - (\sum X_j)^2 / n]$

$= [1991.303 - (348.8)(68.37) / 12] / [10156.66 - (348.8)^2 / 12]$

$= 0.2205$

$\hat{\beta}_0 = \overline{\mathbf{Y}} - \hat{\beta}_1 \overline{\mathbf{X}}$

$= 5.6975 - 0.2205 (29.07)$

$= -0.7124$

Persamaan regresi dugaan adalah $\hat{Y}_j = -0.7128 + 0.2205 X_j$

b. Hipotesis :

H_0 : $\beta_1 = 0$.

H_1 : $\beta_1 \neq 0$

Perhitungan jumlah kuadrat :

JK(Total) $= \sum Y_j^2 - (\sum Y_j)^2 / n$

$= (390.8075)^2 - (68.37)^2 / 12$

$= 1.2694$

JK(Reg) $= \hat{\beta}_1^2 \; [\sum X_i^2 - (\sum X_i)^2 / n]$

$= (0.2205)^2 (10156.66 - (348.8)^2 / 12)$

$= 0.8854$

RJK(Reg) $= 0.8854 / 1$

$= 0.8854$

JK(Sisa) = JK(Total) – JK(Reg)

$= 1.2694 - 0.8854$

$= 0.3840$

RJK(Sisa) $= 0.3840 / 12 - 2 = 0.0384$

F_{hit} = RJK(Reg) / RJK(Sisa)

$= 0.8854 / 0.0384 = 23.0559$

Pada tabel sebaran F diperoleh : $F_{(0.05;1;10)}$-tabel = 4.96

Tabel 5.3 Tabel Analisi Ragam

Sumber Variasi	**Jumlah Kuadrat (JK)**	**Derajat Bebas (Db)**	**Rata-rata Jum. Kuad. (RJK)**	**F-hitung**
Regresi	0.8854	1	0.8854	F= 23.0559
Sisaan	0.3840	10	0.0384	
Total	1.2694	11		

Hasil pengujian : F-hitung > $F_{(\alpha;1;n-2)}$-tabel, yaitu 23.056 > 4.96 , selanjutnya H_0 ditolak pada taraf signifikan 5 persen. Disimpulkan bahwa koefisien parameter tidak semua nol. Ini berarti persamaan regresi dugaan $\hat{Y}_j = -0.7128 + 0.2205 X_j$ merupakan model hubungan antara rata-rata pemakaian bensin dengan tingkat penjualan mobil yang signifikan.

c. Koefisien determinasi :

R^2 = JKReg / JKTotal

= 0.8854 / 1.2694

= 0.6975

Kesimpulan : 69.75 % variasi dari pemakaian bensin dapat dijelaskan oleh hubungan liniernya terhadap tingkat penjualan mobil di Kalimantan timur.

d. Hipotesis :

H_0 : $\beta_1 = 0$

H_1 : $\beta_1 \neq 0$

Ragam : $\hat{S}_{\beta 1} = \sqrt{RJKS / (\sum X_j^2 - (\sum X_j)^2 / n)}$

$= \sqrt{0.0384 / [10156.66 - (348.8)2 \; / 12]}$ = 0.0459

t-hitung = $(\hat{\beta}_1 - 0) / \hat{S}_{\beta 1}$

= 0.2250 / 0.0459 = 4.80

Pada tabel t diperoleh : $t_{0.05/2;10}$ = 2.228

Hasil pengujian : $|$ t-hitung $|$ > $t_{\alpha/2;Db}$-tabel, yaitu 4.80 > 2.228, selanjutnya H_0 ditolak pada taraf signifikan 5 persen. Ini berarti β_1 signifikan tidak sama dengan nol. Disimpulkan bahwa tingkat penjualan mobil signifikan berpengaruh terhadap rata-rata pemakaian bensin.

<u>Pemecahan pengerjaan 5.1 (Komputerisasi):</u>

Hipotesis yang diuji:

H_0 : $\beta_1 = 0$.

H_1 : $\beta_1 \neq 0$

Taraf signifikan pengujian yang digunakan α = 0,05. Statistik yang digunakan untuk menguji keberartian model adalah statistik F, sedangkan untuk menguji keberartian koefisien variabel bebas digunakan statistik T.

Tabel 5.4 *Output* Hasil Analisis Regresi Linier Sederhana Menggunakan SPSS

ANOVA[a]

	Model	Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	.885	1	.885	23.056	.001[b]
	Residual	.384	10	.038		
	Total	1.269	11			

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients		
	Model	B	Std. Error	Beta	t	Sig.
1	(Constant)	-.712	1.336		-.533	.606
	X	.221	.046	.835	4.802	.001

a. Dependent Variable: Y

b. Predictors: (Constant), X

Hasil analisis pada Tabel 5.4, menunjukkan bahwa nilai statistik F = 23,056 dengan taraf signifikan 0,001. Oleh karena taraf signifikan statististik F kurang dari taraf signifikan pengujian, yaitu 0,001 < 0,05 maka dinyatakan H_0 ditolak pada taraf signifikan 5 persen.

Demikian pula pada Tabel 5.4, menunjukkan bahwa nilai statistik T = 4,802 dengan taraf signifikan 0,001. Oleh karena taraf signifikan statististik T kurang dari taraf signifikan pengujian, yaitu 0,001 < 0,05 maka dinyatakan H_0 ditolak pada taraf signifikan 5 persen. Ini berarti koefisien variabel penjualan mobil β_1 signifikan tidak sama dengan nol. Disimpulkan bahwa tingkat penjualan mobil signifikan berpengaruh terhadap rata-rata pemakaian bensin.

Sebagaimana diutarakan sebelumnya bahwa statistik yang digunakan untuk menguji keberartian model adalah statistik F, sedangkan untuk menguji keberartian koefisien variabel bebas digunakan statistik T. Keberartian suatu model regresi lenier dinyatakan oleh koefisien variabel bebas yang signifikan berbeda dengan nol. Analisis regresi linier sederhana hanya menggunakan satu variabel bebas, dan tentunya hanya satu koefisien variabel yang diuji. Oleh karena itu, pengujian keberartin model dengan uji F dan pengujian koefisien dengan uji T akan memberikan hasil yang sama. Pada analisis regresi linier sederhana, kita dapat memilih hanya salah satu dari dua pengujian tersebut untuk mendapatkan kesimpulan apakah variabel bebas berpengaruh terhadap variabel terikat.

5.3 Regresi Linier Ganda

spasi

5.3.1Model dan Pendugaan Regresi Linier Ganda

Regresi ganda memodelkan hubungan antara suatu variabel terikat (Y) dengan beberapa variabel bebas (Xi). Model aditip linier bagi regresi ganda adalah:

$$Y_j = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \ldots + \beta_p X_{pj} + \varepsilon_j \quad \text{(G.1)}$$ ( )

Disingkat,

spasi $$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \text{(G.2)}$$

Atau

$$\underset{(n \times 1)}{\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \cdot \\ \cdot \\ \cdot \\ Y_n \end{pmatrix}} = \underset{(n \times (p+1))}{\begin{pmatrix} 1 & X_{11} & X_{12} & X_{13} & \cdots & X_{1p} \\ 1 & X_{21} & X_{22} & X_{23} & \cdots & X_{2p} \\ 1 & X_{31} & X_{32} & X_{33} & \cdots & X_{3p} \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & \cdot \\ 1 & X_{n1} & X_{n2} & X_{n3} & \cdots & X_{np} \end{pmatrix}} \underset{(n \times 1)}{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_p \end{pmatrix}} + \underset{(n \times 1)}{\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{pmatrix}}$$

Pada pemodelan hubungan dengan regresi, variabel terikat $\boldsymbol{Y}$ sering disebut variabel respon, dan variabel bebas $\mathbf{X}$ disebut variabel penjelas atau regressor. Parameter $\boldsymbol{\beta}$ disebut koefisien regresi, sedangkan beda antara nilai harapan $\boldsymbol{Y}$ pada model ($E(\boldsymbol{Y}) = \mathbf{X}\boldsymbol{\beta}$) dengan nilai $\boldsymbol{Y}$ sesungguhnya , yaitu $\boldsymbol{\varepsilon}$ disebut error model.

1. Asumsi Pada Regresi Ganda

Untuk keperluan pendugaan, pada model di atas diasumsikan bahwa vektor acak $\boldsymbol{\varepsilon}$ menyebar normal multivariat, dengan vektor rata-rata $\mathbf{0}$ dan matriks ragam-peragam $\mathbf{I}\sigma^2$. ditulis dengan singkat $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I}\sigma^2)$.

Vektor rata-rata $\mathbf{0}$ adalah suatu vektor berukuran (n x 1), dengan semua elemennya 0. Matrik ragam-peragam $\mathbf{I}\sigma^2$ adalah matrik berukuran (n x n), dengan elemen diagonal σ^2_{jj} adalah ragam (variance) masing-masing variabel acak ε_j. Sedangkan elemen nondiagonal (k,l) adalah koragam (covariance) antara ε_k dan ε_l.

Tinjau model $\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, olehkarena $\mathbf{X}$ dan $\boldsymbol{\beta}$ konstan, maka suku $\mathbf{X}\boldsymbol{\beta}$ pada model adalah konstan. Oleh penambahan vektor dari error acak $\boldsymbol{\varepsilon}$, menyebabkan $\boldsymbol{Y}$ merupakan suatu vektor acak, dengan vektor rata-rata $\mathbf{X}\boldsymbol{\beta}$, dan matriks ragam-peragam $\mathbf{I}\sigma^2$, ditulis " $\boldsymbol{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{I}\sigma^2)$ ".

2. Pendugaan terhadap $\boldsymbol{\beta}$ atau Model

Pendugaan bagi model $\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta}$, dilakukan melalui pendugaan parameter $\boldsymbol{\beta}$. Penduga bagi $\boldsymbol{\beta}$ adalah $\hat{\boldsymbol{\beta}}$. Dengan menggunakan metode jumlah kuadrat terkecil, yaitu minimum $\sum(Y - \mathbf{X}\hat{\beta})^2$, diperoleh persamaan normal berikut

spasi

$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\boldsymbol{Y}$ (G.3) Dengan, ( )

$$
\mathbf{X'X} = \begin{pmatrix}
n & \sum X_{i1} & \sum X_{i2} & \cdots & \sum X_{ip} \\
\sum X_{i1} & \sum X_{i1}^2 & \sum X_{i1}X_{i2} & \cdots & \sum X_{i1}X_{ip} \\
\sum X_{i2} & \sum X_{i1}X_{i2} & \sum X_{i2}^2 & \cdots & \sum X_{i2}X_{ip} \\
\vdots & \vdots & \vdots & & \vdots \\
\sum X_{ip} & \sum X_{i1}X_{ip} & \sum X_{i2}X_{ip} & \cdots & \sum X_{ip}^2
\end{pmatrix}
$$

$$
\mathbf{X'}\boldsymbol{Y} = \begin{pmatrix}
\sum Y_i \\
\sum X_{i1}Y_i \\
\sum X_{i2}Y_i \\
\vdots \\
\sum X_{ip}Y_i
\end{pmatrix}
$$

Solusi unik bagi persamaan normal (jika ada) adalah

$\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}(\mathbf{X'}\boldsymbol{Y})$ (G.4)

Solusi unik dari persamaan normal ada, jika invers dari **X'X** ada (*nonsingular*). *Nonsingular* tercapai apabila matrik **X** berpangkat penuh, yaitu tidak terdapat ketergantungan linier diantara variabel bebas (antara vektor kolom bebas linier).

Setelah diperoleh dugaan terhadap koefisien regresi atau dugaan bagi vektor $\boldsymbol{\beta}$, yaitu $\hat{\boldsymbol{\beta}}$ maka persamaan regresi dugaan diperoleh, yaitu $\hat{\boldsymbol{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$.

Contoh pengerjaan 5.2:

Berikut pada Tabel 5.5 adalah data tentang hasil studi lingkungan di New York. Studi ini mengukur 4 variabel, yaitu konsentrasi ozone (Y), radiasi matahari (X_1), temperatur (X_2), dan kecepatan angin berhembus (X_3).

Tabel 5.5 Data Hasil Studi Lingkungan

NO	Ozone	Radiation	Temperature	Wind
1	3.45	190	67	7.4
2	3.30	118	72	8.0
3	2.29	149	74	12.6
4	2.62	313	62	11.5
5	2.84	299	65	8.6
6	2.67	99	59	13.8
7	2.00	19	61	20.1
8	2.52	256	69	9.7
9	2.22	290	66	9.2
10	2.41	274	68	10.9
11	2.62	65	58	13.2
12	2.41	334	64	11.5
13	3.24	307	66	12.0
14	1.82	78	57	18.4
15	3.11	322	68	11.5
16	2.22	44	62	9.7
17	1.00	8	59	9.7
18	2.22	320	73	16.6
19	1.59	25	61	9.7
20	3.17	92	61	12.0
21	2.84	13	67	12.0
22	3.56	252	81	14.9
23	4.86	223	79	5.7
24	3.33	279	76	7.4
25	3.07	127	82	9.7
26	4.14	291	90	13.8
27	3.39	323	87	11.5
28	2.84	148	82	8.0
29	2.76	191	77	14.9
30	3.33	284	72	20.7

luruskan tanda titik kebawah

contoh Lajur 2

Sumber : Contoh data pada Splus-2000.

Data hasil studi lingkungan dianalisis menggunakan model (G.1), dengan variabel respon adalah Ozone. Pendugaan koefisien regresi menggunakan persamaan (G.4), dengan rincian perhitungan menggunakan hasil komputerisasi sebagai berikut.

Matriks **X'X** adalah

disusun rapi

	[,1]	[,2]	[,3]	[,4]
[1,]	30.0	5733.0	2085	354.70
[2,]	5733.0	1463179.0	411076	67085.70
[3,]	2085.0	411076.0	147243	24523.00
[4,]	354.7	67085.7	24523	4583.39

Matriks **X'*Y*** adalah

	[,1] → [ ]
[1,]	83.840
[2,]	17097.380
[3,]	5953.470
[4,]	973.641

luruskan titiknya

Invers dari matriks **X'X**, yaitu matriks $(\mathbf{X'X})^{-1}$ adalah

	[,1]	[,2]	[,3]	[,4]
[1,]	2.8559225951	6.173356e-004	-0.03535068677	-4.090973e-002
[2,]	0.0006173356	3.341412e-006	-0.00001807261	1.409908e-008
[3,]	-0.0353506868	-1.807261e-005	0.00053385411	1.439104e-004
[4,]	-0.0409097295	1.409908e-008	0.00014391044	2.613921e-003

Pendugaan terhadap koefisien regresi menggunakan (G.4), yaitu

$\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}(\mathbf{X'}\boldsymbol{Y})$

Sehingga vektor koefisien regresi dugaan $\hat{\boldsymbol{\beta}}$ adalah

```
         [,1]
[1,] -0.295271027
[2,]  0.001305816
[3,]  0.045605744
[4,] -0.027843496
```

Selanjutnya diperoleh persamaan regresi dugaan berikut

$$\hat{Y} = -0{,}2953 + 0{,}0013\,\boldsymbol{X_1} + 0.0456\,\boldsymbol{X_2} - 0{,}0278\,\boldsymbol{X_3}$$

Hasil komputerisasi, menggunakan Splus 2000, menampilkan *output* seperti pada Tabel 5.6. Kolom ke dua dari table menyajikan hasil pendugaan koefisien dari masing-masing variabel penjelas (variabel bebas).

→ copy paste aslinya

Tabel 5.6 Tampilan *Output* Komputerisasi dari Hasil Analisis Regresi Ganda Menggunakan Perangkat Lunak Splus 2000

*** Linear Model ***

lihat tabel asal

Coefficients:

	Value	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.2949	0.9986	-0.2953	0.7701
radiation	0.0013	0.0011	1.2080	0.2379
temperature	0.0456	0.0137	3.3429	0.0025
wind	-0.0280	0.0302	-0.9268	0.3626

? ?

Residual standard error: 0.5909 on 26 degrees of freedom

Multiple R-Squared: 0.4583

F-statistic: 7.332 on 3 and 26 degrees of freedom, the p-value is 0.00102

3. Bentuk Fungsi Linier dari Y

Berikut dapat ditunjukkan bahwa statistik (penduga) dalam regresi ganda ini adalah fungsi linier dari **Y**. Statistik tersebut meliputi $\hat{\beta}$ (penduga koefisien), $\hat{Y}$ (prediksi), dan **e** (sisaan).

$\hat{\beta} = (\mathbf{X'X})^{-1}(\mathbf{X'}\boldsymbol{Y}) = [(\mathbf{X'X})^{-1}\mathbf{X'}]\,\boldsymbol{Y}$ (G.5)

Vektor $\hat{\beta}$ adalah fungsi linier dari ***Y***, dengan koefisien $[(\mathbf{X'X})^{-1}\mathbf{X'}]$.

$$\begin{aligned}\hat{\boldsymbol{Y}} &= \mathbf{X}\hat{\beta}\\ &= \mathbf{X}(\mathbf{X'X})^{-1}(\mathbf{X'}\boldsymbol{Y})\\ &= [\mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'}]\,\boldsymbol{Y}\\ &= \mathbf{H}\,\boldsymbol{Y} \quad \text{(G.6)}\end{aligned}$$

( )

Vektor $\hat{\boldsymbol{Y}}$ adalah fungsi linier dari ***Y***, dengan koefisien $\mathbf{H} = [\mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'}]$. Matriks **H** dibaca " matriks hat " adalah matriks yang ditentukan oleh matriks **X**. Matriks ini sangat penting peranannya dalam analisis regresi. Matriks **H** adalah matriks simetri dan idempoten, yaitu $\mathbf{H'} = \mathbf{H}$ dan $\mathbf{H\,H} = \mathbf{H}$.

$$\begin{aligned}\mathbf{e} &= \boldsymbol{Y} - \hat{\boldsymbol{Y}}\,.\\ &= \boldsymbol{Y} - [\mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'}]\,\boldsymbol{Y}\\ &= \boldsymbol{Y} - \mathbf{H}\,\boldsymbol{Y}\\ &= [\mathbf{I} - \mathbf{H}]\,\boldsymbol{Y} \quad \text{(G.7)}\end{aligned}$$

( )

Vektor **e** adalah fungsi linier dari ***Y***, dengan koefisien **[I - H]** . Matriks **[I - H]** adalah juga matriks simetri dan idempoten.

4. Sebaran dari Statistik $\hat{\beta}$, $\hat{Y}$, dan e .

Jika model benar, maka nilai harapan Y adalah $\mathbf{X}\boldsymbol{\beta}$. Sejak statistik $\hat{\beta}$, $\hat{Y}$, dan **e** adalah fungsi linier dari ***Y***, dan diketahui ***Y*** adalah vektor acak, maka $\hat{\beta}$, $\hat{Y}$, dan **e** adalah juga vektor-vektor acak. Sifat-sifat masing-masing statistik sebagai fungsi linier dari ***Y*** dapat dikemukakan sebagai berikut.

a. Nilai Harapan

$$E(\hat{\boldsymbol{\beta}}) = E((\mathbf{X'X})^{-1}\mathbf{X'Y})$$
$$= [(\mathbf{X'X})^{-1}\ \mathbf{X'}]\ E(\mathbf{Y})$$
$$= [(\mathbf{X'X})^{-1}\ \mathbf{X'}]\ (\mathbf{X}\boldsymbol{\beta})$$
$$= [(\mathbf{X'X})^{-1}(\mathbf{X'X})]\boldsymbol{\beta}$$
$$= \boldsymbol{\beta}$$

Dengan demikian $\hat{\boldsymbol{\beta}}$ adalah penduga tak bias bagi $\boldsymbol{\beta}$, jika model adalah benar. Jika model tak benar, maka $(\mathbf{X'X})^{-1}(\mathbf{X'X}) \neq \mathbf{I}$.

$(\ )^{-1}(\ ) \neq I$ ← contoh

$$E(\hat{\boldsymbol{Y}}) = E(\mathbf{H}\,\boldsymbol{Y})$$
$$= \mathbf{H}\ E(\boldsymbol{Y})$$
$$= [\mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'}]\ \mathbf{X}\boldsymbol{\beta}$$
$$= \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'X}\,\boldsymbol{\beta}$$
$$= \mathbf{X}\,\boldsymbol{\beta}$$

Dengan demikian jika model adalah benar, $\hat{\boldsymbol{Y}}$ adalah penduga tak bias bagi rata-rata $\boldsymbol{Y}$,

$$E(\mathbf{e}) = E([\mathbf{I} - \mathbf{H}]\ \boldsymbol{Y})$$
$$= [\mathbf{I} - \mathbf{H}]\ E(\boldsymbol{Y})$$
$$= [\mathbf{I} - \mathbf{H}]\ \mathbf{X}\boldsymbol{\beta}$$
$$= [\mathbf{I}\,\mathbf{X} - \mathbf{H}\,\mathbf{X}]\ \boldsymbol{\beta}$$
$$= [\mathbf{X} - \mathbf{X}]\ \boldsymbol{\beta}$$
$$= \mathbf{0}$$

Sisaan pengamatan, **e** adalah vektor acak dengan nilai harapan nol.

b. Ragam (Variance)

Pembahasan tentang ragam bagi fungsi linier dari $\boldsymbol{Y}$, dilakukan dengan terlebih dahulu melakukan peninjauan terhadap gagasan notasi dan aljabar matriks. Untuk $\boldsymbol{Y}$ sebagai vektor acak, misalkan memiliki matriks ragam-peragam yang ditulis **Var (Y)**, dan misalkan suatu fungsi linier **u** , ditulis $\mathbf{u} = \mathbf{a'}\,\boldsymbol{Y}$.

Var (u) = **a'** [**Var (*Y*)**] **a**

Spasi

Sebagaimana asumsi pada pendugaan dengan metode kuadrat terkecil biasa, **Var (Y)** = **I**σ^2, sehingga

Var (u) = a' (Iσ^2 **) a = a'a** σ^2

Notasi **a'a** menyatakan jumlah kuadrat dari koefisien-koefisien fungsi linier, yaitu $\sum a_i^2$.

Bentuk fungsi linier **u** = **a'** ***Y*** , diperluas atas beberapa vektor koefisien **a** secara simultan, yaitu dengan notasi matriks **A** berukuran *k x n* , menjadi **U** = **A** ***Y***. Selanjutnya definisi matriks ragam-peragam bagi vektor acak ***Y*** adalah

Var (*Y*)= E([***Y*** - E(***Y***)][***Y*** - E(***Y***)]')

Matriks hasil perkalian [***Y*** - E(***Y***)][***Y*** - E(***Y***)]' berukuran n x n, dengan elemen diagonal utama (Y_i - E(Y_i))^2 dan elemen nondiagonal (Y_i - E(Y_i))(Y_j - E(Y_j))'. Nilai harapan bagi dua kelompok elemen berturut-turut merupakan ragam (*variance*) dan koragam (*covariance*). Jika definisi matriks ragam-peragam diterapkan bagi ~~U~~

U = **A** ***Y***, maka **Var (U)** = E([**U** - E(**U**)][**U** - E(**U**)]')

Spasi

= E([**A** ***Y*** - E(**A** ***Y***)][**A** ***Y*** - E(**A** ***Y***)]')

= E(**A** [***Y*** - E(***Y***)][***Y*** - E(***Y***)]' **A'**)

= **A** E([***Y*** - E(***Y***)][***Y*** - E(***Y***)]') **A'**

= **A** [**Var (*Y*)**] **A'**(G.8) Untuk

Lihat asalnya spy teratur

Var (*Y*) = **I**σ^2, enter

Var (U) = **A** [**I**σ^2]

A' = **A A'** σ^2 (G.9)

Keterangan : Elemen diagonal matriks **A A'** adalah jumlah kuadrat dari koefisien fungsi linier ke i, selanjutnya hasil perkaliannya dengan σ^2 merupakan ragam bagi fungsi linier ke i. Elemen nondiagonal ke (i,j) adalah perkalian silang antara koefisien fungsi linier ke i dengan ke j, selanjutnya hasil perkaliannya dengan σ^2

merupakan koragam antara kedua fungsi linier tersebut. Ragam masing-masing statistik $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{Y}}$, dan **e** adalah

$\hat{\boldsymbol{\beta}}$ = [(**X'X**)$^{-1}$(**X'**)] $\boldsymbol{Y}$, sehingga

Var($\hat{\boldsymbol{\beta}}$ **)** = [(**X'X**)$^{-1}$(**X'**)] [**Var (*Y*)**] [(**X'X**)$^{-1}$(**X'**)]'

= (**X'X**)$^{-1}$(**X' X**)(**X'X**)$^{-1}$ σ^2 =(**X'X**)$^{-1}\sigma^2$ (G.10)

??

$\hat{\boldsymbol{Y}}$ = **H** $\boldsymbol{Y}$, sehingga

Var($\hat{\boldsymbol{Y}}$ **)** = **H** [**Var (*Y*)**] **H'**

= **H I H'** σ^2

= **H H'** σ^2

=**H**σ^2 (G.11)

diedit

Lihat asalnya

Elemen diagonal merupakan ragam bagi nilai prediksi $\hat{\boldsymbol{Y}}_{\boldsymbol{i}}$, i = 1, 2, ...n. Nilai prediksi ini digunakan untuk menduga rata-rata **Y** atas berbagai kombinasi variabel bebas yang diberikan. Untuk nilai prediksi kedepan (nilai ramalan) atas berbagai kombinasi nilai variabel bebas yang diberikan, ditulis $\hat{\boldsymbol{Y}}_{\boldsymbol{i}\ pred}$, maka setiap ragam meningkat menurut σ^2. Matriks ragam- peragam bagi prediksi ini adalah

Var ($\hat{\boldsymbol{Y}}_{pred}$ **)** = (**I + H**) σ^2 (G.12) ( )

e = [**I - H**] $\boldsymbol{Y}$, sehingga

Var (e) = [**I - H**] [**Var (*Y*)**] [**I - H**]'

= [**I - H**] **I** [**I - H**]' σ^2

= [**I - H**] [**I - H**]' σ^2

= [**I - H**] σ^2 (G.13) ( )

Rangkuman dari sebaran masing-masing vektor acak dapat dikemukakan sebagai berikut :

$\boldsymbol{Y}$ ~ N(**X*β*, I**σ^2)

$\hat{\boldsymbol{\beta}}$ ~ N(***β***, (**X'X**)$^{-1}$ σ^2)

$\hat{\boldsymbol{Y}}$ ~ N(**X*β*, H**σ^2)

e ~ N(**0,** [**I - H**] σ^2)

spasi jelek

? $\hat{Y}_{pred} \sim N(\mathbf{X}\beta, [\mathbf{I} + \mathbf{H}]\, \sigma^2)$

5.3.2 Partisi Jumlah Kuadrat

Untuk menyesuaiakan dengan model aditif linier, vektor pengamatan bagi variabel terikat $\boldsymbol{Y}$, dipartisi menjadi vektor dugaan $\hat{\boldsymbol{Y}}$ ditambah vektor sisaan $\mathbf{e}$. Yaitu

$$\boldsymbol{Y} = \hat{\boldsymbol{Y}} + \mathbf{e}$$

Partisi yang serupa digunakan terhadap jumlah kuadrat bagi variabel terikat $\boldsymbol{Y}$.

$$\begin{aligned} \boldsymbol{Y}'\boldsymbol{Y} &= (\hat{\boldsymbol{Y}} + \mathbf{e})'(\hat{\boldsymbol{Y}} + \mathbf{e}) \\ &= \hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}} + \hat{\boldsymbol{Y}}'\mathbf{e} + \mathbf{e}'\hat{\boldsymbol{Y}} + \mathbf{e}'\mathbf{e} \end{aligned}$$

contoh

Subtitusi $\hat{\boldsymbol{Y}} = \mathbf{H}\boldsymbol{Y}$, dan $\mathbf{e} = [\mathbf{I} - \mathbf{H}]\boldsymbol{Y}$ menyebabkan

$Y'Y =$

$$\begin{aligned} \boldsymbol{Y}'\boldsymbol{Y} &= (\mathbf{H}\boldsymbol{Y})'(\mathbf{H}\boldsymbol{Y}) + (\mathbf{H}\boldsymbol{Y})'([\mathbf{I} - \mathbf{H}]\boldsymbol{Y}) + ([\mathbf{I} - \mathbf{H}]\boldsymbol{Y})'(\mathbf{H}\boldsymbol{Y}) \\ &\quad + ([\mathbf{I} - \mathbf{H}]\boldsymbol{Y})'([\mathbf{I} - \mathbf{H}]\boldsymbol{Y}) = \boldsymbol{Y}'\mathbf{H}'\mathbf{H}\boldsymbol{Y} + \boldsymbol{Y}'(\mathbf{H}'[\mathbf{I} - \mathbf{H}])\boldsymbol{Y} \\ &\quad + \boldsymbol{Y}'([\mathbf{I} - \mathbf{H}]'\mathbf{H})\boldsymbol{Y} + \boldsymbol{Y}'([\mathbf{I} - \mathbf{H}]'[\mathbf{I} - \mathbf{H}])\boldsymbol{Y} \end{aligned}$$

Kedua matriks $\mathbf{H}$ dan $[\mathbf{I} - \mathbf{H}]$ adalah simetri dan idempoten, sehingga $\mathbf{H}'\mathbf{H} = \mathbf{H}$ dan $[\mathbf{I} - \mathbf{H}]'[\mathbf{I} - \mathbf{H}] = [\mathbf{I} - \mathbf{H}]$. Dua suku tengah adalah nol, disebabkan oleh dua bentuk kuadrat yang saling orthogonal, yaitu $\mathbf{H}'[\mathbf{I} - \mathbf{H}] = [\mathbf{H} - \mathbf{H}] = \mathbf{0}$. Selanjutnya

$$\boldsymbol{Y}'\boldsymbol{Y} = \boldsymbol{Y}'\mathbf{H}\boldsymbol{Y} + \boldsymbol{Y}'[\mathbf{I} - \mathbf{H}]\boldsymbol{Y} = \hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}} + \mathbf{e}'\mathbf{e} \qquad \text{(G.14)}$$

(1)

$$\text{JK(Total)} = \text{JK(Model)} + \text{JK (Sisa)} \qquad \text{(G.15)}$$

(2)

Jumlah kuadrat total dipartisi atas dua jumlah kuadrat, yaitu jumlah kuadrat model dan jumlah kuadrat sisa. Kedua jumlah kuadrat berturut-turut menyatakan komponen terjelaskan dan komponen tak terjelaskan dari model. JK(Model) $= \hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}} = \boldsymbol{Y}'\mathbf{H}\boldsymbol{Y}$ memiliki matriks pendefinisian $\mathbf{H}$, dan JK(Sisa) $= \mathbf{e}'\mathbf{e} = \boldsymbol{Y}'[\mathbf{I} - \mathbf{H}]\boldsymbol{Y}$, memiliki matriks pendefinisian $[\mathbf{I} - \mathbf{H}]$.

Jika dua matriks pendefinisian diperkalikan, **H [I - H]** = **0**, sehingga dua jumlah kuadrat adalah saling orthogonal, selanjutnya membentuk partisi yang aditip. Derajat bebas bagi kedua jumlah kuadrat ditentukan oleh masing-masing pangkat dari matriks pendefinisiannya. Oleh karena kedua matriks pendefinisian adalah matriks idempoten, pangkatnya akan sama dengan *trace*nya, yaitu

lihat asalnya

KACAU

Rank (**H**) = tr(**H**) = tr(**X** (**X'X**)$^{-1}$ **X'**) = tr(**X'X**)$^{-1}$ **X' X**) (teori dalam aljabar matriks tr(**ABC**) = tr(**BCA**)) **=** tr ($\mathbf{I}_{p+1}$) = p +1 (p + 1 = jumlah kolom matriks **X**)

Rank(**I - H**) = tr(**I - H**) = tr(**I**) – tr(**H**) = n-(p+1)
Derajat bebas bagi JK(Model) adalah p+1, dan derajat bebas JK(sisa) adalah n-(p+1).
Cara menghitung jumlah kuadrat melalui notasi matriks adalah

JK(Total) = $\boldsymbol{Y'}\boldsymbol{Y}$

JK(Model) = $\hat{\boldsymbol{Y}}'\hat{\boldsymbol{Y}}$ = $(\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'}\boldsymbol{Y})$= $\hat{\boldsymbol{\beta}}'\mathbf{X'}\boldsymbol{Y}$

JK(Sisa) = $\boldsymbol{Y'}\boldsymbol{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X'}\boldsymbol{Y}$.(G.16)

Pada analisis regresi, kita memerlukan pengetahuan tentang seberapa besar kontribusi sekelompok variabel bebas terhadap variasi ***Y*** disekitar nilai tengahnya. Ukuran informasi ini dapat dilihat dari selisih antara JK(Model) yang memuat variabel bebas dengan JK(Model) tanpa variabel bebas. JK(Model) tanpa variabel bebas disebut faktor koreksi, dan ditulis JK(μ). Selisih antara JK(Model) dengan JK(μ) disebut jumlah kuadrat regresi, ditulis JK(Reg).

JK(Reg) = JK(Model) – JK(μ)

Untuk mendapatkan JK(μ), pada model aditip linier $\boldsymbol{Y} = \mathbf{X}^* \boldsymbol{\beta} + \boldsymbol{\varepsilon}$, dengan matriks $\mathbf{X}^* = \mathbf{1}$. Matriks **1** hanya suatu vektor kolom dengan semua elemen 1 atau kolom pertama dari matriks **X**.

$$\hat{\boldsymbol{\beta}} = ((\mathbf{X}^*)'\mathbf{X}^*)^{-1} (\mathbf{X}^*)'\boldsymbol{Y} = (\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\boldsymbol{Y} = (1/n)\,\mathbf{1}'\boldsymbol{Y} = \overline{\mathrm{Y}}$$

sehingga

$$\mathrm{JK}(\mu) = \hat{\boldsymbol{\beta}}'(\mathbf{X}^*)'\boldsymbol{Y} = (1/n)(\mathbf{1}'\boldsymbol{Y})'(\mathbf{1}'\boldsymbol{Y})$$

$$= (1/n)\ \boldsymbol{Y}'(\mathbf{1}\,\mathbf{1}')\boldsymbol{Y} \qquad \text{(G.17)}$$

Oleh karena $\mathbf{1}'\boldsymbol{Y} = \sum \mathbf{Y}_i$, maka $\mathrm{JK}(\mu) = n\,\overline{\mathrm{Y}}^2$.

Jika dimisalkan $\mathbf{1}\ \mathbf{1}' = \mathbf{J}$, dengan **J** adalah matriks berukuran (nxn) dengan semua elemennya 1, maka

$$\mathrm{JK(Reg)} = \mathrm{JK(Model)} - \mathrm{JK}(\mu)$$

$$= \boldsymbol{Y}'\mathbf{H}\boldsymbol{Y} - \boldsymbol{Y}'(\mathbf{J}/\mathbf{n})\boldsymbol{Y}$$

$$= \boldsymbol{Y}'(\mathbf{H} - \mathbf{J}/\mathbf{n})\boldsymbol{Y} \qquad \text{(G.18)}$$

Derajat bebas bagi JK(μ) adalah 1, sehingga derajat bebas bagi JK(Reg) adalah *p*.

Partisi jumlah kuadrat pada regresi linier ganda diperlihatkan melalui Tabel 5.7 .

Tabel 5.7 Analisis Ragam

Sumber variasi	Derajat Bebas	Formula Jumlah Kuadrat	Formula Perhitungan
$\mathrm{Total}_{\mathrm{korks}}$	n-1	$\boldsymbol{Y}'(\mathbf{I} - \mathbf{J})\boldsymbol{Y}$	$\boldsymbol{Y}'\boldsymbol{Y} - n\overline{\mathrm{Y}}^2$
Model	p +1	$\boldsymbol{Y}'\mathbf{H}\boldsymbol{Y}$	$\hat{\boldsymbol{\beta}}'\mathbf{X}'\boldsymbol{Y}$
Rataan	1	$(1/n)\ \boldsymbol{Y}'(\mathbf{1}\,\mathbf{1}')\boldsymbol{Y}$	$n\overline{\mathrm{Y}}^2$
Regresi	p	$\boldsymbol{Y}'[\mathbf{H} - \mathbf{J}/n]\boldsymbol{Y}$	$\hat{\boldsymbol{\beta}}'\mathbf{X}'\boldsymbol{Y} - n\overline{\mathrm{Y}}^2$
Sisaan	n-(p+1)	$\boldsymbol{Y}'[\mathbf{I} - \mathbf{H}]\boldsymbol{Y}$	$\boldsymbol{Y}'\boldsymbol{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\boldsymbol{Y}$.

spasi

5.3.3 Uji Keberartian Model Regresi

Suatu model regresi dinyatakan berarti jika terdapat paling kurang satu variabel bebas di dalam model yang signifikan berpengaruh atau berhubungan terhadap variabel terikat. Oleh karena pengaruh atau hubungan yang signifikan suatu variabel bebas, X_i terhadap variabel terikat Y ditentukan oleh koefisiennya yang signifikan berbeda dengan nol, $\beta_j \neq 0$, maka suatu model regresi dinyatakan berarti apabila terdapat paling kurang satu koefisien variabel dalam model yang signifikan berbeda dengan nol. Untuk menguji keberartian dari model regresi atau apakah sekelompok variabel bebas dapat memberikan informasi terhadap variasi ***Y*** disekitar nilai tengahnya, dirumuskan hipotesis uji berikut.

spasi

$H_0 : \beta_1 = \beta_2 = \ldots = \beta_p = 0$

$H_1 : \beta_j \neq 0$,Untuk paling kecil satu dari j.

$j = 1, 2, \ldots p$ (G.19)

( )

Jika H_0 benar, maka ratio antara rata-rata JK(Reg) dengan rata-rata JK(Sisa) akan menyebar F dengan derajat bebas pembilang p dan derajat bebas penyebut n-p-1 (bukti dapat dijumpai pada Searle, 1971, dan Kshirsagar, 1983). Sehingga statistik yang digunakan untuk menguji hipotesis di atas adalah

spasi

$$F = \frac{JK(Reg)/p}{JK(Sisa)/n-(p+1)} \qquad (G.20)$$

( )

Kriteria pengujian :

Untuk taraf signifikan pengujian α tertentu yang digunakan, H_0 masih diterima jika

$F_{hit} < F(\alpha, p, n\text{-}p\text{-}1)$ atau $P(F > F_{hit}) > \alpha$, sebaliknya H_0 ditolak.

spasi

5.3.4 Uji Koefisien Regresi Parsial

Uji parsial digunakan untuk mempelajari kontribusi secara tunggal suatu variabel bebas X_j , terhadap variasi variabel respon ***Y*** pada model regresi yang memuat semua variabel bebas lainnya. Hasil dari pengujian parsial adalah mengetahui apakah suatu variabel bebas berpengaruh atau tidak terhadap variabel terikat pada saat variabel lain hadir dalam model. Kehadiran variabel-variabel lainnya dalam model bertindak sebagai variabel kontrol bagi suatu variabel yang sedang diuji. Pengujian dilakukan terhadap koefisien variabel, yaitu β_j. Besaran koefisien β_j diartikan sebagai perubahan rata-rata variabel respon akibat perubahan perunit variabel bebas j, *dengan variabel bebas lain dipertahankan konstan*.

Pada penguraian jumlah kuadrat secara parsial dan secara sekuensial, JK(Reg) ditulis JK($\beta_1, \beta_2, \ldots, ,\beta_p \mid \beta_0$), sehingga jumlah kuadrat parsial ke j ditulis

JK($\beta_j \mid \beta_0, \beta_1, \ldots, \beta_{j-1}, \beta_{j+1}, \ldots ,\beta_p$)

Atau

$\boldsymbol{R}(\beta_j \mid \beta_0, \beta_1, \ldots, \beta_{j-1}, \beta_{j+1}, \ldots ,\beta_p)$

Uji partial dapat menggunakan uji-F-partial, dan dapat juga menggunakan uji- t dua arah. Sebagaimana ditunjukkan sebelumnya, bahwa penduga ragam dari koefisien regresi ke j adalah $S^2.c_{jj}$, dengan c_{jj} adalah elemen diagonal ke j = 0, 1, ..., p dari matriks $(\mathbf{X'X})^{-1}$.

Rumusan hipotesis untuk pengujian koefisien parsial adalah :

H_0 : $\beta_j = 0$.

H_1 : $\beta_j \neq 0$, j = 1, 2, ... p (G.21) ( )

Pengujian hipotesis di atas dapat dilakukan dengan uji-F arah kanan, yaitu statistik

sparse

$$F = \frac{\hat{\beta}_j^2}{c_{jj}\, S^2} \qquad (G.22)$$ ( )

menyebar F dengan derajat bebas pembilang 1 dan penyebut n-p-1.

Kriteria pengujian (F) :

Untuk taraf signifikan pengujian α tertentu yang digunakan, H_0 masih diterima jika

$F_{hit} < F(\alpha, 1, n\text{-}p\text{-}1)$ atau $P(F > F_{hit}) > \alpha$, sebaliknya H_0 ditolak.

Pengujian hipotesis di atas dapat juga dan lebih sering menggunakan uji-T dua arah dengan derajat bebas n-(p+1).

$$t = \frac{\hat{\beta}_j}{\sqrt{c_{jj} S^2}} \qquad (G.23)$$ ( )

Kriteria pengujian (t) :
Jika pada pengujian ini digunakan taraf signifikan sebesar α , maka

H_0 diterima jika $|\text{t-hitung}| \leq t_{\alpha/2;n-p-1}$, atau
$Pr = [\ P\ (\ t \leq -|t_{\text{hitung}}|\ ) + P\ (\ t \geq |t_{\text{hitung}}|\ )\ ] \geq \alpha$, sebaliknya H_0 ditolak.

Pada uji parsial ini perlu dikemukakan bahwa, jika hasil pengujian suatu variabel adalah signifikan secara statistika pada model yang melibatkan semua variabel bebas, maka belum tentu signifikan pada model dengan hanya subkelompok variabel bebas.

Contoh pengerjaan 5.3 :
Contoh soal menggunakan data hasil studi lingkungan pada Tabel 5.5 pada contoh pengerjaan 5,2. Dengan menggunakan perangkat lunak statistika *S-Plus*, diperoleh berturut-turut matriks $(\mathbf{X'X})^{-1}$ dan *output* hasil analisis regresi sebagai berikut.

Matriks $(\mathbf{X'X})^{-1}$:

	[,1]	[,2]	[,3]	[,4]
[1,]	2.8559225951	6.173356e-004	-0.03535068677	-4.090973e-002
[2,]	0.0006173356	3.341412e-006	-0.00001807261	1.409908e-008
[3,]	-0.0353506868	-1.807261e-005	0.00053385411	1.439104e-004
[4,]	-0.0409097295	1.409908e-008	0.00014391044	2.613921e-003

a. Pengujian Keberartian Model Regresi

Untuk menguji keberartian dari model regresi atau rumusan hipotesis uji berikut.

$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$

$H_1 : \beta_j \neq 0$,Untuk paling kecil satu dari j. j = 1, 2, 3

Pada Tabel 5.6 diperlihatkan nilai statistik F, yaitu
$F = \dfrac{JK(Reg)/3}{JK(Sisa)/26}$ = 7,332 dengan taraf signifikan F adalah Pr = 0,00102.

Jika pengujian menggunakan taraf signifikan α = 0,05, maka α > Pr , sehingga hipotetis H_0 ditolak. Disimpulkan bahwa pada model ini, kelompok variabel bebas (radiasi, temperatur, dan angin) signifikan berkontribusi terhadap variasi variabel terikat (Ozone).
Penjelasan :
JK(Sisa) = 9.079074 dan

$$JK(Reg) = JK(\beta_1 \mid \beta_0) + JK(\beta_2 \mid \beta_0, \beta_1) + JK(\beta_3 \mid \beta_0, \beta_1, \beta_2)$$
$$= 3{,}148149 + 4{,}233164 + 0{,}299932$$
$$= 7{,}681245$$

b Uji Koefisien parsial

Pengujian parsial bagi setiap koefisien regresi adalah menguji hipotesis berikut.

spasi

H_0 : $\beta_j = 0$.

H_1 : $\beta_j \neq 0$, $\quad$ j = 1, 2, 3

Pengujian hipotesis di atas dapat menggunakan uji-T dua arah dengan derajat bebas
n-(p+1) = 30-3-1 = 26.

$$t = \frac{\hat{\beta}_j}{\sqrt{c_{jj} S^2}}$$

Pada Tabel 5.6 , yaitu pada bagian koefisien diperlihatkan bahwa untuk pengujian hipotesis :

1) $H_0 : \beta_1 = 0$, nilai t = 1,208 dengan taraf signifikan Pr = 0,2379. Jika pengujian menggunakan taraf signifikan α = 0,05, maka α < Pr , sehingga bukti belum cukup untuk menolak hipotesis H_0. Disimpulkan bahwa variabel radiasi pada model linier yang memuat (variabel temperatur dan angin) tidak signifikan berpengaruh terhadap variabel ozone.
2) $H_0 : \beta_2 = 0$, nilai t = 3,3429 dengan taraf signifikan Pr = 0,0025. Jika pengujian menggunakan taraf signifikan α = 0,05, maka α > Pr , sehingga diperoleh petunjuk yang kuat untuk menolak hipotesis H_0. Disimpulkan bahwa variabel temperatur pada model linier yang memuat (variabel radiasi dan angin) adalah signifikan berpengaruh terhadap variabel

ozone. Rata-rata variabel ozone berubah (meningkat) sebesar 0.0456 akibat perubahan (peningkatan) per unit variabel temperatur, dengan variabel lain dipertahankan konstan.

3) H_0 : β_3 = 0, nilai t = -0,9268 dengan taraf signifikan Pr = 0,3626. Nilai t ini diperoleh dari perhitungan berikut

$$t = \frac{\hat{\beta}_3}{\sqrt{c_{33} RJK(Sisa)}} = \frac{-0{,}028}{\sqrt{2{,}613921e-003(0{,}349195)}} = -0{,}9268$$

dengan c_{33} diperoleh dari elemen diagonal baris ke 4 kolom ke 4 dari matriks $(\mathbf{X'X})^{-1}$. Jika pengujian menggunakan taraf signifikan α = 0,05, maka α < Pr , sehingga bukti belum cukup untuk menolak hipotesis H_0. Disimpulkan bahwa variabel angin pada model linier yang memuat (variabel radiasi dan temperatur) tidak signifikan berpengaruh terhadap variabel ozone.

<u>Contoh pengerjaan 5.4</u>:

Data tentang berat ikan cumi dan lima variabel sebagai hasil pengukuran panjang untuk 5 bagian tertentu pada badan ikan cumi, disajikan pada Tabel 5.8. Lakukan analisis regresi dengan komputerisasi untuk :

a. Menduga persamaan regresi dari model

$$Y_j = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \beta_3 X_{3j} + \beta_4 X_{4j} + \beta_5 X_{5j} + \varepsilon_j$$

b. Menguji keberartian model dan menyimpulkan, dengan terlebih dahulu menuliskan hipotesis ujinya

c. Melakukan uji partial terhadap setiap koefisien variabel bebas, melalui pengujian hipotesis hipotesis

H_0 : β_j = 0.

H_1 : $\beta_j \neq 0$, j = 1, 2, 3, 4, 5

<u>Pemecahan pengerjaan 5.3 :</u>

a. Hasil komputerisasi dengan SPSS pada Table 5.8, menyajikan koefisien masing masing variabel (Unstandardized Coefficients), sehingga dugaan persaamaan regresi adalah

Y_j = -6.512 + 1.999 X_1 - 3.675 X_2 + 2.524 X_3 + 5.158 X_4 + 14.401 X_5

b. Rumusan hipotesis untuk menguji keberartian model regresi yang kita peroleh ini adalah:
$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$
$H_1 : \beta_j \neq 0$,Untuk paling kecil satu dari j. j = 1, 2, ... 5
Pada Table 5.8 diperoleh nilai statistik F = 84.070 dengan taraf sigifikan 0,000. Jika pada pengujian ini digunakan taraf signifikan pengujian α = 0,05, maka taraf signifikan statististik F kurang dari taraf signifikan pengujian, yaitu 0,000 < 0,05, (Nilai statististik F-hitung lebih dari $F_{(5;16)}$-tabel, yaitu 84,070 > 2,120), sehingga H_0 ditolak. Ini berarti tidak semua koefisien variabel bebas sama dengan nol atau model regresi dinyatakan berarti.

Tabel 5.8 Berat ikan cumi dan hasil pengukuran panjang dari 5 bagian badan ikan cumi

Rostral X1	Wing X2	Rostral to Notch X3	Notch to Wing X4	Width X5	Weigh Y
1.31	1.07	0.44	0.75	0.35	1.95
1.55	1.49	0.53	0.90	0.47	2.90
0.99	0.84	0.34	0.57	0.32	0.72
0.99	0.83	0.34	0.54	0.27	0.81
1.05	0.90	0.36	0.64	0.30	1.09
1.09	0.93	0.42	0.61	0.31	1.22
1.08	0.90	0.40	0.51	0.31	1.02
1.27	1.08	0.44	0.77	0.34	1.93
0.99	0.85	0.36	0.56	0.29	0.64
1.34	1.13	0.45	0.77	0.37	2.08
1.30	1.10	0.45	0.76	0.38	1.98
1.33	1.10	0.48	0.77	0.38	1.90
1.86	1.47	0.60	1.01	0.65	8.56
1.58	1.34	0.52	0.95	0.50	4.49
1.97	1.59	0.67	1.20	0.59	8.49
1.80	1.56	0.66	1.02	0.59	6.17
1.75	1.58	0.63	1.09	0.59	7.54
1.72	1.43	0.64	1.02	0.63	6.36
1.68	1.57	0.72	0.96	0.68	7.63

Rostral X1	Wing X2	Rostral to Notch X3	Notch to Wing X4	Width X5	Weigh Y
1.75	1.59	0.68	1.08	0.62	7.78
2.19	1.86	0.75	1.24	0.72	10.15
1.73	1.67	0.64	1.14	0.55	6.88

Sumber : Contoh data analisis regresi dari Raymond H. Myers, 1990

c. Berdasarkan penyajian hasil analisis dengan SPSS pada Tabel 5.9, untuk pengujian partial:
   (1) Koefisien β_1, diperoleh nilai statistik T = 0.777, dengan taraf signifikan 0,449. Jika pada pengujian ini digunakan taraf signifikan pengujian α = 0,05, maka taraf signifikan statististik T lebih dari taraf signifikan pengujian, yaitu 0,449 > 0,05, (Nilai statististik |T-hitung| kurang dari $T_{(0,05;16)}$, yaitu 0,777 < 2,120), sehingga dinyatakan belum cukup bukti untuk menolak H_0. Ini berarti bahwa koefisien β_1 tidak signifikan berbeda dengan 0 dan varibel panjang rostral, X_2 berpengaruh terhadap berat ikan cumi, Y.
   (2) Koefisien β_2 , diperoleh nilai statistik T = -1,325, dengan taraf signifikan 0, 204. Jika pada pengujian ini digunakan taraf signifikan pengujian α = 0,05, maka taraf signifikan statististik T lebih dari taraf signifikan pengujian, yaitu 0, 204 > 0,05, (Nilai statististik |T-hitung| kurang dari $T_{(0,05;16)}$, yaitu 1,325 < 2,120), sehingga belum cukup bukti untuk menolak H_0. Ini berarti bahwa koefisien β_2 tidak signifikan berbeda dengan 0 dan varibel panjang *wing*, X_2 berpengaruh terhadap berat ikan cumi, Y.

Tabel 5.9 Analisis Ragam dari Analisis Regresi Ganda

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	208.007	5	41.601	84.070	.000[b]
	Residual	7.918	16	.495		
	Total	215.925	21			

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-6.512	.934		-6.976	.000
	X1	1.999	2.573	.223	.777	.449
	X2	-3.675	2.774	-.371	-1.325	.204
	X3	2.524	6.347	.105	.398	.696
	X4	5.158	3.660	.366	1.409	.178
	X5	14.401	4.856	.671	2.966	.009

a. Dependent Variable: Y

b. Predictors: (Constant), X5, X4, X3, X2, X1

(3) Koefisien β_3, diperoleh nilai statistik T = 0,398, dengan taraf signifikan 0,696. Jika pada pengujian ini digunakan taraf signifikan pengujian α = 0,05, maka taraf signifikan statististik T lebih dari taraf signifikan pengujian, yaitu 0,696 > 0,05, (Nilai statististik |T-hitung| kurang dari $T_{(0,05;16)}$, yaitu 0,398 < 2,120), sehingga belum cukup bukti untuk menolak H_0. Ini berarti bahwa koefisien β_3 tidak signifikan berbeda dengan 0 dan varibel panjang *rostral to Notch* , X_3 berpengaruh terhadap berat ikan cumi, Y.

(4) Koefisien β_4, diperoleh nilai statistik T = 1,409, dengan taraf signifikan 0,178. Jika pada pengujian ini digunakan taraf signifikan pengujian α = 0,05, maka taraf signifikan statististik T lebih dari taraf signifikan pengujian, yaitu 0,178 > 0,05, (Nilai statististik |T-hitung| kurang dari $T_{(0,05;16)}$, yaitu 1,409 < 2,120), sehingga belum cukup bukti untuk menolak H_0. Ini berarti bahwa koefisien β_4 tidak signifikan berbeda dengan 0 dan varibel panjang *rostral to wing*, X_4 berpengaruh terhadap berat ikan cumi, Y.

(5) Koefisien β_5, diperoleh nilai statistik T = 2,966, dengan taraf signifikan 0,009. Jika pada pengujian ini digunakan taraf signifikan pengujian α = 0,05, maka taraf signifikan

statististik T lebih dari taraf signifikan pengujian, yaitu 0,009 < 0,05, (Nilai statististik |T-hitung| kurang dari $T_{(0,05;16)}$, yaitu 2,966 < 2,120), sehingga belum cukup bukti untuk menolak H_0. Ini berarti bahwa koefisien β_5 tidak signifikan berbeda dengan 0 dan varibel Width, X_5 berpengaruh terhadap berat ikan cumi, Y.

spasi

5.3.5Uji Koefisien Regresi Sekuensial

Uji koefisien sekuensial (*sequential*) digunakan untuk mempelajari kontribusi suatu variabel pada model yang melibatkan variabel bebas sebelumnya. Uji sekuensial memperhatikan urutan pemasukan variabel bebas ke dalam model. Misalkan varabel ke 4 , yang terkoreksi oleh variabel 1, 2, dan 3 dinyatakan berkontribusi terhadap JK(Reg), maka belum tentu kontribusi ini sama jika variabel 4 hanya terkoreksi oleh variabel 1 atau hanya terkoreksi oleh variabel 1 dan 2.

Jumlah kuadrat yang dibicarakan pada Uji parsial, tidak membentuk kontribusi secara aditip terhadap JK(Reg), yaitu

supaya suku tdk terpotong maka Font dikecilkan saja

$$JK(\beta_1, \beta_2, \dots, \beta_p \mid \beta_0) \neq JK(\beta_1 \mid \beta_0, \beta_2, \dots, \beta_p) + JK(\beta_2 \mid \beta_0, \beta_1, \beta_3, \dots, \beta_p) + \dots + JK(\beta_j \mid \beta_0, \beta_1, \dots, \beta_{j-1}, \beta_{j+1}, \dots, \beta_p) + \dots + JK(\beta_p \mid \beta_0, \beta_1, \beta_2, \dots, \beta_{p-1}).$$

enter

Tetapi jumlah kuadrat sekuensial membentuk partisi aditip terhadap JK(Reg), yaitu

$$JK(\beta_1, \beta_2, \dots, \beta_p \mid \beta_0) = JK(\beta_1 \mid \beta_0) + JK(\beta_2 \mid \beta_0, \beta_1) + JK(\beta_3 \mid \beta_0, \beta_1, \beta_2) + \dots + JK(\beta_p \mid \beta_0, \beta_1, \beta_2, \dots, \beta_{p-1}). \quad (G.24)$$

enter ( )

JK(. .) menyatakan " regresi terjelaskan oleh ...", dengan garis pertikal adalah notasi yang menyatakan " dalam kehadiran dari ...". Sebagai contoh $JK(\beta_2 \mid \beta_0, \beta_1)$ adalah peningkatan pada jumlah kuadrat regresi, ketika regressor X_2 ditambahkan kedalam model yang hanya melibatkan X_1 dan suku konstan.

Partisi jumlah kuadrat secara sekuensial ini sangat bermanfaat jika kita membutuhkan informasi tentang harga bagi suatu subkelompok regressor (variabel penjelas). Misalkan pada model regresi dengan p = 4 regressor, kita dapat menuliskan persamaan

$$JK(\beta_1, \beta_2, \beta_3, \beta_4 \mid \beta_0) = JK(\beta_1, \beta_2 \mid \beta_0) + JK(\beta_3, \beta_4 \mid \beta_0, \beta_1, \beta_2)$$

Dua suku pada ruas kanan mewakili partisi dua derajat bebas. $JK(\beta_1, \beta_2 \mid \beta_0)$ kurang bermanfaat bagi pendugaan terhadap β_1 dan β_2, karena tidak terkoreksi oleh X_3 dan X_4. Tetapi $JK(\beta_3, \beta_4 \mid \beta_0, \beta_1, \beta_2)$ adalah sangat vital dalam menjelaskan kepentingan β_4 bersama β_3. Jika kita tertarik untuk melakukan pendugaan terhadap β_4 bersama β_3, maka $JK(\beta_3, \beta_4 \mid \beta_0, \beta_1, \beta_2)$ dapat dihitung dengan

$$JK(\beta_3, \beta_4 \mid \beta_0, \beta_1, \beta_2) = JK(\beta_3 \mid \beta_0, \beta_1, \beta_2) + JK(\beta_4 \mid \beta_0, \beta_1, \beta_2, \beta_3) \quad \text{(G.25)}$$

Selanjutnya dapat digunakan statistik uji

$$F = \frac{JK(\beta_3, \beta_4 \mid \beta_0, \beta_1, \beta_2)/2}{RJK(Sisa)} \quad \text{(G.26)}$$

dengan derajat bebas pembilang 2 dan penyebut n-p-1. Satistik ini digunakan untuk menguji hipotesis,

$H_0 : \beta_3 = \beta_4 = 0$

$H_1 : \beta_3 \neq 0$ atau $\beta_4 \neq 0$

Jika kita ingin menguji kontribusi suatu variabel X_j, pada model yang hanya melibatkan variabel sebelumnya, yaitu $X_1, X_2, .., X_{j-1}$, statistik uji adalah

$$F = \frac{JK(\beta_j \mid \beta_0, \beta_1, \ldots, \beta_{j-1})}{RJK(Sisa)} \quad \text{(G.27)}$$

Statistik uji ini menyebar F dengan derajat bebas pembilang 1 dan penyebut n-p-1.

Contoh pengerjaan 5.5:

Contoh soal menggunakan data hasil studi lingkungan pada Tabel 5.5 pada contoh pengerjaan 5.2. Hasil analisis regresi ganda menggunakan perangkat lunak statistika *S-Plus*, diperlihatkan pada Tabel 5.10. Lakukan pengujian sekuensial terhadap:

a) Pengaruh variabel radiasi X_1 terhadap variabel ozone Y pada model yang memuat hanya konstanta.
b) Pengaruh variabel temperatus X_2 terhadap variabel ozone Y pada model yang memuat konstanta dan variabel radiasi.
c) Pengaruh variabel angin X_3 terhadap variabel ozone Y pada model yang memuat konstanta, variabel radiasi dan temperatur.
d) Pengaruh subkelompok variabel angin X_3 dan temperatur X_2 terhadap variabel ozone Y pada model yang memuat konstanta dan variabel radiasi.

Pemecahan pengerjaan 5.5:

Tabel 5.10 *Output* Komputerisasi dari Hasil Analisis Regresi Linier

```
*** Linear Model ***

Call: lm(formula = ozone ~ radiation + temperature + wind,
       data = air10, na.action = na.exclude)


Coefficients:
                Value Std. Error t value Pr(>|t|)
(Intercept) -0.2949  0.9986    -0.2953  0.7701
  radiation  0.0013  0.0011     1.2080  0.2379
temperature  0.0456  0.0137     3.3429  0.0025
       wind -0.0280  0.0302    -0.9268  0.3626

Residual standard error: 0.5909 on 26 degrees of freedom
Multiple R-Squared: 0.4583
F-statistic: 7.332 on 3 and 26 degrees of freedom, the p-value is 0.00102

Correlation of Coefficients:
            (Intercept) radiation temperature
  radiation  0.1998
temperature -0.9053     -0.4279
       wind -0.4735      0.0002    0.1218

Analysis of Variance Table

Response: ozone

Terms added sequentially (first to last)

            Df Sum of Sq  Mean Sq  F Value     Pr(F)
  radiation  1  3.148149 3.148149  9.01544 0.0058488
temperature  1  4.233164 4.233164 12.12263 0.0017766
       wind  1  0.299932 0.299932  0.85892 0.3625634
  Residuals 26  9.079074 0.349195
```

Pengujian kontribusi suatu variabel X_j terhadap JK(Reg), pada model yang hanya melibatkan variabel sebelumnya, yaitu X_1, X_2, .., X_{j-1} , menggunakan statistik uji

$$F = \frac{JK(\beta_j | \beta_0, \beta_1, \ldots, \beta_{j-1})}{RJK(Sisa)}$$

Dari nilai-nilai yang diperoleh pada Tabel 5.10 pada bagian penambahan suku secara sekuensial (*Terms added sequentially*), dapat dilakukan perhitungan berikut :

a. $F = \dfrac{JK(\beta_1 | \beta_0)}{RJK(Sisa)} = \dfrac{3{,}148149}{0{,}349195} = 9{,}01544$, dengan Pr = 0,0058488.

   Jika pengujian menggunakan taraf signifikan α = 0,05, maka α > Pr , sehingga diperoleh petunjuk yang kuat untuk menolak hipotesis H_0. Disimpulkan bahwa pada model regresi linier sederhana ini, variabel radiasi signifikan berpengaruh terhadap variasi ozone.

b. $F = \dfrac{JK(\beta_2 | \beta_0, \beta_1)}{RJK(Sisa)} = \dfrac{4{,}233164}{0{,}349195} = 12{,}12263$, dengan Pr = 0,0058488

   Jika pengujian menggunakan taraf signifikan α = 0,05, maka α < Pr , sehingga diperoleh petunjuk yang kuat untuk menolak H_0. Disimpulkan bahwa variabel temperatur pada model linier yang memuat (variabel radiasi) signifikan berpengaruh terhadap variasi ozone.

c. $F = \dfrac{JK(\beta_3 | \beta_0, \beta_1, \beta_2)}{RJK(Sisa)} = \dfrac{0{,}299932}{0{,}349195} = 0{,}85892$, dengan Pr = 0,3625634

   Jika pengujian menggunakan taraf signifikan α = 0,05, maka α < Pr , sehingga bukti belum cukup untuk menolak H_0. Disimpulkan bahwa variabel angin pada model linier yang memuat (variabel radiasi dan temperatur) tidak signifikan berpengaruh terhadap variasi ozone.

d. Pengujian kontribusi subkelompok terakhir variabel bebas, yang terkoreksi oleh subkelompok variabel sebelumnya, terhadap JK(Reg) dapat dilakukan dengan uji koefisien sekuensial. Pengujian kontribusi 2 regressor terakhir, yaitu temperatur dan angin pada model yang memuat variabel radiasi (X_1), digunakan statistik uji

$$F = \frac{JK(\beta_2, \beta_3 \mid \beta_0 \ , \beta_1)/2}{RJK(Sisa)}$$

Rumusan hipotesis yang diuji adalah

$H_0 : \beta_2 = \beta_3 = 0$
$H_1 : \beta_2 \neq 0$ atau $\beta_3 \neq 0$

Dari nilai-nilai yang diperoleh pada Tabel 5.10, dapat dilakukan perhitungan berikut :

$$\begin{aligned} JK(\beta_2, \beta_3 \mid \beta_0, \beta_1) &= JK(\beta_2 \mid \beta_0, \beta_1) + JK(\beta_3 \mid \beta_0, \beta_1, \beta_2) \\ &= 4{,}233164 + 0{,}299932 \\ &= 4{,}533096 \end{aligned}$$

enter

$$F = \frac{JK(\beta_2, \beta_3 \mid \beta_0 \ , \beta_1)/2}{RJK(Sisa)} = \frac{4{,}533096/2}{0{,}349195} = 6{,}491$$

F > F(0,05;2;26), yaitu 6,491 > 3,37 , sehingga diperoleh petunjuk yang kuat untuk menolak H_0. Disimpulkan bahwa pada model dengan 3 regressor berturut-turut radiasi, temperatur dan angin, subkelompok 2 variabel temperatur dan angin signifikan berkontribusi terhadap variasi variabel respon ozone.

spasi

5.3.6 Pengujian Hipotesis Linier Gabungan

Uji koefisien parsial maupun uji terhadap subkelompok koefisien, sesungguhnya dapat dilakukan melalui pembentukan hipotesis linier gabungan (*General Linear Hipotesys*). Hipotesis linier gabungan didefinisikan sebagai berikut.

spasi

$H_0 : \mathbf{K'} \boldsymbol{\beta} = \mathbf{m}$
$H_1 : \mathbf{K'} \boldsymbol{\beta} \neq \mathbf{m}$ (G.28)

( )

K' adalah matriks berukuran (k x (p+1)), yang mendefinisikan sebanyak k fungsi linier dari $\boldsymbol{\beta}$ untuk diuji. Setiap baris dari matriks **K'** terdiri dari koefisien-koefisien suatu fungsi linier, dan **m** adalah vektor konstan berukuran (kx1).

Misalkan $\boldsymbol{\beta} = (\beta_0 , \beta_1, \beta_2, \beta_3)$ dan kita ingin menguji komposisi hipotesis nol bahwa $\beta_1 = \beta_2$, $\beta_1 + \beta_2 = 2\ \beta_3$, dan $\beta_0 = 20$. Hipotesis ekuivalen dengan penulisan

$$
\left.\begin{aligned}
H_0 : \beta_1 - \beta_2 &= 0 \\
\beta_1 + \beta_2 - 2\beta_3 &= 0 \\
\beta_0 &= 20
\end{aligned}\right\}
$$

lihat asalnya dirapikan

Hipotesis di atas dapat ditulis dalam bentuk $\mathbf{K'}\ \boldsymbol{\beta} = \mathbf{m}$, dengan mendefinisikan

$$
\mathbf{K'} = \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 1 & -2 \\ 1 & 0 & 0 & 0 \end{pmatrix} \text{ dan } \mathbf{m} = \begin{pmatrix} 0 \\ 0 \\ 20 \end{pmatrix}
$$

Penduga jumlah kuadrat terkecil bagi $\mathbf{K'}\ \boldsymbol{\beta} - \mathbf{m}$ diperoleh dengan menggunakan penduga $\hat{\boldsymbol{\beta}}$ bagi $\boldsymbol{\beta}$, sehingga diperoleh $\mathbf{K'}\hat{\boldsymbol{\beta}} - \mathbf{m}$. Oleh asumsi $\boldsymbol{Y}$ menyebar normal, dan persamaan (G.5) bahwa $\hat{\boldsymbol{\beta}}$ fungsi linier dari $\boldsymbol{Y}$, maka

$E(\mathbf{K'}\hat{\boldsymbol{\beta}} - \mathbf{m}) = \mathbf{K'}\, E(\hat{\boldsymbol{\beta}}) - m = \mathbf{K'}\,\boldsymbol{\beta} - \mathbf{m}$

Jika H_0 benar, maka $\mathbf{K'}\ \boldsymbol{\beta} - \mathbf{m} = \mathbf{0}$, dan matriks ragam-peragam adalah

$$
\begin{aligned}
Var(\mathbf{K'}\hat{\boldsymbol{\beta}} - \mathbf{m}) &= Var\,(\mathbf{K'}\hat{\boldsymbol{\beta}}) - 0 \\
&= \mathbf{K'}\, Var\,(\hat{\boldsymbol{\beta}})\, \mathbf{K} \\
&= \mathbf{K'}\,(\mathbf{X'X})^{-1}\,\mathbf{K}\,\sigma^2. \quad (G.29)
\end{aligned}
$$

Ragam dapat diperoleh dengan menerapkan aturan tentang ragam fungsi linier pada persamaan (G.8)

Menurut (Searle, 1971), jumlah kuadrat bagi hipotesis H_0 : $\mathbf{K'}\,\boldsymbol{\beta} = \mathbf{m}$ adalah

$$
Q = (\mathbf{K'}\hat{\boldsymbol{\beta}} - \mathbf{m})'\ (\mathbf{K'}\,(\mathbf{X'X})^{-1}\,\mathbf{K})^{-1}\,(\mathbf{K'}\hat{\boldsymbol{\beta}} - \mathbf{m}) \quad (G.30)
$$

Matriks pendefinisian dari bentuk kuadrat ini, yaitu $\mathbf{A} = (\mathbf{K'}\,(\mathbf{X'X})^{-1}\,\mathbf{K})^{-1}$ adalah invers dari matriks ragam-peragam fungsi linier $\mathbf{K'}\hat{\boldsymbol{\beta}} - \mathbf{m}$, tanpa $1/\sigma^2$. Oleh karena itu, tr(**AV**) = tr($\mathbf{I}_k$) = kσ^2 . Selanjutnya nilai harapan Q

$$E(Q) = k\sigma^2 + (\mathbf{K}'\boldsymbol{\beta} - \mathbf{m})'\ (\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K})^{-1}(\mathbf{K}'\boldsymbol{\beta} - \mathbf{m}) \qquad \text{(G.31)}$$

Dengan asumsi normalitas, Q/σ^2 adalah variabel acak yang menyebar Chi-kuadrat nonsentral, dengan derajat bebas k. Menurut Rawlings (1988), Uji-F bagi hipotesis

H_0 : $\mathbf{K}'\boldsymbol{\beta} = \mathbf{m}$ adalah

$$F = \frac{Q/k}{s^2} \qquad \text{(G.32)}$$

1. Hipotesis Linier Gabungan Sederhana

Jika pada pendefinisian $\mathbf{K}'\boldsymbol{\beta} = \mathbf{m}$, dengan **K'** adalah hanya suatu vektor tunggal atau hanya suatu fungsi linier bagi $\boldsymbol{\beta}$, maka $\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}$ adalah skalar, inversnya adalah $1/\ \mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}$. Jumlah kuadrat dapat ditulis sebagai

$$Q = \frac{(\mathbf{K}'\hat{\boldsymbol{\beta}} - \mathbf{m})^2}{\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}} \qquad \text{(G.33)}$$

Pembilang adalah kuadrat bagi fungsi linier dari $\hat{\beta}$, dan penyebut adalah ragamnya, tanpa σ^2. Selanjutnya Statistik uji-F adalah

$$F = \frac{(\mathbf{K}'\hat{\boldsymbol{\beta}} - \mathbf{m})^2}{(\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K})\,s^2} \qquad \text{(G.34)}$$

Dengan derajat bebas pembilang 1 dan penyebut n-p-1.

Uji-F bagi hipotesis ini dapat ditulis sebagi uji-T dua arah, yaitu

$$t = \frac{(\mathbf{K}'\hat{\boldsymbol{\beta}} - \mathbf{m})}{\sqrt{(\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K})\,s^2}} \qquad \text{(G.35)}$$

2. Hipoteisis Linier Gabungan Untuk Subkelompok Koefisien

Pada hipotesis $\mathbf{K}'\boldsymbol{\beta} = \mathbf{m}$, pengujian terhadap subkelompok koefisien regresi dapat dilakukan dengan mendefinisikan $\mathbf{K}'\boldsymbol{\beta} = \mathbf{0}$. Pada kasus seperti ini **K'** adalah matriks berukuran (k x (p+1)) yang terdiri dari elemen 0, kecuali terdapat elemen 1 pada setiap baris yang mengidentifikasi β_j. Sebagai contoh untuk menguji $\beta_2 = \beta_4 = 0$, pada $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$.

$$\mathbf{K}'\boldsymbol{\beta} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Oleh pembentukan **K'** seperti ini, perkalian matriks $\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{K}$ akan mengekstraksi dari matriks $(\mathbf{X}'\mathbf{X})^{-1}$ menjadi suatu submatriks berukuran (k x k). Elemen submatriks terdiri dari koefisien-koefisien bagi ragam dan koragam dari kβ_j yang ingin diuji. Pada contoh ini, submatriks berukuran (2x2) dan jumlah kuadrat Q berbentuk

$$Q = \begin{pmatrix} \hat{\beta}_2 & \hat{\beta}_4 \end{pmatrix} \begin{pmatrix} c_{22} & c_{24} \\ c_{42} & c_{44} \end{pmatrix}^{-1} \begin{pmatrix} \hat{\beta}_2 \\ \hat{\beta}_4 \end{pmatrix} \qquad \text{(G.36)}$$

Dengan c_{ij} adalah elemen dari baris (i+1) dan kolom (j+1) dari matriks $(\mathbf{X}'\mathbf{X})^{-1}$.

Uji-F bagi hipotesis H_0 : $\mathbf{K}'\boldsymbol{\beta} = \mathbf{0}$ sama dengan (G.), yaitu

$$F = \frac{Q/k}{s^2} \qquad \text{(G.37)}$$

Statistik ini menyebar F dengan derajat bebas pembilang k (pada contoh di atas k = 2) dan derajat bebas penyebut n-p-1.

Contoh pengerjaan 5.6:

Pengujian kontribusi terhadap JK(Reg), oleh subkelompok variabel bebas, tanpa urutan yang sekuensial, tetapi tetap terkoreksi oleh subkelompok variabel lainnya, dapat dilakukan dengan uji hipotesis linier gabungan. Sebagai contoh pada data hasil studi lingkungan di contoh pengerjaan 5.2.

(a) Lakukan pengujian kontribusi 2 regressor, yaitu radiasi (X_1) dan angin (X_3) pada model regresi yang memuat variabel temperatur (X_2), dengan merumuskan hipotesis uji sebagai hipotesis linier gabungan.

(b) Lakukan pengujian pada rumusan hipotesis uji liner gabungan bagian (a) dengan tambahan β_0 = 0,5, yaitu H_0 : β_0 = 0,5 dan $\beta_1 = \beta_3 = 0$.

Pemecahan pengerjaan 5.6:

(a) Untuk pengujian kontribusi 2 regressor, yaitu radiasi (X_1) dan angin (X_3) pada model yang memuat variabel temperatur (X_2). Hipotesis yang diuji adalah

H_0 : $\beta_1 = \beta_3 = 0$. Hipotesis dalam bentuk linier gabungan dituliskan sebagai berikut.

$$\mathbf{K'}\,\boldsymbol{\beta} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Oleh perkalian matriks **K'** (**X'X**)$^{-1}$ **K**, diperoleh submatriks (2x2) hasil ekstraksi dari matriks (**X'X**)$^{-1}$, yaitu

	[,1]	[,2]	[,3]	[,4]
[1,]				
[2,]			3.341412e-006	1.409908e-008
[3,]				
[4,]			1.409908e-008	2.613921e-003

→ what aslinya

Dari hasil *output* komputer pada Tabel 5.10, diperoleh koefisien dugaan $\hat{\beta}_1$ = 0.0013 dan $\hat{\beta}_3$= -0.0280. Selanjutnya jumlah kuadrat Q dapat dihitung sebagai berikut

$$Q = (\mathbf{K'}\hat{\boldsymbol{\beta}})'\,(\mathbf{K'}\,(\mathbf{X'X})^{-1}\,\mathbf{K})^{-1}\,(\mathbf{K'}\hat{\boldsymbol{\beta}})$$

$$= \begin{pmatrix} 0{,}0013 \\ -0{,}0280 \end{pmatrix}^{T} \begin{pmatrix} 3{,}341412e-006 & 1{,}409908e-008 \\ 1{,}409908e-008 & 2{,}613921e-003 \end{pmatrix}^{-1} \begin{pmatrix} 0{,}0013 \\ -0{,}0280 \end{pmatrix}$$

$$= 0{,}807$$

$$F = \frac{Q/k}{s^2} = \frac{0{,}807/2}{0{,}349195} = 1{,}1555$$

F < F(0,05;2;26), yaitu 1,156 < 3,37 , sehingga belum cukup bukti untuk menolak H_0. Disimpulkan bahwa pada model dengan 3 regressor berturut-turut radiasi, temperatur dan angin, subkelompok 2 variabel radiasi dan angin tidak signifikan berkontribusi terhadap variasi variabel respon ozone. Demikian pula hasil pengujian langsung pada output computer, diperoleh nilai statistik F = 1,1562 dengan taraf signifikan 0,3303. Oleh karena taraf signifikan statistik F lebih besar dari taraf signifikan pengujian, yaitu 0,3303 > 0,05, maka dinyatakan belum cukup bukti untuk menolak H_0.
Output menggunakan perangkat lunak *Statistical Analysis System* (*SAS*) menampilkan hasil berikut, pada Tabel 5.11.

Tabel 5.11 *Output* Hasil Pengujian Hipotesis Linier Gabungan dengan *SAS*

```
                              The SAS System

                    L Ginv(X'X) L'                Lb-c

          3.3414119E-6      1.4099083E-8      0.0013058159
          1.4099083E-8      0.002613921      -0.027843496

                 Inv(L Ginv(X'X) L')              Inv()(Lb-c)

          299274.69475     -1.614241062       390.84258745
          -1.614241062     382.56703747      -10.65411167

Dependent Variable: Y

Test:          Numerator:      0.4035   DF:    2   F value:   1.1562
               Denominator:  0.348994   DF:   26   Prob>F:    0.3303
```

(b) Pengujian hipotesis linier gabungan yang kedua adalah hipotesis linier gabungan dengan **m** $\neq$ 0. Dengan penambahan β_0 = 0,5 dari hipotesis (a), hipotesis dapat ditulis dalam bentuk berikut.

$$\mathbf{K'\,\beta - m} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} - \begin{pmatrix} 0{,}5 \\ 0 \\ 0 \end{pmatrix}$$

Dari hasil *output* komputer pada Tabel 5.12, diperoleh koefisien dugaan bagi **K' β- m**

$$\mathbf{K'}\hat{\boldsymbol{\beta}} - \mathbf{m} = \begin{pmatrix} -0{,}2949 - 0{,}5 \\ 0{,}0013 \\ -0{,}0280 \end{pmatrix} = \begin{pmatrix} -0{,}7949 \\ 0{,}0013 \\ -0{,}0280 \end{pmatrix}$$

Dan matriks pendefinisiannya adalah

$$(\mathbf{K'}\,(\mathbf{X'X})^{-1}\,\mathbf{K})^{-1} = \begin{pmatrix} 2.8559225951 & 6.173356\text{e}-004 & -4.090973\text{e}-002 \\ 0.0006173356 & 3.341412\text{e}-006 & 1.409908\text{e}-008 \\ -0.0409097295 & 1.409908\text{e}-008 & 2.613921\text{e}-003 \end{pmatrix}^{-1}$$

Selanjutnya diperoleh jumlah kuadrat Q

$$Q = (\mathbf{K'}\hat{\boldsymbol{\beta}} - \mathbf{m})'\ (\mathbf{K'}\,(\mathbf{X'X})^{-1}\,\mathbf{K})^{-1}\,(\mathbf{K'}\hat{\boldsymbol{\beta}} - \mathbf{m})$$
$$= 1{,}8387$$

spasi

$$F = \frac{Q/k}{s^2} = \frac{1{,}8387/3}{0{,}349195} = 1{,}755$$

F < F(0,05;3;26), yaitu 1,755 > 2,98, sehingga belum cukup bukti untuk menolak H_0. Disimpulkan bahwa pada model dengan 3 regressor berturut-turut radiasi, temperatur dan angin, subkelompok vektor koefisien [β_0, β_1 , β_3] tidak signifikan berbeda dengan [0.5, 0, 0]. Demikian pula hasil pengujian langsung pada *output* komputer, diperoleh nilai statistik F = 1,7561 dengan taraf signifikan 0,1803. Oleh karena taraf signifikan statistik F lebih besar dari taraf signifikan pengujian, yaitu 0,1803 > 0,05, maka dinyatakan belum cukup bukti untuk menolak H_0.

Output menggunakan perangkat lunak *Statistical Analysis System* (*SAS*) menampilkan hasil berikut, pada Tabel 5.12

Tabel 5.12 *Output* Hasil Pengujian Hipotesis Linier Gabungan dengan *SAS*

```
                         The SAS System

                  L Ginv(X'X) L'                          Lb-c

  2.8559225951     0.0006173356     -0.04090973      -0.795271027
  0.0006173356     3.3414119E-6      1.4099083E-8     0.0013058159
 -0.04090973       1.4099083E-8      0.002613921     -0.027843496

                Inv(L Ginv(X'X) L')                     Inv()(Lb-c)

  0.4758460504    -87.94537601      7.4478046495     -0.700639963
-87.94537601       315528.6684     -1378.109811      520.33414234
  7.4478046495    -1378.109811      499.1379201      -21.62032556

Dependent Variable: Y
Test:          Numerator:      0.6129   DF:    3   F value:   1.7561
               Denominator:  0.348994   DF:   26   Prob>F:    0.1803
```

Contoh pengerjaan 5.7:

Data pada Tabel 5.13, terdiri dari 26 subyek yang dipilih untuk mempelajari pengaruh aktivitas latihan (lari dan angkat berat),dan berat badan terhadap kadar kolestrol HDL. Subyek terdiri dari 8 orang ditempatkan sebagai kelompok kontrol, 8 orang sebagai kelompok perlakuan latihan lari intensif, dan 10 orang sebagai kelompok latihan lari dikombinasi dengan angkat berat. Jika model regresi linier diterapkan pada setiap kelompok, akan terdapat tiga persamaan regresi linier sederhana. Lakukan pengujian terhadap kesaman tiga regresi tersebut, melalui pengujian kesamaan koefisien regresi.

Tabel 5.13 Berat Badan dan Kadar Kolestrol Menurut Metode Latihan

	Kelompok	Berat (lb)	Kolestrol HDL (mg/decaliter)
Kontrol	0	163.5	75.0
	0	180.0	72.5
	0	178.5	62.0
	0	161.5	60.0
	0	127.0	53.0
	0	161.0	53.0
	0	165.0	65.0
	0	144.0	63.5
Lari	1	141.0	49.0
	1	162.0	53.5
	1	134.0	30.0
	1	121.0	40.5
	1	145.0	51.5
	1	106.0	57.5
	1	134.0	49.0
	1	216.5	74.0
Lari dan Angkat berat	2	136.5	54.5
	2	142.5	79.5
	2	145.0	64.0
	2	165.0	69.0
	2	226.0	50.5
	2	122.0	58.0
	2	193.0	63.5
	2	163.5	76.0
	2	154.0	55.5
	2	139.0	68.0

Sumber : Contoh data analisis regresi dari Myers,1990.

Pemecahan pengerjaan 5.7:
Tiga model persamaan regresi linier dapat disusun sebagai berikut:

$Y_j = \beta_{01} + \beta_{1.1} X_{1j} + \varepsilon_j$ $\quad j = 1, 2, \ldots 8$ (kelp. Kontrol)
$Y_j = \beta_{02} + \beta_{1.2} X_{1j} + \varepsilon_j$ $\quad j = 9, 10, \ldots 16$ (kelp. Lari)
$Y_j = \beta_{03} + \beta_{1.3} X_{1j} + \varepsilon_j$ $\quad j = 17, 18, \ldots 26$ (kelp. Lari dan berat)

Dengan hipotesis uji
Ho : $\beta_{1.1} = \beta_{1.2} = \beta_{1.3}$
H_1 : Terdapat paling kecil sepasang $\beta_{1.k} \neq \beta_{1.s}$

Atau dalam bentuk hipotesis linier gabungan

H_0 : $\beta_{1.1} - \beta_{1.2} = 0$
$\beta_{1.1} - \beta_{1.3} = 0$

Pertama-tama data disusun sebagai model gabungan berikut

$$
\begin{array}{cccccc}
C1 & C2 & C3 & X1 & X2 & X3
\end{array}
$$

$$
\begin{bmatrix}
1 & 0 & 0 & 163.5 & 0.0 & 0.0 \\
1 & 0 & 0 & 180.0 & 0.0 & 0.0 \\
1 & 0 & 0 & 178.5 & 0.0 & 0.0 \\
1 & 0 & 0 & 161.5 & 0.0 & 0.0 \\
1 & 0 & 0 & 127.0 & 0.0 & 0.0 \\
1 & 0 & 0 & 161.0 & 0.0 & 0.0 \\
1 & 0 & 0 & 165.0 & 0.0 & 0.0 \\
1 & 0 & 0 & 144.0 & 0.0 & 0.0 \\
0 & 1 & 0 & 0.0 & 141.0 & 0.0 \\
0 & 1 & 0 & 0.0 & 162.0 & 0.0 \\
0 & 1 & 0 & 0.0 & 134.0 & 0.0 \\
0 & 1 & 0 & 0.0 & 121.0 & 0.0 \\
0 & 1 & 0 & 0.0 & 145.0 & 0.0 \\
0 & 1 & 0 & 0.0 & 106.0 & 0.0 \\
0 & 1 & 0 & 0.0 & 134.0 & 0.0 \\
0 & 1 & 0 & 0.0 & 216.5 & 0.0
\end{bmatrix}
$$

$$
\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ . \\ . \\ . \\ Y_{16} \end{pmatrix} = \begin{vmatrix} 0 & 0 & 1 & 0.0 & 0.0 & 136.5 \\ 0 & 0 & 1 & 0.0 & 0.0 & 142.5 \\ 0 & 0 & 1 & 0.0 & 0.0 & 145.0 \\ 0 & 0 & 1 & 0.0 & 0.0 & 165.0 \\ 0 & 0 & 1 & 0.0 & 0.0 & 226.0 \\ 0 & 0 & 1 & 0.0 & 0.0 & 122.0 \\ 0 & 0 & 1 & 0.0 & 0.0 & 193.0 \\ 0 & 0 & 1 & 0.0 & 0.0 & 163.5 \\ 0 & 0 & 1 & 0.0 & 0.0 & 154.0 \\ 0 & 0 & 1 & 0.0 & 0.0 & 139.0 \end{vmatrix} \begin{pmatrix} \beta_{01} \\ \beta_{02} \\ \beta_{03} \\ \beta_{1.1} \\ \beta_{1.2} \\ \beta_{1.3} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ . \\ . \\ . \\ \varepsilon_{16} \end{pmatrix}
$$

Pendugaan terhadap koefisien regresi menggunakan $\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}(\mathbf{X'}\boldsymbol{Y})$

$\hat{\boldsymbol{\beta}} = (23.054,\ 14.255,\ 76.880,\ 0.2495,\ 0.250,\ -0.082)$

Hipotesis linier gabungan di atas dapat ditulis dalam bentuk $\mathbf{K'}\boldsymbol{\beta} = \mathbf{m}$, dengan mendefinisikan

$$
\mathbf{K'} = \begin{pmatrix} 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \end{pmatrix} \text{ dan } \mathbf{m} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}
$$

$$
Q = (\mathbf{K'}\hat{\boldsymbol{\beta}})' (\mathbf{K'} (\mathbf{X'X})^{-1} \mathbf{K})^{-1} (\mathbf{K'}\hat{\boldsymbol{\beta}}) \text{ dan } F = \frac{Q/k}{S^2}
$$

Output menggunakan perangkat lunak *Statistical Analysis System* (*SAS*) menampilkan hasil berikut, pada Tabel 5.14.

Tabel 5.14 *Output* Hasil Pengujian Hipotesis Linier Gabungan dengan *SAS*

The SAS System

General Linear Models Procedure

Dependent Variable: Y

Source	DF	Sum of Squares	F Value	Pr > F
Model	6	93701.3912362	182.40	0.0001
Error	20	1712.3587638		
Uncorrected Total	26	95413.7500000		

R-Square	C.V.	Y Mean
0.982053	15.54623	59.5192308

NOTE: No intercept term is used: R-square is not corrected for the mean.

Source	DF	Type I SS	F Value	Pr > F
C1	1	31752.0000000	370.86	0.0001
C2	1	20503.1250000	239.47	0.0001
C3	1	40768.2250000	476.16	0.0001
X1	1	132.5179843	1.55	0.2278
X2	1	488.3993696	5.70	0.0269
X3	1	57.1238823	0.67	0.4237

Parameter	Estimate	T for H0: Parameter=0	Pr > |T|	Std Error of Estimate
C1	23.05431286	0.71	0.4833	32.27434845
C2	14.25500103	0.92	0.3710	15.57527400
C3	76.88001895	4.74	0.0001	16.21826497
X1	0.24956306	1.24	0.2278	0.20059739
X2	0.25093574	2.39	0.0269	0.10506483
X3	-0.08213060	-0.82	0.4237	0.10054917

L Ginv(X'X) L'		Lb-c
0.0005989157	0.0004699869	-0.001372682
0.0004699869	0.0005880712	0.3316936544

Inv(L Ginv(X'X) L')		Inv()(Lb-c)
4478.2428763	-3579.014663	-1193.283656
-3579.014663	4560.8251109	1517.7095972

Dependent Variable: Y

Jika digunakan taraf signifikan α = 0,05, maka pada Tabel 5.14 , diperoleh Pr > α yaitu 0,0754 > 0,05 , sehingga belum cukup bukti untuk menolak Ho. Disimpulkan bahwa tiga koefisien regressor sama.

Cara pengujian kesamaan koefisien di atas dapat pula dilakukan melalui **reduksi model.**, yaitu dari **Model penuh**

$Y_j = \beta_{01} C_{1j} + \beta_{02} C_{2j} + \beta_{03} C_{3j} + \beta_{1.1} X_{1j} + \beta_{1.2} X_{2j} + \beta_{1.3} X_{3j} + \varepsilon_j$

Direduksi menjadi **Model tereduksi**

$Y_j = \beta_{01} C_{1j} + \beta_{02} C_{2j} + \beta_{03} C_{3j} + \beta_1 X_j + \varepsilon_j$

Sehingga Q = JK($G_{reduksi}$)- JK(G_{penuh}), yaitu selisih antara jumlah kuadrat galat dari model tereduksi dengan jumlah kuadrat galat dari model penuh.

Tabel 5.15 *Output* Hasil Analisis Variansi dari Model Reduksi dengan *SAS*

The SAS System
General Linear Models Procedure

Model : Y=C1 C2 C3 X
Dependent Variable: Y

Source	DF	Sum of Squares	F Value	Pr > F
Model	3	1090.32897921	3.61	0.0295
Error	22	2217.41140541		
Corrected Total	25	3307.74038462		

Analisis regresi dengan metode jumlah kuadrat terkecil pada Tabel 5.14, diperoleh jumlah kuadrat galat model penuh JK(G_{penuh}) = 1712.35876 dengan derajat bebas 26-6 = 20, dan pada Table 5.15, diperoleh jumlah kuadrat galat model tereduksi JK($G_{reduksi}$) = 2217.41141 dengan derajat bebas 26-4 = 22. Selisih kedua jumlah kuadrat galat adalah

JK($G_{reduksi}$)- JK(G_{penuh}) = 2217.41141- 1712.35876 = 252.526 = Q dengan derajat bebas 22-20= 2.

$$F = \frac{Q/k}{s^2} = \frac{252{,}526/2}{85{,}61794} = 2{,}9495 \quad \text{dan} \quad F_{(0,05;2;20)} = 3{,}49$$

$F < F_{(0,05;2;20)}$, yaitu 2,9495 < 3,49, sehingga belum cukup bukti untuk menolak Ho. Disimpulkan bahwa tiga koefisien regressor sama.
Pengujian kesamaan koefisien regresor juga dapat dilakukan melalui pengujian koefisien interaksi pada model regresi dengan variabel kategori. Cara ini akan dibahas pada contoh pengerjaan 5.8 berikut.

Contoh pengerjaan 5.8:
Hasil pengamatan terhadap 18 kali pencucian disajikan pada Tabel 5.16. Respon Y adalah jumlah tanah yang terlepas dari suatu sistem pencucian batubara. Catatan PH terhadap tangki pencucian sebagai variabel regressor kuantitatif X. Operasi pencucian menggunakan 3 jenis polymer, yaitu polymer 1, polymer 2, dan polymer 3. Ketiga jenis polymer ini sebagai variabel kategori (Z1, Z2, Z3). Analisis regresi linier dimodelkan dengan persamaan berikut.
$Y_j = \beta_0 + \beta_1 X_j + \beta_2 Z_{2j} + \beta_3 Z_{3j} + \varepsilon_j$

Tabel 5.16 Data Pencucian Batubara

PH	Jumlah tanah mg/L	Jenis Polymer
6.5	292	polymer 1
6.9	329	polymer 1
7.8	352	polymer 1
8.4	378	polymer 1
8.8	392	polymer 1
9.2	410	polimer 1
6.7	198	polymer 2
6.9	227	polymer 2
7.5	277	polymer 2
7.9	297	polymer 2
8.7	364	polymer 2
9.2	375	polymer 2
6.5	167	polymer 3
7.0	225	polymer 3

7.2	247	polymer 3
7.6	268	polymer 3
8.7	288	polymer 3
9.2	342	polymer 3

Sumber : Contoh analisis regresi oleh Myer, 1990

Akan tetapi pada model di atas, pengaruh PH (X) terhadap jumlah tanah yang terlepas (Y) diasumsikan sama pada setiap level polymer. Asumsi ini dapat saja tidak benar, akibatnya dapat menyesatkan kesimpulan pendugaan model atau hasil analisis. Untuk menyelidiki apakah pengaruh PH sama disetiap penggunaan jenis polymer, dapat dilakukan dengan menyelidiki pengaruh interaksi antara variabel PH dengan Polimer, menggunakan model

$Y_j = \beta_0 + \beta_1 X_j + \beta_2 Z_{1j} + \beta_3 Z_{2j} + \delta_1 X_j Z_{1j} + \delta_2 X_j Z_{2j} + \varepsilon_j$

Lakukan uji apakah pengaruh PH sama disetiap penggunaan jenis polymer, dengan
hipotesis uji berikut

Ho : $\delta_1 = \delta_2 = 0$
H_1 : Ada $\delta_i \neq 0$, i=1,2.

Pemecahan pengerjaan 5.8:
Analisis dengan model regresi linier memuat variabel kategori dan suku interaksi antar variabel pada contoh pengerjaan ini dilakukan dengan *SPSS*. Hasil analisis dengan *SPSS* ditampilkan pada Tabel 5.17.

Tabel 5.17 Hasil Analisi Regresi Linier untuk Model yang Memuat Variabel Kategori dan Suku Interaksi dengan *SPSS*

Tests of Between-Subjects Effects

Dependent Variable: Y

Source	Type I Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	82707.777[a]	5	16541.555	77.761	.000
Intercept	1636843.556	1	1636843.556	7694.743	.000
PH	54855.581	1	54855.581	257.874	.000
Z1	23117.846	1	23117.846	108.676	.000
Z2	2208.304	1	2208.304	10.381	.007
PH * Z1	1747.093	1	1747.093	8.213	.014
PH * Z2	778.953	1	778.953	3.662	.080
Error	2552.668	12	212.722		
Total	1722104.000	18			
Corrected Total	85260.444	17			

a. R Squared = .970 (Adjusted R Squared = .958)

Jumlah kuadrat dapat dihitung menggunakan jumlah kuadrat sekuensial, yaitu

JK($\delta_1, \delta_2 \mid \beta_0, \beta_1, \beta_2\, \beta_3$)

= JK($\delta_1 \mid \beta_0, \beta_1, \beta_2\, \beta_3$)+ JK($\delta_4 \mid \beta_0, \beta_1, \beta_2, \beta_3\ \delta_1$)

= 1747,093 + 778,953

= 2526,045

$$F = \frac{JK(\delta_1, \delta_2 \mid \beta_0\ , \beta_1, \beta_2, \beta_3\ )/2}{RJK(Sisa)} = \frac{2526{,}045/2}{212{,}722} = 5{,}94$$

$F > F_{(0,05;2;12)}$, yaitu 5,95 > 3,89, sehingga menolak Ho ditolak pada taraf signifikan 5 persen, koefisien regresi pada tiga jenis polymer tidak sama. Disimpulkan bahwa pengaruh PH terhadap jumlah tanah adalah tidak sama diantara tiga penggunaan jenis polymer. Oleh karena itu untuk menyelidiki pengaruh PH terhadap jumlah tanah yang terlepas, analisis dengan model regresi linier sederhana harus dilakukan secara terpisah pada tiga jenis polymer.

DAFTAR PUSTAKA

Kshirsagar. A. M., (1983), A Course in Linear Models, New York and Basel, Marcel Dekker Inc.

MathSoft, Inc., (1993), *S-Plus User's Manual*, Seattle, Washington, MathSoft, Inc.

Myers, H.M., (1990), *Classical and Modern Regression With Applications*, 2 edition, Boston, Pws-Kent Publishing Company.

Rawlings, J.O., (1988), *Applied Regression Analysis*, California, Wadsworth and Brooks

SAS Institute,Inc ,(1985f), *SAS User's Guide : Basics*, Version 5 edition, Cary, North Carolina : SAS Institute, Inc.

Searle, S.R., (1971), *Linear Models*, New York, John Wiley and Sons, Inc.

LATIHAN UNTUK BAB 5

1. Data dibawah ini adalah hasil ujian dua matakuliah, yaitu Statistika dan Ekonomi. Ingin diselidiki apakah rata-rata hasil ujian Ekonomi (Y) dapat dipengaruhi secara linier oleh pencapaian hasil ujian Statistika (X). Dengan model regresi linier ,

   a. Dugalah persamaan regresi liniernya
   b. Lakukan pengujian terhadap keberartian model dugaan melalui uji F.
   c. Hitung koefisien determinasi R^2
   d. Lakukan pengujian terhadap koefisien $\beta_1 = 0$

Tabel 5.18 Data Hasil Ujian Matematika dan Fisika

Mhs	Nilai Statistika	Nilai Ekonomi
1.	65	62
2.	90	71
3.	52	58
4.	44	58
5.	95	64
6.	36	40
7.	48	42
8.	63	66
9.	80	67
10.	15	55

<u>Pemecahan (Hitung Manual):</u>

Σ X= ……. ΣX^2 = ……….. ΣYX =…………
Σ Y= ……. ΣY^2 = ………..

Rata-rata : $\bar{X} = \dfrac{\sum_{j=1}^{10} Xj}{10} = \dfrac{\ldots\ldots\ldots}{10} = \ldots\ldots\ldots$

$\bar{Y} = \dfrac{\sum_{j=1}^{10} Y_j}{10} = \dfrac{\ldots\ldots\ldots}{10} = \ldots\ldots\ldots$

a. Perhitungan koefisien dugaan $\hat{\beta}_1$ dan $\hat{\beta}_0$:

$\hat{\beta}_1 = \sum(X_j - \bar{\mathbf{X}})(Y_j - \bar{\mathbf{Y}}) / \sum (X_j - \bar{\mathbf{X}})^2$

$\hat{\beta}_1 = [\sum X_j Y_j - (\sum X_j)(\sum Y_j) / n] / [\sum X_j^2 - (\sum X_j)^2 / n]$

$= [\ldots\ldots\ldots\ldots - (\ldots\ldots)(\ldots\ldots) / \ldots] / [\ldots\ldots\ldots - (\ldots\ldots)^2 / \ldots]$

$= \ldots\ldots\ldots$

$\hat{\beta}_0 = \bar{\mathbf{Y}} - \hat{\beta}_1 \bar{\mathbf{X}}$

$= \ldots\ldots\ldots - \ldots\ldots\ldots (\ldots\ldots\ldots)$

$= \ldots\ldots\ldots$

Persamaan regresi dugaan adalah $\hat{Y}_j = \ldots\ldots + \ldots\ldots\ldots X_j$

b. Hipotesis :

H_0 : $\beta_1 = 0$.

H_1 : $\beta_1 \neq 0$

Perhitungan jumlah kuadrat :

JK(Total) $= \sum Y_j^2 - (\sum Y_j)^2 / n$

$= (\ldots\ldots\ldots)^2 - (\ldots\ldots\ldots)^2 / \ldots$

$= \ldots\ldots\ldots$

JK(Reg) $= \hat{\beta}_1^2 [\sum X_i^2 - (\sum X_i)^2 / n]$

$= (\ldots\ldots\ldots)^2 (\ldots\ldots\ldots - (\ldots\ldots\ldots)^2 / \ldots)$

$= \ldots\ldots\ldots$

RJK(Reg) $= \ldots\ldots\ldots / 1$

$= \ldots\ldots\ldots$

JK(Sisa) = JK(Total) – JK(Reg)

= ………… – …………..
= …………..
RJK(Sisa) = ……….. /(……. – 2)
= …………

F_{hit} = RJK(Reg) / RJK(Sisa)
= ……….. / ………. = …………

Pada tabel sebaran F diperoleh : $F_{(0.05;1;8)}$-tabel = 5,32

Tabel 5.19 Tabel Analisi Ragam

Sumber Variasi	**Jumlah Kuadrat (JK)**	**Derajat Bebas (Db)**	**Rata-rata Jum. Kuad. (RJK)**	**F-hitung**
Regresi Sisaan	………… …………	1 8	………….. …………..	F= …………
Total	………..	9		

Hasil pengujian : F-hitung …….. $F_{(0.05;1;8))}$-tabel, yaitu ……. . … , selanjutnya H_0 ………….pada taraf signifikan 5 persen. Disimpulkan bahwa ………

c. Koefisien determinasi :

R^2 = JKReg / JKTotal
= ………… / …………
= …………

Kesimpulan : ………..% variasi dari hasil ujian ekonomi dapat dijelaskan oleh hubungan liniernya terhadap hasil ujian statistika.

d. Hipotesis :

H_0 : $\beta_1 = 0$
H_1 : $\beta_1 \neq 0$

Ragam : $S_{\hat{\beta}_1} = \sqrt{\text{RJKS} / (\sum X_j^2 - (\sum X_j)^2 / n)}$

$= \sqrt{\ldots\ldots\ldots\ldots/[\ldots\ldots\ldots\ldots - (\ldots\ldots\ldots\ldots)^2 \quad / \ldots\ldots]}$

= ………..

t-hitung $= (\hat{\beta}_1 - 0) / S_{\hat{\beta}_1}$

$= \ldots\ldots\ldots / \ldots\ldots\ldots\ldots = \ldots\ldots\ldots\ldots$

Pada tabel t diperoleh : $t_{0.05/2;8} = 2{,}306$

Hasil pengujian : $|$ t-hitung $|$ ..... $t_{\alpha/2;Db}$-tabel, yaitu ………. 2,306, selanjutnya H_0 ………. pada taraf signifikan 5 persen. Disimpulkan bahwa……………….

2. Data tentang ukuran beberapa organ tubuh disajikan pada Tabel 5.20. Dengan model regresi linier, ingin selidiki pengaruh sekelompok variabel bebas, yaitu ukuran dari paru-paru (X_1), jantung (X_2), liver (X_3), limpa (X_4) terhadap ukuran Ginjal (Y). Pada Tabel 5.21 dan Tabel 5.22 berturut-turut disajikan hasil dari matriks ($\mathbf{X'X}$)$^{-1}$, dan *output* hasil analisis regresi menggunakan S-Plus. Pada Tabel 5.23 disajikan *output* hasil pengujian hipotesis linier gabungan menggunakan SAS.

   a. Tuliskan persamaan regresi dugaannya
   b. Ujilah keberartian dari model regresi melalui pengujian rumusan hipotesis uji berikut.

      $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$
      $H_1 : \beta_j \neq 0$,Untuk paling kecil satu dari j. j = 1, 2, 3,4
      Hitung jumlah kuadrat regresi menggunakan menggunakan jumlah kuadrat sekuensial pada Tabel 5.22, yaitu

      $$\text{JK(Reg)} = \text{JK}(\beta_1 \mid \beta_0) + \text{JK}(\beta_2 \mid \beta_0, \beta_1) + \text{JK}(\beta_3 \mid \beta_0, \beta_1, \beta_2) + \text{JK}(\beta_4 \mid \beta_0, \beta_1, \beta_2, \beta_3)$$

   c. Lakukan pengujian parsial bagi setiap koefisien regresi dengan menggunakan *output* komputer pada Tabel 5.22, yaitu menguji hipotesis berikut.
      $H_0 : \beta_j = 0.$
      $H_1 : \beta_j \neq 0$, j = 1, 2, 3, 4

d. Lakukan pengujian sekuensial terhadap urutan variabel bebas X_1, X_2, X_3 , dan X_4 dengan menggunakan *output* komputer pada Tabel 5.22.

Tabel 5.20 Data Hasil Pengukuran Organ Tubuh : Paru-paru (*lung*), Jantung (*heart*),Liver (*liver*), Limpa (*spleen*) , dan Ginjal (*kidney*).

	lung	heart	liver	spleen	kidney
1	0.9695	1.3520	8.1855	0.2255	1.3480
2	1.1410	1.4994	8.5445	0.2705	1.3221
3	0.8720	1.2685	6.8230	0.0500	0.8970
4	0.7399	1.2702	9.0675	0.1130	1.2780
5	1.1393	1.4962	9.6220	0.1198	1.4150
6	1.1256	1.1914	9.7432	0.1393	1.3405
7	0.3672	0.6986	2.7266	0.0478	0.8199
8	0.3227	0.4193	1.7442	0.0448	0.8224
9	0.4972	0.4134	2.6617	0.0568	1.0697
10	0.5366	0.5298	3.2621	0.0751	0.9601
11	0.5219	0.6020	3.6231	0.0701	1.1635
12	0.4291	0.4230	2.0776	0.0383	0.8199
13	0.5354	0.5129	2.8608	0.0745	0.9825
14	0.4529	0.5715	2.8093	0.0559	1.0074
15	0.6532	0.5187	3.6025	0.0604	1.0751
16	0.9191	0.5479	2.8636	0.0687	0.9596
17	0.5620	0.5001	2.9257	0.0780	0.9589
18	0.5362	0.9423	2.6974	0.0543	1.0797
19	0.5516	0.7950	2.7466	0.0657	1.0101
20	0.5789	0.5051	3.1271	0.0755	1.0798
21	0.6553	0.5999	3.4968	0.0436	1.1976
22	0.5613	1.0344	3.1108	0.0668	1.2131
23	0.5370	0.7446	2.6780	0.0450	1.0426
24	0.6220	0.9023	8.3071	0.0997	1.3686
25	0.7738	0.7628	3.7972	0.0685	1.0629
26	1.0500	0.8981	4.8977	0.0821	1.0119
27	0.8416	1.5094	7.4928	0.1739	1.3171
28	0.8739	0.6723	2.6366	0.0536	1.0983
29	0.8260	0.9604	12.8040	0.1193	1.5933

30	0.7589	0.9483	5.3909	0.0828	1.1872
31	1.1778	1.3692	8.1259	0.1376	1.3538
32	1.0119	1.2080	7.1872	0.0784	1.4338
33	0.9820	1.2860	7.9245	0.1194	1.6396
34	0.6657	1.0690	4.4023	0.0752	1.2131
35	0.6190	1.0002	8.0259	0.1056	1.5638
36	0.7849	1.1167	5.9199	0.0952	1.2582
37	0.5092	0.8698	3.2649	0.0876	1.3482
38	0.5687	0.8791	3.6023	0.1796	1.1772
39	1.4364	0.9932	8.2075	0.1187	1.3165
40	1.0725	0.9484	6.9736	0.1043	1.4693
41	1.1626	1.2746	11.9148	0.1072	1.4999
42	0.9650	1.5450	6.5425	0.1062	1.2677
43	1.0239	1.0789	8.1096	0.1424	1.4012
44	0.3216	0.4987	1.6264	0.0293	0.8252
45	1.1653	1.6300	7.0897	0.1444	1.2884
46	0.8532	1.1439	4.0120	0.0769	1.2271
47	0.6354	0.8940	2.6877	0.0463	0.9939
48	0.6489	1.2394	8.7761	0.1145	1.4262
49	0.9620	0.8039	3.4576	0.0713	1.1646
50	0.7959	1.0385	8.3850	0.1990	1.3135
51	0.8099	1.1543	7.5313	0.0595	1.3867
52	0.9824	1.4903	7.9353	0.1567	1.3642
53	0.9592	1.0950	15.4923	0.2959	1.4378
54	0.9596	1.2763	10.9988	0.1617	1.3394
55	1.1194	1.4415	5.9829	0.1308	1.1359
56	1.0748	1.1714	8.6132	0.1283	1.2379

Sumber : Contoh analisis regresi oleh Rawlings, 1988.

Tabel 5.21 *Output SAS* bagi matriks $(\mathbf{X'X})^{-1}$

X'X Inverse

	INTERCEP	X1	X2
INTERCEP	0.2143847334	-0.168027136	-0.095712472
X1	-0.168027136	0.6152818699	-0.225741783
X2	-0.095712472	-0.225741783	0.3972158583
X3	0.006655296	-0.015950671	-0.011878562
X4	-0.094818419	-0.038820519	-0.428510383

	X3	X4
INTERCEP	0.006655296	-0.094818419
X1	-0.015950671	-0.038820519
X2	-0.011878562	-0.428510383
X3	0.005261954	-0.131875573
X4	-0.131875573	12.971992938

e. Gunakan hasil *output SAS* pada Tabel 5.23, untuk melakukan pengujian kontribusi 3 regressor, yaitu paru-paru (X_1), liver (X_3), dan limpa (X_4) terhadap jumlah kuadrat regresi. Definisikan terlebih dahulu suatu hipotesis linier gabungan untuk pengujian hipotesis $H_0 : \beta_1 = \beta_3 = \beta_4 = 0$.

Tabel .22 Tampilan *Output* Komputerisasi bagi Hasil Analisis Regresi Linier

```
                         *** Linear Model ***

Call: lm(formula = kidney ~ lung + heart + liver + spleen,
      data = hams1, na.action = na.exclude)

Coefficients:
               Value Std. Error  t value Pr(>|t|)
(Intercept)   0.8344   0.0587    14.2107   0.0000
       lung   0.0116   0.0995     0.1167   0.9075
      heart   0.1295   0.0799     1.6200   0.1114
      liver   0.0436   0.0092     4.7429   0.0000
     spleen  -0.1900   0.4567    -0.4160   0.6792

Residual standard error: 0.1268 on 51 degrees of freedom
Multiple R-Squared: 0.6471
F-statistic:23.38 on 4 and 51 degrees of freedom,the p-value is 5.099e-
011


                       Analysis of Variance Table

Response: kidney
Terms added sequentially (first to last)

          Df Sum of Sq   Mean Sq  F Value      Pr(F)
     lung  1 0.7806807 0.7806807 48.54909 0.0000000
    heart  1 0.2770767 0.2770767 17.23089 0.0001258
    liver  1 0.4433743 0.4433743 27.57263 0.0000030
   spleen  1 0.0027822 0.0027822  0.17302 0.6791853
Residuals 51 0.8200918 0.0160802
```

f. Gunakan hasil *output* SAS pada Tabel 5.23 dan definisikan terlebih dahulu suatu hipotesis linier gabungan untuk melakukan hipotesis $H_0 : \beta_0 = 0$ dan $\beta_1 = \beta_3 = \beta_4 = 0$.

Tabel 5.23 *Output* Hasil Pengujian Hipotesis Linier Gabungan dengan *SAS*

```
                          The SAS System

                      L Ginv(X'X) L'    Lb-c

     0.6152818699      -0.015950671      -0.038820519      0.0116106488
     -0.015950671       0.005261954      -0.131875573      0.0436278372
     -0.038820519      -0.131875573      12.971992938      -0.189978157

                 Inv(L Ginv(X'X) L')     Inv()(Lb-c)

     1.8280450862      7.6199527166      0.0829364687      0.3379107288
     7.6199527166       286.7812496      2.9382726476      12.041910651
     0.0829364687      2.9382726476      0.1072083551      0.1087861813

Dependent Variable: Y
Test:          Numerator:      0.1695   DF:    3   F value:  10.5435
               Denominator:   0.01608   DF:   51   Prob>F:    0.0001
```

```
                             The SAS System

                      L Ginv(X'X) L'    Lb-c

             0.2143847334      -0.168027136       0.006655296
             -0.168027136      0.6152818699      -0.015950671
              0.006655296      -0.015950671       0.005261954
             -0.094818419      -0.038820519      -0.131875573

                      -0.094818419      -0.165634253
                      -0.038820519      0.0116106488
                      -0.131875573      0.0436278372
                      12.971992938      -0.189978157

                 Inv(L Ginv(X'X) L')     Inv()(Lb-c)

              5.980322392      1.5806712389      -2.091037155
             1.5806712389      2.2458355332      7.0672664108
             -2.091037155      7.0672664108      287.51238684
             0.0271855275      0.0901219311      2.9287671487

                      0.0271855275      -1.068585698
                      0.0901219311      0.0554706582
                      2.9287671487      12.415544755
                      0.1073319358      0.1039285726

Dependent Variable: Y
Test:          Numerator:      0.1749   DF:    4   F value:  10.8761
               Denominator:   0.01608   DF:   51   Prob>F:    0.0001
```

3. Pengamatan atas hasil reaksi kimia untuk berbagai variasi temperatur dicatat sebagai berikut :

Temperatur (Co) : 125 125 125 150 150 150 175 175 175 200 200 200 Reaksi (Y %) : 77 76 78 84 84 83 88 88 89 94 94 95

a. Dugalah garis regresi liniernya melalui pendugaan koefisien β_0 , β_1.
b. Untuk melihat sejauhmana ketepatan model dugaan regresi linier dalam menjelaskan hubungan antara kedua variabel, dapat dilihat dari besarnya koefisien determinasi (R^2). Koefisien determinasi (R^2) memiliki nilai antara 0 sampai 1. Jika nilai R^2 semakin dekat 1 ,maka model regresi dugaan semakin dapat menjelaskan hubungan kedua variabel. *Hitunglah koefisien determinasi (R^2) dari regresi dugaan pada bagian (a).*
c. Besar perubahan variabel respon (Y) sebagai akibat dari perubahan satu unit variabel penjelas (X), dapat dilihat dari besarnya koefisien β_1. *Ujilah hipotesis berikut untuk menyimpulkan apakah* β_1 *sama dengan nol atau tidak*
   $H_0 : \beta_1 = 0$
   $H_1 : \beta_1 \neq 0$.

4. Data pada Table 5.24 memuat satu variabel terikat Y, dan 3 variabel bebas. Pada model regresi linier $Y_j = \beta_0 + \beta_1 X_j + \beta_2 X_{2j} + \beta_3 X_{3j} + \varepsilon_j$. Dengan terlebih dahulu merumuskan hipotesis uji, lakukan pengujian terhadap keberartian model regresi dan uji parsial terhadap masing-masing koefisien.

Tabel 5.24 Data dengan Tiga Variabel Bebas

No.	y	x1	x2	x3
1	17.6	8.8	83.1	158.2
2	10.9	8.5	24.2	96.2
3	9.2	7.7	4.5	31.8
4	16.2	4.9	9.1	95.0
5	10.1	9.6	158.2	407.2
6	11.7	10.0	132.2	404.6

7	17.9	12.5	501.5	1180.6
8	21.1	14.6	904.0	1807.5
9	14.7	11.2	227.6	470.0
10	7.7	10.7	66.6	151.4
11	8.4	10.0	43.4	93.8
12	32.8	12.8	1253.0	3293.4

5. Pada Tabel 5.25 disajikan jumlah kuadrat sekuensial (type I) menggunakan komputerisasi dengan SPSS terhadap data pada pengejaan 5.3 , di Table 5.8. Lakukan pengujian sekuensial terhadap Demikian juga dengan mereduksi model regresi menjadi $Y_j = \beta_0 + \beta_1 X_j + \beta_2 X_{2j} + \beta_3 X_{3j} + \varepsilon_j$., gunakan jumlah kuadrat dari model tereduksi tersebut untuk menguji hipotesis $H_0 : \beta_4 = \beta_5 = 0$.

Tabel 5.25 Hasil Analisis Regresi Linier dengan *SPSS* yang menyajikan Jumlah Kuadrat Sekuensial.

Tests of Between-Subjects Effects

Dependent Variable: Y

Source	Type I Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	208.007[a]	5	41.601	84.070	.000
Intercept	387.157	1	387.157	782.379	.000
X1	199.145	1	199.145	402.440	.000
X2	.127	1	.127	.256	.620
X3	4.120	1	4.120	8.325	.011
X4	.263	1	.263	.532	.476
X5	4.352	1	4.352	8.795	.009
Error	7.918	16	.495		
Total	603.081	22			
Corrected Total	215.925	21			

a. R Squared = .963 (Adjusted R Squared = .952)

Tabel 5.26 Data Hasil Pembelajaran Menggunakan Model Pembelajaran dan Struktur Tugas sebagai Variabel Kategori.

Model pembelajaran	Struktur Tugas					
	I		II		III	
	Kesbel	Hasbel	Kesbel	Hasbel	Kesbel	Hasbel
Inquiry	2.5	51.0	3.0	65.0	2.2	54.0
	2.2	55.0	2.8	52.0	1.8	59.0
	3.1	45.0	2.8	41.0	1.6	66.0
	4.3	42.0	2.7	51.0	2.1	54.0
	2.5	53.0	2.6	41.0	3.3	45.0
	4.3	50.0	2.8	45.0	3.8	49.0
	3.8	50.0	2.6	51.0	3.2	49.0
	4.3	52.0	2.6	45.0	3.6	55.0
	1.7	56.0	2.6	61.0	4.2	49.0
	3.1	49.0	3.5	42.0	1.6	68.0
Discovery	2.0	58.0	4.0	52.0	1.5	78.0
	2.4	55.0	2.8	70.0	1.4	75.0
	1.9	67.0	3.1	57.0	1.7	70.0
	2.8	61.0	4.2	58.0	1.3	84.0
	1.7	67.0	3.7	47.0	1.7	71.0
	3.2	68.0	3.0	56.0	1.6	72.0
	2.0	58.0	2.2	72.0	1.4	62.0
	2.2	63.0	2.3	63.0	1.0	68.0
	2.2	56.0	3.8	54.0	1.5	66.0
	2.2	72.0	2.0	60.0	1.6	72.0

Sumber : Modifikasi contoh analisis regresi Oleh Rawlings, 1988

6. Data hasil pembelajaran menggunakan 2 model pembelajaran (*inquiry* dan *discovery learning*) dan 3 struktur tugas sebagai variabel kategori, disajikan pada Tabel 5.26. Penyelidikan pengaruh kesulitan belajar X terhadap hasil belajar Y menggunakan model

   $Y_j = \beta_0 + \beta_1 X_j + \beta_2 W_{1j} + \beta_3 Z_{1j} + \beta_4 Z_{2j} + \beta_5 W_{1j} Z_{1j} + \beta_6 W_{1j} Z_{2j} + \lambda_1 X_j W_{1j} + \delta_1 X_j Z_{1j} + \delta_2 X_j Z_{2j} + \gamma_1 X_j W_{1j} Z_{1j} + \gamma_2 X_j W_{1j} Z_{2j}\ \varepsilon_j.$

Dengan kategori pembelajaran W1(*inquiry*) dan W2(*discovery*), sedangkan kategori struktur tugas Z1(struktur I), Z2(struktur II), dan Z3(struktur III). Untuk menyelidiki pengaruh apakah pengaruh kesulitan belajar adalah tidak berbeda diantara 6 kombinasi metode pembelajaran dengan struktur tugas, dirumuskan hipotesis uji

$H_0: \lambda_1 = \delta_1 = \delta_2 = \gamma_1 = \gamma_2 = 0$

Lakukan pengujian hipotesis ini dengan 2 cara, yaitu:

a. Lakukan reduksi bagi 5 suku interaksi dengan X untuk memperoleh jumlah kuadrat regresi tereduksi dalam pengujian hipotesis di atas.
b. Gunakan jumlah kuadrat sekuensial dari hasil analisis regresi dengan SPSS pada Tabel 5.27 untuk melakukan pengujian hipotesis di atas.

Tabel 5.27 Hasil Analsis Regresi Linier dengan *SPSS* yang Menyajikan Jumlah Kuadrat Sekuensial

Tests of Between-Subjects Effects

Dependent Variable: Y

Source	Type I Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4193.614[a]	11	381.238	9.906	.000
Intercept	201492.150	1	201492.150	5235.726	.000
X	2630.533	1	2630.533	68.354	.000
W	1145.096	1	1145.096	29.755	.000
Z1	21.122	1	21.122	.549	.462
Z2	238.312	1	238.312	6.192	.016
W * Z1	30.436	1	30.436	.791	.378
W * Z2	.299	1	.299	.008	.930
X * W	.172	1	.172	.004	.947
X * Z1	95.313	1	95.313	2.477	.122
X * Z2	7.549	1	7.549	.196	.660
X * W * Z1	7.466	1	7.466	.194	.662
X * W * Z2	17.316	1	17.316	.450	.506
Error	1847.236	48	38.484		
Total	207533.000	60			
Corrected Total	6040.850	59			

a. R Squared = .694 (Adjusted R Squared = .624)

PROFIL PENULIS