UKRIDA
Universitas Kristen Krida Wacana

Regresi Linier dan Korelasi

Dr. Budi Marpaung, ST., MT.

Being Transformed To Win The Future

Regresi (1)

Suatu persamaan yang mengekpresikan hubungan dua unsur penting dalam masalah statistik

Diperkenalkan pertama sekali oleh **Sir Francis Galton** (1822 – 1911), seorang antropolog dan pakar meteorologi terkenal dari Inggris.

Banyak digunakan dalam bidang rekayasa, ekonomi dan sosial, untuk memprediksi output berdasarkan perkiraan input tertentu

Regresi (2)

Dikenal dua variabel, yaitu variabel penjelas ***(explanatory variable)*** atau sering disebut variabel bebas ***(independent variable),*** dan variabel respon ***(response variable)*** atau sering juga disebut variabel tidak bebas/terikat ***(dependent variabel).***

Nilai variabel terikat ditentukan oleh nilai variabel bebas

Regresi sebagai alat perkiraan difokuskan pada persamaan garis regresi

Regresi sebagai alat untuk menjelaskan sistem, difokuskan pada upaya untuk memperoleh pemahaman yang komprehensif tentang sistem yang menjadi objek penelitian.

Persamaan Regresi Linier Sederhana

$$Y_i = \alpha + \beta X_i + \varepsilon_i \quad i = 1, 2, \ldots\ldots, n$$

$$\Rightarrow \quad \hat{y} = a + bx$$

$$b = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}, \text{ dan } a = \bar{y} - b\bar{x}$$

$$JKT = \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2; \quad JKR = \sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2; \quad JKG = \sum_{i=1}^{n}\left(y_i - \hat{y}\right)^2 = \sum_{i=1}^{n}\varepsilon_i^2$$

JKT = JKR + JKG

Standar Deviasi Galat, Koef. Determinasi & Korelasi

Standar Deviasi Galat

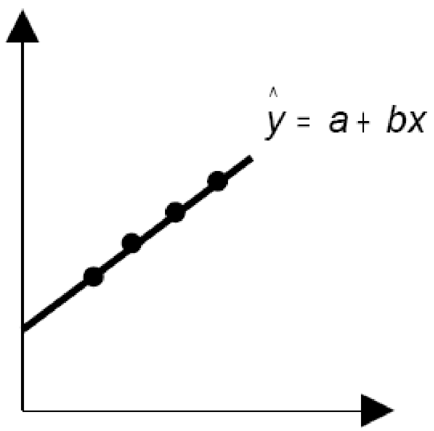
$$S_\varepsilon = \sqrt{\frac{JKG}{n-2}}$$

Koefisien Determinasi

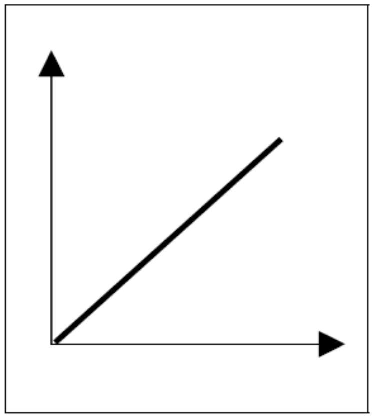
$$r^2 = \frac{JKR}{JKT} = 1 - \frac{JKG}{JKT}$$

Korelasi

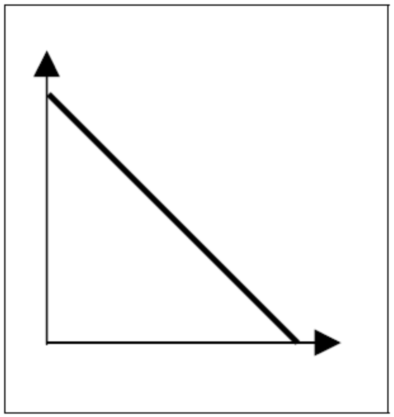
$$r = \sqrt{\frac{JKR}{JKT}} = \sqrt{1 - \frac{JKG}{JKT}}$$



Kurva Regresi untuk JKG = 0



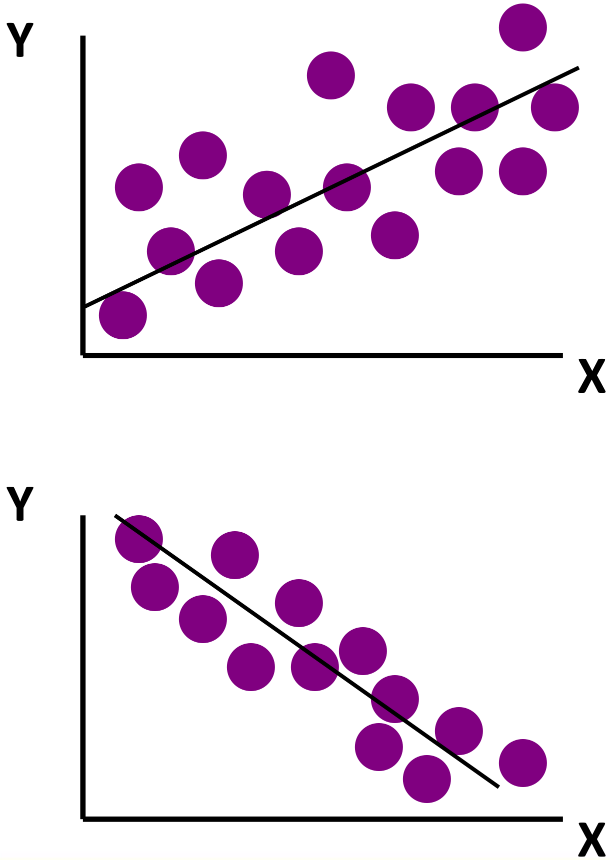
Korelasi Positif



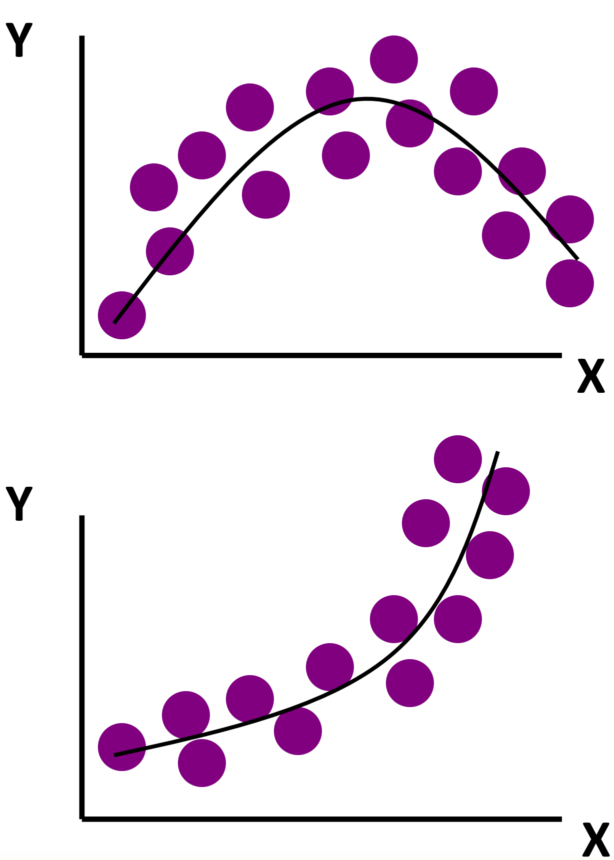
Korelasi Negatif

Korelasi

Hubungan Linier

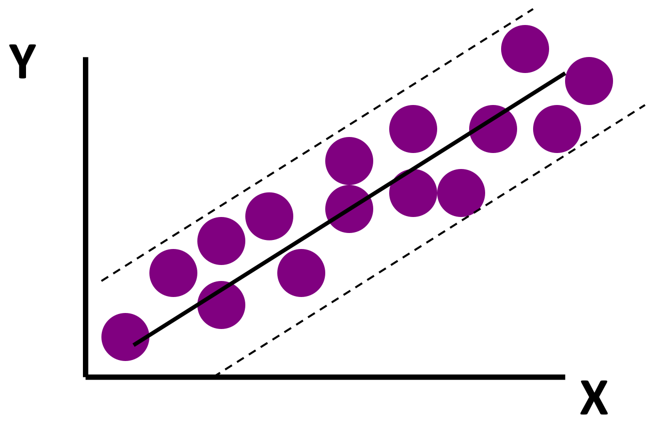


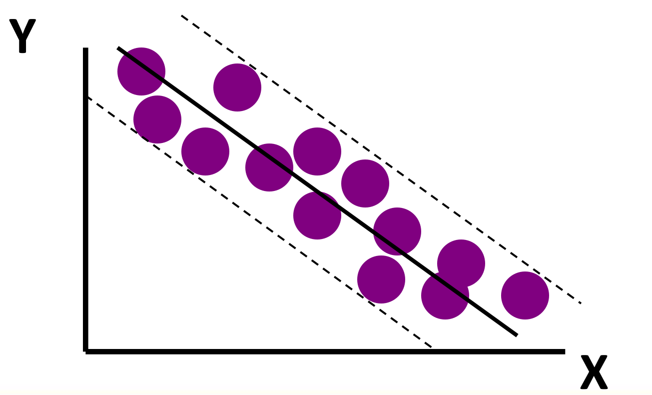
Hubungan Curvilinear



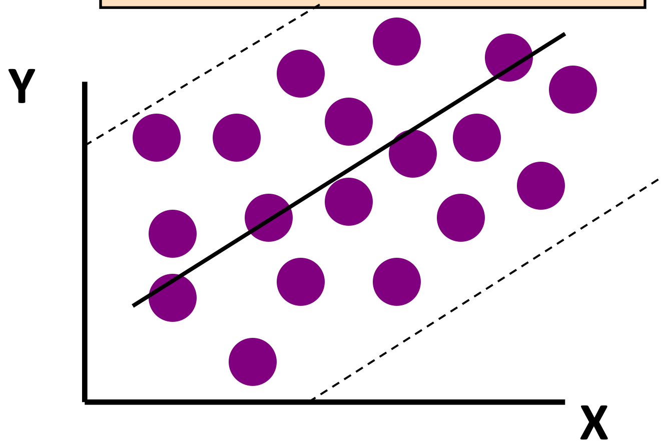
Jenis Hubungan

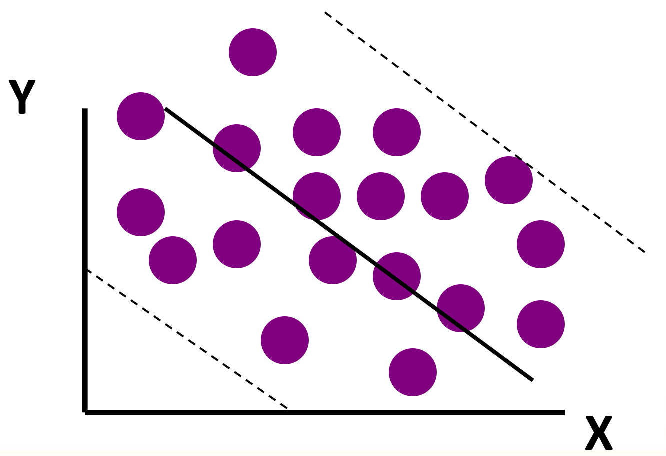
Hubungan kuat





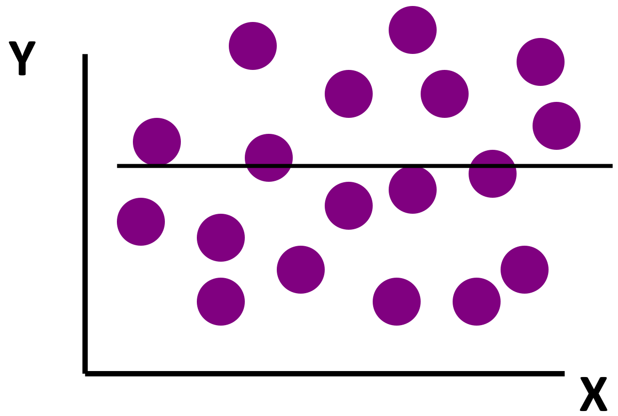
Hubungan lemah

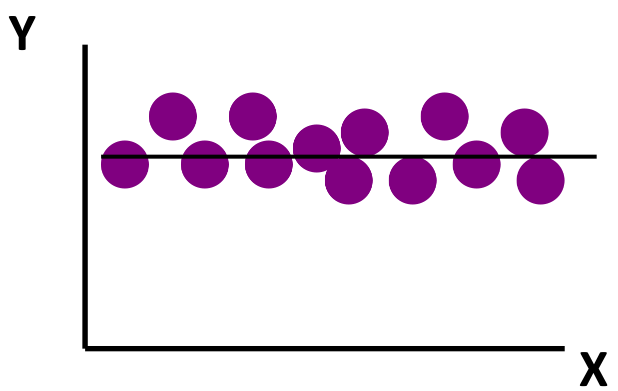




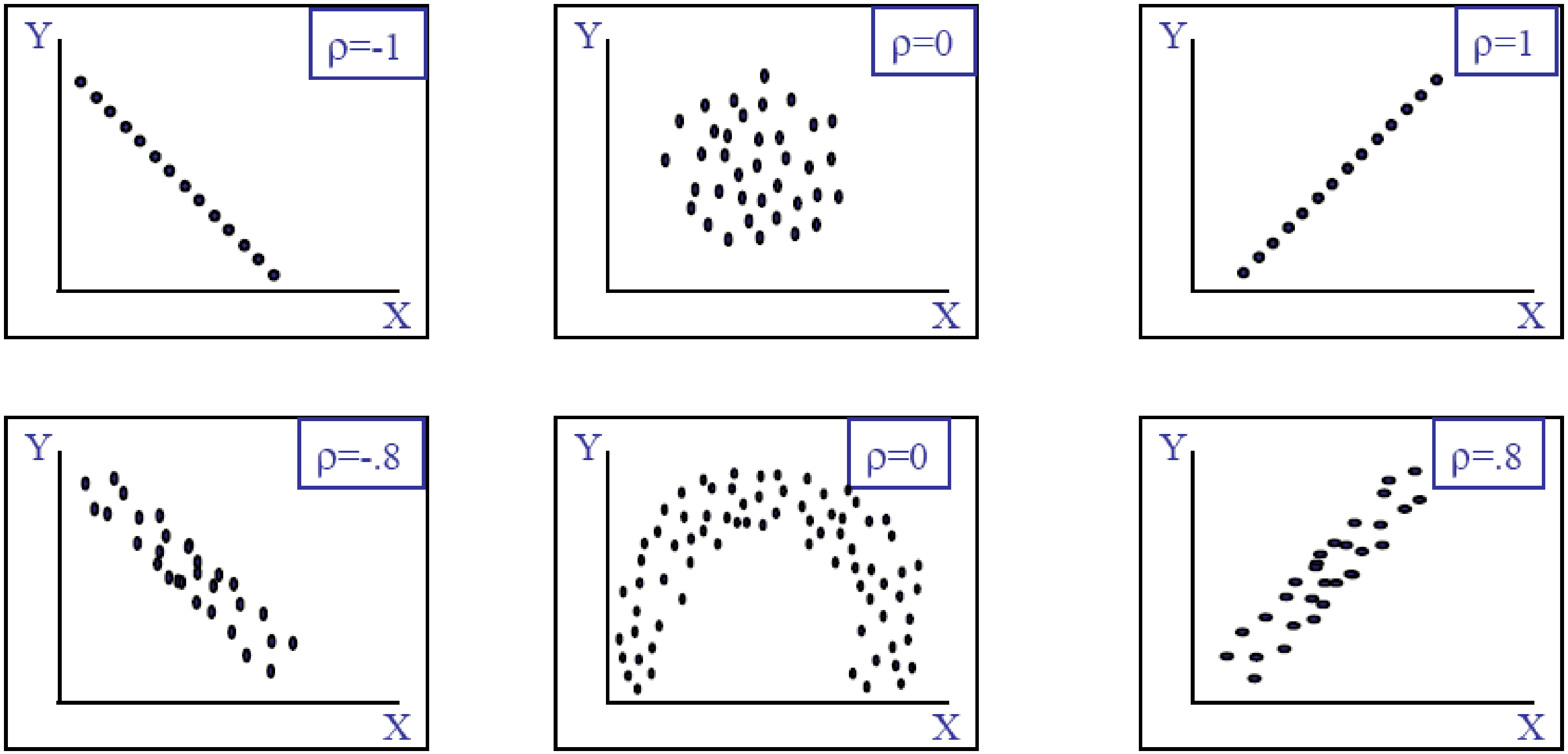
Jenis Hubungan

Tidak ada Hubungan





Konsep Korelasi



Nilai Korelasi

Nilai r berkisar antara -1 s/d +1

Interval Korelasi (r)	Tingkat Hubungan
0.000 – 0.1999	Sangat Rendah
0.200 – 0.3999	Rendah
0.400 – 0.5999	Sedang
0.600 – 0.7999	Kuat
0.800 – 1.0000	Sangat Kuat

Contoh Soal Persamaan R.L. Sederhana

Data berikut ini adalah data tentang banyaknya jam kerja (dalam jam) dan hasil produksi (dalam ton).

Tentukan persamaan regresi yang dapat digunakan untuk memperkirakan hasil produksi.

No	Jam Kerja	Hasil Produksi
1	78	18
2	57	15
3	75	17
4	84	21
5	67	16
6	73	17
7	69	15
8	71	14
9	82	15
10	74	12

Solusi Soal Persamaan R.L. Sederhana

Bulan	x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1	78	18	5	2	10	25	4
2	57	15	-16	-1	16	256	1
3	75	17	2	1	2	4	1
4	84	21	-11	5	55	121	25
5	67	16	-6	0	0	36	0
6	73	17	0	1	0	0	1
7	69	15	-4	-1	4	16	1
8	71	14	-2	-2	4	4	4
9	82	15	9	-1	-9	81	1
10	74	12	1	-4	-4	1	16
Jumlah	730	160	0	0	78	544	54
Mean	73	16					

$b = \dfrac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} = \dfrac{78}{544} = 0.1434$; dan $a = \bar{y} - b\bar{x} = 16 - (0.1434)(73) = 5.5331$.

$$\hat{y} = 5.5331 + 0.1434x$$

Solusi Soal Persamaan R.L. Sederhana

$$S_\varepsilon = \sqrt{\frac{JKG}{n-2}} = \sqrt{\frac{42{,}816}{10-2}} = 2{,}313$$

$$r^2 = \frac{JKR}{JKT} = \frac{11{,}187}{54{,}00} = 0{,}207 \text{, atau } r^2 = 1 - \frac{JKG}{JKT} = 1 - \frac{42{,}816}{54{,}00} = 0{,}207$$

Sedangkan koefisien korelasinya :

$$r = \sqrt{\frac{JKR}{JKT}} = \sqrt{\frac{11{,}187}{54{,}00}} = 0.455 \text{, atau } r = \sqrt{1 - \frac{JKG}{JKT}} = \sqrt{1 - \frac{42{,}816}{54{,}00}} = 0.455\ .$$

Tugas 10 Nomor 1 (20%)

PT. Rembulan adalah sebuah perusahaan yang memproduksi meja. Berikut ini adalah data banyaknya karyawan di bagian produksi perusahaan dan banyaknya meja yang dihasilkan sejumlah karyawan tersebut.

Jumlah Karyawan (orang)	56	59	42	38	51	46	37	44
Jumlah Meja (unit)	204	215	199	187	208	201	186	182

a) Tentukan persamaan regresi linier sederhana berdasarkan data tersebut.
b) Tentukan prediksi jumlah meja yang dihasilkan bila jumlah karyawan 60 orang.
c) Tentukan standar deviasi galat, dan jelaskan artinya.
d) Tentukan koefisien determinasi dan korelasi, lalu jelaskan artinya.

Tugas 10 Nomor 2 (20%)

Massa material yang hilang (y-dalam gram) untuk setiap durasi pengeringan (x-dalam jam), dinyatakan sebagai berikut.

X (jam)	Y (gram)	
4.4	13.1	14.2
4.5	9.0	11.5
4.8	10.4	11.5
5.5	13.8	14.8
5.7	12.7	15.1
5.9	9.9	12.7
6.3	13.8	16.5
6.9	16.4	15.7
7.5	17.6	16.9
7.8	18.3	17.2

a) Tentukan persamaan regresi linier sederhana yang menggambarkan hubungan durasi pengeringan dengan jumlah massa material yang hilang.
b) Tentukan prediksi jumlah material yang hilang bila dilakukan pengeringan selama 5.0 jam.
c) Tentukan standar deviasi galat, dan jelaskan artinya.
d) Tentukan koefisien determinasi dan korelasi, lalu jelaskan artinya.

Regresi Linier Ganda

$$\hat{Y}_i = a + b_{1i} X_1 + b_{2i} X_{2i} + \ldots\ldots\ldots\ldots + b_k X_{ki}$$

dimana : i = 1, 2, ..., n

$\hat{y}$ = nilai regresi/variabek terikat

X = variable bebas

$$na + b_1 \sum X_1 + b_2 \sum X_1 = \sum Y$$

$$a \sum X_1 + b_1 \sum X_1^2 + b_2 \sum X_1 X_2 = \sum X_1 Y$$

$$a \sum X_2 + b_1 \sum X_1 X_2 + b_2 \sum X_2^2 = \sum X_2 Y$$

Contoh Soal Persamaan R.L. Ganda

Dua belas orang mahasiswa ditanyakan tentang berat badan (Y-dalam kg), tinggi (X_1-dalam cm), dan umur mereka (X_2-dibulatkan dalam tahun).

Y	64	71	53	67	55	58	77	57	56	51	76	58
X_1	157	159	149	162	151	150	155	148	152	142	161	157
X_2	18	20	16	21	18	19	20	21	20	22	21	19

a) Tentukan persamaan regresi yang dapat digunakan untuk memperkirakan berat badan seorang mahasiswa.
b) Tentukan perkiraaan berat badan seorang mahasiswa yang memiliki tinggi 160 cm dan berumur 23 tahun.

Solusi Soal Persamaan R.L. Ganda

Y	X_1	X_2	X_1Y	X_2Y	X_1X_2	Y^2	X_1^2	X_2^2
64	157	18	10048	1152	2826	4096	24649	324
71	159	20	11289	1420	3180	5041	25281	400
53	149	16	7897	848	2384	2809	22201	256
67	162	21	10854	1407	3402	4489	26244	441
55	151	18	8305	990	2718	3025	22801	324
58	150	19	8700	1102	2850	3364	22500	361
77	155	20	11935	1540	3100	5929	24025	400
57	148	21	8436	1197	3108	3249	21904	441
56	152	20	8512	1120	3040	3136	23104	400
51	142	22	7242	1122	3124	2601	20164	484
76	161	21	12236	1596	3381	5776	25921	441
58	157	19	9106	1102	2983	3364	24649	361
$\sum Y$	$\sum X_1$	$\sum X_2$	$\sum X_1Y$	$\sum X_2Y$	$\sum X_2X_2$	$\sum Y^2$	$\sum X_1^2$	$\sum X_2^2$
=743	=1843	=235	=114560	=14596	=36096	=46879	=283443	=4633

Solusi Soal Persamaan R.L. Ganda

$na + b_1\sum X_1 + b_2\sum X_1 = \sum Y$ → $12a + 1843b_1 + 235b_2 = 743$

$a\sum X_1 + b_1\sum X_1^2 + b_2\sum X_1X_2 = \sum X_1Y$ → $1843a + 283443b_1 + 36096b_2 = 114560$

$a\sum X_2 + b_1\sum X_1X_2 + b_2\sum X_2^2 = \sum X_2Y$ → $235a + 36096b_1 + 4633b_2 = 14596$

$a = -104.1$; $b_1 = 1.16$; dan $b_2 = -0.59$

$$\hat{y} = -104.1 + 1.16x_1 - 0.59x_2$$

$$\hat{y}(X_1 = 160; X_2 = 23) = -104.1 + 1.16(160) - 0.59(23) = 67.93 \sim 68 \text{ kg}.$$

Korelasi Berganda

$$r_{x_1y} = \frac{\sum X_1Y}{\sqrt{\sum X_1^2}\sqrt{\sum Y^2}} \qquad r_{x_2y} = \frac{\sum X_2Y}{\sqrt{\sum X_2^2}\sqrt{\sum Y^2}} \qquad r_{x_1x_2} = \frac{\sum X_1X_2}{\sqrt{\sum X_1^2}\sqrt{\sum X_2^2}}$$

$$r_{Y.X_1X_2} = r_{Y.12} = \sqrt{\frac{r_{x_1y}^2 + r_{x2y}^2 - 2r_{x_1y}r_{x2y}}{1 - r_{x_1x_2}}}$$

Korelasi Berganda

Contoh Soal Sebelumnya:

$$r_{x_1y} = \frac{\sum X_1Y}{\sqrt{\sum Y^2}\sqrt{\sum x_1^2}} = \frac{114560}{\sqrt{283443)}\sqrt{46879}} = 0.9938$$

$$r_{x_2y} = \frac{\sum X_2Y_i}{\sqrt{\sum x_2^2}\sqrt{\sum Y^2}} = \frac{14596}{\sqrt{(4633)(46879)}} = 0.9904$$

$$r_{x_1x_2} = \frac{\sum X_1X_2}{\sqrt{\sum X_1^2}\sqrt{\sum x_2^2}} = \frac{36096}{(\sqrt{283443}\sqrt{4633}} = 0.9961$$

$$r_{Y.X_1X_2} = r_{Y.12} = \sqrt{\frac{r_{x_1y}^2 + r_{x2y}^2 - 2r_{x_1y}r_{x2y}}{1 - r_{x_1x_2}}}$$

$$= \sqrt{\frac{(0.9938)^2 + (0.9904)^2 - 2(0.9938)(0.9904)}{1 - 0.9960}} = 0.85$$

$$r_{Y.X_1X_2}^2 = (0.85)^2 = 0.7225 = 72.25\%$$

kontribusi X_1 dan X_2 terhadap naik-turun nilai Y sebesar 72,25%, ada 27.75% nilai Y dipengaruhi oleh faktor lain di luar X_1 dan X_2.

Korelasi Parsial

$$r_{x_1y.x_2} = \frac{r_{x_1y} - r_{x_2y}r_{x_1x_2}}{\sqrt{1-r_{x_2y}^2}\sqrt{1-r_{x_1x_2}^2}}$$

Korelasi X_1 dengan Y saat X_2 konstan

$$r_{x_2y.x_1} = \frac{r_{x_2y} - r_{x_1y}r_{x_1x_2}}{\sqrt{1-r_{x_1y}^2}\sqrt{1-r_{x_1x_2}^2}}$$

Korelasi X_2 dengan Y saat X_1 konstan

$$r_{x_1x_2.Y} = \frac{r_{x_1x_2} - r_{x_1y}r_{x_2Y}}{\sqrt{1-r_{x_1y}^2}\sqrt{1-r_{x_2Y}^2}}$$

Korelasi X_1 dengan X_2 saat Y konstan

Korelasi Parsial

Contoh Soal Sebelumnya:

$$r_{x_1y.x_2} = \frac{r_{x_1y} - r_{x_2y}r_{x_1x_2}}{\sqrt{1-r^2_{x_2y}}\sqrt{1-r^2_{x_1x_2}}} = \frac{0.9938-(0.9904)(0.9961)}{\sqrt{1-(0.9904)^2}\sqrt{1-(0.9961)^2}} = 0.5955$$

$$r_{x_2y.x_1} = \frac{r_{x_2y} - r_{x_1y}r_{x_1x_2}}{\sqrt{1-r^2_{x_1y}}\sqrt{1-r^2_{x_1x_2}}} = \frac{0.9904-(0.9938)(0.9961)}{\sqrt{1-(0.9938)^2}\sqrt{1-(0.9961)^2}} = 0.0485$$

$$r_{x_1x_2.Y} = \frac{r_{x_1x_2} - r_{x_1y}r_{x_2Y}}{\sqrt{1-r^2_{x_1y}}\sqrt{1-r^2_{x_2Y}}} = \frac{0.9961-(0.9938)(0.9904)}{\sqrt{1-(0.9938)^2}\sqrt{1-(0.9904)^2}} = 0.7704$$

Tugas 10 Nomor 3 (20%)

Suatu percobaan dilakukan untuk menentukan apakah berat seekor binatang dapat diprediksikan setelah jangka waktu tertentu berdasarkan berat awal dan jumlah makanan yang dikonsumsi. Data yang diperoleh (dalam kg), sbb.

Berat Akhir Y	Berat Awal X1	Makanan X2
95	42	272
77	33	226
80	33	259
100	45	292
97	39	311
70	36	183
50	32	173
80	41	236
92	40	230
84	38	235

Tugas 10 Nomor 3 (Samb....)

a) Tentukan persamaan regresi linier ganda
b) Tentukan prediksi bobot akhir seekor binatang yang berbobot awal 35 kg dan menghabiskan makanan 250 kg
c) Tentukan nilai korelasi ganda dan korelasi parsial, dan artinya
d) Tentukan koefisien determinasi dan artinya.

Tugas 10 Nomor 4 (20%)

12.1 A set of experimental runs was made to determine a way of predicting cooking time y at various values of oven width x_1 and flue temperature x_2. The coded data were recorded as follows:

y	x_1	x_2
6.40	1.32	1.15
15.05	2.69	3.40
18.75	3.56	4.10
30.25	4.41	8.75
44.85	5.35	14.82
48.94	6.20	15.15
51.55	7.12	15.32
61.50	8.87	18.18
100.44	9.80	35.19
111.42	10.65	40.40

Estimate the multiple linear regression equation

$$\mu_{Y|x_1,x_2} = \beta_0 + \beta_1 x_1 + \beta_2 x_2.$$

Tugas 10 Nomor 5 (25%)

12.2 In *Applied Spectroscopy*, the infrared reflectance spectra properties of a viscous liquid used in the electronics industry as a lubricant were studied. The designed experiment consisted of the effect of band frequency x_1 and film thickness x_2 on optical density y using a Perkin-Elmer Model 621 infrared spectrometer. (*Source*: Pacansky, J., England, C. D., and Wattman, R., 1986.)

y	x_1	x_2
0.231	740	1.10
0.107	740	0.62
0.053	740	0.31
0.129	805	1.10
0.069	805	0.62
0.030	805	0.31
1.005	980	1.10
0.559	980	0.62
0.321	980	0.31
2.948	1235	1.10
1.633	1235	0.62
0.934	1235	0.31

Estimate the multiple linear regression equation

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2.$$

SEKIAN

Ringkasan Materi Perkuliahan

Statistik Deskriptif
- Penyajian Data
- Ukuran pemusatan data
- Ukuran penyebaran data

Teori Probabilitas

• Konsep probabilitas	• Ekspektasi matematis
• Probabilitas bersyarat	• Distribusi peluang diskrit
• Variabel random	• Distribusi peluang kontinu

Statistik Inferensial

Teori Penaksiran
- Rataan dan selisih rataan
- Proporsi dan selisih proporsi
- Varians dan rasio varians

Pengujian Hipotesis

• Rataan	• Uji *Chi Square*, terdiri dari:
• Proporsi	- *Good of fit-test* - Uji kebebasan
• Varians	- Uji keseragaman

ANOVA
- Kesamaan Varians
- Uji Bartlett
- ANOVA One Way
- ANOVA Two Way

Statistik Non-parametrik
- Uji Tanda
- Uji Peringkat Wilcoxon
- Uji Kruskal-Wallis

Regresi Linier dan Korelasi
- Korelasi Sederhana & Ganda
- Regresi Linier Sederhana
- Regresi Linier Ganda