MODUL PERKULIAHAN

Statistika dan Probabilitas

Modul Standar untuk digunakan dalam Perkuliahan di Universitas Mercu Buana

Fakultas	Program Studi	Tatap Muka	Kode MK	Disusun Oleh
Ilmu Komputer penerbit Modul	Teknik Informatika Studi	06	87006	Yulius Eka Agung Seputra,ST,MSi

Abstract

Matakuliah statistik Menjadi Dasar dari Pemikiran penelitian seorang yang akan Mempelajari statistik.Statistik di sangat Penting dalam Membangun sebuah Aplikasi Program Mata Kuliah ini merupakan prayarat bagi Mata kuliah Algoritma dan Stuktur Data

Kompetensi

Mahasiswa dapat Memahami operasi dasar himpunan, dan penyajian himpuan

KATA PENGANTAR

Para mahasiswa/i pada saat ini tidak asing lagi dengan teknologi, karena sudah merupakan bagian dari kehidupan mereka sehari-hari. Mulai mereka menginjakkan kaki di sekolah dasar, mereka sudah terbiasa melihat komputer seperti melihat peralatan elektronik biasa baik dirumah maupun di lingkungan mereka. Modul ini dibuat untuk dapat cocok dengan apa yang telah mereka ketahui tentang komputer, dan apa yang kami percayai harus diketahui oleh mereka mengenai komputer dan peralatan lainnya.

Isi dari modul ini sedemikian rupa kami susun sehingga kami harapkan tidak ada pengetahuan yang terpisah, semua menjadi kesatuan pengetahuan yang menyatu dan berkesinambungan. Pada modul ini juga dibahas mengenai komunikasi dengan dan tanpa kabel pada peralatan komputer. Komputasi enterprise atau perusahaan besar juga menjadi bagian pengetahuan dari modul ini untuk memperluas wawasan para mahasiswa/i untuk dapat siap menghadapi dunia kerja yang terbentang di masa depan mereka.

Untuk mendukung pengetahuan mereka, mata kuliah juga akan dilengkapi dengan modul-modul laboratorium, yang akan mengembangkan kemampuan mahasiswa/i dalam memakai aplikasi komputer khususnya suite software: *Microsoft Office XP 2005*, kemampuan dan keahlian ini dikenal juga dengan istilah *"soft-skill"*.

Kami harapkan modul ini dapat menjadi pegangan untuk memahami dan juga aplikasi dari teknologi komputer, atau lebih luasnya lebih dikenal dengan istilah baru yaitu: Telematika. Akhir kata kami tim penyusun dengan rendah hati mohon maaf apabila ada kekurangan di sana sini, dan dengan hati terbuka kami dengan senang hati akan menerima semua jenis masukan, terutama kritik-kritik yang membangun untuk menjadikan modul ini menjadi lebih baik di masa mendatang.

Penulis modul,
Yulius Eka Agung Seputra,ST,MSi

DAFTAR ISI

ANALISIS KORELASI DAN REGRESI LINEAR BERGANDA

Regresi artinya peramalan, penaksiran, atau pendugaan pertama kali di perkenalkan pada tahun 1877 oleh Sir Francis Galton (1822 – 1911). Sehubungan dengan penelitiannya terhadap tinggi manusia. Penelitian tersebut membandingkan antara tinggi anak laki-laki dan tinggi badan ayahnya.

A. Rumusan Masalah

Adapun rumusan masalah dalam makalah ini yaitu :

1. Bagaimana hubungan linear lebih dari dua variabel
2. Bagaimana menggunakan persamaan linear berganda

B. Tujuan Penulisan

Adapun tujuan penulisan makalah ini yaitu :

1. Mengetahui penggunaan rumus-rumus yang berlaku pada regresi dan kolinear berganda
2. Pengaplikasikan rumus-rumus regresi dan kolinear berganda pada suatu penelitian

BAB II PEMBAHASAN

REGRESI DAN KORELASI LINEAR BERGANDA

A. Regresi Linear Berganda

1. Hubungan liniear lebih dari dua variabel

Regresi artinya peramalan penaksiran atau pendugaan pertama kali diperkenalkan pada tahun 1877 oleh Sir Francis Galtoon (1822-1911). Analisis regresi digunakan untuk menentukan bentuk dari hubungan antar variabel. Tujuan utama dalam penggunaan analisis itu adalah untuk meramalkan atau memperkirakan nilai dari suatu variabel dalam hubungannya dengan variabel yang lain. Disamping hubungan linear dua variabel, hubungan linear dari dua variabel bisa juga terjadi misalnya; hubungan antara hasil penjualan dengan harga dan daya beli.

Hubungan linear lebih dari dua variabel bila dinyatakan dalam bentuk persamaan matematis adalah :

$$Y = a + b_1x_1 + b_2x_2 + \ldots\ldots\ldots\ldots\ldots b_kx_k +$$

Keterangan :

$x, x_{1,} x_2 \ldots\ldots\ldots x_k$ = variabel-variabel

$a, b_{1,} b_2 \ldots\ldots\ldots b_k$ = bilangan konstan (konstanta) koefisien variabel

2. Persamaan regresi linear berganda

Regresi linear berganda adalah regresi dimana variabel terikatnya (Y) dihubungkan atau dijelaskan lebih dari satu variabel, mungkin dua, tiga dan seterusnya variabel bebas (x, $x_{1,}$ $x_2 \ldots\ldots\ldots x_n$) namun masih menunjukkan diagram hubungan yang linear.

Penambahan variabel bebas ini diharapkan dapat lebih menjelaskan karakteristik hubungan yang ada walaupun masih saja ada variabel yang terabaikan.

Bentuk umum dari persamaan linear berganda dapat ditulis sebagai berikut:

a. Bentuk stokastik

$\hat{y} = a + b_1x_1 + b_2x_2 + b_3x_3 \ldots\ldots\ldots\ldots b_kx_k + c$

b. Bentuk non stokastik

$\hat{y} = a + b_1x_1 + b_2x_2 + b_3x_3 \ldots\ldots\ldots\ldots b_kx_k$

Keterangan

$\hat{y}$: Variabel terikat (nilai duga y)

$a, b_1, b_2\, b_3 \ldots\ldots b_k$: koefisien regresi

$x_1, x_2\, x_3 \ldots\ldots x_k$: variabel bebas

e : kesalahan pengganggu

B. Pendugaan dan Pengujian Koefisien Regresi

1. Kesalahan baku regresi dan koefisien regresi berganda

Kesalahan baku atau selisih taksir standar regresi adalah nilai menyatakan seberapa jauh menyimpangnya nilai regresi tersebut terhadap nilai sebenarnya. Nilai ini digunakan untuk mengukur tingkat ketepatan suatu pendugaan dalam menduga nilai. Jika nilai ini sama dengan nol maka penduga tersebut memiliki tingkat ketepatan 100%.

Kesalahan baku atau selisih taksir standar regresi berganda dirumuskan

$$S_e = \sqrt{\frac{\sum y^2 - b_1\left(\sum x_1 y\right) + b_2\left(\sum x_2 y\right)}{n - m}}$$

Keterangan

Se : Kesalahan baku regresi berganda

n : Jumlah pasangan observasi

m : jumlah konstant dalam persamaan regresi berganda.

Untuk koefisien b_1 dan b_2 kesalahan bakunya dirumuskan

$$Sb_1 = \frac{Se}{\sqrt{\left(\sum x_1^2 - n\bar{x}_1^2\right)\left(1 - r^2 y_1\right)}}$$

$$Sb_2 = \frac{Se}{\sqrt{\left(\sum x_2^2 - n\bar{x}_2^2\right)\left(1 - r^2 y_1\right)}}$$

2. Pendugaan interval koefisien regresi berganda (parameter B_1 dan B_2)

Parameter B_1 dan B_2 sering juga disebut sebagai koefisien regresi parsial. Pendugaan parameter B_1 dan B_2 menggunakan distribusi t dengan derajat bebas db = n – m secara umum pendugaan parameter B_1 dan B_2 adalah :

$b_1 - t_{a/2n-m}\, Sb_i \leq B_i \leq b_i + t_{a/2n-m}\, Sb_i$

i = 2,3

3. Pengujian hipotesis koefisien regresi berganda (parameter B_1 dan B_2)

Pengujian hipotesis bagi koefisien regresi berganda atau regresi parsial parameter B_1 dan B_2 dapat dibedakan menjadi 2 bentuk, yaitu pengujian hipotesis serentak dan pengujian hipotesis individual.

Pengujian hipotesis individual yaitu merupakan pengujian hipotesis koefisien regresi berganda dengan hanya satu B (B_1 dan B_2) yang mempunyai pengaruh Y. pengujian hipotesis serentak merupakan pengujian hipotesis koefisien regresi berganda dengan B_1 dan B_2 serentak atau bersama-sama mempengaruhi Y.

C. Peramalan dengan Regresi Linear Berganda

Peramalan terhadap nilai Y dengan menggunakan regresi linear berganda, dapat dilakukan apabila persamaan garis regresinya sudah diestimasi dan nilai variabel bebas x_1, x_2 sudah diketahui.

Suatu persamaan garis regresi linear berganda dapat dipakai dalam peramalan dengan terlebih dahulu melakukan pengujian hipotesis terhadap koefisien-koefisien regresi parsialnya. Tujuan ialah mengetahui variabel-variabel bebas yang digunakan itu memiliki pengaruh yang nyata atau tidak terhadap y tersebut. Variabel bebas x_1 dan x_2 disebut memiliki pengaruh yang nyata apabila dalam pengujian hipotesis koefisien parsialnya $H_0 : B_1 = B_2 = 0$ ditolak atau $H_1 : B_1 \neq B_2 \neq 0$ diterima, khususnya pada taraf nyata 1%

Kelebihan peramalan y dengan menggunakan regresi linear berganda adalah dapat diketahui besarnya pengaruh secara kuantitatif setiap variabel bebas (x_1 atau x_2) apabila pengaruh variabelnya dianggap konstan. Misalnya sebuah persamaan regresi berganda

$$y = a + b_1x_1 + b_2x_2$$

Keterangan :

y : Nilai statistik mahasiswa

x_1 : Nilai inteligensi mahasiswa

x_2 : Frekuensi membolos mahasiswa

b_1 : Pengaruh x_1 terhadap y jika x_2 konstan

b_2 : Pengaruh x_2 terhadap y jika x_1 konstan jika a = 17,547; b_1 = 0,642; b_2 = - 0,284 maka persamaan regresi linear bergandanya menjadi

$$\hat{y} = 17{,}547 + 0{,}624\ (75) - 0{,}284\ (4)$$

Dengan persamaan regresi linear berganda tersebut, nilai y (nilai statistik maha siswa) dapat diramalkan dengan mengetahui nilai x_1 (nilai inteligensi mahasiswa) dan x_2 (frekuensi membolos mahasiswa) misalkan, nilai x_1 = 75 dan x_2 = 24 maka ramalan nilai y adalah

$$\hat{y} = 17{,}547 + 0{,}624\ (75) - 0{,}284\ (4)$$

$$= 63.211$$

Penulisan persamaan garis regresi linear berganda biasanya disertai dengan kesalahan baku masing-masing variabel bebas dan koefisien determinasi berganda r^2, sebagai ukuran tepat atau tidaknya garis tersebut sehingga pendekatan.

D. Korelasi Linear Berganda

Korelasi linear berganda merupakan alat ukur mengenai hubungan yang terjadi antara variabel yang terikat. (variabel Y) dan dua atau lebih variabel bebas (x_1, x_2……x_k). Analisis korelasinya menggunakan tiga koefisien korelasi yaitu koefisien determinasi berganda, koefisien korelasi berganda, dan koefisien korelasi parsial.

1. Korelasi linear berganda dengan dua variabel bebas

   a. Koefisien penentu berganda atau koefisien determinasi berganda

   Koefisien determinasi berganda, disimbolkan KPB y.12 atau R^2 merupakan ukuran kesusaian garis regresi linear berganda terhadap suatu data. Rumus

$$KPB_{y.12} = \frac{b_1 \sum x_1 y + b_2 \sum x_2 y}{\sum y^2}$$

   b. Koefisien korelasi berganda

   Koefisien korelasi berganda disimbolkan r_{y12} merupakan ukuran keeratan hubungan antara variabel terikat dan semua variabel bebas. Secara bersama-sama. Rumus :

$$R_{y.12} = \sqrt{\frac{b_1 \sum x_1 y + b_2 \sum x_2 y}{\sum y^2}}$$

   c. Koefisien korelasi parsial

   Koefisien korelasi parsial merupakan koefisien korelasi antara dua variabel. Jika variabel lainnya konstan, pada hubungan yang melibatkan lebih dari dua variabel.

   Ada 3 koefisien korelasi parsial untuk hubungan yang melibatkan 3 variabel yaitu sebagai berikut :

   1) Koefisien korelasi parsial antara y dan x_1, apabila x_2 konstan dirumuskan

$$r_{y.12} = \frac{r_{y1} - r_{y2}.r_{12}}{\sqrt{\left(I - r^2_{y1}\right)\left(I - r^2_{12}\right)}}$$

2) Koefisien korelasi parsial antara y dan x_2, apabila x_1 konstan dirumuskan

$$r_{y.12} = \frac{r_{y2} - r_{y1}.r_{12}}{\sqrt{\left(I - r^2_{y1}\right)\left(I - r^2_{y2}\right)}}$$

3) Koefisien korelasi parsial antara x_1 dan x_2 apabila y konstan dirumuskan

$$R_{12}y = \frac{r_{12} - r_{y1}.r_{12}}{\sqrt{\left(I - r^2_{y1}\right)\left(I - r^2_{y2}\right)}}$$

2. Korelasi linear berganda dengan 3 variabel bebas

a. Koefisien penentu berganda

$$KPB = \frac{b_1 \sum x_1 y + b_2 \sum x_2 y + b_3 \sum x_3 y}{\sum y^2}$$

$$\sum y^2 = \sum y^2 - \frac{\left(\sum y\right)^2}{n}$$

b. Koefisien korelasi berganda

$$ry_{123} = \sqrt{\frac{b_1 \sum x_1 y + b_2 \sum x_2 y + b_3 \sum x_3 y}{\sum y^2}}$$

HUBUNGAN LIBIH DARI DUA VARIABEL REGRESI LINEAR BERGANDA :

Apabila terdapat lebih dari dua variabel, maka hubungan linear dapat dinyatakan dalam persamaan regresi linear berganda sebagai berikut :

$Y' = b_0 + b_1X_1 + b_2X_2 + \ldots + b_kX_k$

Disini ada satu variabel tidak bebas, yaitu Y' dan ada k varibel bebas, yaitu $X_1, \ldots, X_k$.

Untuk menghitung $b_0, b_1, b_2, \ldots, b_k$ kita gunakan metode kuadrat terkecil yang menghasilkan persamaan normal sebagai berikut :

$$
\begin{array}{l}
b_0\, n + b_1 \Sigma X_1 + b_2 \Sigma X_2 + \ldots + b_k \Sigma X_k = \Sigma Y \\
b_0 \Sigma X_1 + b_1 \Sigma X_1 X_1 + b_2 \Sigma X_1X_2 + \ldots + b_k \Sigma X_1X_k = \Sigma X_1Y \\
b_0 \Sigma X_2 + b_1 \Sigma X_1 X_2 + b_2 \Sigma X_2X_2 + \ldots + b_k \Sigma X_2X_k = \Sigma X_2Y \\
\quad . \qquad . \qquad . \qquad . \qquad . \\
\quad . \qquad . \qquad . \qquad . \qquad . \\
\quad . \qquad . \qquad . \qquad . \qquad . \\
b_0 \Sigma X_k + b_1 \Sigma X_1 X_k + b_2 \Sigma X_2X_k + \ldots + b_k \Sigma X_kX_k = \Sigma X_kY
\end{array}
$$

Kalau persamaan ini dipecahkan, kita akan memperoleh nilai $b_0, b_1, b_2, \ldots, b_k$. Kemudian dapat dibentuk persamaan regresi linear berganda. Apabila persamaan regresi itu telah diperoleh, barulah kita dapat meramalkan nilai Y dengan syarat kalau nilai $X_1, X_2, \ldots, X_k$ sebagai variabel bebas sudah diketahui.

Misalkan: k =2, maka $Y' = b_0 + b_1X_1 + b_2X_2$, satu variabel tak bebas(Y), dan dua variabel bebas (X_1 dan X_2), maka b_0, b_1, dan b_2 dihitung dari persamaan normal berikut :

$$
\begin{array}{l}
b_0\, n + b_1 \Sigma X_1 + b_2 \Sigma X_2 = \Sigma Y \\
b_0 \Sigma X_1 + b_1 \Sigma X_1 X_1 + b_2 \Sigma X_1X_2 = \Sigma X_1Y \\
b_0 \Sigma X_2 + b_1 \Sigma X_1 X_2 + b_2 \Sigma X_2X_2 = \Sigma X_2Y
\end{array}
$$

Persamaan diatas dapat dinyatakan dalam persamaan matriks berikut :

$$\underbrace{\begin{bmatrix} n & \sum X_1 & \sum X_2 \\ \sum X_1 & \sum X_1^2 & \sum X_1X_2 \\ \sum X_2 & \sum X_1X_2 & \sum X_2^2 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}}_{b} = \underbrace{\begin{bmatrix} \sum Y \\ \sum X_1Y \\ \sum X_2Y \end{bmatrix}}_{H}$$

Variabel b dapat diselesaikan dengan cara sebagai

berikut :

$$b_0 = \frac{\det(A_0)}{\det(A)}, \quad b_1 = \frac{\det(A_1)}{\det(A)}, \quad b_2 = \frac{\det(A_2)}{\det(A)}$$

Dimana :

$$A = \begin{bmatrix} n & \sum X_1 & \sum X_2 \\ \sum X_1 & \sum X_1^2 & \sum X_1X_2 \\ \sum X_2 & \sum X_1X_2 & \sum X_2^2 \end{bmatrix}$$

det(A) = (n) (ΣX_1X_1) (ΣX_2X_2) + (ΣX_1) (ΣX_1X_2) (ΣX_2) + (ΣX_2) (ΣX_1) (ΣX_1X_2)

– (ΣX_2) (ΣX_1X_1) (ΣX_2) – (ΣX_1X_2) (ΣX_1X_2) (n) – (ΣX_2X_2) (ΣX_1) (ΣX_1)

$$A_0 = \begin{bmatrix} \sum Y & \sum X_1 & \sum X_2 \\ \sum X_1Y & \sum X_1^2 & \sum X_1X_2 \\ \sum X_2Y & \sum X_1X_2 & \sum X_2^2 \end{bmatrix}$$

$\det(A_0) = (\Sigma Y)\,(\Sigma X_1X_1)\,(\Sigma X_2X_2) + (\Sigma X_1)\,(\Sigma X_1X_2)\,(\Sigma X_2Y) + (\Sigma X_2)\,(\Sigma X_1Y)\,(\Sigma X_1X_2)$

$- (\Sigma X_2Y)\,(\Sigma X_1X_1)\,(\Sigma X_2) - (\Sigma X_1X_2)\,(\Sigma X_1X_2)\,(\Sigma Y) - (\Sigma X_2X_2)\,(\Sigma X_1Y)\,(\Sigma X_1)$

$$A_1 = \begin{bmatrix} n & \sum Y & \sum X_2 \\ \sum X_1 & \sum X_1Y & \sum X_1X_2 \\ \sum X_2 & \sum X_2Y & \sum X_2^2 \end{bmatrix}$$

$\det(A_1) = (n)\,(\Sigma X_1Y)\,(\Sigma X_2X_2) + (\Sigma Y)\,(\Sigma X_1X_2)\,(\Sigma X_2) + (\Sigma X_2)\,(\Sigma X_1)\,(\Sigma X_2Y)$

$- (\Sigma X_2)\,(\Sigma X_1Y)\,(\Sigma X_2) - (\Sigma X_2Y)\,(\Sigma X_1X_2)\,(n) - (\Sigma X_2X_2)\,(\Sigma X_1)\,(\Sigma Y)$

$$A_2 = \begin{bmatrix} n & \sum X_1 & \sum Y \\ \sum X_1 & \sum X_1^2 & \sum X_1Y \\ \sum X_2 & \sum X_1X_2 & \sum X_2Y \end{bmatrix}$$

$\det(A_2) = (n)\,(\Sigma X_1X_1)\,(\Sigma X_2Y) + (\Sigma X_1)\,(\Sigma X_1Y)\,(\Sigma X_2) + (\Sigma Y)\,(\Sigma X_1)\,(\Sigma X_1X_2)$

$- (\Sigma X_2)\,(\Sigma X_1X_1)\,(\Sigma Y) - (\Sigma X_1X_2)\,(\Sigma X_1Y)\,(n) - (\Sigma X_2Y)\,(\Sigma X_1)\,(\Sigma X_1)$

TREND PARABOLA :

Garis trend pada dasarnya adalah garis regresi di mana variabel bebas X merupakan variabel waktu. Baik garis regresi maupun trend dapat berupa garis lurus maupun tidak lurus. Persamaan garis trend parabola adalah sebagai berikut : $Y' = a + bX + cX^2$

Perhatikan bahwa bentuk persamaa seperti persamaan garis regresi linear berganda adalah $Y' = b_0 + b_1X_1 + b_2X_2$, di mana $b_0 = a$, $b_1 = b$, $b_2 = c$, $X_1 = X$, dan $X_2 = X^2$.

Dengan demikian cara menghitung koefisien a, b, dan c sama seperti menghitung b_0, b_1, dan b_2, yaitu menggunakan persamaan normal sebagai berikut :

$$a\,n + b\,\Sigma X + c\,\Sigma X^2 = \Sigma Y$$

$$a\,\Sigma X + b\,\Sigma X^2 + c\,\Sigma X^3 = \Sigma XY$$

$$a\,\Sigma X^2 + b\,\Sigma X^3 + c\,\Sigma X^4 = \Sigma X^2Y$$

TREND EKSPONENSIAL (LOGARITMA) :

Ada beberapa jenis trend yang tidak linear tetapi dapat dibuat linear dengan jalan melakukan transformasi (perubahan bentuk). Misalnya, trend eksponensial :

$Y' = ab^x$ dapat diubah menjadi trend semi log: $\log Y' = \log a + (\log b)X$;

$\log Y' = Y'_0$; $\log a = a_0$ dan $\log b = b_0$. Dengan demikian, $Y'_0 = a_0 + b_0X$, dimana koefisien a_0 dan b_0 dapat dicari berdasarkan persamaan normal.