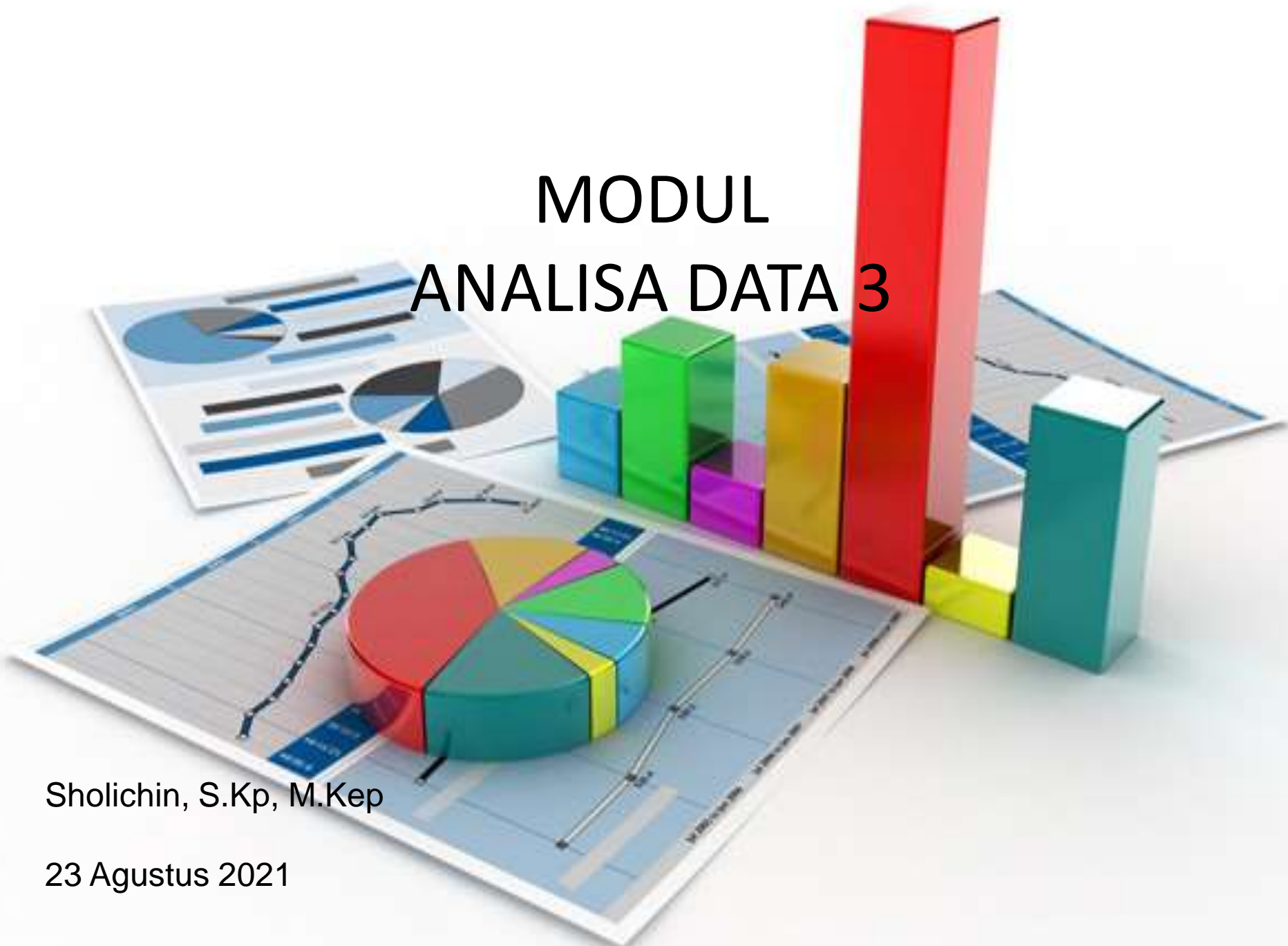


MODUL ANALISA DATA 3

Sholichin, S.Kp, M.Kep

23 Agustus 2021



ANALISIS REGRESI LINIER GANDA

Sholichin

- Disebut juga analisis multiple regression linier
- Mrp perluasan analisis simple linier regression
- Ingat Simple linier regression hanya satu variabel independen dihubungkan dg satu variabel dependen.
- Multiple regression linier ----- beberapa variabel independen dihubungkan dg satu variabel dependen, ex : faktor-faktor yang berhub dengan tekanan darah ----- maka variabel independennya : umur, bb, jk, dll

- Pada RLG variabel dependennya hrs numerik sedang independennya boleh semuanya numerik atau campuran (numerik dan katagorik)
- Model persamaannya :
$$Y = a + b_1X_1 + b_2X_2 + b_3X_3\dots + e$$
- Kegunaan :
 1. Prediksi (perkiraan), ex : seorang individu dpt diperkirakan Tdnya pada umur, bb, dll
 2. Assosiasi : dpt diketahui variabel independen yg paling besar/dominan mempengaruhi variabel dependen.

- Langkah Pemodelan :

1. Lakukan analisis bivariat untuk kandidat model ----- bila hasil uji bivariat nilai $P < 0,25$ maka variabel tsb masuk dalam pemodelan
2. Lakukan analisis RLG scr bersamaan ($P < 0.25$)---- (klik analyze---regression---linier---isikan dep---indep scr bersamaan---method isikan backward (scr otomatis variabel indep yg memp $P > 0,05$ akan dikeluarkan---ok). Jika pada method diisikan **enter** maka variabel yang $p > 0,05$ dikeluarkan dari value yg terbesar.
3. Diagnostik regresi linier : 5 asumsi dan Diagnostik Multicollinearity
4. Penilaian reliabilitas model (mebut regresi linier dg beberapa model dan bandingkan), jika hasilnya sama/hampir sama maka model regresi reliable.

- Agar inferensi valid mk dalam RLG dianjurkan memenuhi asumsi/syarat sbb :
 1. Eksistensi (random)---(nilai residual menunjukkan adanya mean mendekati nol berarti asumsi terpenuhi)
 2. Independensi ----- (jika nilai durbin watson antara -2 s/d 2 berarti asumsi terpenuhi)
 3. Linieritas ----- (jika nilai pd uji anova $p < \alpha$ berarti asumsi terpenuhi)
 4. Homoscedasticity ---- (jika titik tebaran tdk berpola ttt dan menyebar merata berarti asumsi terpenuhi)
 5. Normalitas ---- (data menyebar disekitar garis diagonal dan mengikuti arah garis diagonal berarti asumsi terpenuhi)
 6. Diagnostik Multicollinearity)----tdk boleh terjadi korelasi scr kuat antar variabel independen (multicollinearity)----(jk nilai VIF > 10 berarti tjd collinearity)

- Langkah uji asumsi :
 1. Regression----linier
 2. Mskkan dependen dan independen scr bersamaan
 3. Klik statistik---klik 'collinearity diagnostik' dan covariance matrix (uji multicolinearity), klik durbin watson (uji independensi)
 4. Klik tombol 'plot'
 5. Masukkan 'SRESID' ke kotak Y dan masukkan 'ZPRED' ke kotak X (uji asumsi homoscedasity), Klik kotak 'histogram' dan 'normal probability' (uji normality)
 6. Continue----ok

Latihan

- Latihan.....gunakan data V di halaman berikutnya :
- “faktor-fator yang mempengaruhi kadar kolesterol”

variabel independen : umur, jk, pekerjaan, tb, bb, diit.

variabel dependen : kadar kolesterol

- Latihan.....gunakan data V di halaman berikutnya :
- “faktor-faktor yang mempengaruhi pengeluaran jumlah sputum pada pasien COPD”

variabel independen : umur, jk, alat bantu nafas, suksion, batuk, napas dalam, humidifikasi dan pengaturan temperatur, terapi, perkusi dada, vibrasi dada, postural drainage.

variabel dependen : jumlah sputum

DATA V

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	no	umur	jk	jumlah	abn	suksn	btk	nd	hum	tx	perkusi	vibrasi	postural	nafas
2	1	25	1	120	1	2	1	1	1	2	2	2	2	2
3	2	34	1	98	1	2	2	2	2	2	1	2	2	2
4	3	56	1	90,9	2	1	1	2	1	1	2	2	2	1
5	4	60	1	100,23	1	2	2	2	2	2	2	3	3	2
6	5	45	2	110	1	1	2	2	2	2	2	3	3	1
7	6	47	1	100	2	1	1	2	1	1	2	2	2	1
8	7	62	2	123,04	1	1	2	1	1	1	1	3	3	1
9	8	27	2	96,89	1	2	2	1	2	2	3	2	2	2
10	9	44	2	80	2	1	1	1	1	1	1	1	1	1
11	10	38	1	75	2	1	1	1	1	1	1	1	1	1
12	11	49	1	145	1	1	1	2	1	1	1	3	3	1
13	12	39	1	138	1	2	1	2	2	2	2	2	2	2
14	13	56	1	200	1	1	1	2	1	1	1	3	3	1
15	14	62	1	188,75	1	2	2	2	1	2	1	3	3	2
16	15	45	1	160	1	2	1	2	2	2	2	3	3	2
17	16	35	1	89	2	1	1	2	1	1	1	1	1	1
18	17	44	2	74	2	1	2	2	1	1	1	1	1	1
19	18	56	2	120,25	1	2	2	2	2	2	2	2	2	2
20	19	35	1	90	1	2	2	1	2	2	1	2	2	2
21	20	47	2	75,06	2	1	1	1	1	1	1	1	1	1
22	21	68	1	86	2	1	2	2	1	1	1	1	1	1
23	22	22	2	76	2	1	1	2	1	1	1	1	1	1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
24	23	42	2	100,02	1	2	1	2	2	2	2	2	2	2
25	24	36	2	90,54	2	1	1	2	1	1	1	1	1	1
26	25	48	1	110	1	2	2	1	2	2	2	2	2	2
27	26	53	1	90	2	1	2	1	1	1	1	1	1	1
28	27	21	1	100	1	2	2	2	2	2	2	3	3	2
29	28	42	1	100	2	1	2	2	1	1	2	3	3	1
30	29	38	1	84	2	1	2	2	1	1	1	2	2	1
31	30	46	1	74,68	2	1	1	2	1	1	1	1	1	1
32	31	56	1	68,9	2	1	2	2	1	1	1	1	1	1
33	32	46	2	156,09	1	2	1	1	2	2	3	3	3	2
34	33	43	1	88	2	1	1	1	1	1	1	1	1	1
35	34	42	1	90,9	1	2	2	1	2	2	2	2	2	2
36	35	35	2	56	2	1	2	2	1	1	1	1	1	1
37	36	55	2	89	1	2	2	2	2	2	1	1	1	2
38	37	56	2	100	2	1	1	1	1	1	3	3	3	1
39	38	70	2	96,78	2	1	1	1	1	1	2	2	2	1
40	39	44	2	59,99	2	1	2	1	1	1	1	1	1	1
41	40	33	1	89	1	2	2	2	2	2	2	2	2	2
42	41	32	1	79	1	2	2	2	2	2	3	3	3	2
43	42	26	1	100,06	2	1	1	1	1	1	2	2	2	1
44	43	51	2	98	2	1	2	1	1	1	2	2	2	1
45	44	32	1	87,09	1	2	2	2	2	2	1	1	1	2
46	45	48	1	100	2	1	1	1	1	1	3	3	3	1



REGRESI LOGISTIK

Sholichin

- Dependen katagorik yg dichotomus dan independen dpt numerik atau katagorik
- Fungsi : probabilitas kejadian suatu penyakit berdasarkan faktor risiko ttt, ex : probabilitas kejadian jantung pada umur ttt
- Bentuk modelnya :

$$Z = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots \quad X = \text{independen}, Z = \text{dep}$$

$$F(Z) = \frac{1}{1 + e^{-(\alpha + \beta_1 X_1 + \beta_2 X_2 + \dots)}}$$

- Contoh kasus : hubungan antara kejadian infeksi (nama variabel PI) dengan tinggi rendahnya leukosit (nama variabel LS). Pemberian kode nilai variabel adalah sbb :
 - Variabel PI ---- 1=timbul infeksi, 0=tidak ada infeksi
 - Variabel LS ---- 1=leukosit tinggi, 0=leukosit rendah

Pertanyaan :

1. Berapa peluang mereka yang leukositnya tinggi memp risiko untuk terjadi infeksi ?
2. Berapa peluang mereka yang leukositnya rendah memp risiko untuk terjadi infeksi ?
3. Bandingkan risiko terjadi infeksi antara mereka yang leukositnya tinggi dengan yang rendah.

JAWAB :

- Misalkan didapatkan hasil analisis dg paket program statistik : $\alpha = 3,911$ dan $\beta = 0,652$, maka :

a.
$$P(X) = \frac{1}{1 + e^{-(3,911+0,652 \cdot 1)}} = 0,037 \text{ atau sekitar } 4 \%$$

jadi mereka dg leukosit tinggi memp risiko untuk tjdnnya infeksi sebesar 4%

b.
$$P(X) = \frac{1}{1 + e^{-(3,911+0,652 \cdot 0)}} = 0,019$$

jadi mereka dg leukosit rendah memp risiko untuk tjdnnya infeksi sebesar 2%

c.
$$\frac{P1(X)}{P2(X)} = \frac{0,037}{0,019} = 1,947 = 2$$

jadi mereka yang leukositnya tinggi memp risiko tjdn infeksi dua kali lebih tinggi dibanding mereka yg leukositnya rendah.

Latihan

Diketahui :

0 = usia > 30 th

1 = usia < 30 th;

0 =wanita

1 = laki-laki

0 = lingkungan bersih

1 = lingkungan kumuh

didapatkan hasil analisis regresi logistik :

variabel	β
usia	0,987
jk	0,781
Lingkungan	-0,091
Constanta	2,313

Pertanyaan :

1. Berapa peluang mereka yang berjenis kelamin laki-laki memp risiko untuk terjadi GE ?
2. Berapa peluang mereka yang berjenis kelamin wanita, tinggal di lingkungan bersih dan berusia dibawah 30 th memp risiko untuk terjadi GE
3. Berapa peluang mereka yang usianya dibawah 30 th dan tinggal didaerah yang kumuh memp risiko untuk terjadi GE ?
4. Bandingkan risiko terjadi GE antara mereka yang tinggal di lingk bersih dan kumuh.

REGRESI LOGISTIK GANDA

- Beberapa variabel independen (numerik/kata-gorik), sebaiknya katagorik karena lebih mudah menginterpretasikan.
- Variabel dependen : katagorik dichotom
- Mencakup 2 hal :

1. Model prediksi

memperoleh model yg tdd beberapa variabel independen yg dianggap terbaik untuk memprediksi kejadian variabel dependen.

Bentuk kerangka konsepnya :



Tahapan pemodelan :

1. Analisis bivariat ($P < 0,25$ masuk pemodelan)----
-jika v indep > 2 katagorik lakukan dummy.
2. Pemodelan multivariat---keluarkan v indep jk $p > 0,05$ (lihat di omnibus test/chi-square....sig) satu persatu dan tetap masuk pemodelan jk nilai OR {Exp(B)} yg baru ada $> 10\%$
3. Penilaian Interaksi variabel

Latihan Faktor Prediksi Gunakan data V (latihan 1a)

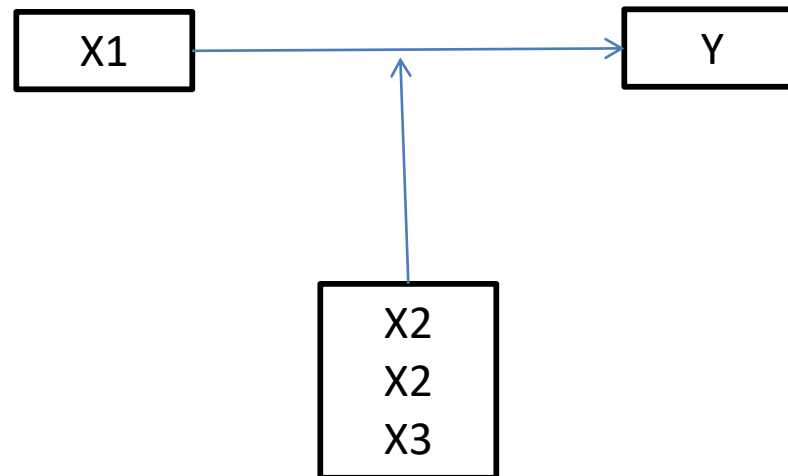
- Suatu Penelitian ingin mengetahui hubungan antara umur, JK, suksion, napas dalam , perkusi dada dengan pernafasan
- Langkahnya :
 1. Seleksi bivariat ($P < 0,25$ msk tahap multivariat namun bila scr substansi penting mk variabel tsb tetap dimasukkan. Gunakan uji regresi logistik sederhana satu persatu

Analisis bivariat umur dg nafas

1. Analyze—regression---binary logistik
2. Dependen mskkan pernafasan dan covariates mskkan umur
3. Klik option---klik CI for $\text{Exp}(B)$ ---continue---ok

2. Model Faktor Risiko

- ✚ Mengestimasi secara valid hubungan satu variabel utama dengan variabel dependen dg mengontrol beberapa variabel confounding
- ✚ Bentuk kerangka konsepnya :



Tahapan pemodelan :

1. Analisis bivariat ($P < 0,25$ masuk pemodelan)
2. Pemodelan lengkap (semua variabel baik dependen, independen maupun confounding)
3. Penilaian Interaksi
4. Penilaian confounding

Latihan Faktor Risiko

- Gunakan file data V (latihan 1a)
- “hubungan nd dengan pernafasan”
- V dep : pernafasan; V indep : napas dalam; V konfounding : umur, JK, suksion
- Cara :
 1. Langkah Pertama (semua varabel dan variabel interaksi :
 - analyze---regression----binary logistik
 - Masukkan pernafasan di kotak dependen
 - Masukkan (nd, umur, jk, suksion, nd*umur, nd*jk, nd*suksion)
 - Keluarkan $p > 0,05$ (lihat di variabel in the equation) satu persatu untuk variabel interaksi (nd*umur, nd*jk, nd*suksion)
 2. Langkah kedua (uji counfounding).....

2. Langkah kedua (uji counfounding)

keluarkan satu persatu variabel konfounding dan lihat perubahan pada $OR\{exp(B)\}$ ---(dalam hal ini OR nya variabel independen yaitu nd, caranya : OR_{nd} sesudah dikurangi OR_{nd} sebelumnya dibagi OR_{nd} sebelumnya. Bila perubahan $> 10\%$ maka vaiabel tsb sbg confounding dan bila $< 10\%$ maka hrs dikeluarkan dari model.

3. Model terakhir

- Latihan model prediksi.....gunakan data V
“hubungan antara umur, jk, alat bantu nafas, suksion, batuk, napas dalam, humidifikasi dan pengaturan temperatur, terapi, perkusi dada, vibrasi dada, postural drainage dengan pernafasan”
- Latihan model prediksi.....gunakan data V
“hubungan suksion dengan pernafasan”
Variabel konfounding : umur, jk, alat bantu nafas, batuk, napas dalam, humidifikasi dan pengaturan temperatur, terapi, perkusi dada, vibrasi dada, postural drainage