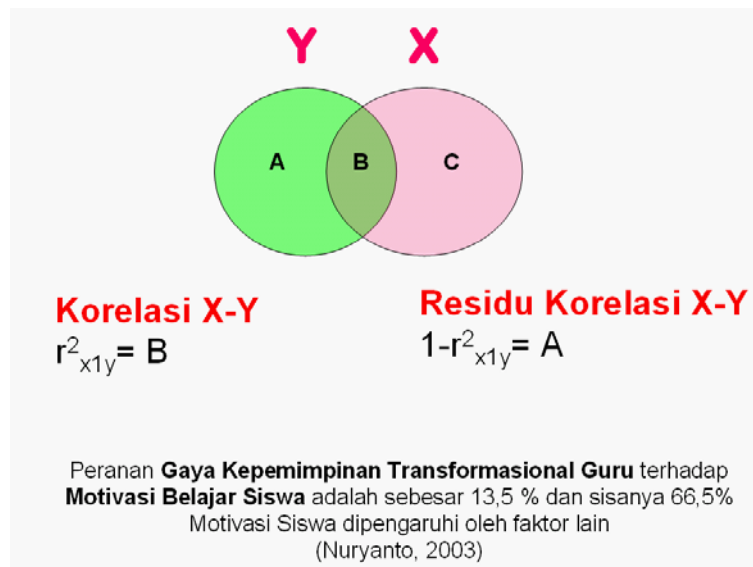


Korelasi dan Regresi Secara Visual

Oleh : Wahyu Widhiarso | Fakultas Psikologi UGM
Tahun 2009

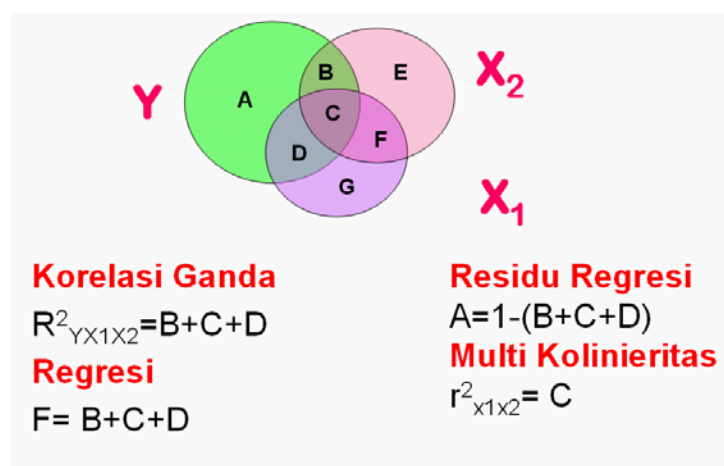
Korelasi dan Regresi memiliki keterkaitan. Akan lebih mudah memahaminya jika menggunakan teknik visualisasi. Saya mencoba menjelaskannya dengan menggunakan gambar. Materi yang dimuat adalah visualisasi korelasi, korelasi berganda, korelasi semi parsial dan parsial, regresi dan multikolinieritas.



Gambar 1. Visualisasi Korelasi dua Variabel

KORELASI BIASA

Gambar 1 menunjukkan bahwa ada hubungan antara X dan Y yang terlihat dari adanya irisan antara wilayah X dan wilayah Y. Lingkaran menunjukkan keragaman skor sehingga dapat kita lihat bahwa keragaman skor Y dapat dijelaskan oleh keragaman skor X. Misalnya Y adalah Kepemimpinan Guru, maka sebagian skor yang kepemimpinan guru yang bervariasi dapat dijelaskan oleh Motivasi Siswa (area B). Namun ada juga keragaman skor guru yang tidak bisa dijelaskan oleh Motivasi siswa (area A), artinya tinggi rendahnya Kepemimpinan guru dijelaskan oleh variabel lain selain Motivasi Siswa.



Gambar 2. Visualisasi Korelasi tiga Variabel

KORELASI GANDA (MAJEMUK)

Gambar 2 menunjukkan bahwa ada 3 variabel yang berhubungan, yaitu 2 variabel independen (X_1 , X_2) dan 1 variabel dependen (Y). Korelasi ganda adalah seberapa besar ketiga variabel tersebut saling menjelaskan dengan lainnya. Korelasi ganda digambarkan dengan area $B+C+D$. Area A adalah area Y yang tidak bisa dijelaskan oleh variabel independen X_1 dan X_2 .

REGRESI

Dari satu sisi regresi sama seperti dengan korelasi berganda, nilai F regresi menunjukkan seberapa jauh kedua variabel independen mampu menjelaskan variabel dependen dengan baik. Namun demikian regresi juga menunjukkan berapa kontribusi masing-masing variabel independen terhadap dependen. Kontribusi X_1 adalah area D (bisa juga ditambah area C), sedangkan Kontribusi X_2 adalah area B (bisa juga ditambah area C).

Gimana sih kok kedua variabel tersebut berebutan area C ? Tergantung pada jenis metode regresinya. Jika menggunakan metode ENTER, maka area C akan dimiliki oleh prediktor yang memiliki kontribusi terbesar. Misalnya prediktor yang memiliki kontribusi terbesar adalah X_1 , maka wilayah C akan diklaim oleh X_1 , sedangkan X_2 harus puas mendapatkan area C saja.

Ada juga metode regresi yang dinamakan dengan REGRESI BERJENJANG. Regresi berjenjang adalah proses memasukkan variabel independen/prediktor secara bertahap, misalnya memasukkan X_2 dulu baru memasukkan X_1 ke dalam lingkaran Y . Nah, dalam regresi berjenjang siapa yang masuk duluan akan mendapatkan wilayah irisan tersebut.

Model Summary									
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.657 ^a	.431	.360	2.140	.431	6.067	1	8	.039
2	.818 ^b	.669	.574	1.745	.238	5.028	1	7	.060

a. Predictors: (Constant), x1
b. Predictors: (Constant), x1, x2

Gambar 3. Hasil Analisis Regresi Berjenjang (X_1 masuk dulu baru diikuti X_2)

Regresi Berjenjang – Cara 1

Gambar 3 menjelaskan hasil keluaran analisis regresi berjenjang dengan dua prediktor (X_1 dan X_2) dan satu kriteria (Y). Regresi berjenjang cara pertama dilakukan dengan memasukkan X_1 terlebih dahulu baru diikuti oleh X_2 . Terlihat ada dua model, Model 1 adalah $X_1 \rightarrow Y$ sedangkan model kedua X_2 dimasukkan sehingga menjadi $X_1 \& X_2 \rightarrow Y$. Terlihat bahwa sumbangan ketika X_1 masuk, dia mampu menjelaskan variasi skor Y sebesar 43.1% ($R\text{-Square} = 0.431$).

Setelah X_2 masuk, semakin besar wilayah Y yang bisa dijelaskan yaitu sebesar 66.9% ($R\text{-Square} = 0.669$). Jadi sumbangan X_2 terhadap Y adalah selisih $66.9 - 43.1 = 23.8$. Atau anda bisa lihat pada kolom $R\text{-Square Change}$. Sumbangan X_1 terhadap Y adalah 43.1% sedangkan X_2 terhadap Y adalah 23.8%.

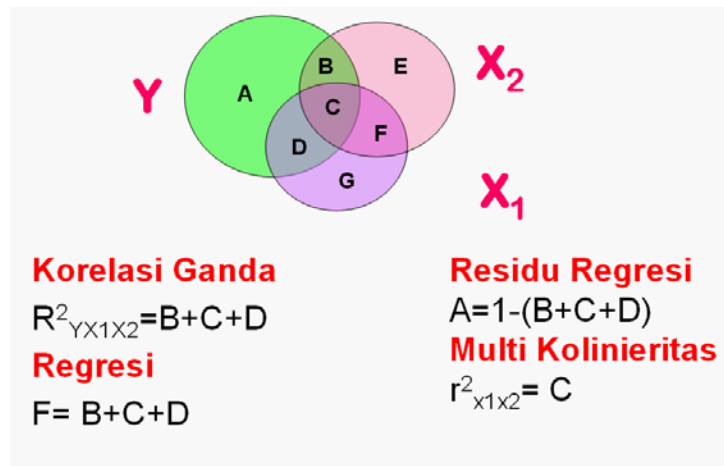
Model Summary									
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.785 ^a	.616	.568	1.758	.616	12.842	1	8	.007
2	.818 ^b	.669	.574	1.745	.053	1.118	1	7	.326

a. Predictors: (Constant), x2
b. Predictors: (Constant), x2, x1

Gambar 4. Hasil Analisis Regresi Berjenjang (X_2 masuk dulu baru diikuti X_1)

Regresi Berjenjang – Cara 2

Sama seperti membaca tabel di atas, terlihat bahwa sumbangan X_2 adalah 61.6% sedangkan sumbangan X_1 adalah 0,53%. Terlihat ada perbedaan dengan hasil analisis di atas. Mengapa kok berbeda? Seperti yang saya jelaskan di atas: SIAPA YANG MASUK DULUAN AKAN MENDAPATKAN WILAYAH IRISAN KETIGA VARIABEL. Kali ini wilayah irisan sudah dikuasai oleh X_2 sehingga sumbangan dia lebih banyak dibanding dengan sumbangan X_1 .



Gambar 5. Visualisasi Korelasi tiga Variabel

MULTIKOLINIERITAS

Saya menampilkan lagi Gambar 2, biar tidak usah melihat lagi ke atas. Multikolinieritas adalah kondisi ketika wilayah irisan antar prediktor (X) dengan kriteria (Y) terlalu tinggi. Dalam hal ini terlihat bahwa wilayah tersebut adalah wilayah C, WILAYAH KONFLIK yang menjadi perebutan dua prediktor. Bayangkan jika kedua prediktor tidak memiliki hubungan, maka wilayah konflik itu tidak ada, atau mungkin nilainya kecil.

Saya akan mendemonstrasikan dampak dari adanya multikolinieritas. Di bawah ini terlihat korelasi antara dua prediktor (X_1 dan X_2) dengan Y. Korelasi antara X_1 dan X_2 sangat tinggi ($r_{xy} = 0.955$). Kedua korelasi prediktor dengan Y sama-sama positif. X_1 -Y sama dengan 0.552 sedangkan X_2 -Y sama dengan 0.721.

Correlation Matrix			
	y	x1	x2
Correlation y	1.000	.552	.721
x1	.552	1.000	.955
x2	.721	.955	1.000

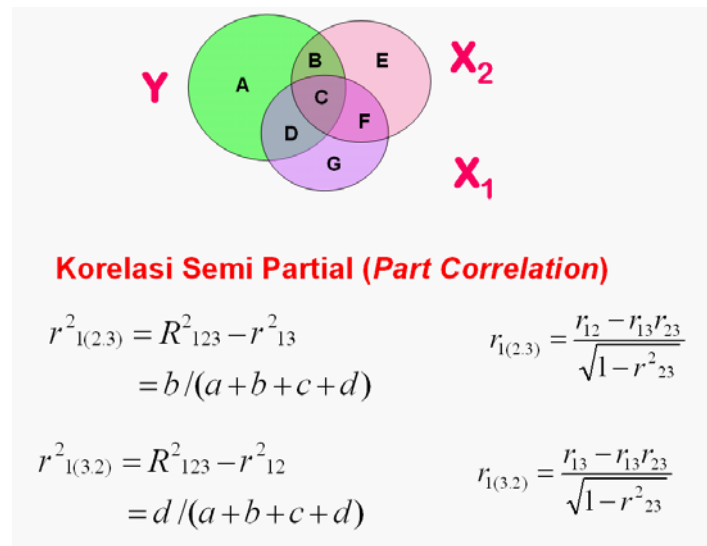
Gambar 6. Korelasi antara Prediktor (X_1 dan X_2) dengan Kriteria (Y)

Prediksi kedua prediktor terhadap Y melalui regresi menunjukkan hasil sebagai berikut :

Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	3.497	1.221		2.864
	x1	-1.536	.655	-1.547	2.344
	x2	2.172	.652	2.198	3.331

a. Dependent Variable: y

Gambar 7. Prediksi dua Prediktor (X1 dan X2) terhadap Kriteria (Y)
 Hasil tersebut menunjukkan bahwa meningkatnya nilai X1 akan menurunkan nilai Y, kok bisa? Padahal sebelumnya korelasi X1 dengan Y adalah positif, kok sekarang berubah jadi negatif? Ya, ini karena WILAYAH KONFLIK terlalu besar, wilayah itu direbut oleh X2 sehingga X1 tidak memiliki apa2 akhirnya prediksinya menjadi negatif. Inilah salah satu dampak adanya MULTIKOLINIERITAS.

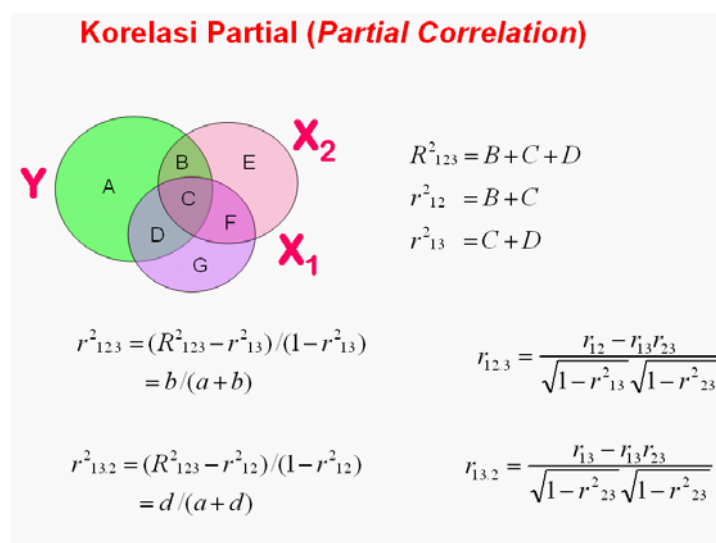


Gambar 7. Visualisasi Korelasi Semi Parsial

KORELASI SEMI PARSIAL

Korelasi semi parsial menunjukkan kontribusi unik dari prediktor setelah kontribusi dari prediksi lainnya dikendalikan hanya dari prediksi yang bersangkutan. Terlihat bahwa korelasi semi parsial antara X1 dan Y dengan mengendalikan X2 adalah area D dibagi jumlah area Y secara keseluruhan (A+B+C+D).

Kita tahu bahwa A+B+C+D=100 persen, karena merupakan wilayah variabel Y. X1 sebenarnya memiliki dua wilayah yang mampu menjelaskan Y, yaitu D dan C. Namun karena X2 dilibatkan sebagai kontrol, maka wilayah C harus direlakan dibuang, X1 harus ikhlas mendapatkan D saja. Dengan demikian korelasi semi parsial didefinisikan sebagai porsi wilayah unik X1 terhadap Y dibagi jumlah porsi Y secara keseluruhan.



Gambar 8. Visualisasi Korelasi Semi Parsial

KORELASI PARSIAL

Korelasi parsial menunjukkan kontribusi unik X_2 dibagi dengan porsi unik Y . Terlihat di gambar, misalnya kita menguji korelasi parsial antara X_1 dan Y dengan mengendalikan X_2 maka hasilnya adalah area D dibagi dengan $(D+A)$. Jadi yang dilibatkan di sini hanya wilayah unik saja antara X_1 dan Y . Jadi wilayah yang dimakan oleh variabel kontrol (X_2) adalah semua wilayah terkait dengannya yaitu wilayah B dan C . X_1 dan Y hanya sebagian wilayah D dan A .

Yogyakarta, 2009
Wahyu Widhiarso