

BAHAN AJAR

**ANALISIS KOLERASI DAN
REGRESI**

Disusun oleh:
Budi Subandriyo, S.ST, M.Stat

Diklat Statistisi Tingkat Ahli BPS Angkatan XXI
Badan Pusat Statistik
Tahun 2020

1. PENDAHULUAN

1.1. Pengertian Statistika Deskriptif

Statistika deskriptif adalah metode-metode yang berkaitan dengan pengumpulan dan penyajian suatu gugus data sehingga memberikan informasi yang berguna. Pengklasifikasian menjadi statistika deskriptif dan statistika inferensia dilakukan berdasarkan aktivitas yang dilakukan.

Statistika deskriptif hanya memberikan informasi mengenai data yang dipunyai dan sama sekali tidak menarik inferensia atau kesimpulan apapun tentang gugus induknya yang lebih besar. Contoh statistika deskriptif yang sering muncul adalah, tabel, diagram, grafik, dan besaran-besaran lain di majalah dan koran-koran. Dengan Statistika deskriptif, kumpulan data yang diperoleh akan tersaji dengan ringkas dan rapi serta dapat memberikan informasi inti dari kumpulan data yang ada. Informasi yang dapat diperoleh dari statistika deskriptif ini antara lain ukuran pemusatan data, ukuran penyebaran data, serta kecenderungan suatu gugus data

1.2. Pengertian Korelasi

Korelasi merupakan teknik analisis yang termasuk dalam salah satu teknik pengukuran asosiasi atau hubungan (*measures of association*). Pengukuran asosiasi merupakan istilah umum yang mengacu pada sekelompok teknik dalam statistik bivariat yang digunakan untuk mengukur kekuatan hubungan antara dua variabel. Diantara sekian banyak teknik-teknik pengukuran asosiasi, terdapat dua teknik korelasi yang sangat populer sampai sekarang, yaitu **Korelasi Pearson Product Moment dan Korelasi Rank Spearman**. Selain kedua teknik tersebut, terdapat pula teknik-teknik korelasi lain, seperti Kendal, *Chi-Square*, *Phi Coefficient*, *Goodman-Kruskal*, *Somer*, dan *Wilson*.

Pengukuran asosiasi mengenakan nilai numerik untuk mengetahui tingkatan asosiasi atau kekuatan hubungan antara variabel. Dua variabel dikatakan berasosiasi jika perilaku variabel yang satu mempengaruhi variabel yang lain. Jika tidak terjadi pengaruh, maka kedua variabel tersebut disebut *independen*.

Korelasi bermanfaat untuk mengukur kekuatan hubungan antara dua variabel (kadang lebih dari dua variabel) dengan skala-skala tertentu, misalnya Pearson data harus berskala interval atau rasio; Spearman dan Kendal menggunakan skala ordinal; *Chi Square* menggunakan data

nominal. Kuat lemah hubungan diukur diantara jarak (range) 0 sampai dengan 1. Korelasi mempunyai kemungkinan pengujian hipotesis dua arah (*two tailed*). Korelasi searah jika nilai koefisien korelasi ditemukan positif; sebaliknya jika nilai koefisien korelasi negatif, korelasi disebut tidak searah. Yang dimaksud dengan koefisien korelasi ialah suatu pengukuran statistik kovariansi atau asosiasi antara dua variabel. Jika koefisien korelasi ditemukan tidak sama dengan nol (0), maka terdapat ketergantungan antara dua variabel tersebut. Jika koefisien korelasi ditemukan +1, maka hubungan tersebut disebut sebagai korelasi sempurna atau hubungan linear sempurna dengan kemiringan (*slope*) positif.

Jika koefisien korelasi ditemukan -1, maka hubungan tersebut disebut sebagai korelasi sempurna atau hubungan linear sempurna dengan kemiringan (*slope*) negatif.

Dalam korelasi sempurna tidak diperlukan lagi pengujian hipotesis, karena kedua variabel mempunyai hubungan linear yang sempurna. Artinya variabel X mempengaruhi variabel Y secara sempurna. Jika korelasi sama dengan nol (0), maka tidak terdapat hubungan antara kedua variabel tersebut. Dalam korelasi sebenarnya tidak dikenal istilah variabel bebas dan variabel tergantung. Biasanya dalam penghitungan digunakan simbol X untuk variabel pertama dan Y untuk variabel kedua. Dalam contoh hubungan antara variabel remunerasi dengan kepuasan kerja, maka variabel remunerasi merupakan variabel X dan kepuasan kerja merupakan variabel Y.

1.3. Pengertian Regresi

Analisis regresi pada dasarnya adalah studi mengenai ketergantungan satu variabel *dependen* (terikat) dengan satu atau lebih variabel *independen* (variabel penjelas/bebas), dengan tujuan untuk mengestimasi dan/ atau memprediksi rata-rata populasi atau nilai rata-rata variabel *dependen* berdasarkan nilai variabel *independen* yang diketahui. Pusat perhatian adalah pada upaya menjelaskan dan mengevaluasi hubungan antara suatu variabel dengan satu atau lebih variabel *independen*.

Hasil analisis *regresi* adalah berupa koefisien *regresi* untuk masing-masing variabel *independen*. Koefisien ini diperoleh dengan cara memprediksi nilai variabel *dependen* dengan suatu persamaan. Koefisien regresi dihitung dengan dua tujuan sekaligus : Pertama, meminimumkan penyimpangan antara nilai aktual dan nilai estimasi variabel *dependen*; Kedua, mengoptimalkan korelasi antara nilai aktual dan nilai estimasi variabel *dependen* berdasarkan

data yang ada. **Teknik estimasi variable *dependen* yang melandasi analisis regresi disebut Ordinary Least Squares (pangkat kuadrat terkecil biasa)**

2. Studi Kasus

Studi kasus yang akan dibahas oleh praktikan adalah menghitung nilai statistika deskriptif masing-masing variabel, apakah ada korelasi antara kedua data, dan menginterpretasikan model regresinya menggunakan data berikut.

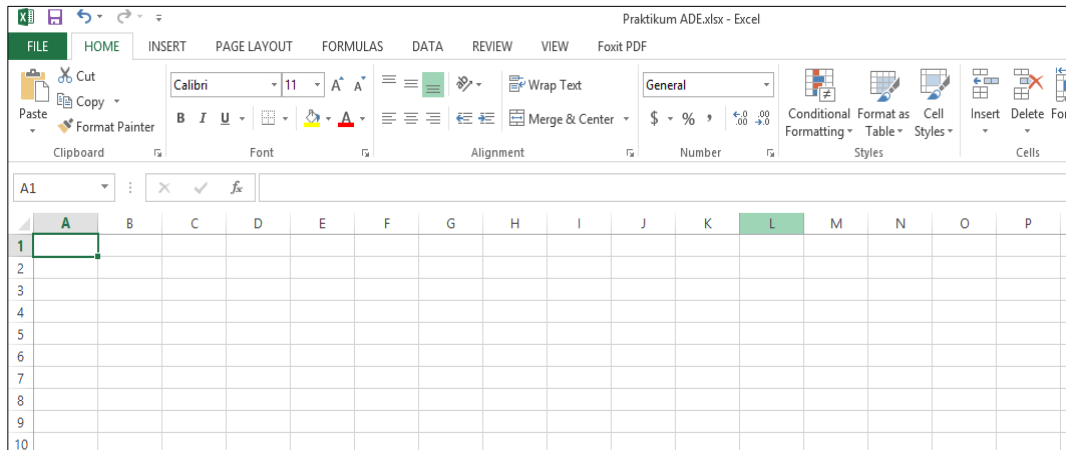
Provinsi	Jumlah Penduduk Miskin	Luas Daerah
Aceh	876.6	57956
Sumatera Utara	1378.4	72981.23
Sumatera Barat	397.9	42012.89
Riau	481.3	87023.66
Jambi	270.1	8201.72
Sumatera Selatan	1042	50058.16
Bengkulu	310.5	91592.43
Lampung	1219	16424.06
Kepulauan Bangka Belitung	70.2	19919.33
Kepulauan Riau	131.2	34623.8
DKI Jakarta	366.8	664.01
Jawa Barat	4421.5	35377.76
Jawa Tengah	4863.4	9662.92
DI Yogyakarta	562.1	32800.69
Jawa Timur	4960.5	3133.15
Banten	648.3	47799.75
Bali	161	5780.06
Nusa Tenggara Barat	828.3	18572.32
Nusa Tenggara Timur	1000.3	48718.1
Kalimantan Barat	355.7	147307
Kalimantan Tengah	141.9	153564.5
Kalimantan Selatan	189.2	38744.23
Kalimantan Timur	246.1	204534.34
Sulawesi Utara	177.5	13851.64
Sulawesi Tengah	409.6	11257.07
Sulawesi Selatan	805.9	61841.29
Sulawesi Tenggara	304.3	46717.48
Gorontalo	187.7	16787.18
Sulawesi Barat	160.6	38067.7
Maluku	338.9	46914.03
Maluku Utara	88.3	31982.5
Papua Barat	223.2	319036.05
Papua	976.4	97024.27

Gambar 2.1 Data Provinsi, Jumlah Penduduk Miskin dan Luas Daerah

2.1 Langkah Kerja

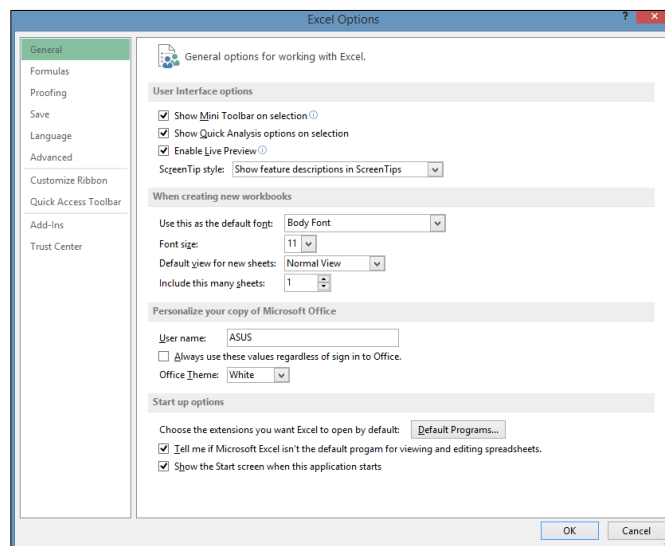
Setelah praktikan membuat kerangka konsep dan studi kasus, praktikan mengaplikasikannya pada *Microsoft excel* untuk menyelesaikan studi kasus. Adapun langkah kerja yang digunakan oleh praktikan adalah sebagai berikut:

2.1.1 Langkah pertama yang dilakukan adalah membuka lembar kerja *Microsoft excel*. Berikut adalah tampilan utama atau lembar kerja pada *Microsoft excel*.



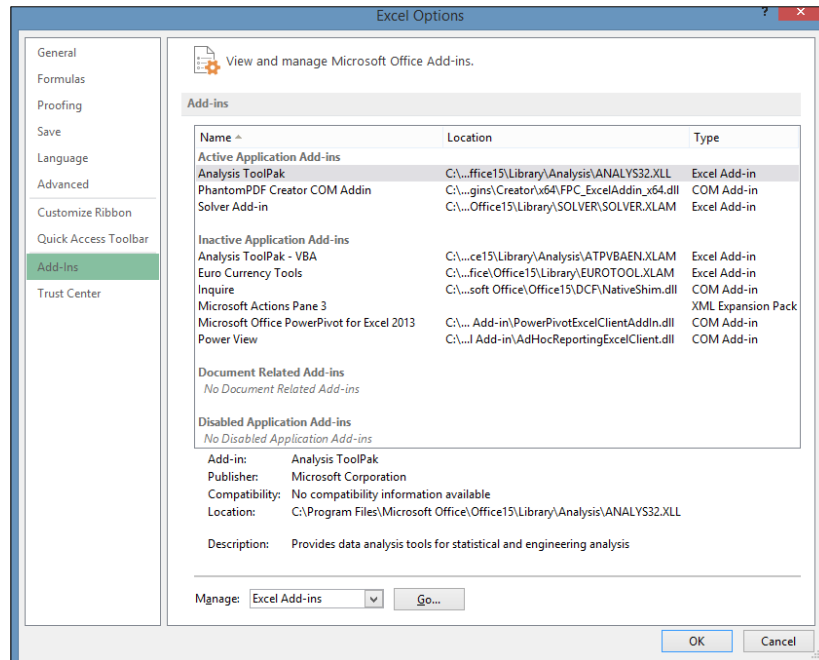
Gambar 2.1 Lembar Kerja *Microsoft excel*

2.1.2 Sebelum menyelesaikan persoalan terlebih dahulu praktikan mengaktifkan data analysis. Pada *menu file* mengklik *excel options*. Maka akan tampak seperti gambar di bawah ini



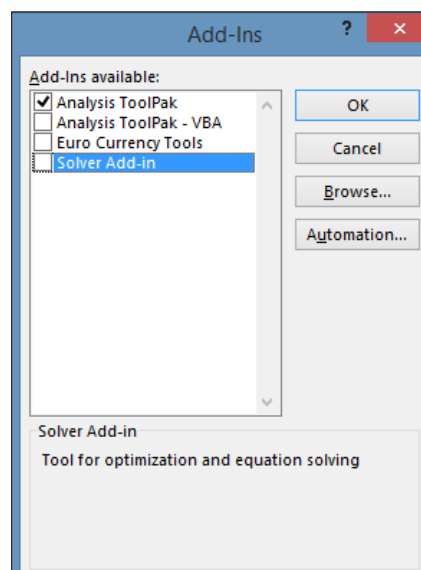
Gambar 2.2 Tampilan Excel Options

2.1.3 Kemudian memilih *Add-In* pada *Inactive Application Add-Ins*, lalu mengklik *Go*. Maka akan tampak gambar seperti di bawah ini.



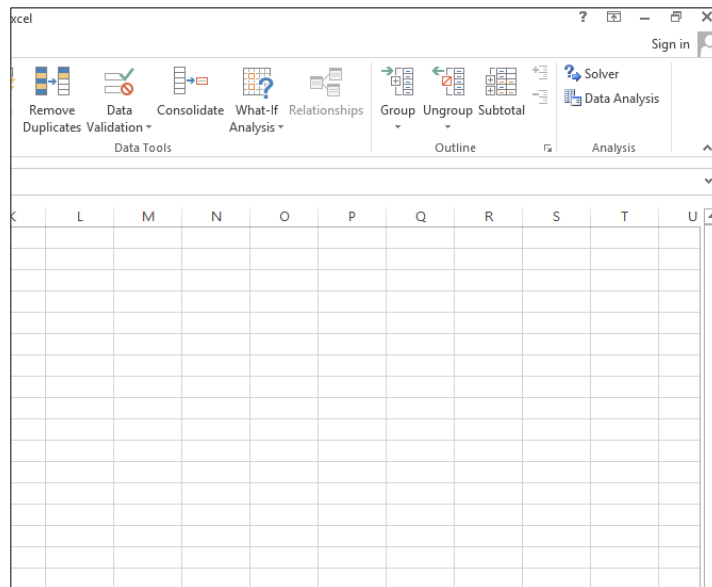
Gambar 2.3 Tampilan Add In pada Excel Options

2.1.4 Pada tampilan *Add-Ins available* memilih *Analysis Toolpak*, kemudian mengklik *OK* seperti gambar di bawah ini.



Gambar 2.4 Tampilan Add-Ins

2.1.5 Penambahan *menu Analysis* untuk menyelesaikan persoalan deskriptif, korelasi dan regresi sudah terselesaikan. Menu *Data Analysis* tersebut ada dalam *Data* pada *Menu Bar* seperti gambar di bawah ini.



Gambar 2.5 Tampilan *Data Analysis*

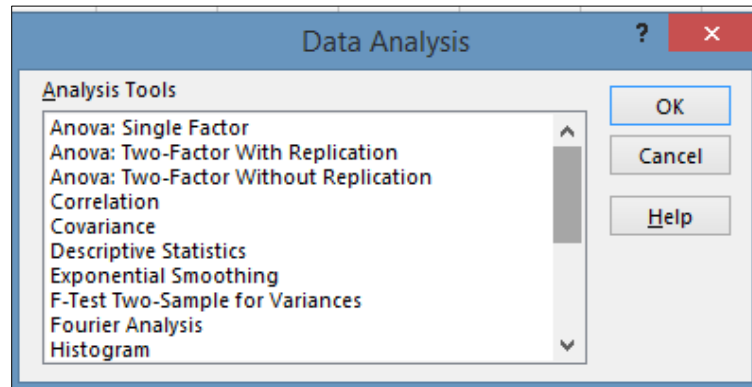
2.1.6 Setelah mengaktifkan *Data Analysis*, maka untuk menyelesaikan persoalan terlebih dahulu memasukkan data yang dibutuhkan pada *MS. Excel* seperti gambar di bawah ini.

The image shows a screenshot of an Excel spreadsheet with the following data:

Provinsi	Jumlah Penduduk Miskin	Luas Daerah
Aceh	876.6	57956
Sumatera Utara	1378.4	72981.23
Sumatera Barat	397.9	42012.89
Riau	481.3	87023.66
Jambi	270.1	8201.72
Sumatera Selatan	1042	50058.16
Bengkulu	310.5	91592.43
Lampung	1219	16424.06
Kepulauan Bangka Belitung	70.2	19919.33
Kepulauan Riau	131.2	34623.8
DKI Jakarta	366.8	664.01
Jawa Barat	4421.5	35377.76
Jawa Tengah	4863.4	9662.92
DI Yogyakarta	562.1	32800.69
Jawa Timur	4960.5	3133.15
Banten	648.3	47799.75
Bali	161	5780.06
Nusa Tenggara Barat	828.3	18572.32
Nusa Tenggara Timur	1000.3	48718.1

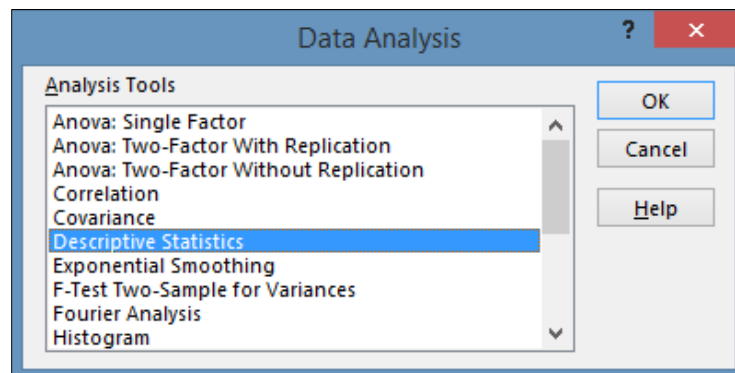
Gambar 2.6 Tahap Input Data pada *MS.Excel*

2.1.7 Untuk menyelesaikan persoalan yang pertama yaitu menghitung statistika deskriptif, maka memilih *Data analysis* yang ada pada menu data seperti yang sudah dijelaskan sebelumnya seperti gambar di bawah ini.



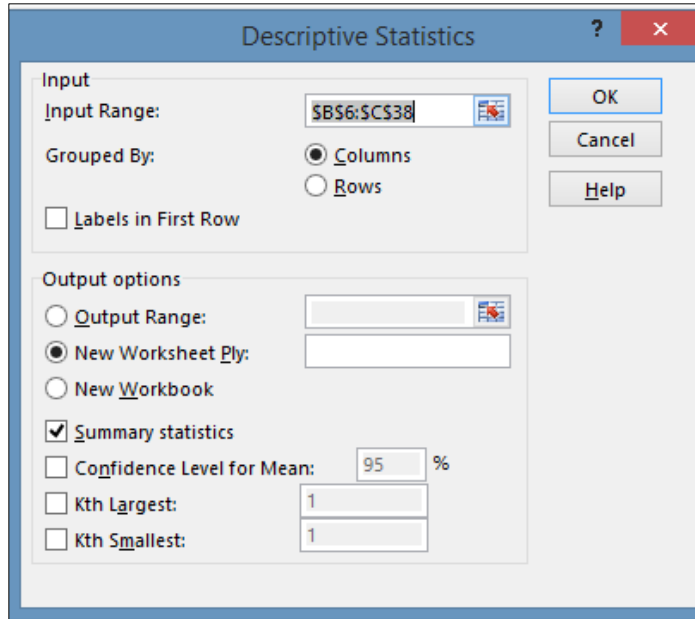
Gambar 2.7 Tahap Data Analysis

2.1.8 Selanjutnya pada lembar kerja *Data Analysis* di *Analysis Tools* memilih *Descriptive Statistics*, kemudian mengklik *OK* seperti gambar di bawah ini.



Gambar 2.8 Tahap Memilih *Descriptive Statistics*

2.1.9 Setelah Mengklik *OK* maka akan tampil kotak dialog *Descriptive Statistics*. Pada *Input Range* diisi dengan *cells* data dari jumlah penduduk miskin dan luas daerah, untuk *Grouped By* memilih *columns* dan untuk *output options* memilih *new worksheet by* dimana hasil output akan dicantumkan pada sheet baru di *MS. Excel* dan *summary statistics*, kemudian mengklik *OK* seperti gambar di bawah ini



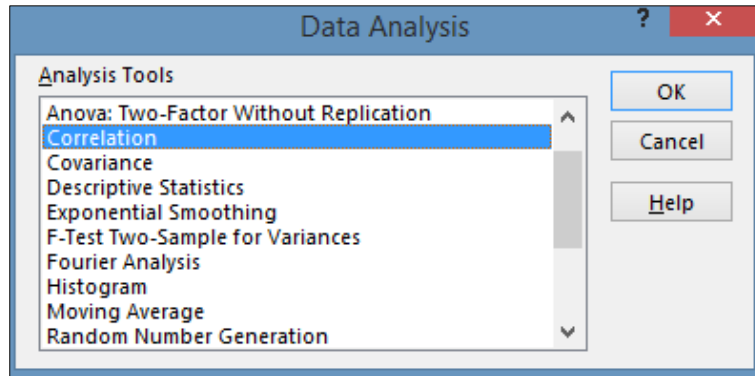
Gambar 2.9 Tahap Mengisi Descriptive Statistics

2.1.10 Setelah mengklik *OK*, maka akan didapatkan hasil seperti gambar di bawah ini.

Column1		Column2	
Mean	866.5060606	Mean	57907.0097
Standard Error	225.7303682	Standard Error	11462.93024
Median	366.8	Median	38744.23
Mode	#N/A	Mode	#N/A
Standard Deviation	1296.722242	Standard Deviation	65849.52088
Sample Variance	1681488.572	Sample Variance	4336159400
Kurtosis	6.022675669	Kurtosis	7.370532821
Skewness	2.63042541	Skewness	2.495654864
Range	4890.3	Range	318372.04
Minimum	70.2	Minimum	664.01
Maximum	4960.5	Maximum	319036.05
Sum	28594.7	Sum	1910931.32
Count	33	Count	33

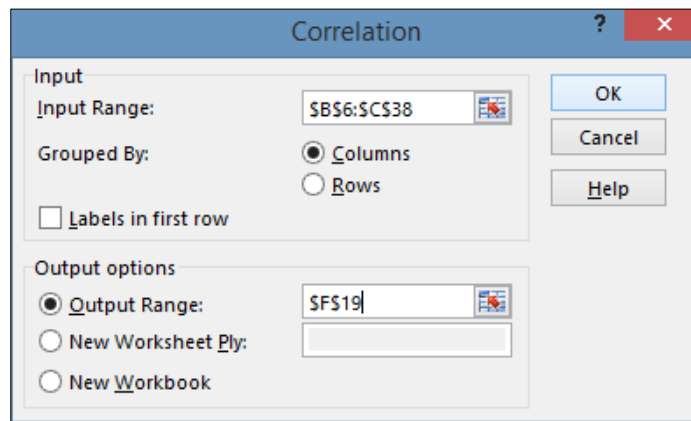
Gambar 2.10 Hasil Analisis pada *MS. Excel*

2.1.11 Setelah mendapatkan hasil dari *descriptive statistics*, kemudian menghitung korelasi yang bertujuan untuk mengetahui ada tidaknya hubungan antara jumlah penduduk miskin dan luas daerah. Hal tersebut dihitung dengan cara yang sama seperti sebelumnya. Namun, pada kotak *Data Analysis* dalam *analysis tools* memilih *Correlation*, kemudian *OK* seperti gambar di bawah ini.



Gambar 2.11 Tahap Memilih Correlation pada Data Analysis

2.1.12 Selanjutnya akan tampil kotak dialog *Correlation*. Dimana pada *Input Range* diisi dengan *cells* data jumlah penduduk miskin dan luas daerah. Sedangkan pada *group by* diisi dengan *columns* dan pada *output options* dipilih *output range* yang berarti hasil *output* akan ada pada *sheet* yang sama dengan data awal dan pada range tertentu, kemudian mengklik *OK* seperti gambar di bawah ini.



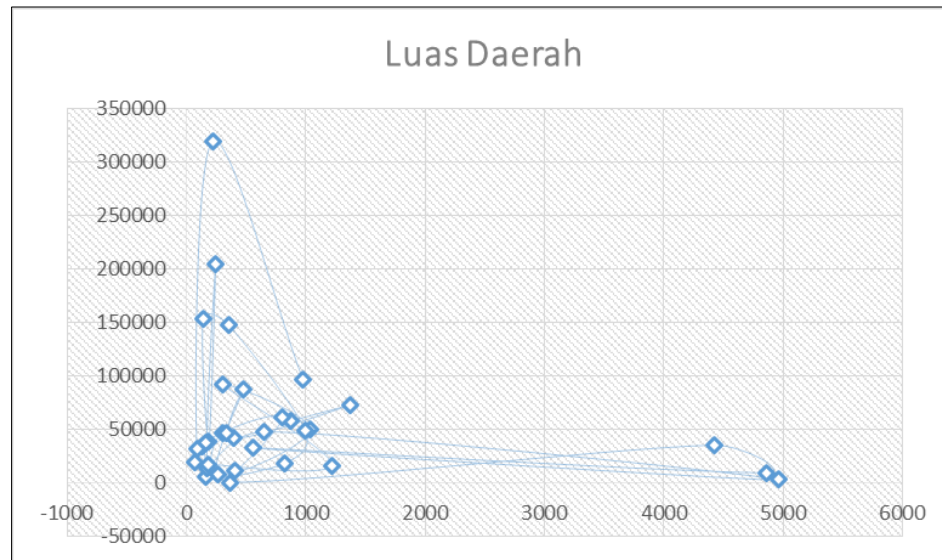
Gambar 2.12 Tahap Mengisi Lembar Kerja Correlation

2.1.13 Setelah mengklik *OK*, maka akan didapatkan hasil seperti gambar di bawah ini

	<i>Column 1</i>	<i>Column 2</i>
<i>Column 1</i>	1	
<i>Column 2</i>	-0.22195	1

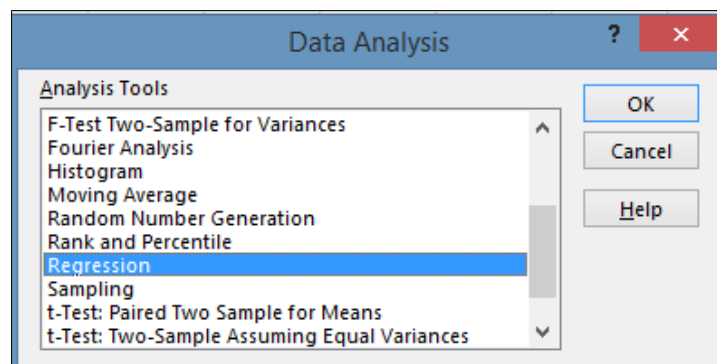
Gambar 2.13 Hasil Analisis Correlation pada *MS. Excel*

2.1.14 Untuk melihat ada tidaknya hubungan antara jumlah penduduk miskin dengan luas daerah dapat juga digambarkan dengan membuat *scatterplot* dari kedua data tersebut dengan cara memilih *menu insert* dan pada kumpulan charts terdapat *scatterplot* seperti di bawah ini.



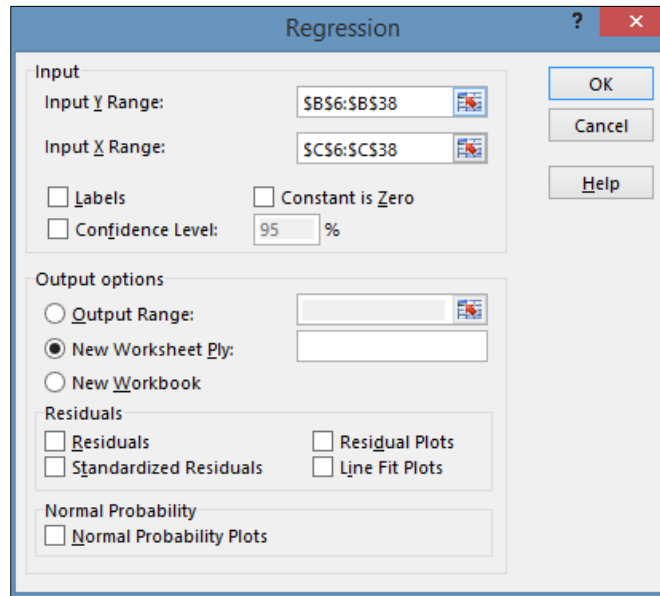
Gambar 2.14 Scatterplot dari Jumlah Penduduk Miskin dan Luas Daerah

2.1.15 Setelah mendapatkan hasil dari korelasi, kemudian menghitung regresi yang bertujuan untuk mengetahui besarnya hubungan antara jumlah penduduk miskin dan luas daerah. Hal tersebut dihitung dengan cara yang sama seperti sebelumnya. Namun, pada kotak dialog *Data Analysis* dalam *analysis tools* memilih *regression*, kemudian *OK* seperti gambar di bawah ini.



Gambar 2.15 Tahap Memilih Regression pada Lembar Kerja Data Analysis

2.1.16 Kemudian akan tampil kotak dialog *Regression*. Dimana pada *input Y range* diisi dengan *cells* data jumlah penduduk miskin dan pada *input X range* diisi dengan *cells* data luas daerah. Hal tersebut disebabkan oleh luas daerah mempengaruhi (*Independen*) jumlah penduduk miskin (*dependen*). Sedangkan *confidence level* diisi dengan 95% dan pada *output options* dipilih *new worksheet by* yang berarti hasil *output* akan ada pada *worksheet* baru atau berbeda dengan data awal, kemudian mengklik *OK* seperti gambar di bawah ini.



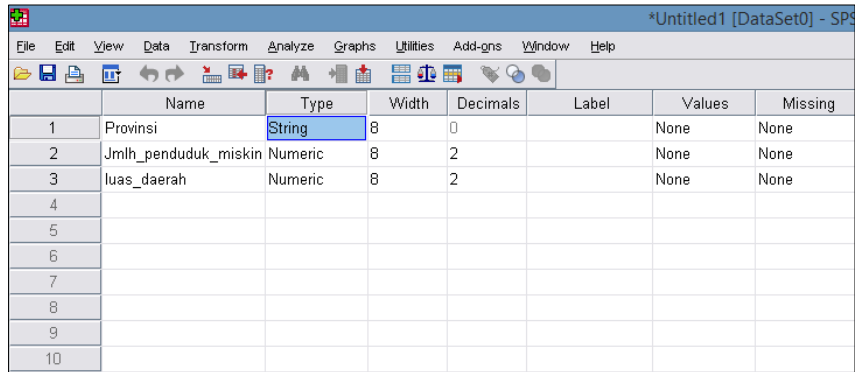
Gambar 2.16 Tahap Mengisi Lembar Kerja Regression

2.1.17 Setelah mengklik *OK*, maka akan didapatkan hasil seperti gambar di bawah ini

1	SUMMARY OUTPUT								
2									
3	Regression Statistics								
4	Multiple R	0.22195192							
5	R Square	0.04926265							
6	Adjusted R Square	0.01859371							
7	Standard Error	1284.61024							
8	Observations	33							
9									
10	ANOVA								
11		df	SS	MS	F	Significance F			
12	Regression	1	2650706.899	2650707	1.606271	0.214454864			
13	Residual	31	51156927.4	1650223					
14	Total	32	53807634.3						
15									
16		Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
17	Intercept	1119.60154	299.8104102	3.734365	0.00076	508.1341747	1731.0689	508.134175	1731.068901
18	X Variable 1	-0.00437072	0.003448608	-1.26739	0.214455	-0.011404204	0.0026628	-0.0114042	0.002662759
19									

Gambar 2.17 Hasil Analisis Correlation pada MS. Excel

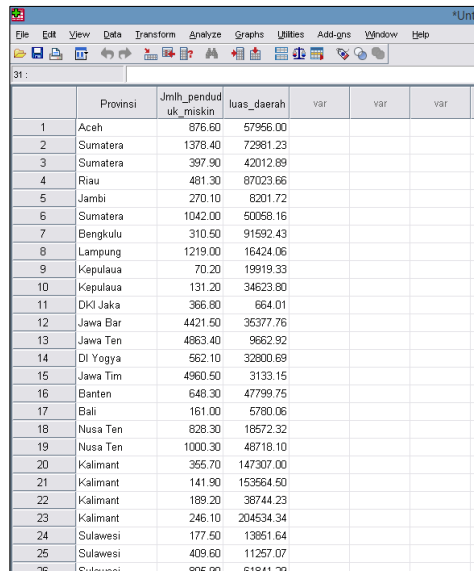
2.1.18 Permasalahan tersebut dapat diselesaikan dengan aplikasi lain yaitu *SPSS*. Langkah pertama membuka *SPSS*, pada variabel *view* menulis variabel yang dibutuhkan seperti gambar di bawah ini



	Name	Type	Width	Decimals	Label	Values	Missing
1	Provinsi	String	8	0		None	None
2	Jmlh_penduduk_miskin	Numeric	8	2		None	None
3	luas_daerah	Numeric	8	2		None	None
4							
5							
6							
7							
8							
9							
10							

Gambar 2.18 Tampilan Variable View

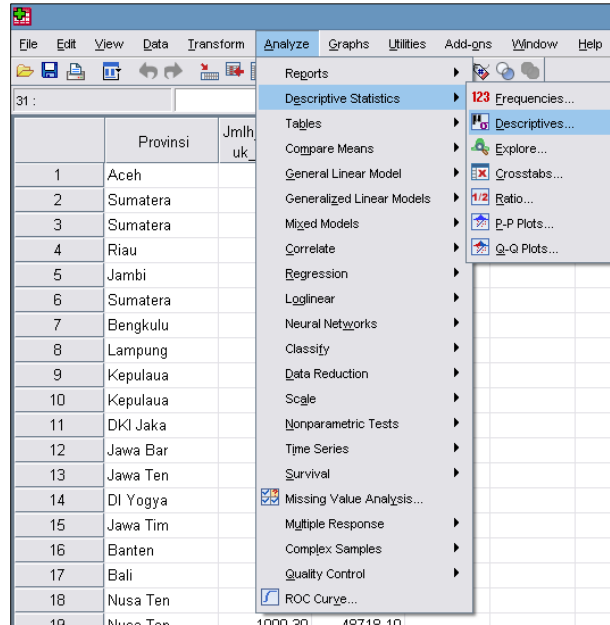
2.1.19 Setelah itu, pada lembar kerja data *view*, menuliskan data yang sudah dibuat di *Ms. Excel* seperti gambar di bawah ini.



	Provinsi	Jmlh_penduduk_miskin	luas_daerah	var	var	var
1	Aceh	876.60	57956.00			
2	Sumatera	1378.40	72981.23			
3	Sumatera	397.90	42012.89			
4	Riau	481.30	87023.66			
5	Jambi	270.10	8201.72			
6	Sumatera	1042.00	50058.16			
7	Bengkulu	310.50	91592.43			
8	Lampung	1219.00	16424.06			
9	Kepulauan	70.20	19919.33			
10	Kepulauan	131.20	34623.80			
11	Dki Jaka	366.80	664.01			
12	Jawa Bar	4421.50	36377.76			
13	Jawa Ten	4863.40	9662.92			
14	Di Yogya	562.10	32800.69			
15	Jawa Tim	4960.50	3133.15			
16	Banten	648.30	47799.75			
17	Bali	161.00	5780.06			
18	Nusa Ten	828.30	18572.32			
19	Nusa Ten	1000.30	48718.10			
20	Kalimant	355.70	147307.00			
21	Kalimant	141.90	153564.50			
22	Kalimant	189.20	38744.23			
23	Kalimant	246.10	204534.34			
24	Sulawesi	177.50	13851.64			
25	Sulawesi	409.60	11257.07			
26	Sulawesi	806.90	61841.79			

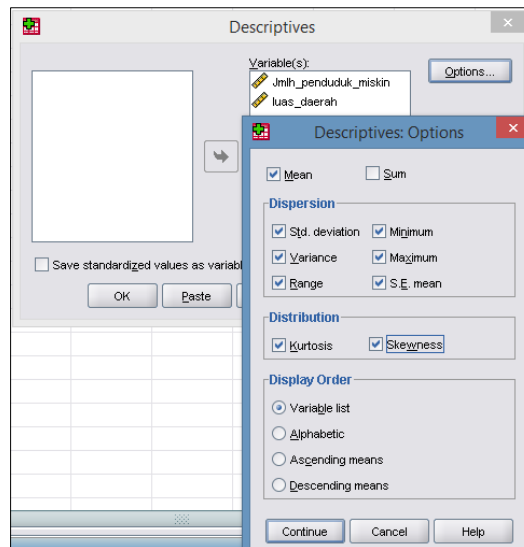
Gambar 2.19 Tampilan Data View

2.1.20 Kemudian untuk menghitung *Descriptive Statistics* dapat dilakukan dengan mengklik *Analyze* pada menu bar seperti gambar di bawah ini.



Gambar 2.19 Tampilan Analyze Descriptive Statistics

2.1.21 Maka akan muncul lembar kerja *Descriptive Statistic*, memasukkan variabel yang dibutuhkan ke kolom variable(s) seperti gambar di bawah ini.



Gambar 2.20 Tampilan Descriptive dan Descriptive Options

2.1.22 Setelah mengklik ok maka akan didapatkan hasil sebagai berikut.

Descriptives

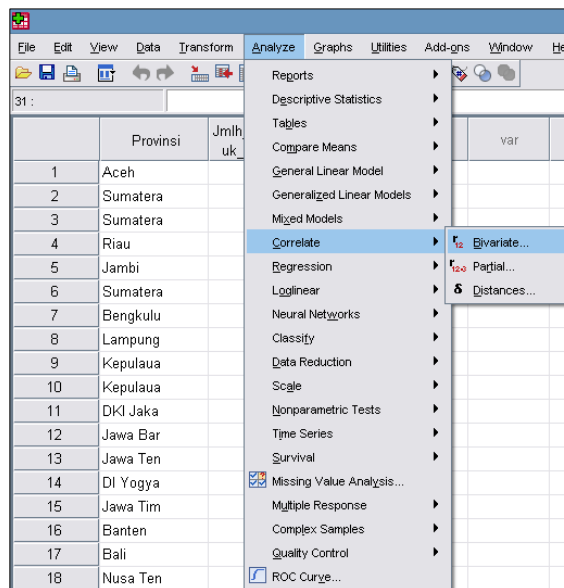
[DataSet0]

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean		Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Jmlh_penduduk_miskin	33	4890.30	70.20	4960.50	8.6651E2	2.2573E2	1296.72224	1.681E6	2.630	.409	6.023	.798
luas_daerah	33	3.18E5	664.01	3.19E5	5.7907E4	1.1462E4	65849.52088	4.336E9	2.496	.409	7.371	.798
Valid N (listwise)	33											

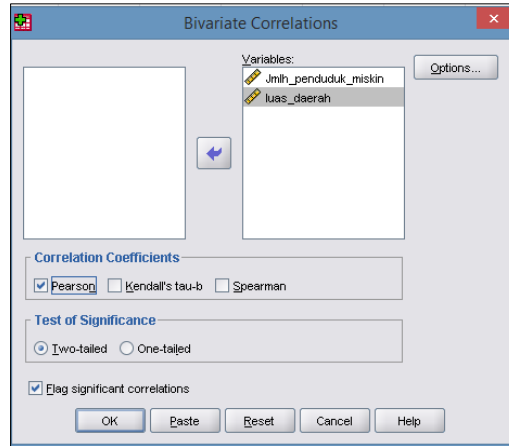
Gambar 2.21 Tampilan Hasil Descriptive Statistics

2.1.23 Dengan langkah yang sama untuk mencari korelasinya dengan mengklik *Analyze* seperti gambar di bawah ini.



Gambar 2.21 Tampilan Correlation Bivariate

2.1.24 Setelah itu, maka akan muncul lembar kerja *correlation-Bivariate*, memasukkan variabel yang dibutuhkan ke kolom variable(s) seperti gambar di bawah ini.



Gambar 2.21 Tampilan Lembar Kerja Correlation Bivariate

2.1.25 Setelah mengklik ok maka akan didapatkan hasil dari *correlation* sebagai berikut.

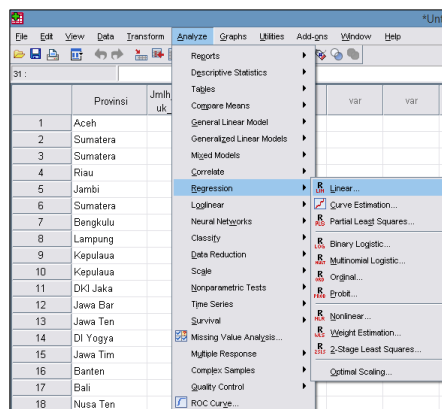
Correlations

[DataSet0]

		Jmih_penduduk_miskin	luas_daerah
Jmih_penduduk_miskin	Pearson Correlation	1	-.222
	Sig. (2-tailed)		.214
	N	33	33
luas_daerah	Pearson Correlation	-.222	1
	Sig. (2-tailed)	.214	
	N	33	33

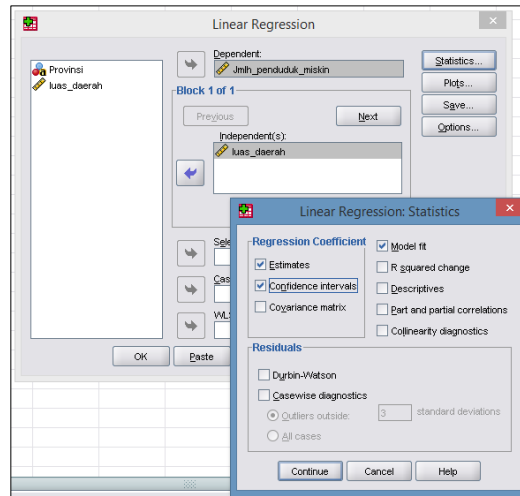
Gambar 2.22 Tampilan Hasil Correlation Bivariate

2.1.26 Analisis yang terakhir adalah *regression-Linear*, seperti langkah sebelumnya seperti gambar di bawah ini.



Gambar 2.23 Tampilan Analyze untuk Regression Linier

2.1.27 Setelah itu akan muncul lembar kerja *regression*, maka masukkan jumlah penduduk miskin pada kolom *dependen* dan luas daerah pada kolom *independent* seperti gambar di bawah ini.



Gambar 2.23 Tampilan Linier Regression dan Statisticnya

2.1.28 Maka setelah mengklik ok didapatkan hasil regresi seperti di bawah ini.

Variables Entered/Removed^a

Mode	Variables Entered	Variables Removed	Method
1	luas_daerah*	.	Enter

a. All requested variables entered.
b. Dependent Variable: Jmlh_penduduk_miskin

Model Summary

Mode	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.222 ^a	.049	.019	1284.61024

a. Predictors: (Constant), luas_daerah

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2650706.899	1	2650706.899	1.606	.214 ^a
	Residual	5.116E7	31	1650223.464		
	Total	5.381E7	32			

a. Predictors: (Constant), luas_daerah
b. Dependent Variable: Jmlh_penduduk_miskin

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	1119.602	299.810		3.734	.001	508.134	1731.069
	luas_daerah	-.004	.003	-.222	-1.267	.214	-.011	.003

a. Dependent Variable: Jmlh_penduduk_miskin

Gambar 2.24 Tampilan Hasil Linier Regression

3. PEMBAHASAN

3.1 Hasil Analisis Deskriptif

Berikut adalah hasil output dari analisis deskriptif yang diselesaikan dengan menggunakan *MS. Excel*.

Column1		Column2	
Mean	866.5060606	Mean	57907.0097
Standard Error	225.7303682	Standard Error	11462.93024
Median	366.8	Median	38744.23
Mode	#N/A	Mode	#N/A
Standard Deviation	1296.722242	Standard Deviation	65849.52088
Sample Variance	1681488.572	Sample Variance	4336159400
Kurtosis	6.022675669	Kurtosis	7.370532821
Skewness	2.63042541	Skewness	2.495654864
Range	4890.3	Range	318372.04
Minimum	70.2	Minimum	664.01
Maximum	4960.5	Maximum	319036.05
Sum	28594.7	Sum	1910931.32
Count	33	Count	33

Gambar 3.1 Hasil Analisis Descriptive Statistics

Berdasarkan hasil analisis di atas, terdapat dua kolom yaitu *column 1* yang mewakili jumlah penduduk miskin dan *column 2* mewakili luas daerah. Berdasarkan kedua hasil analisis dapat diketahui bahwa jumlah atau *count* data dari jumlah penduduk miskin dan luas daerah adalah 33 data.

Berdasarkan data di atas dapat diketahui bahwa jumlah penduduk miskin dan luas daerah memiliki masing-masing rata-rata 866,506 dan 57907,0097. Dimana rata-rata atau mean merupakan ukuran pemusatan yang sangat sering digunakan. Keuntungan dari menghitung rata-rata adalah angka tersebut dapat digunakan sebagai gambaran atau wakil dari data yang diamati.

Standar error mencerminkan keakuratan sample yang dipilih terhadap populasinya. Semakin kecil nilai *standard error*, semakin mengindikasikan bahwa sampling yang diambil bagus, atau cukup mewakili populasi yang sedang diteliti. Dan sebaliknya, makanya nilai *standard error* akan mengecil saat jumlah *sample* diperbanyak. Pada data di atas *standar error* dari jumlah penduduk miskin dan luas daerah masing-masing adalah 225,730 dan 11462,930. Standar error jumlah penduduk miskin lebih kecil dari luas daerah menunjukkan bahwa tingkat kesalahan pada pengambilan sample di jumlah penduduk miskin kecil.

Selanjutnya adalah nilai tengah atau sering disebut median yang merupakan suatu nilai ukuran pemusatan yang menempati posisi tengah setelah data diurutkan. Dimana ketika data sudah diurutkan dari terkecil ke terbesar akan memberikan median jika genap maka mediannya adalah data ke- n ditambah data ke- $n+1$ dibagi dua. Sedangkan jika data tersebut berjumlah ganjil, maka mediannya tepat ditengan.pada data di atas median dari jumlah penduduk miskin dan luas daerah masing-masing adalah 366,8 dan 38744,23..

Kemudian untuk mencari data yang paling banyak muncul dalam *sample* yang ada dalam statistika disebut dengan *mode*. *Mode* merupakan nilai yang paling sering muncul dari serangkaian data. Mode tidak dapat digunakan sebagai gambaran mengenai data. Dalam data di atas, tidak memiliki nilai mode dikarenakan tidak ada data yang sama.

Simpangan baku (*standar deviation*) dinotasikan sebagai s atau σ , menunjukkan rata-rata penyimpangan data dari harga rata-ratanya. Simpangan baku merupakan akar pangkat dua dari variansi. Simpangan baku dalam data jumlah penduduk miskin dan luas daerah berdasarkan analisis di atas adalah 1296.722242 dan 65849.52088

Variansi (*variance*) atau sample variance dinotasikan sebagai S^2 atau σ^2 adalah ukuran penyebaran data yang mengukur rata-rata kuadrat jarak seluruh titik pengamatan dari nilai tengah (*meannya*). Variansi ini menunjukkan keberagaman pada data yang ada. Dalam analisis data di atas variansi dari jumlah penduduk miskin dan luas daerah adalah 1681488.572 dan 4336159400. Adanya perbedaan nilai menunjukkan kedua data tersebut memiliki keberagaman yang berbeda.

Skewness adalah derajat ketidaksimetrisan suatu distribusi. Jika kurva frekuensi suatu distribusi memiliki ekor yang lebih memanjang ke kanan (dilihat dari meannya) maka dikatakan menceng kanan (positif) dan jika sebaliknya maka menceng kiri (negatif). Secara perhitungan, *skewness* adalah momen ketiga terhadap mean. Distribusi normal (dan distribusi simetris lainnya, misalnya distribusi t atau Cauchy) memiliki *skewness* 0 (nol). Sedangkan *kurtosis* adalah derajat keruncingan suatu distribusi (biasanya diukur relatif terhadap distribusi normal). Kurva yang lebih lebih runcing dari distribusi normal dinamakan *leptokurtik*, yang lebih datar platikurtik dan distribusi normal disebut *mesokurtik*. *Kurtosis* dihitung dari momen keempat terhadap mean. Distribusi normal memiliki *kurtosis* = 3, sementara distribusi yang *leptokurtik* biasanya *kurtosisnya* > 3 dan platikurtik $< >$. Data di jumlah penduduk miskin dan luas daerah masing-masing memiliki nilai *kurtosis* yaitu 6.022675669 dan 7.370532821. sedangkan, nilai *skeweness* dari jumlah penduduk miskin dan luas daerah adalah 2.63042541 dan 2.495654864.

Kemudian jumlah penduduk miskin dan luas daerah memiliki range masing-masing adalah 4890.3 dan 318372.04. Rentang (*Range*) yang biasanya dinotasikan sebagai R, menyatakan ukuran yang menunjukkan selisih nilai antara maksimum dan minimum. Rentang cukup baik digunakan untuk mengukur penyebaran data yang simetrik dan nilai datanya menyebar merata. Kemudian kedua data tersebut memiliki nilai maksimum dan minimum. Dimana nilai maksimum merupakan nilai terbesar yang ada dalam sekumpulan data tersebut. Sedangkan nilai minimum merupakan data terkecil yang ada pada data tersebut. Berdasarkan hasil di atas nilai maksimum jumlah penduduk miskin dan luas daerah secara berturut-turut adalah 4960.5 dan 319036.05. Sedangkan nilai minimum dari jumlah penduduk miskin dan luas daerah yaitu 70.2 dan 664.01.

Kemudian kedua data tersebut dapat dicari jumlah data secara keseluruhan. Berdasarkan perhitungan di atas jumlah data jumlah penduduk miskin adalah 28594.7 dan jumlah luas daerah adalah 1910931.32. Jumlah ini biasanya dalam statistika deskriptif disebut *sum*. Selain jumlah, rata-rata juga dapat dicari dengan menggunakan rumus dimana jumlah data dibagi dengan banyaknya data..

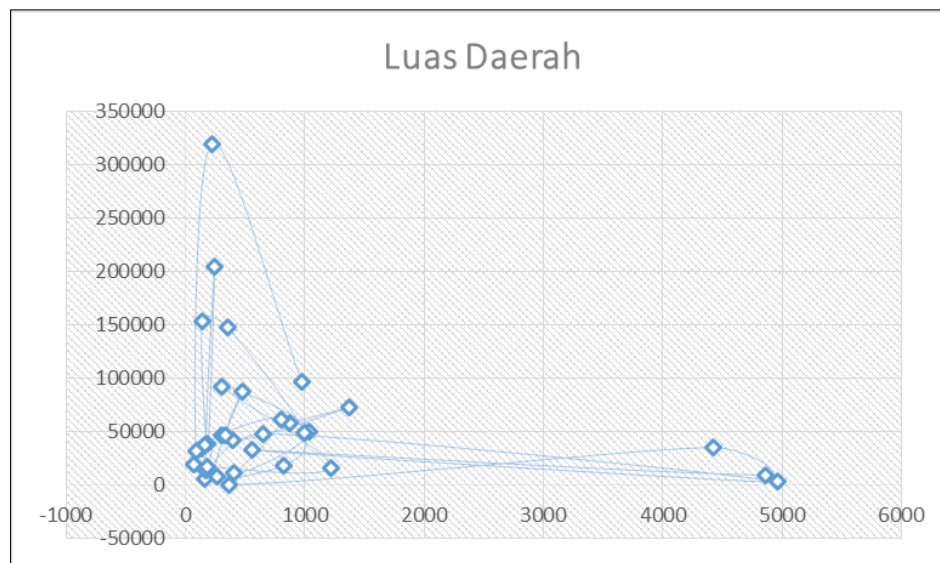
3.2 Hasil Analisis Korelasi

Di bawah ini merupakan hasil analisis korelasi antara jumlah penduduk miskin dengan luas daerah. Pada dasarnya korelasi merupakan sebuah analisis yang berfungsi untuk mengetahui hubungan antara variabel yang satu dengan variabel yang lainnya, yang berarti ketika satu variabel terjadi variabel yang lain dapat mempengaruhinya. Nilai korelasi (r) berkisar antara 1 sampai -1, nilai semakin mendekati 1 atau -1 berarti hubungan antara dua variabel semakin kuat, sebaliknya nilai mendekati 0 berarti hubungan antara dua variabel semakin lemah. Nilai positif menunjukkan hubungan searah (X naik maka Y naik) dan nilai negatif menunjukkan hubungan terbalik (X naik maka Y turun). Berikut adalah hasil analisis korelasi jumlah penduduk miskin dengan luas daerah yang dihitung menggunakan *MS. Excel*.

	<i>Column 1</i>	<i>Column 2</i>
<i>Column 1</i>	1	
<i>Column 2</i>	-0.22195	1

Gambar 3.2 Hasil Analisis Correlation

Menurut Sugiyono (2007) pedoman untuk memberikan interpretasi koefisien korelasi adalah 0,00-0,199 untuk sangat rendah, 0,20-0,399 untuk rendah, 0,40-0,599 untuk sedang, 0,60-0,799 untuk kuat dan 0,80 -1,000 untuk sangat kuat. Jadi berdasarkan keterangan di atas, maka dapat disimpulkan bahwa korelasi antara jumlah penduduk miskin dengan luas daerah memiliki hubungan yang rendah yaitu -0,22195. Pada hasil korelasi tersebut memiliki hasil negatif. Hal itu disebut Korelasi negatif yang berarti korelasi antara dua variabel dalam hal ini jumlah penduduk miskin dan luas daerah berjalan dengan arah yang berlawanan, bertentangan atau sebaliknya. Korelasi negatif terjadi jika antara dua variabel berjalan berlawanan yang berarti jika variabel X mengalami kenaikan maka variabel Y mengalami penurunan atau sebaliknya. Untuk melihat hubungan yang jelas antara kedua variabel tersebut, berikut adalah scatterplot dari kedua variabel.



Gambar 3.3 Scatterplot Jumlah Penduduk Miskin dan Luas Daerah

Scatterplot di atas menunjukkan bahwa semakin tinggi nilai luas daerah (variabel y) maka jumlah penduduk miskin (x) cenderung mengalami penurunan. Kedua variabel tersebut berlawanan satu sama lain, namun memiliki hubungan meskipun tidak terlalu signifikan.

3.3 Hasil Analisis Regresi

Berikut adalah hasil *output* dari analisis regresi antara variabel jumlah penduduk miskin dengan variabel luas daerah..

1	SUMMARY OUTPUT				
2					
3	<i>Regression Statistics</i>				
4	Multiple R	0.22195192			
5	R Square	0.04926265			
6	Adjusted R Square	0.01859371			
7	Standard Error	1284.61024			
8	Observations	33			
9					

Gambar 3.4 Hasil Summary Output untuk Regression Statistics

Tabel *Summary output* ini melaporkan kekuatan hubungan antara model (variabel bebas) dengan variabel terikat.

Pada *regression statistics*, *multiple R* (R majemuk) adalah suatu ukuran untuk mengukur tingkat (keeratan) hubungan linear antara variabel terikat dengan seluruh variabel bebas secara bersama-sama. Pada kasus dua variabel (satu variabel terikat dan satu variabel bebas), besaran *r* (biasa dituliskan dengan huruf kecil untuk dua variabel) dapat bernilai positif maupun negatif (antara -1 – 1), tetapi untuk lebih dari dua variabel, besaran *R* selalu bernilai positif (antara 0 – 1). Nilai *R* yang lebih besar (+ atau -) menunjukkan hubungan yang lebih kuat. Pada hasil analisis dua variabel di atas bahwa nilai *R* adalah 0. 22 yang menunjukkan bahwa kedua hubungan variabel tersebut tidak terlalu kuat.

R Square (R^2) sering disebut dengan koefisien determinasi, adalah mengukur kebaikan suai (*goodness of fit*) dari persamaan regresi yaitu memberikan proporsi atau persentase variasi total dalam variabel terikat yang dijelaskan oleh variabel bebas. Nilai R^2 terletak antara 0% sampai 100%, dan kecocokan model dikatakan lebih baik kalau R^2 semakin mendekati 100%. Berdasarkan hasil di atas nilai dari koefisien determinasinya adalah 0,049 atau 4.9%. Artinya 4.9% keragaman *y* mampu dijelaskan oleh *x* dalam model 4.9%, sedangkan sisanya dijelaskan oleh peubah lain yang diluar model.

Adjusted R Square merupakan suatu sifat penting R^2 adalah nilainya merupakan fungsi yang tidak pernah menurun dari banyaknya variabel bebas yang ada dalam model. Oleh karenanya, untuk membandingkan dua R^2 dari dua model, harus memperhitungkan banyaknya variabel bebas yang ada dalam model. Hal ini dapat dilakukan dengan menggunakan *adjusted R square*. Istilah penyesuaian berarti nilai R^2 sudah disesuaikan dengan banyaknya variabel (derajat bebas) dalam model. Pada dasarnya R^2 yang disesuaikan ini juga akan meningkat bersamaan

meningkatnya jumlah variabel, tetapi peningkatannya relatif kecil. Pada hasil analisis di atas, nilai dari *adjusted R square* sangat kecil yaitu 0,01859.

Standard Error merupakan standar error dari estimasi variabel terikat (dalam kasus ini adalah jumlah penduduk miskin). Angka ini dibandingkan dengan standar deviasi dari jumlah penduduk miskin. Semakin kecil angka *standar error* ini dibandingkan angka standar deviasi dari jumlah penduduk miskin maka model regresi semakin tepat dalam memprediksi jumlah penduduk miskin. Nilai *standar error* regresi adalah 1284,6

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	2650706.899	2650707	1.606271	0.214454864
Residual	31	51156927.4	1650223		
Total	32	53807634.3			

Gambar 3.5 Hasil Analisis ANOVA

Tabel *ANOVA (Analysis of Variance)* menguji penerimaan (*acceptability*) model dari perspektif statistik dalam bentuk analisis sumber keragaman. *ANOVA* ini sering juga diterjemahkan sebagai analisis ragam.

Dari tabel *ANOVA* tersebut diungkapkan bahwa keragaman data aktual variabel terikat (Jumlah penduduk miskin) bersumber dari model regresi dan dari residual. Dalam pengertian sederhana untuk kasus ini adalah variasi (turun-naiknya atau besar kecilnya) jumlah penduduk miskin disebabkan oleh variasi dari luas daerah (model regresi) serta dari faktor-faktor lainnya yang mempengaruhi jumlah penduduk miskin yang tidak dimasukkan dalam model regresi (*residual*).

Degree of Freedom (df) atau derajat bebas dari total adalah $n-1$, dimana n adalah banyaknya observasi. Dalam hal ini banyaknya observasi adalah 33 maka derajat bebas total adalah 32. Derajat bebas dari model regresi adalah 1, karena ada satu variabel bebas dalam model ini (luas daerah). Derajat bebas untuk residual adalah sisanya yaitu derajat bebas total – derajat bebas regresi = $33 - 1 = 32$.

Kolom *SS (Sum of Square)* atau jumlah kuadrat untuk *regression* diperoleh dari penjumlahan kuadrat dari prediksi variabel terikat (Jumlah penduduk miskin) dikurangi dengan nilai rata-rata jumlah penduduk miskin dari data sebenarnya. Jadi secara manual mencari terlebih dahulu rata-

rata permintaan dari data asli. Kemudian masing-masing prediksi jumlah penduduk miskin dikurangi dengan rata-rata tersebut kemudian dikuadratkan. Selanjutnya, seluruh hasil perhitungan tersebut dijumlahkan. Pada data di atas nilai *dari SS regression* adalah 2650706.899.

Kolom *SS* untuk residual diperoleh dari jumlah pengkuadratan dari residual. Pada hasil output di atas dapat diketahui bahwa nilai *SS* untuk residual adalah 51156927.4. Kolom *SS* untuk total adalah penjumlahan dari *SS* untuk regresi dengan dengan *SS* untuk residual. Sebenarnya *SS* total ini adalah variasi (besar-kecil,naik-turun) dari jumlah penduduk miskin. Hal ini diukur dengan mengurangi nilai masing-masing permintaan aktual dengan rata-ratanya, kemudian dikuadratkan. Hasil perhitungan tersebut kemudian dijumlahkan. Berdasarkan output di atas maka nilai *SS* total adalah 53807634.3

Hasil ketika *SS* tersebut memiliki arti dimana apanila *SS* total yang diperoleh adalah 53807634.3 yang memiliki arti, variasi dari jumlah penduduk miskin yang dikuadratkan adalah sebesar nilai tersebut. Bervariasinya jumlah penduduk miskin disebabkan oleh sebagian berasal dari variabel bebas (luas daerah) yaitu sebesar 2650706.899 (regresi). Kemudian sisanya sebesar 51156927.4 disebabkan oleh variabel lain yang juga mempengaruhi jumlah penduduk miskin tetapi tidak dimasukkan dalam model (residual).

Jika membandingkan (bagi) antara *SS* regresi dengan *SS* total, maka akan didapatkan proporsi dari total variasi jumlah penduduk miskin yang disebabkan oleh variasi luas daerah. Peraktikan mecoba membagi antara nilai *SS* regresi dengan *SS* total yaitu $2650706.899 / 53807634.3 = 0.049$. hasil tersebut sama dengan hasil dari R^2 atau koefisien determinasi yang telah dibahas di atas.

Selanjutnya kolom berikutnya dari *ANOVA* adalah kolom *MS (Mean of Square)* atau rata-rata jumlah kuadrat. Ini adalah hasil bagi antara kolom *SS* dengan kolom *df*. Berdasarkan hasil di atas besarnya *MS* adalah regresi dan residual berturut-turut adalah 2650707 dan 1650223. Dari perhitungan *MS* ini, selanjutnya dengan membagi antara *MS* Regresi dengan *MS* Residual didapatkan nilai *F*. Nilai *F* ini yang dikenal dengan *F* hitung dalam pengujian hipotesa dibandingkan dengan nilai *F* tabel. Jika *F* hitung > *F* tabel, maka dapat dinyatakan bahwa secara simultan (bersama-sama) luas daerah berpengaruh signifikan terhadap jumlah penduduk miskin. Setelah melakukan pembagian antara *MS* regresi dengan residul didapatkan hasil nilai *F* hitung yaitu 1.606271 seperti pada hasil analisis. Selain itu, dapat juga membandingkan antara taraf nyata dengan *p-value* (dalam istilah Excel adalah *Significance F*). Jika taraf nyata > dari *p-value* maka kesimpulannya sama dengan di atas

15									
16		<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
17	Intercept	1119.60154	299.8104102	3.734365	0.00076	508.1341747	1731.0689	508.134175	1731.068901
18	X Variable 1	-0.00437072	0.003448608	-1.26739	0.214455	-0.011404204	0.0026628	-0.0114042	0.002662759
19									
20									
21									
22									

Gambar 3.6 Hasil Analisis Intercept dan X Variabel

Tabel berikutnya dari *output Excel* seperti tabel di atas menampilkan nilai-nilai koefisien, *standard error*, *t Stat*, *P-value* dan selang kepercayaan. Dalam pengujian hipotesis regresi, tahap berikutnya setelah pengujian secara simultan (uji F seperti yang telah disampaikan sebelumnya) adalah pengujian koefisien regresi secara parsial. Pengertian pengujian secara parsial ini dalam kasus yang dibahas praktikan adalah apakah luas daerah mempengaruhi jumlah penduduk miskin.

Dalam uji parsial, digunakan uji t, yaitu membandingkan antara t-hitung (*t Stat*) dengan t tabel. Jika t hitung > t tabel pada taraf nyata tertentu, maka dapat disimpulkan variabel tersebut berpengaruh secara signifikan. T hitung ditampilkan pada kolom 4, yang merupakan hasil bagi antara kolom 2 (*coefficients*) dengan kolom 3 (Standard Error).

Selain membandingkan dengan nilai t-tabel, dapat juga ditarik kesimpulan signifikansinya dengan membandingkan taraf nyata dengan p-value (kolom 5). Jika dimisalkan dengan menggunakan taraf nyata 5 %, maka variabel dengan p-value sama atau lebih kecil dari 5 %, dapat dinyatakan sebagai variabel yang secara parsial berpengaruh signifikan.

Berdasarkan hal tersebut, terlihat bahwa luas daerah (x) dengan nilai v-palve sebesar 0.2 lebih besar dari 0.05 yang berarti bahwa luas daerah tidak terlalu berpengaruh pada jumlah penduduk miskin.

Selanjutnya, kolom 6 dan 7 memberikan selang kepercayaan untuk koefisien. Pada hasil *output* judulnya tertulis *Lower 95%* dan *Upper 95%* dimana angka 95% adalah penetapan pada waktu pengolahan dengan Excel dan bisa dirubah sesuai keinginan arti dari selang kepercayaan tersebut adalah nilai koefisien yang diberikan pada *output* regresi merupakan dugaan titik (*point estimate*) dari parameter koefisien regresi. Tingkat kepercayaan sebesar 95% dengan tingkat kesalahan sebesar 5% memberikan dugaan selang (*confidence interval*) kepercayaan sebesar 508.1341747 -1731.0689, dimana nilai paramater sebenarnya diharapkan berada dalam selang tersebut dengan tingkat kepercayaan tertentu.

Berdasarkan hal tersebut, dari *output Excel* terlihat bahwa dengan tingkat kepercayaan 95%, maka koefisien regresi untuk x atau luas daerah adalah -0.00437072. selanjutnya adalah membuat persamaan regresi atau model regresi dari permasalahan di atas.

Tabel 3.1 Model Regresi

$$\text{Jumlah penduduk miskin (Y)} = 1119.60154 - 0.00437072X.$$

Berdasarkan model regresi di atas bahwa konstanta yang sebesar 1119.60154 secara matematis berarti bahwa ketika variabel bebas nilainya 0, maka variabel terikat nilainya adalah sebesar konstanta tersebut.

Berdasarkan hasil pembahasan yang sudah dipaparkan oleh praktikan pada bab sebelumnya, maka dapat ditarik kesimpulan sebagai berikut.

1. Pada analisis Deskriptif menggunakan *Ms. Excel* didapatkan *mean, median, sum, standar of error, variance, skweeness, kurtosis* dan lain sebagainya pada jumlah penduduk miskin dengan luas daerah.
2. Berdasarkan analisis korelasi diketahui kedua variabel memiliki hubungan korelasi negatif sebesar -0,22195.
3. Hubungan antar kedua variabel tersebut tergolong dalam kategori sangat rendah.
4. Analisis regresi menunjukkan bahwa kedua variabel saling mempengaruhi dimana variabel luas daerah mempengaruhi jumlah penduduk miskin dengan koefisien determinasi (R^2) sebesar 4,9%.
5. Berdasarkan analisis regresi juga didapatkan model yaitu Jumlah penduduk miskin (Y) = 1119.60154 - 0.00437072X.
6. Kemampuan model yang sudah dibuat dalam menjelaskan permasalahan yang ada hanya 4,9%.

4. TUGAS MANDIRI

Diketahui data ekspor, harga suatu komoditas serta nilai tukar sebagai berikut :

TAHUN	EKS (Y)	HRG (X1)	KURS (X2)
1985	3678.8	248.48	5.65
1986	4065.3	331.48	10.23
1987	8431.4	641.88	13.5
1988	15718	100.8	13.84
1989	11891	536.69	12.66
1990	9349.7	332.25	13.98
1991	14561	657.6	15.69
1992	20148	928.1	16.62
1993	26776	1085.5	18.96
1994	43501	1912.2	22.05
1995	49223	2435.8	22.5
1996	65076	6936.7	20.6
1997	54941	3173.14	43
1998	58097	2107.7	70.67
1999	112871	2935.7	71.2
2000	108280	3235.8	84

Tentukan Persamaan Regresi dengan mengikuti langkah-langkah yang sudah dijelaskan (Asumsi Klasik, H0 dan H1, Kesimpulan Penelitian) sertakan dalam laporan materi yang sudah didapat dari pembelajaran mata diklat ini.

5. Regresi Logistik

Regresi logistik adalah sebuah pendekatan untuk membuat model prediksi seperti halnya regresi linear atau yang biasa disebut dengan istilah *Ordinary Least Squares (OLS) regression*. Perbedaannya adalah pada regresi logistik, peneliti memprediksi variabel terikat yang berskala dikotomi. Skala dikotomi yang dimaksud adalah skala data nominal dengan dua kategori, misalnya: Ya dan Tidak, Baik dan Buruk atau Tinggi dan Rendah.

Apabila pada OLS mewajibkan syarat atau asumsi bahwa error varians (residual) terdistribusi secara normal. Sebaliknya, pada regresi ini tidak dibutuhkan asumsi tersebut sebab pada regresi jenis logistik ini mengikuti distribusi logistik.

Asumsi Regresi Logistik

Asumsi Regresi Logistik antara lain:

1. Regresi logistik tidak membutuhkan hubungan linier antara variabel independen dengan variabel dependen.
2. Variabel independen tidak memerlukan asumsi *multivariate normality*.
3. Asumsi homokedastisitas tidak diperlukan
4. Variabel bebas tidak perlu diubah ke dalam bentuk metrik (interval atau skala ratio).
5. Variabel dependen harus bersifat dikotomi (2 kategori, misal: tinggi dan rendah atau baik dan buruk)
6. Variabel independen tidak harus memiliki keragaman yang sama antar kelompok variabel
7. Kategori dalam variabel independen harus terpisah satu sama lain atau bersifat eksklusif
8. Sampel yang diperlukan dalam jumlah relatif besar, minimum dibutuhkan hingga 50 sampel data untuk sebuah variabel prediktor (independen).
9. Dapat menyeleksi hubungan karena menggunakan pendekatan non linier log transformasi untuk memprediksi odds ratio. Odd dalam regresi logistik sering dinyatakan sebagai probabilitas.

Model Persamaan Regresi Logistik

Model persamaan aljabar layaknya OLS yang biasa kita gunakan adalah berikut: $Y = B_0 + B_1X + e$. Dimana e adalah error varians atau residual. Dengan model regresi ini, tidak menggunakan interpretasi yang sama seperti halnya persamaan regresi OLS. Model Persamaan yang terbentuk berbeda dengan persamaan OLS.

Berikut persamaannya:

$$\ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = B_0 + B_1X$$

Ln: Logaritma Natural. Di mana:

$B_0 + B_1X$: Persamaan yang biasa dikenal dalam OLS.

Sedangkan \hat{p} Aksen adalah probabilitas logistik yang didapat rumus sebagai berikut:

$$\hat{p} = \frac{\exp(B_0 + B_1X)}{1 + \exp(B_0 + B_1X)} = \frac{e^{B_0+B_1x}}{1 + e^{B_0+B_1x}}$$

Di mana:

exp atau ditulis "e" adalah fungsi exponen.

(Perlu diingat bahwa exponen merupakan kebalikan dari logaritma natural. Sedangkan logaritma natural adalah bentuk logaritma namun dengan nilai konstanta 2,71828182845904 atau biasa dibulatkan menjadi 2,72).

Dengan model persamaan di atas, tentunya akan sangat sulit untuk menginterpretasikan koefisien regresinya. Oleh karena itu maka diperkenalkanlah istilah Odds Ratio atau yang biasa disingkat Exp(B) atau OR. Exp(B) merupakan exponen dari koefisien regresi. Jadi misalkan nilai slope dari regresi adalah sebesar 0,80, maka Exp(B) dapat diperkirakan sebagai berikut:

$$2,72^{0,8} = 2,23$$

Nilai Odds Ratio

Besarnya nilai Exp(B) dapat diartikan sebagai berikut:

Misalnya nilai Exp (B) pengaruh rokok terhadap terhadap kanker paru adalah sebesar 2,23, maka disimpulkan bahwa orang yang merokok lebih beresiko untuk mengalami kanker paru

dibandingkan dengan orang yang tidak merokok. Interpretasi ini diartikan apabila pengkodean kategori pada tiap variabel sebagai berikut:

1. Variabel bebas adalah Rokok: Kode 0 untuk tidak merokok, kode 1 untuk merokok.
2. Variabel terikat adalah kanker Paru: Kode 0 untuk tidak mengalami kanker paru, kode 1 untuk mengalami kanker paru.

Pseudo R Square

Perbedaan lainnya yaitu pada regresi ini tidak ada nilai "R Square" untuk mengukur besarnya pengaruh simultan beberapa variabel bebas terhadap variabel terikat. Dalam regresi logistik dikenal istilah *Pseudo R Square*, yaitu nilai **R Square Semu** yang maksudnya sama atau identik dengan R Square pada OLS.

Jika pada OLS menggunakan uji F Anova untuk mengukur tingkat signifikansi dan seberapa baik model persamaan yang terbentuk, maka pada regresi ini menggunakan Nilai Chi-Square. Perhitungan nilai Chi-Square ini berdasarkan perhitungan Maximum Likelihood.

6. Referensi

Draper, N. dan Smith, H. 1992. Analisis Regresi Terapan. Edisi Kedua. Terjemahan Oleh Bambang Sumantri. Jakarta: Gramedia Pustaka Utama.

Gujarati, N.D. 2003. Basic Econometrics. 4th ed. New York: McGraw-Hill Companies, Inc.

Kutner, M.H., C.J. Nachtsheim., dan J. Neter. 2004. Applied Linear Regression Models. 4th ed. New York: McGraw-Hill Companies, Inc.

Santoso, S. 2000. Buku Latihan SPSS Statistik Parametrik. Jakarta: Elex Media Komputindo.