

Statistik Deskriptif dengan Microsoft Office Excel

Junaidi

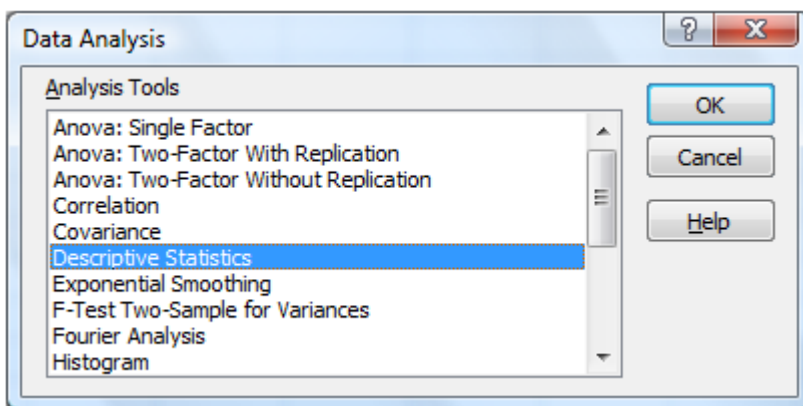
I. Prosedur Statistik Deskriptif pada Excel

Statistik deskriptif adalah statistik yang bertujuan untuk mendeskripsikan atau menggambarkan data. Terdapat berbagai cara dalam mendeskripsikan data, yang salah satunya adalah dalam bentuk ukuran-ukuran numerik dari hasil pengolahan terhadap data tersebut.

Di dalam Excel, terdapat fungsi-fungsi yang berguna untuk penggambaran (pendeskripsian) data, misalnya fungsi AVERAGE untuk rata-rata, STDEV untuk standar deviasi dan lainnya. Tulisan ini memberikan berbagai fasilitas Excel yang bisa memunculkan sekaligus beberapa pengukuran yang berguna untuk pendeskripsian data.

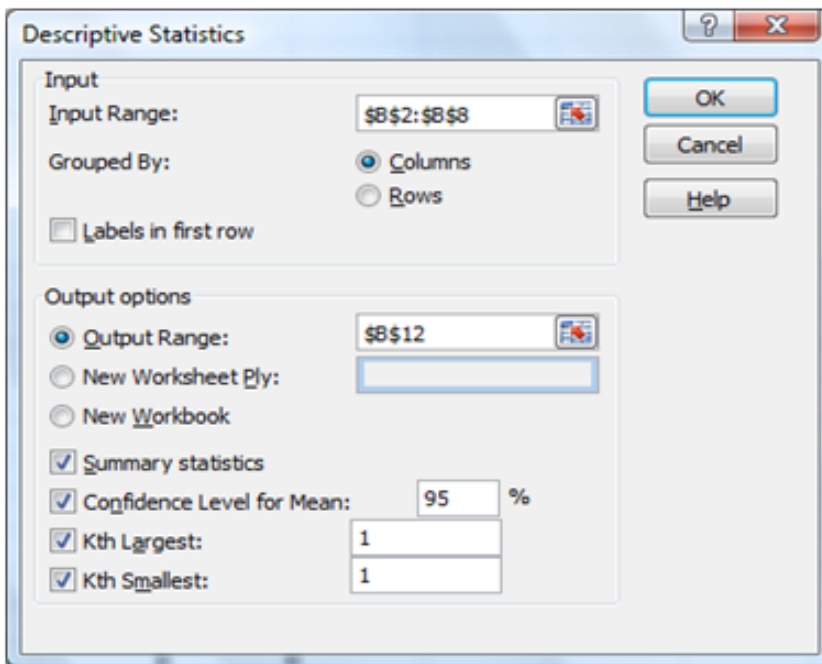
Tahap-Tahapnya sebagai berikut:

1. Klik menu **Tool** kemudian klik **Data Analysis**. (Catatan: jika setelah mengklik **Tool**, ternyata tidak muncul pilihan **Data Analysis**, berarti menu tersebut belum diaktifkan di program Excel Anda. Untuk mengaktifkannya, klik **Tool**, kemudian klik **Add ins**, selanjutnya conteng pada pilihan **Analysis Toolpak**, setelah itu klik **ok**. Lalu ulangi tahap 1 ini).
2. Setelah itu akan muncul tampilan berikut:



Klik Descriptive Statistics, kemudian Ok.

3. Setelah itu akan muncul tampilan berikut



Pada input range, masukkan range data yang akan diolah. Dalam contoh diatas adalah pada range B2:B8. Kemudian klik Output Range, dan masukkan sel awal dimana hasil akan dimunculkan. Dalam contoh ini ditempatkan di sel B12. Anda bisa menempatkan hasil pada halaman yang berbeda dari data dengan mengklik *New Worksheet Ply*.

Selanjutnya, conteng kotak pilihan-pilihan statistik yang ingin dimunculkan. Jika hanya ingin memunculkan summary statistic, maka klik kotak tersebut. Atau klik semua kotak pilihan statistik untuk memunculkan semua perhitungan.

Perhitungan-perhitungan summary statistic yang dikeluarkan adalah:

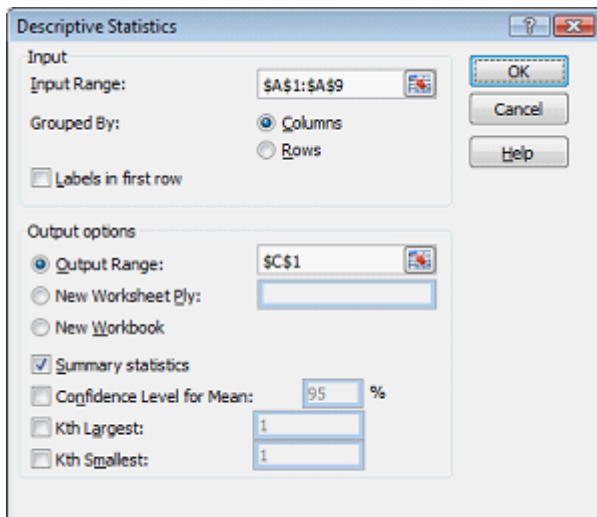
Mean (rata-rata)	Skewness
Standard Error	Range (selisih data terkecil dgn terbesar)
Median (nilai tengah)	Minimum (data terendah)
Mode (modus)	Maximum (data tertinggi)
Standard Deviation	Sum (jumlah data)
Sample Variance	Count (banyaknya data)
Kurtosis	

II. Pengertian dan Interpretasi Output Statistik Deskriptif pada Excel

Sebagai latihan, ketikkan kumpulan data berikut: 20 40 60 80 100 130 130 160 180. Tempatkan di kolom A mulai dari sel A1 sampai A9 (lihat tampilan 2).

Selanjutnya, klik option, Data Analysis maka akan tampilan seperti tampilan berikut:

Tampilan menu Descriptive Statistics



Selanjutnya, isilah input range nya sesuai dengan penempatan data kita tadi. Kemudian isilah output range di C1 , selanjutnya conteng kotak di depan “summary statistics”, kemudian klik Ok. Akan muncul tampilan output seperti tampilan berikut:

Tampilan Output Descriptive Statistics

	A	B	C	D
1	20		Column1	
2	40			
3	60		Mean	100
4	80		Standard Error	18.1812
5	100		Median	100
6	130		Mode	130
7	130		Standard Deviation	54.5436
8	160		Sample Variance	2975
9	180		Kurtosis	-1.1764
10			Skewness	-0.0178
11			Range	160
12			Minimum	20
13			Maximum	180
14			Sum	900
15			Count	9

Terdapat dua bagian pokok dari statistik deskriptif secara numerik, yaitu pengukuran nilai sentral dan pengukuran penyebaran data. Pengukuran nilai sentral adalah penentuan suatu nilai yang mampu menggambarkan/mewakili sekumpulan data. Pengukuran penyebaran data adalah penggambaran bagaimana data tersebut tersebar atau terdistribusi dari nilai sentralnya, atau perbandingan distribusi data tersebut terhadap distribusi normal.

A. Pengukuran Nilai Sentral

1. Mean (Rata-rata)

Dalam output Excel pada menu descriptive statistics, nilai rata-rata yang ditampilkan adalah rata-rata hitung (arithmetic mean). Rata-rata hitung ini adalah pengukuran nilai sentral yang paling umum digunakan. Dalam keseharian kita biasanya mengenal hanya dengan istilah rata-rata.

Rumus untuk menentukan nilai rata-rata hitung:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

Dimana:

\bar{X} = rata-rata hitung

$\sum_{i=1}^n X_i$ = jumlah nilai data dari data pertama sampai data ke n

n = banyaknya data

Nilai rata-rata hitung yang diperoleh sangat dipengaruhi oleh nilai-nilai ekstrim (outlier) dari kumpulan data.

Dari data yang terlihat pada tampilan 2, banyaknya data (count=n) adalah 9 dan jumlah nilai data (sum)

($\sum_{i=1}^n X_i$) adalah 900, sehingga rata-rata hitungnya:

$$\bar{X} = \frac{900}{9} = 100$$

Misalkan lagi, kita punya sekumpulan data lain berikut: 20 40 60 80 100 130 130 160 630.

Kalau kita perhatikan, kumpulan data ini mirip dengan kumpulan data pada tampilan 2, tetapi terdapat satu data (data terakhir) yang bernilai ekstrim (nilainya berada jauh dari kebanyakan nilai data yang ada). Kalau kita hitung rata-ratanya adalah 150. Dengan kata lain, hanya satu nilai ekstrim saja, nilai rata-rata yang kita peroleh 1,5 kali lebih besar dibandingkan nilai rata-rata sebelumnya.

2. Median

Median merupakan ukuran yang kuat (robust) dari nilai sentral. Hal ini dikarenakan nilai median tidak dipengaruhi oleh nilai ekstrim.

Median adalah nilai yang berada di tengah-tengah dari sekumpulan data, jika data tersebut diurutkan baik dari nilai terkecil ke nilai terbesar maupun dari nilai terbesar ke nilai terkecil. Secara rumus, median terletak pada urutan ke $(n+1)/2$, dimana n adalah banyaknya data.

Sebagai contoh, dari tampilan 2, banyaknya data kita adalah 9, maka mediannya terletak pada urutan ke $(n+1)/2 = (9+1)/2 = 5$. Data urutan ke 5 pada kumpulan data kita bernilai 100. Artinya, median data tadi adalah 100. (Catatan: perhatikan, meskipun angka terakhir pada data tampilan 2 kita ganti dengan angka ekstrim (misalnya 630 seperti contoh sebelumnya), median kita nilainya juga 100. Menunjukkan bahwa nilai median tidak dipengaruhi oleh nilai ekstrim).

Dalam kasus lain, misalnya kita punya kumpulan data berikut: 5 7 10 13 20 22. Banyaknya data adalah 6, dengan demikian median terletak pada urutan ke $(n+1)/2 = (6+1)/2 = 3,5$. Ini berarti median terletak ditengah data urutan ketiga (10) dan keempat (13). Dengan demikian, nilai mediannya adalah $(10+13)/2 = 11,5$.

3. Mode (Modus)

Modus (Mode) adalah nilai yang paling sering muncul dari sekumpulan data. Sebagaimana halnya dengan median, nilai modus juga tidak dipengaruhi oleh nilai ekstrim.

Modus dapat digunakan baik pada data numerik maupun data kategorik/kualitatif untuk menggambarkan nilai sentral atau nilai rata-rata. Meskipun demikian, modus lebih umum digunakan untuk data kategorik atau kualitatif.

Dalam keseharian, sebenarnya kita sudah sering menggunakan modus ini. Misalnya pernyataan "sekarang lagi mode perempuan berambut pendek". Ini artinya, banyak ditemukan perempuan berambut pendek. Dalam pengertian lain dapat juga diartikan, **rata-rata** perempuan berambut pendek.

Pada tampilan sebelumnya output Excel menghasilkan modus dengan nilai 130, karena angka tersebut muncul dua kali dalam distribusi data kita, sedangkan yang lainnya hanya satu kali. Pada sekumpulan data, bisa saja tidak terdapat mode, atau bahkan terdapat lebih dari satu mode. Pada kumpulan data berikut: 5 7 8 12 14 18 20, maka tidak ada modus pada data tersebut. Pada kumpulan data berikut: 5 7 7 8 12 14 18 18 20, maka terdapat dua modus yaitu 7 dan 18. (Catatan: pada Excel, ketika tidak ada modus, tampilan yang dihasilkan adalah #N/A. Sebaliknya, jika terdapat lebih dari satu modus, maka yang ditampilkan sebagai modus adalah modus yang berada pada urutan awal)

B. Ukuran Dispersi (Penyebaran) Data

1. Range (Jarak)

Merupakan pengukuran yang paling sederhana untuk dispersi data. Rumus untuk range adalah:

Range = nilai maksimum – nilai minimum

Dalam kasus kita, range nya adalah 160, karena nilai maksimum 180 dan nilai minimum 20.

2. Variance (Varians)

Varians adalah suatu ukuran penyebaran data, yang diukur dalam pangkat dua dari selisih data terhadap rata-ratanya.

Untuk data populasi perumusan varians adalah:

$$S^2 = \frac{1}{n} \sum (X_i - \bar{X})^2$$

Sedangkan untuk data sampel adalah:

$$S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$$

Dalam output Excel, rumus yang digunakan adalah rumus untuk data sampel dalam bentuk sample variance. Misalnya dari data kita di awal, dapat dihitung sampel variansnya sebagai berikut:

X_i	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
20	-80	6400
40	-60	3600
60	-40	1600
80	-20	400
100	0	0
130	30	900
130	30	900
160	60	3600
180	80	6400
Jumlah		23800

$$S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2 = \frac{1}{9-1} (23800) = \frac{1}{8} (23800) = 2975$$

3. Standar Deviasi

Standar deviasi merupakan akar dari varians (ingat, karena pada varians kita mengkuadratkan selisih data dari rata-ratanya, maka dengan mengakarkannya, kita mendapatkan kembali nilai asalnya).

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n} \sum (X_i - \bar{X})^2} \quad (\text{untuk data populasi})$$

atau

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1} \sum (X_i - \bar{X})^2} \quad (\text{untuk data sampel})$$

Sehingga berdasarkan data kita, standar deviasinya menjadi:

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1} \sum (X_i - \bar{X})^2} = \sqrt{2975} = 54,54356$$

4. *Standard error of Mean*

Standard error yang ditampilkan sebagai output Excel (dalam menu descriptive statistics) adalah standard error dari rata-rata (Standard error of Mean). Ini adalah pengukuran untuk mengukur seberapa jauh nilai rata-rata bervariasi dari satu sampel ke sampel lainnya yang diambil dari distribusi yang sama.

Apa perbedaan standard error (of mean) dengan standar deviasi ?

Kalau standar deviasi adalah suatu indeks yang menggambarkan sebaran data terhadap rata-ratanya, maka standard error (of mean) adalah indeks yang menggambarkan sebaran rata-rata sampel terhadap rata-rata dari rata-rata keseluruhan kemungkinan sampel (rata-rata populasi).

Pengukuran ini berguna, terutama untuk menjawab pertanyaan “seberapa baik rata-rata yang kita dapatkan dari data sampel dapat mengestimasi rata-rata populasi ?”

Rumus standard error of mean dan perhitungan berdasarkan data kita sebagai berikut:

$$SE = \frac{S}{\sqrt{n}} = \frac{54,5436}{\sqrt{9}} = 18,1812$$

5. *Skewness dan Kurtosis*

Distribusi normal, atau disebut juga dengan distribusi Gauss, adalah distribusi probabilitas yang paling banyak digunakan dalam berbagai analisis statistika. Distribusi normal memodelkan fenomena kuantitatif pada ilmu alam maupun ilmu sosial, dan kebanyakan estimasi dan pengujian hipotesis statistik mengasumsikan normalitas suatu data.

Oleh karenanya, sebelum menganalisis data lebih jauh, peneliti umumnya terlebih dahulu menyelidiki normalitas datanya. Jika kemudian, data (sampel) menunjukkan distribusi tidak normal, dilakukan penambahan sampel atau transformasi data dengan transformasi matematik seperti logaritma, mengkuadratkan, mengakarkan atau transformasi resiprok (1/x).

Skewness dan kurtosis merupakan dua alat ukur dalam menelusuri distribusi data yang diperbandingkan dengan distribusi normal. Skewness merupakan pengukuran tingkat ketidaksimetrisan (kecondongan) sebaran data di sekitar rata-ratanya. Distribusi normal merupakan distribusi yang simetris dan nilai skewness adalah 0. Skewness yang bernilai positif menunjukkan ujung dari kecondongan menjulur ke arah nilai positif (ekor kurva sebelah kanan lebih panjang). Skewness yang bernilai negatif menunjukkan ujung dari kecondongan menjulur ke arah nilai negatif (ekor kurva sebelah kiri lebih panjang).

Rumus skewness adalah sebagai berikut:

$$Skewness = \frac{n}{(n-1)(n-2)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^3$$

Dalam kasus data kita, di dapatkan nilai skewness sebesar -0,0178. Secara manual, angka tersebut dapat dihitung sebagai berikut:

(X)	$\left(\frac{X_i - \bar{X}}{s}\right)^3$
20	-3.1553
40	-1.3311
60	-0.3944
80	-0.0493
100	0
130	0.1664
130	0.1664
160	1.3311
180	3.1553
Jumlah	-0.1109

$$Skewness = \frac{9}{(9-1)(9-2)}(-0,1109) = \frac{9}{(8)(7)}(-0,1109) = (0,1607)(-0,1109) = -0,0178$$

Selanjutnya, kurtosis menggambarkan keruncingan (peakedness) atau kerataan (flatness) suatu distribusi data dibandingkan dengan distribusi normal. Pada distribusi normal, nilai kurtosis sama dengan 0. Nilai kurtosis yang positif menunjukkan distribusi yang relatif runcing, sedangkan nilai kurtosis yang negatif menunjukkan distribusi yang relatif rata.

Rumus kurtosis adalah:

$$Kurtosis = \left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^4 \right\} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

Dalam kasus data kita, di dapatkan nilai kurtosis sebesar -1,1764. Secara manual, angka tersebut dapat dihitung sebagai berikut:

(X)	$\left(\frac{X_i - \bar{X}}{s}\right)^4$
20	4.6279
40	1.4643
60	0.2892
80	0.0181
100	0
130	0.0915
130	0.0915
160	1.4643
180	4.6279
Jumlah	12.6748

$$Kurtosis = \left\{ \frac{9(9+1)}{(9-1)(9-2)(9-3)} (12,6748) \right\} - \frac{3(9-1)^2}{(9-2)(9-3)}$$

$$= \left\{ \frac{9(10)}{(8)(7)(6)} (12,6748) \right\} - \frac{3(8)^2}{(7)(6)} = \left\{ \frac{90}{336} (12,6748) \right\} - \frac{192}{42} = (0,2679)(12,6748) - 4,5714 = -1,1764$$

REFERENCES

Frye, CD. 2007. Step by Step Microfost Office Excel 2007. Microsoft Press. Washington

Supranto, J. 2000. Statistik: Teori dan Aplikasi. Erlangga. Jakarta