EMT4Statistika_Tiara Jovita_23030130008_PMA23

Nama : Tiara Jovita Wiji Lestari
NIM : 23030130008
Kelas : Pendidikan Matematika A 2023

Subtopik 1

Menyimpan data dalam bentuk matriks

Pada subtopik 1 ini menjelaskan tentang menyimpan data dalam bentuk matriks untuk dianalisis lebih lanjut.
Untuk itu akan dibahas :
- Penjelasan umum mengenai matriks dalam statistika
- Penggunaan operasi matriks pada persoalan statistika
- Penghitungan lanjutan dengan matriks

Array

Dalam statistika, matriks adalah array (daftar) bilangan yang terdiri dari baris-baris dan kolom-kolom. Aljabar matriks adalah aljabar khusus untuk array tersebut. Setiap array diperlakukan sebagai satu entitas yang membuatnya sangat berguna dalam menganalisa data.

EMT hanya memiliki dua dimensi untuk array. Tipe datanya disebut matriks. Akan lebih mudah untuk menulis fungsi untuk dimensi yang lebih tinggi atau library C untuk hal ini.

Dalam EMT, sebuah vektor adalah sebuah matriks dengan satu baris.
Ini bisa dibuat menjadi sebuah matriks dengan redim().

```
>shortformat; X=redim(1:20,4,5)
```

```
            1         2         3         4         5
            6         7         8         9        10
           11        12        13        14        15
           16        17        18        19        20
```

Ekstraksi baris dan kolom, atau sub-matriks

```
>X[,2:3]
```

```
            2         3
            7         8
           12        13
           17        18
```

Namun, dalam R dimungkinkan untuk mengatur daftar indeks tertentu dari vektor ke suatu nilai. Hal yang sama juga dapat dilakukan dalam EMT hanya dengan sebuah perulangan.

```
>
```

Contoh Soal 1 :

Apabila diketahui matriks Gross Income beberapa Negara (US, Canada, Australia, UK) tahun 1981 dan 1982 adalah sebagai berikut.

```
>A=[27,15,18,21;32,14,21,30]

           27        15        18        21
           32        14        21        30
```

dan diketahui matriks pengeluaran sebagai berikut

```
>B=[19,9,11,17;22,10,13,24]

           19         9        11        17
           22        10        13        24
```

Berapakah Gross Profit dalam tahun 1981 dan 1982 untuk keempat negara

=> untuk menghitung Gross Profit adalah mengurangkan matriks Gross Income dengan matriks pengeluaran. Maka di dapat matriks Gross Profit sebagai berikut :

```
>A-B

            8         6         7         4
           10         4         8         6
```

Contoh Soal 2 :

Perhitungan jumlah uang yang diperlukan dalam masalah pembelian tikus, katak, dan kelinci untuk percobaan di departemen biologi dapat menggunakan perkalian matriks.
Misal harga hewan berturut-turut 3, 1, 10 ribu rupiah. Banyak hewan yang diperlukan berturut-turut 50, 100, dan 30 ekor. Berapa jumlah uang yang diperlukan.

```
>a=[3;1;10]

            3
            1
           10


>b=[50;100;30]

           50
          100
           30


>a*b

          150
          100
          300
```

maka uang yang diperlukan adalah 150+100+300 = 550 ribu rupiah.

```
>function setmatrixvalue (M, i, j, v) ...
 loop 1 to max(length(i),length(j),length(v))
    M[i{#},j{#}] = v{#};
 end;
 endfunction
```

Ini untuk menunjukkan bahwa matriks dilewatkan dengan referensi di EMT. Jika tidak ingin mengubah matriks asli M, kita perlu menyalinnya dalam fungsi.

```
>setmatrixvalue(X,1:3,3:-1:1,0); X,
```

```
     1     2     0     4     5
     6     0     8     9    10
     0    12    13    14    15
    16    17    18    19    20
```

Hasil kali luar dalam EMT hanya dapat dilakukan di antara vektor. Hal ini otomatis karena bahasa matriks. Satu vektor harus berupa vektor kolom dan vektor baris.

```
>(1:5)*(1:5)'
```

```
     1     2     3     4     5
     2     4     6     8    10
     3     6     9    12    15
     4     8    12    16    20
     5    10    15    20    25
```

Dalam pengantar PDF untuk R ada sebuah contoh, yang menghitung distribusi ab-cd untuk a, b, c, d yang dipilih dari 0 sampai n secara acak. Solusinya dalam R adalah membentuk sebuah matriks 4 dimensi dan menjalankan table() di atasnya.

Tentu saja, hal ini bisa dicapai dengan perulangan. Tetapi perulangan tidak efektif dalam EMT. Dalam EMT, kita dapat menulis perulangan dalam C dan itu akan menjadi solusi tercepat.

Tetapi kita ingin meniru perilaku R. Untuk itu, kita perlu meratakan perkalian ab dan membuat matriks ab-cd.

```
>a=0:6; b=a'; p=flatten(a*b); q=flatten(p-p'); ...
 u=sort(unique(q)); f=getmultiplicities(u,q);...
 statplot(u,f,"h"):
```

Selain kelipatan yang tepat, EMT dapat menghitung frekuensi dalam vektor.

```
>getfrequencies(q,-50:10:50)
```

```
  [0,  23,  132,  316,  602,  801,  333,  141,  53,  0]
```

Cara yang paling mudah untuk memplot ini sebagai distribusi adalah sebagai berikut.

```
>plot2d(q,distribution=11):
```

Tetapi juga memungkinkan untuk menghitung jumlah dalam interval yang dipilih sebelumnya. Tentu saja, berikut ini menggunakan getfrequencies() secara internal.

Karena fungsi histo() mengembalikan frekuensi, kita perlu menskalakannya sehingga integral di bawah grafik batang adalah 1.

```
>{x,y}=histo(q,v=-55:10:55); y=y/sum(y)/differences(x); ...
 plot2d(x,y,>bar,style="/"):
```

Subtopik 2

Data acak dengan fungsi distribusi

Pada subtopik 2 ini menjelaskan tentang menghasilkan data acak (simulasi) dengan menggunakan fungsi distribusi tertentu
Untuk itu akan dibahas :
- memplot data distribusi eksperimen acak
- distribusi normal
- distribusi diskrit

Memplot data distribusi eksperimen acak

Dengan plot2d, terdapat metode yang sangat mudah untuk memplot sebaran data eksperimen.

```
>p=normal(1,1000); //1000 sampel acak berdistribusi normal p
>plot2d(p,distribution=20,style="\/"); // plot sampel acak p
>plot2d("qnormal(x,0,1)",add=1): // menambahkan plot distribusi normal standar
```

Perhatikan perbedaan antara plot batang (sampel) dan kurva normal (distribusi sesungguhnya). Masukkan kembali ketiga perintah tersebut untuk melihat hasil pengambilan sampel yang lain.

Berikut ini adalah perbandingan 10 simulasi dari 1000 nilai terdistribusi normal dengan menggunakan apa yang disebut plot kotak. Plot ini menunjukkan median, kuartil 25% dan 75%, nilai minimal dan maksimal, serta pencilan (the outliers).

```
>p=normal(10,1000); boxplot(p):
```

Untuk menghasilkan bilangan bulat acak, Euler memiliki intrandom. Misal kita simulasikan pelemparan dadu dan plot distribusinya.

Kita menggunakan fungsi getmultiplicities(v,x), yang menghitung seberapa sering elemen-elemen v muncul di dalam x. Kemudian kita memplot hasilnya menggunakan columnsplot().

```
>k=intrandom(1,6000,6);  ...
 columnsplot(getmultiplicities(1:6,k));...
 ygrid(1000,color=red):
```

Meskipun intrandom(n,m,k) mengembalikan bilangan bulat yang terdistribusi secara seragam dari 1 sampai k, adalah mungkin untuk menggunakan distribusi bilangan bulat yang lain dengan randpint().

Pada contoh berikut, probabilitas untuk 1,2,3 masing-masing adalah 0.4,0.1,0.5.

```
>randpint(1,1000,[0.4,0.1,0.5]); getmultiplicities(1:3,%)
```

```
  [378,  102,  520]
```

Sebagai contoh, kita mencoba distribusi eksponensial. Variabel acak kontinu X dikatakan memiliki distribusi eksponensial, jika PDF-nya diberikan oleh

$$f_X(x) = \lambda e^{-\lambda x}, \quad x > 0, \quad \lambda > 0,$$

dengan parameter

$$\lambda = \frac{1}{\mu}, \quad \mu \text{ adalah rata-rata, dan dilambangkan dengan } X \sim \text{Eksponensial}(\lambda).$$

```
>plot2d(randexponential(1,1000,2),>distribution):
```

Untuk banyak distribusi, Euler dapat menghitung fungsi distribusi dan inversnya.

```
>plot2d("normaldis",-4,4):
```

Distribusi Normal

EMT dapat menghitung banyak distribusi dan inversnya, terutama distribusi normal.

Berikut ini adalah salah satu cara untuk memplot kuantil.

```
>  plot2d("qnormal(x,1,1.5)",-4,6);  ...
 plot2d("qnormal(x,1,1.5)",a=2,b=5,>add,>filled):
```

$$\text{normaldis(x,m,d)} = \int_{-\infty}^{x} \frac{1}{d\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{t-m}{d})^2}\ dt.$$

Probabilitas untuk berada di area hijau adalah sebagai berikut.

```
>normaldis(5,1,1.5)-normaldis(2,1,1.5)
```

```
  0.24866
```

Hal ini dapat dihitung secara numerik dengan integral berikut ini.

$$\int_{2}^{5} \frac{1}{1.5\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-1}{1.5})^2}\ dx.$$

```
>gauss("qnormal(x,1,1.5)",2,5)
```

```
  0.24866
```

CONTOH :

Pada 1000 lemparan koin, jumlah gambar yang diharapkan terdistribusi dengan nilai rata-rata 500 dan deviasi standar

$$\sigma = \sqrt{1000 \times 0.5 \times 0.5}$$

Hitunglah probabilitas untuk mendapatkan lebih dari 520 muncul gambar, dan ketika probabilitasnya kurang dari 0,1% approximating distribusi binomial dengan distribusi normal.

```
>n=1000; p=0.5;...
 m=n*p; s=sqrt(n*p*(1-p));...
 1-normaldis(520,m,s)
```

```
  0.10295
```

```
>ceil(invnormaldis(99.9%,m,s))
```

```
  549
```

Perhatikan bahwa fungsi normaldis dalam skala EMT dengan cara yang berbeda dari fungsi erf, yang juga tersedia.

Semua distribusi dalam EMT diimplementasikan sebagai fungsi distribusi, dari 0 hingga 1.

Perkiraan untuk distribusi binomial juga dapat dihitung

```
>1-bindis(520,1000,0.5)
```

```
  0.097383
```

```
>invbindis(99.9%,1000,0.5)
```

```
  548.35
```

Fungsi invbindis() menyelesaikan interpolasi linier antara nilai bilangan bulat.

Mari kita bandingkan distribusi binomial dengan distribusi normal mean dan deviasi yang sama. Fungsi invbindis() memecahkan interpolasi linier antara nilai bilangan bulat.

```
>invbindis(0.95,1000,0.5), invnormaldis(0.95,500,0.5*sqrt(1000))
```

```
  525.516721219
  526.007419394
```

Fungsi qdis() adalah kepadatan distribusi chi-kuadrat. Seperti biasa, Euler memetakan vektor ke fungsi ini. Jadi kita mendapatkan plot dari semua distribusi chi-kuadrat dengan derajat 5 sampai 30 dengan mudah dengan cara berikut.

```
>plot2d("qchidis(x,(5:5:50)')",0,50):
```

Euler memiliki fungsi yang akurat untuk mengevaluasi distribusi. Mari kita periksa chidis() dengan integral.

Penamaan mencoba untuk konsisten. Misalnya.,

- distribusi chi-kuadrat adalah chidis(),
- fungsi kebalikannya adalah invchidis(),
- densitasnya adalah qchidis().

Pelengkap distribusi (ekor atas) adalah chicdis().

```
>chidis(1.5,2), integrate("qchidis(x,2)",0,1.5)
```

```
  0.527633447259
  0.527633447259
```

Distribusi Diskrit

Untuk menentukan distribusi diskrit, dapat menggunakan metode berikut.

Pertama, tetapkan fungsi distribusinya terlebih dahulu.

```
>wd = 0|((1:6)+[-0.01,0.01,0,0,0,0])/6
```

```
  [0,  0.165,  0.335,  0.5,  0.666667,  0.833333,  1]
```

Artinya, dengan probabilitas wd[i+1]-wd[i] kita menghasilkan nilai acak i.

Ini hampir merupakan distribusi yang seragam. Mari kita definisikan sebuah generator bilangan acak untuk ini. Fungsi find(v,x) menemukan nilai x dalam vektor v. Fungsi ini juga dapat digunakan untuk vektor x.

```
>function wrongdice (n,m) := find(wd,random(n,m))
```

Kesalahan ini sangat halus sehingga kita hanya bisa melihatnya setelah melakukan iterasi yang sangat banyak.

```
>columnsplot(getmultiplicities(1:6,wrongdice(1,1000000))):
```

Berikut ini adalah fungsi sederhana untuk memeriksa distribusi seragam dari nilai 1... K dalam v. Kami menerima hasilnya, jika untuk semua frekuensi

$$\left| f_i - \frac{1}{K} \right| < \frac{\delta}{\sqrt{n}}.$$

```
>function checkrandom (v, delta=1) ...
 K=max(v); n=cols(v);
 fr=getfrequencies(v,1:K);
 return max(fr/n-1/K)<delta/sqrt(n);
 endfunction
```

Memang fungsi ini menolak distribusi seragam.

```
>checkrandom(wrongdice(1,1000000))
```

```
  0
```

Dan ini menerima generator acak bawaan.

```
>checkrandom(intrandom(1,1000000,6))
```

```
  1
```

Kita dapat menghitung distribusi binomial. Pertama, ada binomialsum(), yang mengembalikan probabilitas i atau kurang dari n percobaan.

```
>bindis(410,1000,0.4)
```

```
  0.751401349654
```

Invers Beta function digunakan untuk menghitung interval kepercayaan Clopper-Pearson untuk parameter p. Tingkat defaultnya adalah alpha.

Arti dari interval ini adalah bahwa jika p berada di luar interval, hasil yang diamati dari 410 dalam 1000 jarang terjadi.

```
>clopperpearson(410,1000)
```

```
  [0.37932,  0.441212]
```

Perintah berikut ini adalah cara langsung untuk mendapatkan hasil di atas. Tetapi untuk n yang besar, penjumlahan langsung tidak akurat dan lambat.

```
>p=0.4; i=0:410; n=1000; sum(bin(n,i)*p^i*(1-p)^(n-i))
```

```
  0.751401349655
```

invbinsum() menghitung invers dari binomialsum().

```
>invbindis(0.75,1000,0.4)
```

```
  409.932733047
```

Dalam Bridge, kita mengasumsikan 5 kartu yang terbuka (dari 52 kartu) di dua tangan (26 kartu). Mari kita hitung probabilitas distribusi yang lebih buruk dari 3:2 (misalnya 0:5, 1:4, 4:1, atau 5:0).

```
>2*hypergeomsum(1,5,13,26)
```

```
  0.321739130435
```

Ada juga simulasi distribusi multinomial.

```
>randmultinomial(10,1000,[0.4,0.1,0.5])
```

```
     392      102      506
     394      108      498
     389       92      519
```

```
381            98           521
404            98           498
384           101           515
413           104           483
394           105           501
405            93           502
388           102           510
```

Input dan Output File (Membaca dan Menulis Data)

Euler Math Toolbox (EMT) dapat membaca data yang tersimpan di dalam berkas dengan berbagai format (teks biasa, CSV, dsb.) untuk melakukan pemrosesan atau analisis lebih lanjut. Berikut ini adalah beberapa metode umum:

1. Membaca Data dari File CSV

File CSV (Comma-Separated Values) adalah format yang umum digunakan untuk data tabular, di mana kolom dipisahkan oleh koma atau tanda tertentu.

Pertama, mari kita tulis matriks ke dalam file. Untuk hasilnya, kita menghasilkan file di direktori kerja saat ini.

```
>file="test.csv"; ...
 M=random(3,3); writematrix(M,file);
```

Berikut adalah isi dari file ini.

```
>printfile(file)
```

```
 0.8270473253443011,0.7620605196688064,0.4736470930557711
 0.7706064244273799,0.5909668510929599,0.0638118965663068
 0.1650645844780326,0.7963239116070477,0.718715244015118
```

CSV ini dapat dibuka pada sistem bahasa Inggris ke Exel dengan klik dua kali. Jika mendapatkan file seperti itu di dalam sistem Jerman, maka perlu mengimpor data ke Excel dengan menggunakan titik desimal.

Tetapi titik desimal adalah format default untuk EMT juga. Maka, bisa membaca matriks dari file dengan "readmatrix ()".

```
>readmatrix(file)
```

```
        0.827047      0.762061      0.473647
        0.770606      0.590967     0.0638119
        0.165065      0.796324      0.718715
```

Dimungkinkan untuk menulis beberapa matriks ke satu file. Perintah "open ()" dapat membuka file untuk ditulis dengan parameter "w".
Standarnya adalah "r" untuk membaca.

```
>open(file,"w"); writematrix(M); writematrix(M'); close();
```

Matriks dipisahkan oleh garis kosong. Untuk membaca matriks, buka file dan panggil "readmatrix()" beberapa kali.

```
>open(file); A=readmatrix(); B=readmatrix(); A=B, close();
```

```
        0.827047      0.770606      0.165065
        0.762061      0.590967      0.796324
        0.473647     0.0638119      0.718715
```

Di Excel atau spreadsheet serupa, kita dapat mengekspor matriks sebagai CSV (nilai dipisahkan dengan koma). Di Excel 2007, gunakan "save as" dan "other format", lalu pilih "CSV". Pastikan tabel saat ini hanya berisi data yang ingin diekspor.

Berikut ini contohnya.

```
>reset;
>file="excel-data.csv"; ...
 printfile("excel-data.csv")
```

```
0;1000;1000
1;1051,271096;1072,508181
2;1105,170918;1150,273799
3;1161,834243;1233,67806
4;1221,402758;1323,129812
5;1284,025417;1419,067549
6;1349,858808;1521,961556
7;1419,067549;1632,31622
8;1491,824698;1750,6725
9;1568,312185;1877,610579
10;1648,721271;2013,752707
```

Seperti yang bisa dilihat, dalam sistem Jerman menggunakan titik koma sebagai pemisah dan koma desimal. Kita dapat mengubahnya di pengaturan sistem atau di Excel, tetapi tidak perlu membaca matriks ke EMT.

Cara termudah untuk membaca ini ke dalam Euler adalah "readmatrix ()". Semua koma diganti dengan titik dengan parameter >comma. Untuk CSV bahasa Inggris, cukup abaikan parameter ini.

```
>M=readmatrix("excel-data.csv",>comma)
```

```
                0          1000          1000
                1       1051.27       1072.51
                2       1105.17       1150.27
                3       1161.83       1233.68
                4        1221.4       1323.13
                5       1284.03       1419.07
                6       1349.86       1521.96
                7       1419.07       1632.32
                8       1491.82       1750.67
                9       1568.31       1877.61
               10       1648.72       2013.75
```

Mari kita plot ini.

```
>plot2d(M'[1],M'[2:3],>points,color=[red,green]'):
```

Ada cara yang lebih mendasar untuk membaca data dari sebuah file. Kita dapat membuka file dan membaca angka baris demi baris. Fungsi "getvectorline ()" akan membaca angka dari sebaris data. Secara default, ini mengharapkan titik desimal. Tapi itu juga bisa menggunakan koma desimal, jika kita memanggil "setdecimaldot (",")" sebelum kita menggunakan fungsi ini.

Fungsi berikut adalah contoh untuk ini. Ini akan berhenti di akhir file atau baris kosong.

```
>function myload (file) ...
 open(file);
 M=[];
 repeat
    until eof();
    v=getvectorline(3);
    if length(v)>0 then M=M_v; else break; endif;
 end;
 return M;
 close(file);
 endfunction
```

```
>myload(file)
```

```
                0          1000          1000             0             0
                1          1051        271096          1072        508181
                2          1105        170918          1150        273799
                3          1161        834243          1233         67806
                4          1221        402758          1323        129812
                5          1284         25417          1419         67549
                6          1349        858808          1521        961556
                7          1419         67549          1632         31622
                8          1491        824698          1750          6725
                9          1568        312185          1877        610579
               10          1648        721271          2013        752707
```

Juga dimungkinkan untuk membaca semua angka dalam file itu dengan "getvector ()".

```
>open(file); v=getvector(10000); close(); redim(v[1:9],3,3)
```

```
                0          1000          1000
                1          1051        271096
             1072        508181             2
```

Oleh karena itu, sangat mudah untuk menyimpan sebuah vekor nilai, satu nilai di setiap baris dan membaca kembali vektor ini.

```
>v=random(1000); mean(v)
```

```
     0.499412468751
```

Contoh Lain

Kita akan membaca file csv yang bernama "GOOG"

```
>reset;
>file="GOOG.csv"; ...
 printfile("GOOG.csv")
```

```
    Date,Open,High,Low,Close,Adj Close,Volume
    2019-11-04,1276.449951,1323.739990,1276.354980,1311.369995,1311.369995,7217800
    2019-11-11,1303.180054,1334.880005,1293.510010,1334.869995,1334.869995,5900600
    2019-11-18,1332.219971,1335.529053,1291.150024,1295.339966,1295.339966,6446400
    2019-11-25,1299.180054,1318.359985,1298.130005,1304.959961,1304.959961,3688500
    2019-12-02,1301.000000,1344.000000,1279.000000,1340.619995,1340.619995,6719700
    2019-12-09,1338.040039,1359.449951,1336.040039,1347.829956,1347.829956,6129400
    2019-12-16,1356.500000,1365.000000,1348.984985,1349.589966,1349.589966,9558800
    2019-12-23,1355.869995,1364.530029,1342.780029,1351.890015,1351.890015,2936500
    2019-12-30,1350.000000,1372.500000,1329.084961,1360.660034,1360.660034,4605700
    2020-01-06,1350.000000,1434.928955,1350.000000,1429.729980,1429.729980,8084600
    2020-01-13,1436.130005,1481.295044,1426.020020,1480.390015,1480.390015,8063800
    2020-01-20,1479.119995,1503.213989,1465.250000,1466.709961,1466.709961,6783300
    2020-01-27,1431.000000,1470.130005,1421.199951,1434.229980,1434.229980,8166900
    2020-02-03,1462.000000,1490.000000,1426.300049,1479.229980,1479.229980,11826100
    2020-02-10,1474.319946,1529.630005,1474.319946,1520.739990,1520.739990,6059400
    2020-02-17,1515.000000,1532.105957,1480.439941,1485.109985,1485.109985,4898300
    2020-02-24,1426.109985,1438.140015,1271.000000,1339.329956,1339.329956,14316700
    2020-03-02,1351.609985,1410.150024,1261.050049,1298.410034,1298.410034,11969000
    2020-03-09,1205.300049,1281.150024,1113.300049,1219.729980,1219.729980,16512100
    2020-03-16,1096.000000,1157.969971,1037.280029,1072.319946,1072.319946,19600200
    2020-03-23,1061.319946,1169.969971,1013.536011,1110.709961,1110.709961,18250300
    2020-03-30,1125.040039,1175.310059,1079.810059,1097.880005,1097.880005,11683000
    2020-04-06,1138.000000,1225.569946,1130.939941,1211.449951,1211.449951,9202500
    2020-04-13,1209.180054,1294.430054,1187.598022,1283.250000,1283.250000,10349000
    2020-04-20,1271.000000,1293.310059,1209.709961,1279.310059,1279.310059,9148200
```

```
2020-04-27,1296.000000,1359.989990,1232.199951,1320.609985,1320.609985,13086900
2020-05-04,1308.229980,1398.760010,1299.000000,1388.369995,1388.369995,7156600
2020-05-11,1378.280029,1416.530029,1323.910034,1373.189941,1373.189941,7924600
2020-05-18,1361.750000,1415.489990,1354.250000,1410.420044,1410.420044,7454400
2020-05-25,1437.270020,1441.000000,1391.290039,1428.920044,1428.920044,7276700
2020-06-01,1418.390015,1446.552002,1404.729980,1438.390015,1438.390015,6970600
2020-06-08,1422.339966,1474.259033,1386.020020,1413.180054,1413.180054,8274100
2020-06-15,1390.800049,1460.000000,1387.920044,1431.719971,1431.719971,9501200
2020-06-22,1429.000000,1475.941040,1351.989990,1359.900024,1359.900024,10226400
2020-06-29,1358.180054,1482.949951,1347.010010,1464.699951,1464.699951,7486900
2020-07-06,1480.060059,1543.829956,1472.859985,1541.739990,1541.739990,7551500
2020-07-13,1550.000000,1577.131958,1483.500000,1515.550049,1515.550049,8018100
2020-07-20,1515.260010,1586.989990,1488.400024,1511.869995,1511.869995,6879500
2020-07-27,1515.599976,1540.969971,1454.030029,1482.959961,1482.959961,9166000
2020-08-03,1486.640015,1516.844971,1458.650024,1494.489990,1494.489990,9785200
2020-08-10,1487.180054,1537.250000,1473.079956,1507.729980,1507.729980,6991400
2020-08-17,1514.670044,1597.719971,1507.969971,1580.420044,1580.420044,8219400
2020-08-24,1593.979980,1659.219971,1580.569946,1644.410034,1644.410034,11011800
2020-08-31,1647.890015,1733.180054,1547.613037,1591.040039,1591.040039,11877700
2020-09-07,1533.510010,1584.081055,1497.359985,1520.719971,1520.719971,7601300
2020-09-14,1539.005005,1564.000000,1437.130005,1459.989990,1459.989990,9323100
2020-09-21,1440.060059,1469.520020,1406.550049,1444.959961,1444.959961,8902600
2020-09-28,1474.209961,1499.040039,1449.301025,1458.420044,1458.420044,7750300
2020-10-05,1466.209961,1516.520020,1436.000000,1515.219971,1515.219971,6728000
2020-10-12,1543.000000,1593.859985,1532.569946,1573.010010,1573.010010,8989300
2020-10-19,1580.459961,1642.359985,1525.670044,1641.000000,1641.000000,9226500
2020-10-26,1625.010010,1687.000000,1514.619995,1621.010010,1621.010010,11248500
2020-11-02,null,null,null,null,null,null
2020-11-02,1628.160034,1660.674927,1627.652466,1638.349976,1638.349976,1278888
```

Kita coba lagi untuk membaca file csv bernama "sample".

```
>reset;
>file="sample.csv"; ...
 printfile("sample.csv")
```

```
female,read,write,math,hon,femalexmath
0,57,52,41,0,0
1,68,59,53,0,53
0,44,33,54,0,0
0,63,44,47,0,0
0,47,52,57,0,0
0,44,52,51,0,0
0,50,59,42,0,0
0,34,46,45,0,0
0,63,57,54,0,0
0,57,55,52,0,0
0,60,46,51,0,0
0,57,65,51,1,0
0,73,60,71,0,0
0,54,63,57,1,0
0,45,57,50,0,0
0,42,49,43,0,0
0,47,52,51,0,0
0,57,57,60,0,0
0,68,65,62,1,0
0,55,39,57,0,0
0,63,49,35,0,0
0,63,63,75,1,0
0,50,40,45,0,0
0,60,52,57,0,0
0,37,44,45,0,0
0,34,37,46,0,0
0,65,65,66,1,0
0,47,57,57,0,0
0,44,38,49,0,0
0,52,44,49,0,0
0,42,31,57,0,0
0,76,52,64,0,0
0,65,67,63,1,0
0,42,41,57,0,0
```

```
0,52,59,50,0,0
0,60,65,58,1,0
0,68,54,75,0,0
0,65,62,68,1,0
0,47,31,44,0,0
0,39,31,40,0,0
0,47,47,41,0,0
0,55,59,62,0,0
0,52,54,57,0,0
0,42,41,43,0,0
0,65,65,48,1,0
0,55,59,63,0,0
0,50,40,39,0,0
0,65,59,70,0,0
0,47,59,63,0,0
0,57,54,59,0,0
0,53,61,61,1,0
0,39,33,38,0,0
0,44,44,61,0,0
0,63,59,49,0,0
0,73,62,73,1,0
0,39,39,44,0,0
0,37,37,42,0,0
0,42,39,39,0,0
0,63,57,55,0,0
0,48,49,52,0,0
0,50,46,45,0,0
0,47,62,61,1,0
0,44,44,39,0,0
0,34,33,41,0,0
0,50,42,50,0,0
0,44,41,40,0,0
0,60,54,60,0,0
0,47,39,47,0,0
0,63,43,59,0,0
0,50,33,49,0,0
0,44,44,46,0,0
0,60,54,58,0,0
0,73,67,71,1,0
0,68,59,58,0,0
0,55,45,46,0,0
0,47,40,43,0,0
0,55,61,54,1,0
0,68,59,56,0,0
0,31,36,46,0,0
0,47,41,54,0,0
0,63,59,57,0,0
0,36,49,54,0,0
0,68,59,71,0,0
0,63,65,48,1,0
0,55,41,40,0,0
0,55,62,64,1,0
0,52,41,51,0,0
0,34,49,39,0,0
0,50,31,40,0,0
0,55,49,61,0,0
0,52,62,66,1,0
0,63,49,49,0,0
1,68,62,65,1,65
1,39,44,52,0,52
1,44,44,46,0,46
1,50,62,61,1,61
1,71,65,72,1,72
1,63,65,71,1,71
1,34,44,40,0,40
```

```
>printfile("test.csv")
```

```
0.8270473253443011,0.7620605196688064,0.4736470930557711
0.7706064244273799,0.5909668510929599,0.0638118965663068
```

```
0.1650645844780326,0.7963239116070477,0.718715244015118

0.8270473253443011,0.7706064244273799,0.1650645844780326
0.7620605196688064,0.5909668510929599,0.7963239116070477
0.4736470930557711,0.0638118965663068,0.718715244015118
```

```
>printfile("excel-data.csv")
```

```
0.6554163483485531,0.2009951854518572,0.8936223876466522,0.2818865431288053,0.5250003829714993,0.3141267749950177,0.44461567829937
0.2709059757306277,0.7044190158488305,0.217693385437102,0.4453627564273003,0.3084110353211584,0.9145409086421026,0.193585477981141
0.4311837813121366,0.7286804774486648,0.4651641815616542,0.3230318624794988,0.5251835885646776,0.5022552426400109,0.16860346325267
0.4934533298683972,0.6013438464674292,0.6594608216798724,0.9674676406359064,0.1931506690652142,0.9359208315393285,0.07287525720633
0.5214304734587624,0.4288933665752114,0.168134262921015,0.1827423588992564,0.288048270849559,0.7500422847241136,0.4729350182225353
0.4460696554351229,0.2165448858432077,0.598477020799437,0.7292714193407775,0.9500185385120919,0.791697647093935,0.4186365649030115
0.6741139140444261,0.07526764059334502,0.854693390265544,0.1232223852881709,0.2050061227872597,0.4650491548372331,0.02812689700666
0.2173910972095296,0.9628723938223689,0.8422672849114607,0.3287576902398857,0.5611456347151561,0.9188021822358414,0.99749247859551
0.7208428570402986,0.2478037411641879,0.8340834946823759,0.4437198302318275,0.3718659969086804,0.7757183550559265,0.37467962682608
0.3568662131827455,0.4088660151481681,0.3500908607420154,0.347412353162677,0.1612572002722239,0.5883249626530436,0.838840074511899
```

Membaca dari Web

```
>reset;
>function readversion () ...
```

```
 urlopen("http://www.euler-math-toolbox.de/Programs/Changes.html");
 repeat
   until urleof();
   s=urlgetline();
   k=strfind(s,"Version ",1);
   if k>0 then substring(s,k,strfind(s,"<",k)-1), break; endif;
 end;
 urlclose();
 endfunction
```

```
>readversion
```

```
    Version 2024-01-12
```

Sub Topik 5 : Perhitungan Analisis Data Statistika Deskriptif

Pada sub topik 5 part 1 ini akan dibahas mengenai:
1. Penjelasan umum mengenai statistika deskriptif
2. Ruang lingkup kajian pada analisis statistika deskriptif yang meliputi distribusi frekuensi, mean, median, dan modus.

Penjelasan Umum Mengenai statistika Deskriptif

Statistika deskriptif yaitu statistik yang mempelajari tata cara mengumpulkan, menyusun, menyajikan, dan menganalisis data yang berwujud angka agar dapat memberikan gambaran yang teratur, ringkas, dan jelas mengenai suatu gejala atau keadaan peristiwa. Analisis ini hanya berupa akumulasi data dasar dalam bentuk deskripsi semata dalam arti tidak mencari atau menerangkan saling hubungan, menguji hipotesis, membuat ramalan atau melakukan penarikan kesimpulan. Analisis data yang tergolong statistik deskriptif terdiri dari distribusi frekuensi, ukuran pemusatan data(mean,median,modus), ukuran letak(kuartil,desil,persentil), ukuran dispersi(jangkauan atau rentang,varians,simpangan baku), dan teknis statistik lain yang bertujuan hanya untuk mengetahui gambaran atau kecenderungan data tanpa bermaksud melakukan generalisasi.

Distribusi Frekuensi

Distribusi frekuensi merupakan ruang lingkup kajian pada analisis statistik deskriptif. Distribusi frekuensi adalah alat penyajian data berbentuk kolom dan lajur (tabel)yang di dalamnya dibuat angka yang menggambarkan pancaran frekuensi dari variabel yang sedang menjadi objek. Unsur-unsur distribusi frekuensi yaitu kelas(kelompok nilai data yang ditulis dalam bentuk interval),batas bawah(jangkauan terendah dari kelas),batas atas(jangkauan tertinggi dari kelas),tepi bawah kelas(batas bawah dikurangi ketelitian data),tepi atas kelas(batas atas kelas ditambah ketelitian data),banyak kelas(1+3,3 log n),dan panjang kelas.
Dalam statistik terdapat berbagai macam distribusi frekuensi, diantaranya:
1. Distribusi frekuensi biasa
Distribusi frekuensi biasa adalah distribusi frekuensi yang berisikan jumlah frekuensi dari setiap kelompok data atau kelas.
2. Distribusi frekuensi relatif
Distribusi frekuensi relatif adalah distribusi frekuensi yang dinyatakan dalam bentuk persentase atau desimal. Besarnya frekuensi relatif yaitu frekuensi absolut setiap kelas dibagi total frekuensi dikali 100%.
3. Distribusi frekuensi kumulatif
Menunjukkan seberapa besar jumlah frekuensi pada tingkat kelas tertentu yang diperoleh dengan menjumlahkan atau mengurangkan frekuensi pada kelas tertentu dengan frekuensi kelas sebelumnya. Distribusi frekuensi kumulatif terdiri dari 2 macam yaitu distribusi kumulatif kurang dari dan distribusi kumulatif lebih dari.

Contoh Soal 1:
1.Disajikan data urut yaitu 45,48,49,50,52,52,52,53,53,54,54,54,54,54,56,56,
56,56,57,57,58,58,58,58,58,58,58,59,59,60,60,60, 62,62,62,63,63,64,64,65,67,68,69,70,70,71,73,74.
Buatlah distribusi frekuensi berdasarkan data diatas!
Penyeleaian:
- Menentukan range
range= nilai maks-nilai min
= 74-45
= 29
- Menentukan banyak kelas dengan aturan struges.
= 1+3,3 log n, n banyaknya data
= 1+3,3 log 48
= 6,64
= 7
- Menentukan panjang kelas

$$p = \frac{range}{banyakkelas}$$

$$p = \frac{29}{7}$$

$$p = 4.14 = 5$$

Berdasarkan pertimbangan beberapa unsur dalam data urut diatas yaitu nilai minimum 45, nilai maksimum 74, banyak kelas yaitu 7, dan panjang kelas yaitu 5 maka dapat dibuat tabel distribusi frekuensi dengan batas bawah kelas pertama yaitu 43 dan batas atas kelas ketujuh yaitu 77. Sehingga dapat ditentukan tepi bawah kelas pertama yaitu 43-0.5=42.5 dan tepi atas kelas ketujuh yaitu 77+0.5=77.5.

```
>r=42.5:5:77.5; v=[1,6,13,15,6,5,2];
>T:=r[1:7]' | r[2:8]' | v'; writetable(T,labc=["TB","TA","Frek"])
```

TB	TA	Frek
42.5	47.5	1
47.5	52.5	6
52.5	57.5	13
57.5	62.5	15
62.5	67.5	6
67.5	72.5	5
72.5	77.5	2

Mencari titik tengah

```
>(T[,1]+T[,2])/2 // the midpoint of each interval
```

```
                45
                50
                55
                60
                65
                70
                75
```

Sajian dalam bentuk histogram

```
>plot2d(r,v,a=40,b=80,c=0,d=20,bar=1,style="\/"):
```

Rata-Rata Hitung(Mean)

Rata-Rata hitung biasa juga disebut sebagai rerata atau mean merupakan ruang lingkup kajian pada analisis statistika deskriptif yang termasuk dalam ukuran pemusatan data.

$$\text{Rata-Rata hitung ini disimbolkan dengan } \mu \text{ untuk data populasi } dan$$

$$\bar{X} \text{ untuk data sampel .}$$

Rata-rata hitung atau mean merupakan nilai yang menunjukkan pusat dari nilai data dan dapat mewakili keterpusatan data. Mean dapat diperoleh dengan membagi jumlah nilai-nilai data dengan jumlah individu(cacah data). Perhitungan mean dibagi dua yaitu mean data tunggal dan mean data kelompok.

1. Perhitungan Mean pada Data Tunggal

Pada data tunggal, perhitungannya yaitu dengan cara menjumlahkan semua nilai dan dibagi banyak data. Rumus yang digunakan adalah sebagai berikut:

$$\bar{X} = \frac{\sum x_i}{n}$$

$$\mu = \frac{\sum x_i}{N}$$

Keterangan:

$$\bar{X} = \text{ Rata-Rata hitung atau mean untuk data sampel}$$

$$\mu = \text{ Rata-Rata hitung atau mean untuk data populasi}$$

$$\sum x_i = \text{ jumlah dari nilai data ke-i}$$

```
n = banyaknya data dalam sampel
N - banyaknya data dalam populasi
```

Data tunggal juga dapat disajikan dalam tabel distribusi.
Misalnya diberikan data

$$x_1, x_2, ..., x_n$$

yang memiliki frekuensi berturut-turut

$$f_1, f_2, ..., f_n$$

maka rataan hitung sampel atau rataan hitung populasi dari data yang disajikan dalam daftar distribusi itu ditentukan dengan rumus:

Untuk rata-rata hitung sampel,

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$

Untuk rata-rata hitung populasi,

$$\mu = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$

Kita dapat mengetahui nilai rata-rata hitung(mean)pada data tunggal dengan menggunakan perintah EMT yaitu 'mean(x)' dan 'mean(x,f)'.

Contoh Soal 1:
Seorang pelatih tembak ingin mengevaluasi nilai ketangkasan delapan anak buahnya jenis senapan yang dipakai M-16 dengan jarak 300 meter dan masing-masing mendapat nilai 76,85,70,65,40,70,50,dan 80. Berapakah rata-rata nilai ketangkasan delapan anak tersebut?
Penyelesaian:

```
>x=[76,85,70,65,40,70,50,80]; mean(x),
```

```
 67
```

Diketahui:

$$\sum x_i = 76 + 85 + 70 + 65 + 40 + 70 + 80 = 536$$

```
n = 8
```

$$\bar{X} = \frac{\sum x_i}{n}$$

$$\bar{X} = \frac{536}{8}$$

$$\bar{X} = 67$$

Sehingga rata-rata nilai ketangkasan delapan anak tersebut adalah 67

Contoh Soal 2:
Banyaknya pegawai di 5 apotik adalah 3,5,6,4,dan 6. Dengan memandang data itu sebagai data populasi, hitunglah nilai rata-rata banyaknya pegawai di 5 apotik tersebut!
Penyelesaian:

```
>x=[3,5,6,4,6]; mean(x),
```

```
 4.8
```

Diketahui:

$$\sum x_i = 3 + 5 + 6 + 4 + 6 = 24$$

```
N = 5
```

$$\mu = \frac{\sum x_i}{N}$$

$$\mu = \frac{24}{5}$$

$$\mu = 4.8$$

Sehingga nilai rata-rata banyaknya pegawai di 5 apotik tersebut adalah 4.8

Contoh Soal 3:
Diberikan data berat kambing di suatu peternakan yang memelihara 50 kambing. Kambing dengan berat 45kg terdapat 5 ekor, kambing dengan berat 46kg terdapat 10 ekor, kambing dengan berat 47kg terdapat 6 ekor, kambing dengan berat 48kg terdapat 9 ekor, kambing dengan berat 49kg terdapat 7 ekor, dan kambing dengan berat 50kg terdapat 13 ekor. Tentukan rata-rata berat kambing di peternakan tersebut.
Penyelesaian:

```
>x=[45,46,47,48,49,50], f=[5,10,6,9,7,13]      //Mendeskripsikan data dan frekuensi
```

```
 [45,  46,  47,  48,  49,  50]
 [5,  10,  6,  9,  7,  13]
```

```
>mean(x,f)    //Menghitung rata-rata
```

```
 47.84
```

Jadi, rata-rata berat kambing di peternakan tersebut adalah 47.84 kg

2. Perhitungan Mean pada Data Kelompok

Jika data yang sudah dikelompokkan dalam hitungan distribusi frekuensi maka data tersebut akan berbaur sehingga keaslian data itu akan hilang bercampur dengan data lain menurut kelasnya, hanya dalam perhitungan mean pada data kelompok diambil titik tengahnya untuk mewakili setiap kelas interval.
Adapun rumus yang digunakan untuk menghitung mean pada data kelompok yaitu:

$$\bar{X} = \frac{\sum t_i f_i}{\sum f_i}$$

Keterangan:

$\sum t_i f_i =$ jumlah dari perkalian antara titik tengah tiap kelas dan frekuensi tiap kelas

$\sum f_i =$ jumlah dari frekuensi tiap kelas

Kita dapat mengetahui nilai rata-rata hitung(mean) pada data kelompok dengan menggunakan perintah EMT yaitu 'mean(t,v)'dimana t menunjukkan titik tengah dan v menunjukkan frekuensi.

Misalkan suatu data berkelompok terdiri dari n kelas dengan nilai tengah masing-masing kelas secara berturut-turut adalah

$$t_1, t_2, ..., t_n$$

dan masing-masing frekuensinya adalah

$$f_1, f_2, ..., f_n$$

maka untuk menghitung rata-rata data tabel distribusi seperti ini di EMT, dapat dilakukan dengan cara berikut:
1. Menentukan tepi bawah kelas (Tb), panjang kelas (P), dan tepi atas kelas (Ta) dengan rumus :

$$T_b = a - 0,5$$

$$P = (b - a) + 1$$

$$P = \frac{range}{banyakkelas}$$

$$T_a = b + 0.5$$

dengan a = batas bawah kelas dan b = batas atas kelas

2. Mendeskripsikan data dalam bentuk tabel, dengan perintah

> r=tepi bawah terkecil:panjang kelas:tepi atas terbesar; v=[frekuensi];
> T:=r[1:jumlah kelas]' | r[2:jumlah kelas + 1]' | v'; writetable(T,labc=["tepi bawah","tepi atas","frekuensi"])

3. Menghitung nilai tengah kelas, dengan perintah

> (T[,1]+T[,2])/2

4. Mengubah baris menjadi kolom

> t=fold(r,[0.5,0.5])

5. Menghitung rata-rata, dengan perintah

> mean(t,v)

Contoh Soal 1:
Data berikut menunjukkan nilai yang diperoleh 50 siswa SMP Negeri 1 Gabus pada Ujian Nasional mata pelajaran matematika.
Siswa yang mendapat nilai dalam rentang 61-65 sebanyak 2 orang, dalam rentang 66-70 sebanyak 5 orang, dalam rentang 71-75 sebanyak 8 orang, dalam rentang 76-80 sebanyak 10 orang, dalam rentang 81-85 sebanyak 12 orang, dalam rentang 86-90 sebanyak 9 orang, dan dalam rentang 91-95 sebanyak 4 orang.
Tentukan rata-rata nilai yang diperoleh 50 siswa tersebut!
Penyelesaian:
Menentukan tepi bawah kelas yang terkecil

```
>61-0.5

    60.5
```

Menentukan panjang kelas

```
>(65-61)+1

    5
```

Menentukan tepi atas kelas yang terbesar

```
>95+0.5

    95.5
```

```
>r=60.5:5:95.5; v=[2,5,8,10,12,9,4];
>T:=r[1:7]' | r[2:8]' | v'; writetable(T,labc=["TB","TA","Frek"])

          TB        TA      Frek
        60.5      65.5         2
        65.5      70.5         5
        70.5      75.5         8
        75.5      80.5        10
        80.5      85.5        12
        85.5      90.5         9
        90.5      95.5         4
```

Menentukan titik tengah

```
>(T[,1]+T[,2])/2 // the midpoint of each interval

                63
                68
                73
                78
                83
                88
                93


>t=fold(r,[0.5,0.5])

     [63,  68,  73,  78,  83,  88,  93]
```

Menentukan mean(rata-rata)

```
>mean(t,v)

     79.8
```

Diketahui:

$$\sum t_i f_i = (2)(63)+(5)(68)+(8)(73)+(10)(78)+(12)(83)+(9)(88)+(4)(93) = 3.990$$

$$\sum f_i = 2+5+8+10+12+9+4 = 50$$

$$\bar{X} = \frac{\sum t_i f_i}{\sum f_i}$$

$$\bar{X} = \frac{3.990}{50}$$

$$\bar{X} = 79,8$$

Jadi rata-rata nilai yang diperoleh 50 siswa tersebut adalah 79,8

3. Perhitungan Rata-Rata dari File yang Tersimpan dalam Direktori

Dalam perhitungan rata-rata hitung(mean), kita dapat menggunakan file yang tersimpan dalam direktori.

Contoh 1:
Misalnya kita akan menghitung nilai rata-rata(mean) yang terdapat dalam file "test.dat"

```
>filename="test.dat"; ...
 V=random(3,3); writematrix(V,filename);
>printfile(filename),


    0.6554163483485531,0.2009951854518572,0.8936223876466522
    0.2818865431288053,0.5250003829714993,0.3141267749950177
    0.4446156782993733,0.2994744556282315,0.2826898577756425



>readmatrix(filename)


         0.655416      0.200995      0.893622
         0.281887         0.525      0.314127
         0.444616      0.299474       0.28269
```

```
>mean(V),

          0.583345
          0.373671
           0.34226
```

Contoh 2:
Misalnya akan dihitung nilai rata-rata dari file 'mtcars.csv'

```
>filename="mtcars.csv";  ...
 V=random(4,4); writematrix(V,filename);
>printfile(filename)

     0.8832274852666707,0.2709059757306277,0.7044190158488305,0.217693385437102
     0.4453627564273003,0.3084110353211584,0.9145409086421026,0.1935854779811416
     0.4633868194128757,0.0951529788263157,0.5950170162024192,0.4311837813121366
     0.7286804774486648,0.4651641815616542,0.3230318624794988,0.5251835885646776


>readmatrix(filename)

          0.883227      0.270906      0.704419      0.217693
          0.445363      0.308411      0.914541      0.193585
          0.463387      0.095153      0.595017      0.431184
           0.72868      0.465164      0.323032      0.525184


>mean(V[1])

     0.519061465571


>mean(V[2])

     0.465475044593


>mean(V[3])

     0.396185148938


>mean(V[4])

     0.510515027514


>mean(V)

          0.519061
          0.465475
          0.396185
          0.510515
```

Median

Median merupakah ruang lingkup kajian pada analisis statistika deskriptif yang termasuk dalam ukuran pemusatan data. Median adalah suatu nilai yang berada di tengah data setelah diurutkan dari data yang

terkecil sampai yang terbesar atau sebaliknya. Dalam perhitungan statistik, median dibagi menjadi 2 bagian yaitu median untuk data tunggal dan median untuk data kelompok.

1. Median Data Tunggal

Jika jumlah suatu data(n) berjumlah ganjil maka nilai mediannya adalah sama dengan data yang memiliki nilai di urutan paling tengah yang memiliki nomor urut k, dimana untuk menentukan nilai k dapat dihitung menggunakan rumus:

$$k = \frac{n+1}{2}$$

Jika jumlah suatu data(n) berjumlah genap,maka untuk menghitung mediannya dengan menggunakan rumus:

$$k = \frac{n}{2}$$

$$Me = \frac{1}{2}(X_k + X_{k+1})$$

Kita dapat mengetahui nilai median pada data tunggal dengan menggunakan perintah EMT yaitu 'median(data)'

Contoh Soal 1:
Diketahui sebuah data hasil nilai Ujian Akhir Semester mata kuliah Psikologi Pendidikan 11 mahasiswa sebagai berikut: 85,90,80,95,50,75,30,60,65,40,70.
Tentukan nilai median dari data tersebut!
Penyelesaian:

```
>data=[85,90,80,95,50,75,30,60,65,40,70];
>urut=sort(data)
```

```
  [30,  40,  50,  60,  65,  70,  75,  80,  85,  90,  95]
```

Dalam menentukan median, langkah pertama yang harus dilakukan adalah dengan mengurutkan data tersebut dari yang terkecil sampai yang terbesar dengan fungsi sort(data). Fungsi sort(data) dalam EMT digunakan untuk mengurutkan elemen-elemen dalam suatu vektor atau matriks(dari nilai terkecil ke nilai terbesar).

```
>median(data)
```

```
  70
```

Diketahui bahwa kasus ini merupakan data yang berjumlah ganjil, sehingga nilai median untuk kasus ini adalah sama dengan data yang memiliki nilai di urutan paling tengah yang memiliki nomor urut k.

$$k = \frac{n+1}{2}$$

$$k = \frac{11+1}{2}$$

$$k = 6$$

$$Me = X_6 = 70$$

Jadi nilai median dari data hasil Ujian Akhir Semester(UAS) mata kuliah Psikologi Pendidikan 11 mahasiswa adalah 70

Contoh Soal 2:
Data upah dari 8 karyawan yang dinyatakan dalam rupiah adalah sebagai berikut:
20,80,75,60,50,85,45,90.

Tentukan nilai median dari data tersebut!
Penyelesaian:

```
>data=[20,80,75,60,50,85,45,90];
>urut=sort(data)
```

```
[20,  45,  50,  60,  75,  80,  85,  90]
```

```
>median(data)
```

```
67.5
```

Diketahui bahwa kasus ini merupakan data yang berjumlah genap, sehingga nilai median untuk kasus ini adalah terletak pada data ke-k dan data ke-(k+1).

$$k = \frac{n}{2}$$

$$k = \frac{8}{2}$$

$$k = 4$$

$$Me = \frac{1}{2}(X_k + X_{k+1})$$

$$Me = \frac{1}{2}(X_4 + X_5)$$

$$Me = \frac{1}{2}(60 + 75)$$

$$Me = 67.5$$

Jadi nilai median pada data upah 8 karyawan adalah 67.5

2. Median Data Kelompok

Untuk menghitung median pada data kelompok, dapat menggunakan rumus di bawah ini:

$$Me = Tb + \frac{\frac{1}{2}n - f_{ks}}{f_m}.p$$

```
Keterangan:
   Tb = Tepi bawah kelas median
    n = Total frekuensi
  fks = Frekuensi kumulatif sebelum median
   fm = frekuensi median
    p = panjang kelas
```

Untuk menghitung median data berkelompok di EMT, dapat dilakukan dengan cara berikut:
1. Menentukan tepi bawah kelas (Tb), panjang kelas (P), dan tepi atas kelas (Ta) dengan rumus :

$$T_b = a - 0,5$$

$$P = (b - a) + 1$$

$$T_a = b + 0.5$$

dengan a = batas bawah kelas dan b = batas atas kelas

2. Mendeskripsikan data dalam bentuk tabel, dengan perintah

> r=tepi bawah terkecil:panjang kelas:tepi atas terbesar; v=[frekuensi];
> T:=r[1:jumlah kelas]' | r[2:jumlah kelas + 1]' | v'; writetable(T,labc=["tepi bawah","tepi atas","frekuensi"]))

3. Mendeskripsikan tepi bawah kelas median, panjang kelas median, banyak data,frekuensi kumulatif sebelum median, frekuensi kelas median

> Tb=(tepi bawah kelas median), p=(panjang kelas median), n=(Total frekuensi), Fks=(frekuensi kumulatif sebelum median), fm=(frekuensi kelas median)

4. Menghitung median data dengan perintah:

> Tb+p*(1/2*n-Fks)/fm

Contoh Soal:
1. Berikut adalah data hasil dari pengukuran berat badan 50 siswa SD Negeri Tambakrejo. Dari ke 50 siswa, mayoritas siswa memiliki berat badan yang ideal. Siswa yang mempunyai berat badan dalam rentang 21-26 kg sebanyak 5 orang, yang mempunyai berat badan dalam rentang 27-32 kg sebanyak 10 orang, yang mempunyai berat badan dalam rentang 33-38 kg sebanyak 12 orang, yang mempunyai berat badan dalam rentang 39-44 kg sebanyak 14 orang, yang mempunyai berat badan dalam rentang 45-50 kg sebanyak 7 orang, dan yang mempunyai berat badan 51-56 kg sebanyak 2 orang. Tentukan median dari data hasil pengukuran berat badan 50 siswa di SD tersebut!
Penyelesaian:
Menentukan tepi bawah kelas yang terkecil

```
>21-0.5
```

```
  20.5
```

Menentukan panjang kelas

```
>(26-21)+1
```

```
  6
```

Menentukan tepi atas kelas yang terbesar

```
>56+0.5
```

```
  56.5
```

```
>r=20.5:6:56.5; v=[5,10,12,14,7,2];
>T:=r[1:6]' | r[2:7]' | v'; writetable(T,labc=["TB","TA","frek"])
```

TB	TA	frek
20.5	26.5	5
26.5	32.5	10
32.5	38.5	12
38.5	44.5	14
44.5	50.5	7
50.5	56.5	2

Berdasarkan data, median berada pada urutan ke 25, maka median berada pada kelas 32.5-38.5.

```
>Tb=32.5, p=6, n=50, Fks=15, fm=12
```

```
  32.5
  6
  50
  15
  12
```

```
>Tb+p*(1/2*n-Fks)/fm

  37.5
```

Diketahui bahwa median berada di data ke 25

$$Me = Tb + \frac{\frac{1}{2}n - f_{ks}}{f_m}.p$$

$$Me = 32.5 + \frac{\frac{1}{2}(50) - 15}{12}.6$$

$$Me = 32.5 + 5$$

$$Me = 37.5$$

Jadi median dari data hasil pengukuran berat badan 50 siswa SD Tambakrejo adalah 37.5

Modus

Modus merupakan ruang lingkup kajian pada analisis statistika deskriptif yang termasuk dalam ukuran pemusatan data. Modus adalah nilai yang sering muncul diantara sebaran data atau nilai yang memiliki frekuensi tertinggi dalam distribusi data. Modus terdiri dari 2 jenis yaitu modus untuk data tunggal dan modus untuk data kelompok.
1. Modus data tunggal
Cara menentukan modus untuk data tunggal terbilang cukup mudah yaitu dengan mengurutkan data dari yang terkecil ke yang terbesar sehingga data-data yang memiliki nilai yang sama akan berdekatan satu sama lain, lalu mencari frekuensi dari masing-masing data dan pilih data dengan frekuensi tertinggi.
2. Modus data kelompok
Jika data telah dikelompokkan, maka telah disajikan dalam bentuk distribusi frekuensi. Berikut rumus untuk mencari modus data kelompok:

$$Mo = Tb + \frac{d_1}{d_1 + d_2}.c$$

```
Keterangan:
  Tb = Tepi bawah
  d1 = selisih f modus dengan f sebelumnya
  d2 = selisih f modus dengan f sesudahnya
   c = panjang kelas
```

Untuk menghitung modus data berkelompok di EMT, dapat dilakukan dengan cara berikut:
1. Menentukan tepi bawah kelas (Tb), panjang kelas (P), dan tepi atas kelas (Ta) dengan rumus :

$$T_b = a - 0,5$$

$$P = (b - a) + 1$$

$$T_a = b + 0.5$$

dengan a = batas bawah kelas dan b = batas atas kelas

2. Mendeskripsikan data dalam bentuk tabel, dengan perintah

```
> r=tepi bawah terkecil:panjang kelas:tepi atas terbesar; v=[frekuensi];
> T:=r[1:jumlah kelas]' | r[2:jumlah kelas + 1]' | v'; writetable(T,labc=["tepi bawah","tepi atas","frekuensi"]))
```

3. Mendeskripsikan tepi bawah kelas modus, panjang kelas modus,selisih frekuensi modus dengan frekuensi sebelumnya, selisih frekuensi modus dengan frekuensi sesudahnya

> Tb=(tepi bawah kelas modus), p=(panjang kelas modus), d1=(selisih frekuensi modus dengan frekunsi sebelumnya), d2=(selisih frekuensi modus dengan frekuensi sesudahnya)

4. Menghitung modus dengan perintah:

> Tb+p*d1/(d1+d2)

Contoh Soal:
1. Berikut adalah data hasil dari pengukuran berat badan 30 siswa SD Negeri Tambakrejo. Dari ke 30 siswa, mayoritas siswa memiliki berat badan yang ideal. Siswa yang mempunyai berat badan dalam rentang 21-25 kg sebanyak 2 orang, yang mempunyai berat badan dalam rentang 26-30 kg sebanyak 8 orang, yang mempunyai berat badan dalam rentang 31-35 kg sebanyak 9 orang, yang mempunyai berat badan dalam rentang 36-40 kg sebanyak 6 orang, yang mempunyai berat badan dalam rentang 41-45 kg sebanyak 3 orang, dan yang mempunyai berat badan 46-50 kg sebanyak 2 orang. Tentukan modus dari data hasil pengukuran berat badan 30 siswa di SD tersebut!
Penyelesaian:
Menentukan tepi bawah kelas yang terkecil

```
>21-0.5
```

```
    20.5
```

Menentukan panjang kelas

```
>(25-21)+1
```

```
    5
```

Menentukan tepi atas yang terbesar

```
>50+0.5
```

```
    50.5
```

```
>r=20.5:5:50.5; v=[2,8,9,6,3,2];
>T:=r[1:6]' | r[2:7]' | v'; writetable(T,labc=["TB","TA","frek"])
```

TB	TA	frek
20.5	25.5	2
25.5	30.5	8
30.5	35.5	9
35.5	40.5	6
40.5	45.5	3
45.5	50.5	2

Berdasarkan data, modus berada pada kelas 30.5-35.5.

```
>Tb=30.5, p=5, d1=1, d2=3
```

```
    30.5
    5
    1
    3
```

```
>Tb+p*d1/(d1+d2)
```

```
    31.75
```

Analisis Data Statistika Deskriptif

Menurut Sudjana (2000) statistika adalah pengetahuan yang

berhubungan dengan cara-cara pengumpulan data, pengolahan atau penganalisaannya dan penarikan kesimpulan berdasarkan kumpulan data dan penganalisaan yang dilakukan. Tujuan dari statistika adalah untuk menyajikan data secara ringkas dan mudah dimengerti, sehingga dapat membuat kesimpulan dari suatu populasi dari data yang diambil. Statistika dibagi menjadi statistika deskriptif dan inferensial. Statistika deskriptif merupakakan cabang statistika yang fokus pada pengumpulan, penyajian, dan menganalisa yang berwujud angka-angka agar data tersebut dapat ditampilkan secara ringkas dan informatif. Dalam statistika terdapat empat jenis statistika deskriptif yaitu, distribusi frekuensi, ukuran pemusatan, ukuran letak, dan ukuran penyebaran.
Distribusi frekuensi adalah merangkum data dalam bentuk tabel atau grafik, yang menunjukkan seberapa sering setiap nilai muncul dalam suatu himpunan data. Ukuran pemusatan terdiri dari mean (rata-rata), median (nilai tengah), dan modus (nilai yang sering muncul). Ukuran letak terdiri dari kuartil (membagi data menjadi empat bagian yang sama besar), desil (membagi data menjadi 10 bagian sama besar), dan persentil (membagi data menjadi 100 bagian sama besar). Ukuran penyebaran terdiri dari jangkauan (nilai maksimum dan minimum dalam suatu set data), varians (mengukur sejauh mana nilai tersebut menyebar dari rata-ratanya), simpangan baku (akar kuadrat dari varians, yang memberikan gambaran tentang seberapa dekat nilai-nilai dalam suatu set data terhadap rata-ratanya.)
Dari penjelasan singkat tentang data statistika deskriptif maka pada sub bab ini, saya akan membahas analisis data statistika deskriptif dengan jenis ukuran letak dan ukuran penyebaran.

Ukuran Letak

Kita singgung kembali. Ukuran letak disini terdiri dari kuartil (membagi data menjadi empat bagian yang sama besar), desil (membagi data menjadi 10 bagian sama besar), dan persentil (membagi data menjadi 100 bagian sama besar).

Kuartil

Kuartil merupakan penggambaran pembagian data menjadi empat bagian yang sama besar. Kuartil membagi data menjadi tiga titik tertentu, yang dikenal sebagai kuartil pertama (Q1), kuartil kedua (Q2), dan kuartil ketiga (Q3). Rumus yang digunakan dalam menentukan posisi Q1,Q2,dan Q3 dengan

$$Q_i = i\frac{n+1}{4}$$

dimana i adalah indeks kuartil dan n adalah jumlah total data

Dalam EMT fungsi yang dapat digunakan untuk menentukan kuartil adalah 'quartiles()'.

```
>data=[93,80,52,41,60,77,55,71,79,81,64,83,32,95,75,54,90,80,95];
```

Diketahui sebuah data dari hasil nilai ujian akhir dari siswa

```
>urut=sort(data)
```

```
[32,  41,  52,  54,  55,  60,  64,  71,  75,  77,  79,  80,  80,  81,
83,  90,  93,  95,  95]
```

Dalam menentukan kuartil langkah pertama yang dapat dilakukan dengan mengurutkan data tersebut dengan fungsi sort(data). Fungsi sort(data) dalam Euler Math Toolbox digunakan untuk mengurutkan elemen-elemen dalam suatu vektor atau matriks (dari nilai terkecil ke nilai terbesar).

```
>quartiles(data)
```

```
[32,  55,  77,  83,  95]
```

Dalam output hitung yang dihasilkan dari 'quartiles(data)' dapat diketahui bahwa nilai Q1(kuartil bawah) = 55 , Q2(kuartil tengah(median)) = 77, dan Q3(kuartil atas)= 83. Lalu untuk nilai paling kanan dan paling kiri merupakan minimum dan maximum dari suatu data yang diketahui.

Dengan cara rumus kuartil yang diketahui dapat dihitung nilai kuartil atas, tengah, dan bawah. Dengan cara...

```
>data=[93,80,52,41,60,77,55,71,79,81,64,83,32,95,75,54,90,80,95];
>urut=sort(data)
```

```
[32,  41,  52,  54,  55,  60,  64,  71,  75,  77,  79,  80,  80,  81,
83,  90,  93,  95,  95]
```

```
>a=length(data)
```

```
19
```

Setelah mengurutkan data tersebut, selanjutnya hitung banyaknya data tersebut dengan menggunakan fungsi 'length()'. Fungsi tersebut berfungsi untuk mengetahui banyaknya elemen dalam suatu vektor atau matriks.

```
>Q1=((a+1)/4)
```

```
5
```

Hasil Q1 menunjukkan 5 yang berarti letak kuartil ke-1 ada di data nomer 5 yaitu 55.

```
>Q2=((a+1)/2)
```

```
10
```

Hasil Q2 menunjukkan 10 yang berarti letak kuartil ke-2 ada di data nomer 10 yaitu 77.

```
>Q3=60/4
```

```
15
```

Hasil Q3 menunjukkan 15 yang berarti letak kuartil ke-3 ada di data nomer 15 yaitu 83. Mengapa fungsi ini saya hitung langsung? karena saat percobaan EMT tidak biasa membaca perintah maka dari itu saya menggunakan EMT secara langsung.

Dari beberapa percobaan tersebut dapat diketahui bahwa dengan cara rumus kuartil ataupun menggunakan fungsi 'quartiles()' tersebut akan sama.

Desil dan Persentil

Desil merupakan pembagian data ke dalam sepuluh kelompok sebanding yang disusun berdasarkan urutan nilainya. Desil ke-1 (D1) adalah nilai terendah, desil ke-2 (D2) adalah nilai yang membagi data menjadi 10% terendah, desil ke-3 (D3) membagi data menjadi 20% terendah, dan seterusnya. Rumus dari desil adalah

$$D_i = x_i \frac{(n+1)}{10}$$

Sedangkan persentil suatu nilai atau titik data yang membagi distribusi data menjadi persentase tertentu atau menjadi 100 bagian yang sama besar. Rumus dari persentil adalah

$$P_i = x_i \frac{(n+1)}{100}$$

Dalam EMT fungsi yang digunakan anatata desil dan persentil sama yaitu 'quantile()'. Perbedaan penggunaannya terletak pada nilai yang akan dibaginya.

```
>data=[93,80,52,41,60,77,55,71,79,81,64,83,32,95,75,54,90,80,95,55];
>urut=sort(data)
```

```
 [32,  41,  52,  54,  55,  55,  60,  64,  71,  75,  77,  79,  80,  80,
 81,  83,  90,  93,  95,  95]
```

```
>quantile(urut,0.1)
```

```
 50.9
```

Dari hasil tersebut dapat diketahui bawha nilai desil ke-1 dan persentil ke-10 adalah 50,9

```
>quantile(urut,0.2)
```

```
 54.8
```

```
>quantile(urut,0.5)
```

```
 76
```

```
>quantile(urut,0.7)
```

```
 80.3
```

Dari dua percobaan tersebut diketahui nilai persentil ke-20 dan persentil ke-50 adalah 54,8 dan 76. Hasil tersebut sama dengan hasil desil ke-2 dan desil ke-5.

Karena desil ke-8 (D8) adalah nilai yang membagi data menjadi 80% di bawahnya dan 20% di atasnya. Sedangkan persentil 80% (P80) juga merujuk pada nilai yang membagi data menjadi 80% di bawahnya dan 20% di atasnya.

Ukuran Penyebaran

Kita singgung kembali. Ukuran penyebaran disini terdiri dari jangkauan yang terdiri dari nilai maksimum dan minimum, varians (mengukur sejauh mana nilai tersebut menyebar dari rata-ratanya), simpangan baku(akar kuadrat dari varians yang memberikan gambaran tentang seberapa dekat nilai-nilai dalam suatu set data terhadap rata-ratanya).

Jangkauan

Jangkauan (range) adalah salah satu ukuran penyebaran yang paling sederhana dalam statistika. Jangkauan dapat dihitung dengan mengambil selisih antara nilai maksimum dan minimum dalam suatu set data.

Untuk menemukan jangkauan data tunggal di EMT dapat menggunakan perintah berikut:
> x=[data]; max(x)-min(x)

```
>data=[65,55,70,85,90,75,80,75];
>a=max(data)
```

```
 90
```

```
>b=min(data)
```

```
 55
```

Permisalan dengan a = max(data) dan b = min(data) memudahkan untuk menghitung nilai jangkauan.

```
>jangkauan = a-b
```

```
  35
```

Dalam menentukan nilai jangkauan antara nilai maximum dan minimum dapat dilakukan dengan cara menentukan nilai maximum dan minimum dari data tersebut, lalu kurangi antara hasil dari nilai maximum dan minimum tersebut.

Untuk data berkelompok dapat menggunakan rumus;
> max(transpose(T[,2]))-min(transpose(T[,1]))

Dimisalkan diketahui sebuah tabel distribusi frekuensi

```
>r=39.5:5:69.5; v=[5,18,42,20,9,6];
>T:=r[1:6]' |r[2:7]'| v'; writetable(T,labc=["TB","TA","Frek"])
```

TB	TA	Frek
39.5	44.5	5
44.5	49.5	18
49.5	54.5	42
54.5	59.5	20
59.5	64.5	9
64.5	69.5	6

```
>max(transpose(T[,2]))-min(transpose(T[,1]))
```

```
  30
```

Jadi dapat diketahui bahwa jangkauan dari tabel tersebut adalah 30 orang.

Varians

Varians adalah nilai statistik yang sering kali dipakai dalam menentukan kedekatan sebaran data yang ada di dalam sampel dan seberapa dekat titik data individu dengan mean atau rata-rata nilai dari sampel itu sendiri.

Pada EMT, dalam menentukan suatu varians dapat menggunakan fungsi 'dev()^2'. Fungsi dev disini merupakan kepanjangan dari deviations yang berarti suatu verktor atau array yang berisi deviasi dari setiap nilai dalam data.

$$dev = x_i - \bar{x}$$

```
>data=[65,55,70,85,90,75,80,75];
>urut=sort(data)
```

```
  [55,  65,  70,  75,  75,  80,  85,  90]
```

```
>a = mean(urut)
```

```
  74.375
```

Langkah pertama dalam menentukan varians adalah menentukan mean (rata-rata) dari suatu data tersebut.

```
>dev = urut-a
```

```
  [-19.375,  -9.375,  -4.375,  0.625,  0.625,  5.625,  10.625,  15.625]
```

Setelah menentukan rata-rata, selanjutnya menentukan deviations dari nilai data tersebut dikurangi dengan mean

```
>varians = mean(dev^2)
```

```
  108.984375
```

Lalu diperoleh nilai varians 108.984375 dengan langkah mean(dev^2)

Varians juga dapat menentukan data kelompok. Dengan menggunakan fungsi 'mean()' dan 'sum(). Sum disini digunakan untuk menjumlahkan data tersebut.

```
>r=499.5:100:1099.5; v=[4,6,12,15,10,3];
>T:=r[1:6]' | r[2:7]' | v'; writetable(T,labc=["tepi bawah","tepi atas","frekuensi"])
```

```
  tepi bawah tepi atas frekuensi
       499.5     599.5         4
       599.5     699.5         6
       699.5     799.5        12
       799.5     899.5        15
       899.5     999.5        10
       999.5    1099.5         3
```

```
>(T[,1]+T[,2])/2; t=fold(r,[0.5,0.5])
```

```
  [549.5,  649.5,  749.5,  849.5,  949.5,  1049.5]
```

Mencari titik tengah antar interval tersebut dan mencari nilai mean dari data tersebut.

```
>m = mean(t,v)
```

```
  809.5
```

Menghitung rata-rata dari data tersebut.

```
>sum(v*(t-m)^2)/(sum(v)-1)
```

```
  17551.0204082
```

Melalui rumus tersebut maka dapat dihasilkan nilai varians dari data table tersebut dengan nilai 17551,0204082.

Simpangan Baku

Simpangan baku atau deviasi standar adalah ukuran seberapa jauh nilai-nilai dalam satu set data tersebar dari nilai rata-ratanya. Ini memberikan gambaran tentang sejauh mana nilai-nilai dalam data "berkumpul" atau "menyebar" di sekitar rata-ratanya.

$$\sigma = \sqrt{varians}$$

Pada sub bab sebelumnya telah dijelaskan bagaiaman cara mencari nilai suatu varians. Mengulang kembali varians menggunakan fungsi

$$mean * dev()^2$$

sedangkan simpangan baku menggunakan

$$\sigma = \sqrt{mean * dev()^2}$$

```
>data=[65,55,70,85,90,75,80,75];
>urut=sort(data)
```

```
     [55,  65,  70,  75,  75,  80,  85,  90]
```

```
>a = mean(urut)
```

```
     74.375
```

```
>dev = urut-a
```

```
     [-19.375,  -9.375,  -4.375,  0.625,  0.625,  5.625,  10.625,  15.625]
```

```
>varians = mean(dev^2)
```

```
     108.984375
```

Langkah dalam menentukan simpangan baku sama dengan cara menentukan varians dari sebuah data tersebut. Dari hasil mean, dev, dan hasil varians maka dapat diketahui nilai simpangan baku tersebut.

```
>simpanganBaku = sqrt(varians)
```

```
     10.4395581803
```

Nilai simpangan baku dari data tersebut adalah 10.4395581803

Tidak hanya data tunggal saja, dalam menentukan simpangan baku dapat dilakukan juga apabila data tersebut berkelompok.

```
>r=499.5:100:1099.5; v=[4,6,12,15,10,3];
>T:=r[1:6]' | r[2:7]' | v'; writetable(T,labc=["tepi bawah","tepi atas","frekuensi"])
```

```
      tepi bawah tepi atas frekuensi
           499.5     599.5         4
           599.5     699.5         6
           699.5     799.5        12
           799.5     899.5        15
           899.5     999.5        10
           999.5    1099.5         3
```

```
>(T[,1]+T[,2])/2; t=fold(r,[0.5,0.5])
```

```
     [549.5,  649.5,  749.5,  849.5,  949.5,  1049.5]
```

```
>m=mean(t,v)
```

```
     809.5
```

```
>sqrt(sum(v*(t-m)^2)/(sum(v)-1))
```

```
     132.480264221
```

Cara tersebut juga sama seperti cara menentukan varians yang sudah dijelaskan sebelumnya. Hanya saja perbedaan disini ditambah kata 'sqrt' yang berarti akar, sesuai dengan rumus simpangan baku tersebut.

SUB TOPIK 6 : Menggambar Grafik Statistika

Diagram adalah suatu representasi simbolis informasi dalam bentuk geometri 2 dimensi sesuai teknik visualisasi. Kadang teknik yang dipakai memanfaatkan visualisasi tiga dimensi yang kemudian diproyeksikan ke permukaan dua dimensi. Kata grafik dan bagan biasa dipakai sebagai sinonim kata diagram.

Grafik juga diartikan sebagai suatu kombinasi angka, huruf, simbol, gambar, lambang, perkataandan lukisan yang disajikan dalam suatu media untuk menggambarkan informasi dari data.

Diagram dapat digunakan untuk alasan yang berbeda, seperti untuk menunjukkan bagian dari keseluruhan, langkah-langkah dari suatu proses, dan hubungan. Sebuah bantuan grafis akan menampilkan informasi secara visual sehingga pembaca dapat lebih memahami dan mengingat ide-ide.

Jenis-Jenis Diagram ataupun kurva pada EMT:
- Diagram kotak,
- Diagram batang,
- Diagram lingkaran,
- Diagram bintang,
- Diagram impuls,
- Histogram,
- Kurva fungsi kerapatan probabilitas,
- Kurva fungsi distribusi kumulatif,
- Diagram titik,
- Diagram garis,
- Kurva regresi.

Diagram Kotak

Dalam statistika deskriptif, diagram kotak garis atau boxplot adalah metode grafis untuk menggambarkan kumpulan data numerik berdasarkan nilai kuartilnya. Diagram kotak garis bersifat nonparametrik, artinya diagram ini menampilkan variasi sampel populasi statistik tanpa membuat asumsi apa pun tentang distribusi statistik yang mendasarinya. Jarak antara bagian-bagian kotak yang berbeda menunjukkan derajat dispersi (sebaran), kemiringan, dan pencilan dari data tersebut.

Boxplot merupakan ringkasan distribusi sampel yang disajikan secara grafis yang bisa menggambarkan bentuk distribusi data (skewness), ukuran tendensi sentral dan ukuran penyebaran (keragaman) data pengamatan.
Terdapat 5 ukuran statistik yang bisa kita baca dari boxplot, yaitu:
- nilai minimum: nilai observasi terkecil
- Q1: kuartil terendah atau kuartil pertama
- Q2: median atau nilai pertengahan
- Q3: kuartil tertinggi atau kuartil ketiga
- nilai maksimum: nilai observasi terbesar.
- Selain itu, boxplot juga dapat menunjukkan ada tidaknya nilai outlierdan nilai ekstrim dari data pengamatan.

Contoh penggunaan diagram kotak di EMT

Misalkan ada sebuah data M sebagai berikut:
Maka, akan divisualisasikan data M dengan diagram kotak

```
>M=[1000,1004,998,997,1002,1001,998,1004,998,997]; ...
 mean(M), dev(M),
```

```
  999.9
  2.72641400622
```

$$mean = \frac{sum(data)}{len(data)}$$

Jumlah data dibagi banyak data.

Sedangkan dev() merupakan standart deviasi nya.
Rumusnya:

$$deviasi = [(x - mean)^2 forxindata]$$

$$variasi = \frac{sum(deviasi)}{len(data)}$$

$$standartdeviasi = \sqrt{variasi}$$

Pertama-tama data M akan diurut berdasarkan nilainya, dari terendah hingga tertinggi(terbesar).

$$M = [997, 997, 998, 998, 998, 1000, 1001, 1002, 1004, 1004]$$

Mencari Median:
Median=kuartil tengah(Q3)
Median terletak pada data ke-

$$Median = \frac{1}{2}.(datake - (\frac{n}{2}) + datake - (\frac{n}{2}) + 1)$$

Dikarenakan data berjumlah genapa maka, digunakan rumus seperti itu.

$$= \frac{1}{2}.(datake - (\frac{10}{2}) + datake - (\frac{10}{2}) + 1)$$

$$= \frac{1}{2}.(datake - (5) + datake - (5) + 1)$$

$$= \frac{1}{2}.(998 + 1000)$$

$$= 999$$

Ditemukan median-nya adalah 999

Mencari Kuartil atas dan bawah

$$Q1 = datake - \frac{1}{4}.n$$

$$= datake - \frac{1}{4}.10$$

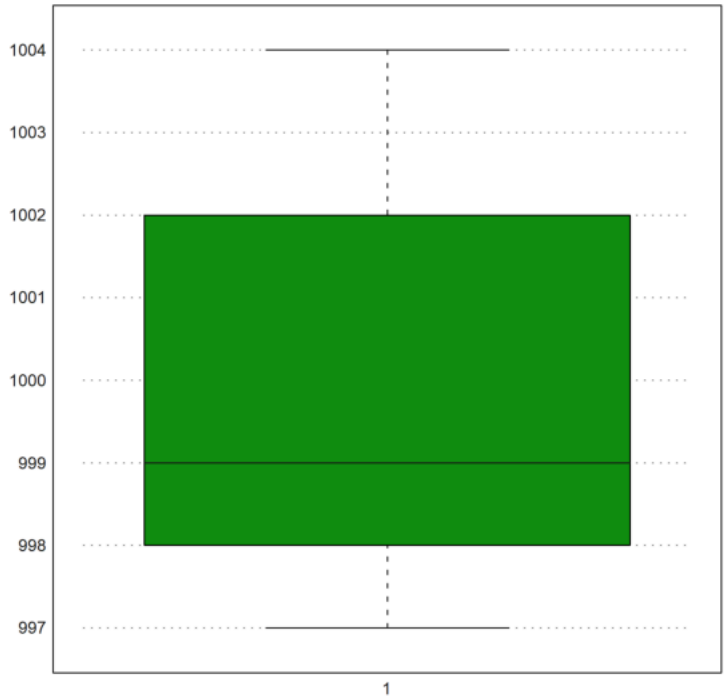
$$= 998$$

$$Q3 = datake - \frac{3}{4}.n$$

$$= datake - \frac{3}{4}.10$$

$$= 1002$$

```
>boxplot(M, outliers=1.5):
```



Plot diatas merupakan diagram kotak, Kotak yang berwarna hijau tersebut merupakan hasil dari perintah EMT yang diberikan dimana kotak tersebut dibatasi dengan Kuartil terendah dan kuartil tertinggi, dan terdapat garis hitam di dalam kotak merupakan Median atau nilai tengah.

image: BoxPlot.png

Di bawah ini diperlihatkan rincian detail boxplot beserta cara penentuan batas-batasnya.

```
# Bagian utama boxplot adalah kotak berbentuk persegi (Box) yang
```

merupakan bidang yang menyajikan interquartile range (IQR), dimana 50 % dari nilai data pengamatan terletak di sana.
- Panjang kotak sesuai dengan jangkauan kuartil dalam (inner Quartile Range, IQR) yang merupakan selisih antara Kuartil ketiga (Q3) dengan Kuartil pertama (Q1). IQR menggambarkan ukuran penyebaran data. Semakin panjang bidang IQR menunjukkan data semakin menyebar. Pada Gambar, IQR = UQ – LQ = Q3 – Q1
- Garis bawah kotak (LQ) = Q1 (Kuartil pertama), dimana 25% data pengamatan lebih kecil atau sama dengan nilai Q1
- Garis tengah kotak = Q2 (median), dimana 50% data pengamatan lebih kecil atau sama dengan nilai ini
- Garis atas kotak (UQ) = Q3 (Kuartil ketiga) dimana 75% data pengamatan lebih kecil atau sama dengan nilai Q1

```
# Garis yang merupakan perpanjangan dari box(baik ke arah atas
```

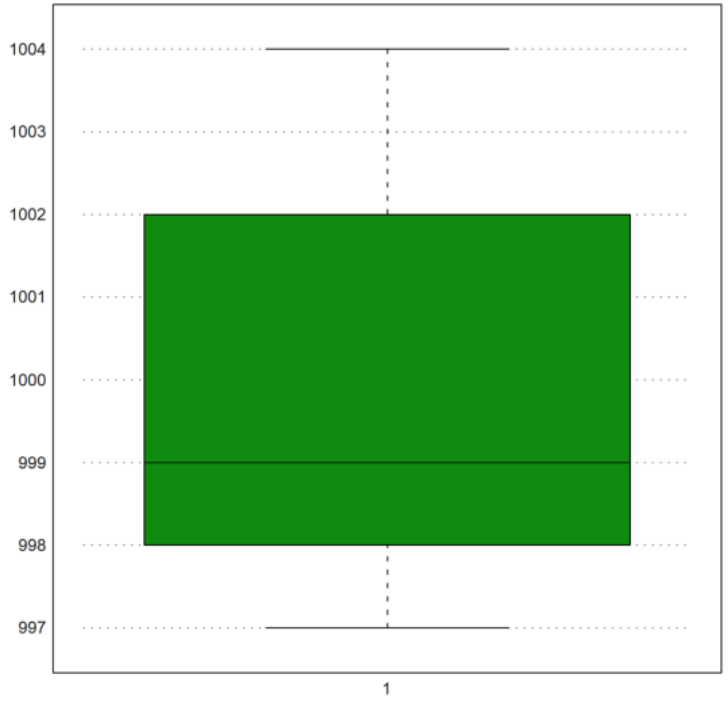
ataupun ke arah bawah) dinamakan dengan whiskers.
- Whiskers bawah menunjukkan nilai yang lebih rendah dari kumpulan data yang berada dalam IQR
- Whiskers atas menunjukkan nilai yang lebih tinggi dari kumpulan data yang berada dalam IQR
- Panjang whisker = 1.5 x IQR. Masing-masing garis whisker dimulai dari ujung kotak IQR, dan berakhir pada nilai data yang bukan dikategorikan sebagai outlier (Pada gambar, batasnya adalah garis UIF dan LIF). Dengan demikian, nilai terbesar dan terkecil dari data pengamatan (tanpa termasuk outlier) masih merupakan bagian dari Boxplot yang terletak tepat di ujung garis tepi whiskers.

```
# Nilai yang berada di atas atau dibawah whiskerdinamakan nilai
```

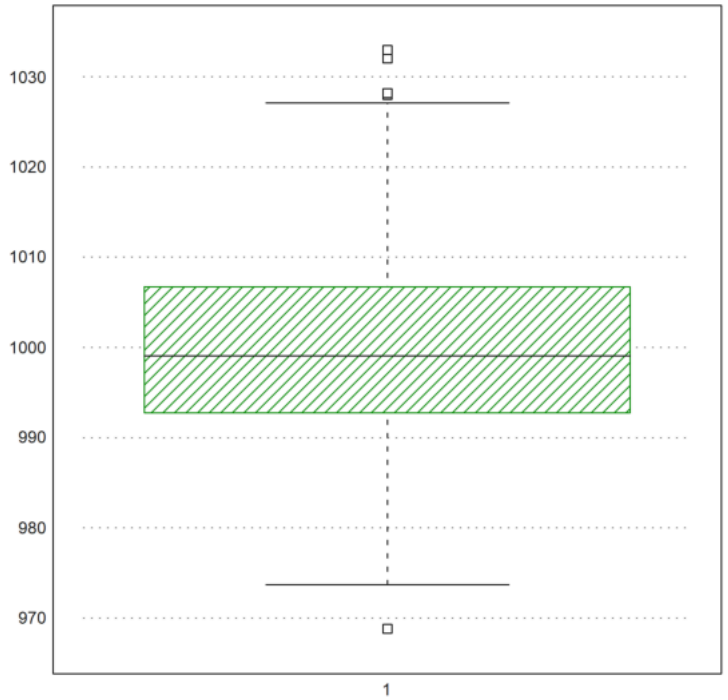
outlier atau ekstrim.
- Nilai outlier adalah nilai data yang letaknya lebih dari 1.5 x panjang kotak (IQR), diukur dari UQ (atas kotak) atau LQ (bawah kotak). Pada Gambar di atas, ada 2 data pengamatan yang merupakan outlier, yaitu data pada case 33 dan case 55 (ada pada baris ke 33 dan baris 35)
=> Q3 + (1.5 x IQR) < outlier atas = Q3 + (3 x IQR)
=> Q1 – (1.5 x IQR) > outlier bawah = Q1 – (3 x IQR)
- Nilai ekstrim adalah nilai-nilai yang letaknya lebih dari 3 x panjang kotak (IQR), diukur dari UQ (atas kotak) atau LQ (bawah kotak). Pada gambar di atas, ada 1 data yang merupakan nilai ekstem, yaitu data pada case 15.
=> Ekstrim bagian atas apabila nilainya berada di atas Q3 + (3 x IQR) dan
=>Ekstrim bagian bawah apabila nilainya lebih rendah dari Q1 – (3 x IQR)

Berikut penggunaan boxplot pada distribusi normal:

```
>boxplot(M, lab=none, style="0#", textcolor=red, outliers=1.5, pointstyle="o", range=none):
```

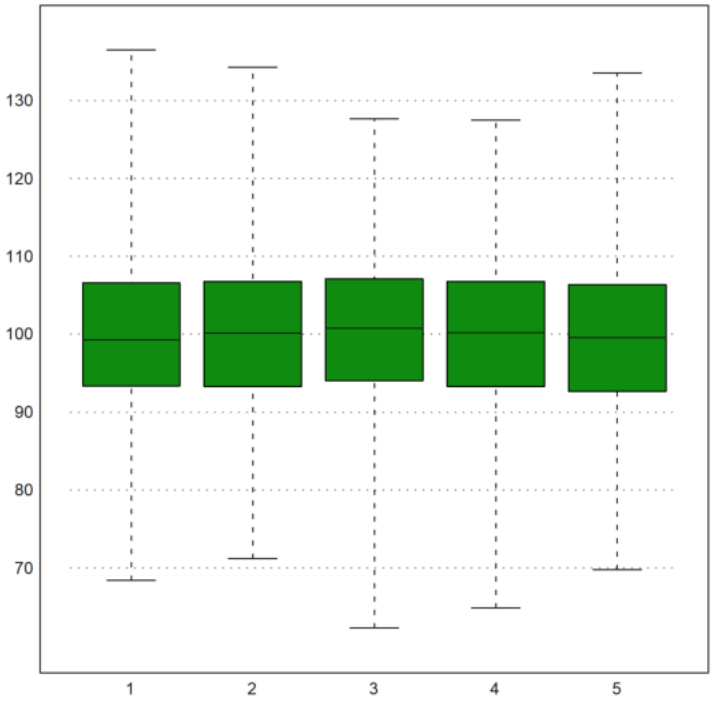



```
>x=normal(1000)*10+1000; boxplot(x, lab=none, style="/", textcolor=red, outliers=1.5, pointstyle="=",range=none):
```



Penjelasan:
- "lab" disini digunakan untuk melabelkan atau menamai kotak
- "style" disini digunakan untuk mengatur style atau gaya yang digunakan pada kotak, bisa menggunakan 0, =, +, /, |
- "textcolor' digunakan untuk mewarnai text pada "lab"
- "outliers" Parameter outliers dalam konteks boxplot menentukan apakah dan bagaimana outlier (nilai-nilai ekstrem yang jauh dari kebanyakan nilai) ditampilkan dalam boxplot. Dalam beberapa fungsi pembuat boxplot, parameter ini digunakan untuk mengontrol penampilan dan penanganan.
- "pointstyle" digunakan untuk mengatur style atau gaya pada point-point yang ada
- "range" digunakan untuk mengatur jarak

```
>x=randnormal(5,1000,100,10); boxplot(x,outliers=none):
```



Boxplots dapat membantu kita dalam memahami karakteristik dari distribusi data. Selain untuk melihat derajat penyebaran data (yang dapat dilihat dari tinggi/panjang boxplot) juga dapat digunakan untuk menilai kesimetrisan sebaran data. Panjang kotak menggambarkan tingkat penyebaran atau keragaman data pengamatan, sedangkan letak median dan panjang whisker menggambarkan tingkat kesimetrisannya.

Diagram Batang

```
Diagram batang merupakan jenis grafik yang digunakan untuk
```

menunjukkan dan membandingkan kuantitas data dalam kategori yang berbeda. Diagram batang umumnya digunakan untuk menggambarkan perkembangan nilai suatu objek penelitian dalam kurun waktu tertentu. Diagram batang menunjukkan keterangan-keterangan dengan batang-batang tegak atau mendatar dan sama lebar dengan batang-batang terpisah

```
Diagram batang memiliki kelebihan, yaitu diagram batang merupakan
```

diagram paling sederhana dan paling umum digunakan. Namun diagram batang juga memiliki kekurangan, yaitu diagram batang hanya disajikan data yang telah dikelompokkan atas atribut dan kategori. Dan Diagram batang tidak dapat menampilkan datum dari tiap orang atau benda yang dicatat(sebut saja data individual).

Contoh penggunaan diagram batang di EMT

Untuk membuat suatu diagram atau plot kita memerlukan sebuah data yang nantinya akan diolah. Data yang kita gunakan merupakan hasil pemilu Jerman sejak tahun 1990, diukur dalam kursi.

```
>BW := [ ...
 1990,662,319,239,79,8,17; ...
 1994,672,294,252,47,49,30; ...
 1998,669,245,298,43,47,36; ...
```

```
 2002,603,248,251,47,55,2; ...
 2005,614,226,222,61,51,54; ...
 2009,622,239,146,93,68,76; ...
 2013,631,311,193,0,63,64];
```

Untuk P disini dimasukkan sebagai nama-nama partai pada data sebelumnya.

```
>P:=["CDU/CSU","SPD","FDP","Gr","Li"];
```

Kolom pertama = Tahun Terjadinya Pemilu
Kolom kedua = jumlah kursi keseluruahan pada tahun tertentu
kolom ketiga sampai ketujuh = jumlah kursi tiap partai

```
>BT:=BW[,3:7]; BT:=BT/sum(BT); YT:=BW[,1]';
```

Fungsi BW[,3:7] disini mengartikan bahwa fungsi bw akan dipakai pada kolom 3 sampai 7

```
>writetable(BT*100,wc=6,dc=0,>fixed,labc=P,labr=YT)
```

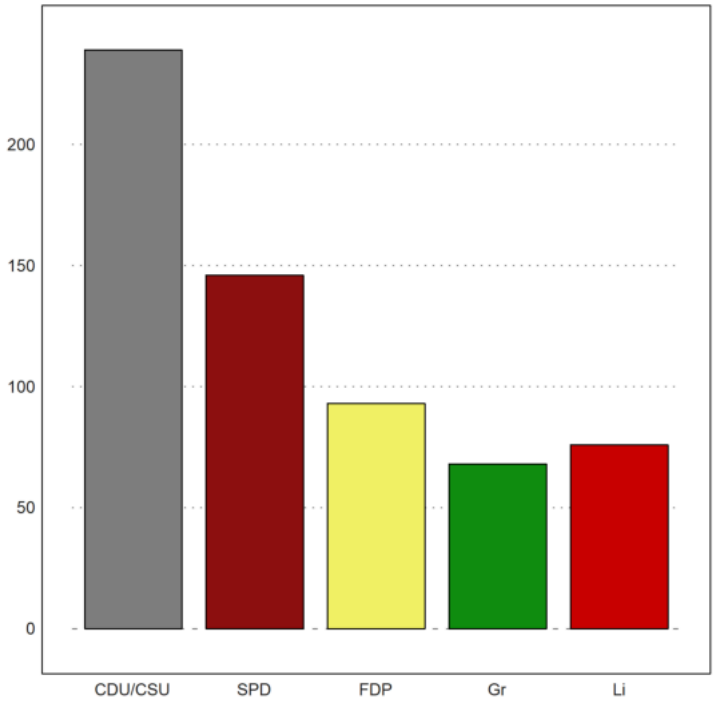
	CDU/CSU	SPD	FDP	Gr	Li
1990	48	36	12	1	3
1994	44	38	7	7	4
1998	37	45	6	7	5
2002	41	42	8	9	0
2005	37	36	10	8	9
2009	38	23	15	11	12
2013	49	31	0	10	10

Memmbaca dari data sebelumnya yang merupakan matriks menjadi sebuah tabel yang tiap kolom nya sudah dinamai.

```
>CP:=[rgb(0.5,0.5,0.5),red,yellow,green,rgb(0.8,0,0)];
```

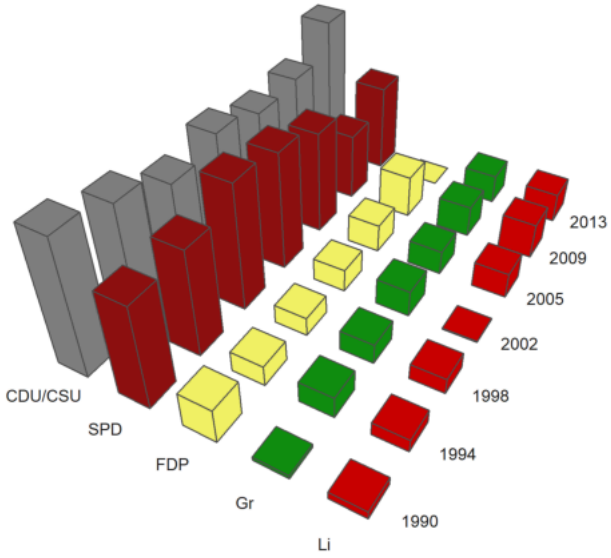
Fungsi CP disini digunakan nantinya untuk membuat perbedaan warna pada setiap batang yang menyesuaikan nama partai

```
>columnsplot(BW[6,3:7],P,color=CP):
```



Berikut merupakan diagram batang, hasil dari memplot data sebelumnya. Ini merupakan diagram batang dalam 2 dimensi.

```
>columnsplot3d(BT,scols=P,srows=YT, ...
   angle=30°,ccols=CP):
```



Ini adalah perintah untuk membuat diagram batang(kolom) dalam 3 Dimensi . Plot ini akan menggunakan data dari BT, dengan kolom-kolom yang dipilih pada data BW, yaitu nama-nama partai (scols=p). Selain itu, parameter angle=30° mengatur sudut pandang plot dalam ruang 3D, ccols=CP mungkin mengatur warna kolom.

Diagram Lingkaran

```
Diagram lingkaran merupakan suatu diagram yang difungsikan untuk
```

menyajikan data dalam bentuk lingkaran baik menggunakan data absolut maupun relatif. Untuk membuat diagram lingkaran pertama-tama kita harus membuat lingkaran terlebih dahulu lalu dibagi-bagi menjadi beberapa sektor. Tiap sektor melukiskan kategori data yang terlebih dahulu diubah kedalam derajat.

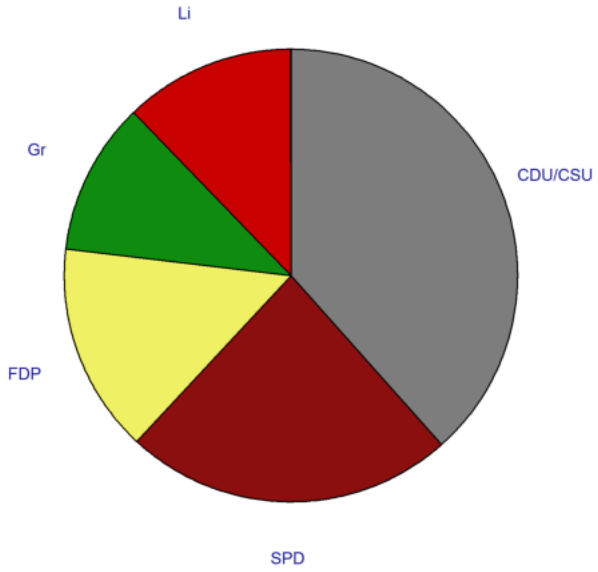
```
Kelebihan diagram lingkaran, yaitu Tempat untuk membuat diagram
```

lingkaran tidak terlalu besar. Dan Diagram lingkaran sangat berguna untuk menunjukkan dan membandingkan proporsi dari data. Namun diagram lingkaran juga memiliki Kekurangan, yaitu Diagram lingkaran tidak dapat menunjukkan frekuensinya.

Contoh penggunaan diagram lingkaran di EMT

Data yang dipakai Masih menggunakan data sebelumnya, yaitu data pemilu jerman dari tahun.

```
>i=[1,2,3,4,5]; piechart(BW[6,3:7][i], style="O#", color=CP[i],lab=P[i], textcolor=blue, r=1.5):
```



"i" digunakan untuk iterasi kolom, atau mengurutkan kolom sesuai potongan pada diagram lingkaran
"textcolor" digunakan untuk mewarnai text
"color" digunakan untuk mewarnai potongan lingkaran
"style" digunakan untuk mengatur gradasi style atau gaya pada diagram lingkaran
"lab" digunakan untuk melabelkan atau menamai setiap potongan lingkaran
"r" digunakan untuk mengatur radius pada diagram lingkaran

```
>i=[1,5,3,2,4]; piechart(BW[6,3:7][i], style="/", color=CP[i],lab=P[i], textcolor=green, r=2.5):
```

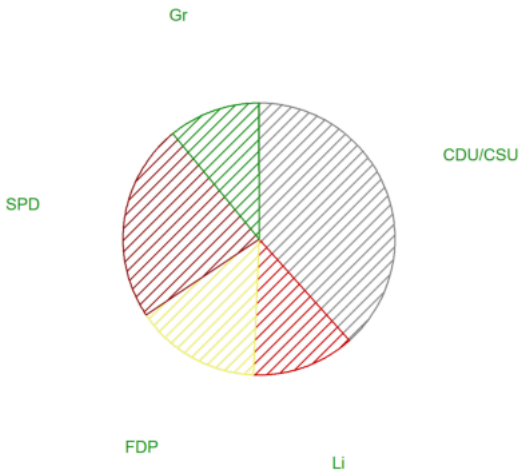


Diagram Bintang

```
Diagram bintang adalah jenis diagram yang terdiri dari titik-titik
```

yang terhubung oleh garis-garis sehingga membentuk pola seperti bintang.Grafik ini memiliki serangkaian garis yang menjulur dari titik pusat, menyerupai bentuk bintang. Setiap lengan dari bintang mewakili dimensi atau variabel yang berbeda, dan panjang garis atau area yang diisi dapat mencerminkan nilai atau kontribusi relatif dari setiap dimensi.

```
Diagram ini sering digunakan dalam matematika, statistik, dan ilmu
```

komputer untuk memvisualisasikan hubungan antara titik-titik atau entitas dalam sebuah sistem. Diagram bintang dapat memiliki berbagai bentuk dan ukuran tergantung pada jumlah titik yang digunakan.

```
Diagram bintang paling sederhana adalah diagram bintang dengan lima
```

titik yang membentuk pola bintang lima sudut. Namun, diagram bintang juga dapat memiliki lebih banyak titik dan membentuk pola yang lebih kompleks. Diagram bintang dapat digunakan untuk berbagai tujuan, seperti memvisualisasikan hubungan antara entitas dalam jaringan sosial, menggambarkan hubungan antara variabel dalam analisis multivariat, atau menggambarkan struktur hierarki dalam organisasi.

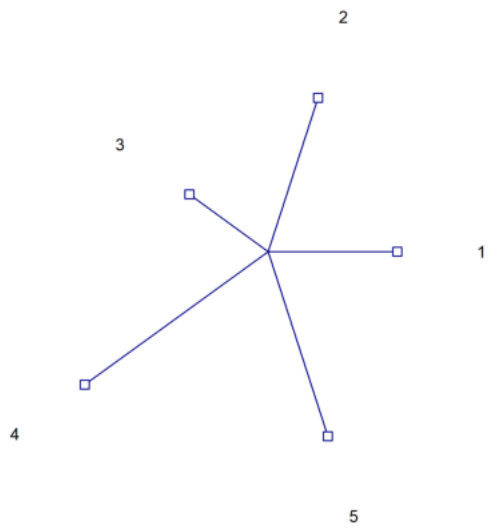
Contoh penggunaan diagram bintang di EMT

Misalkan ada sebuah data V:

```
>v = [4, 5, 3, 7, 6];
```

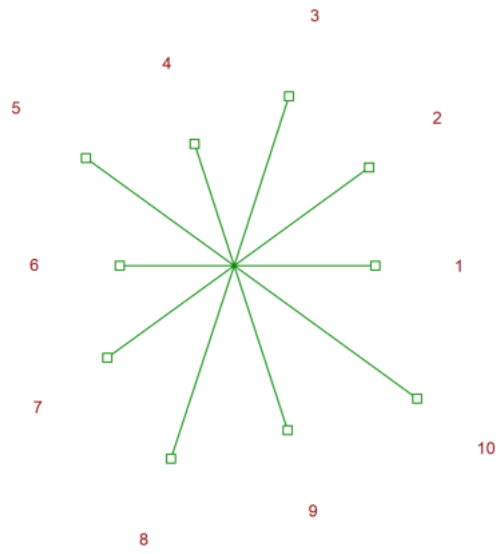
Selanjutnya akan di ilustrasikan data V kebentuk diagram bintang, sebagai berikut:

```
>starplot(v, style="+", color=blue, lab=1:5, rays=10, pstyle="=+", textcolor=black, r=1.5):
```



Penjelasan:
- "lab" disini digunakan untuk melabelkan atau menamai kotak
- "style" disini digunakan untuk mengatur style atau gaya yang digunakan pada garis bintang.
- "textcolor" digunakan untuk mewarnai text pada "lab".
- "rays" digunakan untuk menunjukkan jumlah "rays" atau "spokes" pada radar plot.
- "pstyle" digunakan untuk mengatur style atau gaya pada point-point yang ada.
- "r" digunakan untuk mengatur radius.

```
      #Contoh penggunaan diagram bintang pada distribusi normal statistika

>starplot(normal(1,10)+4,lab=1:10,>rays):
```



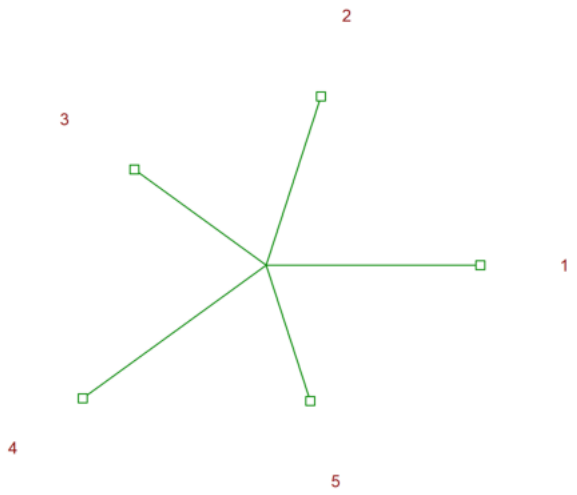

```
#Contoh soal lainnya
```

Misalkan ada sebuah data nilai matematika dari latihan UTBK di suatu tempat les. Data tersebut berisikan 5 orang siswa dengan nilai 87 untuk siswa pertama, 72 untuk siswa kedua, 66 untuk siswa ketiga, 92 untuk siswa keempat, dan 58 untuk siswa kelima. Ilustrasikan data tersebut dengan diagram bintang!

```
>A=[87,72,66,92,58]
```

```
[87,  72,  66,  92,  58]
```

```
>starplot(A,lab=1:5,>rays):
```



#Contoh soal lainnya

PT SukaMaju memiliki tim penjualan yang terdiri dari lima anggota.

Perusahaan tersebut ingin mengevaluasi kinerja masing-masing anggota tim berdasarkan Penjualan tiap bulan. Berikut adalah data evaluasi kinerja untuk setiap anggota tim pada bulan september:

- John: Penjualan (85)
- Maria: Penjualan (78)
- David: Penjualan (92)
- Lisa: Penjualan (85)
- Kevin: Penjualan (80)

Akan diilustrasikan dengan diagram bintang, sebagi berikut:

```
>Nama=["Johan","Maria","David","Lisa","Kevin"];
>Penjualan=[85,78,92,85,80];
>starplot(Penjualan, style="+", color=red, lab=Nama, rays=10, pstyle="=+", textcolor=blue, r=1.5):
```

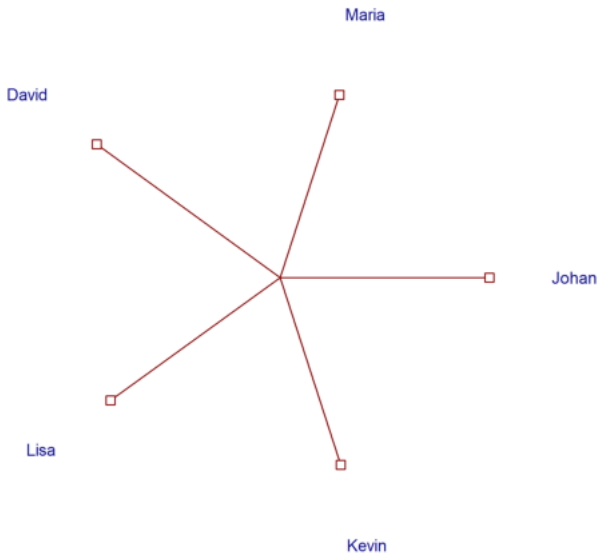


Diagram Impuls

Dalam statistik, istilah "grafik impuls" mungkin tidak umum digunakan. Namun, jika Anda merujuk pada grafik distribusi fungsi massa probabilitas diskrit, di mana probabilitas terkonsentrasi pada nilai tertentu, maka histogram atau diagram batang mungkin sesuai dengan konsep ini.

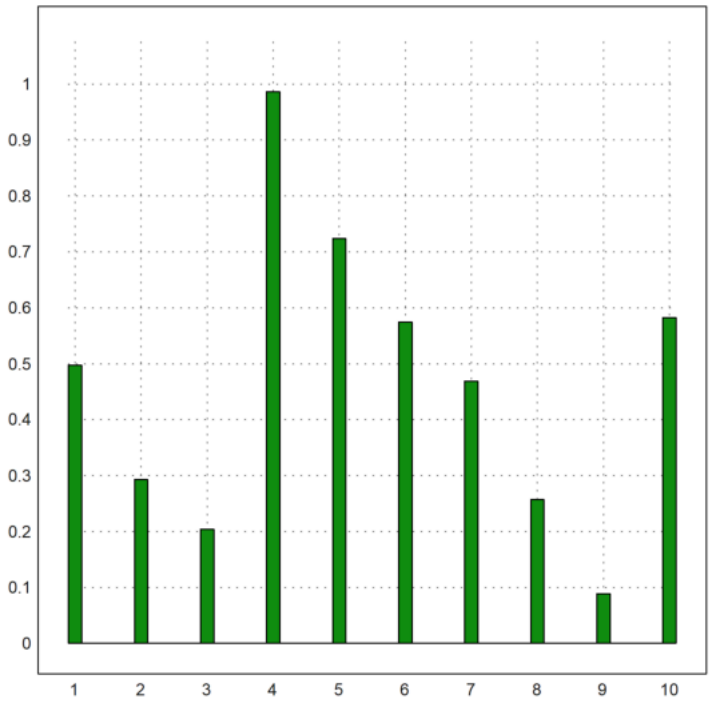
Sebagai contoh, jika Anda memiliki distribusi probabilitas diskrit seperti distribusi Poisson atau distribusi binomial, Anda dapat membuat grafik yang menunjukkan probabilitas pada setiap nilai yang mungkin. Ini akan menunjukkan "puncak" pada nilai-nilai tertentu, yang mungkin tampak mirip dengan bentuk grafik impuls.

Dalam Euler Math Toolbox, dalam membuat grafik statistik, termasuk grafik impuls, kita dapat menggunakan fungsi perencanaan yang dibangun. Berikut salah satu contoh bagaimana anda dapat membuat bagan impuls untuk distribusi probabilitas diskrit

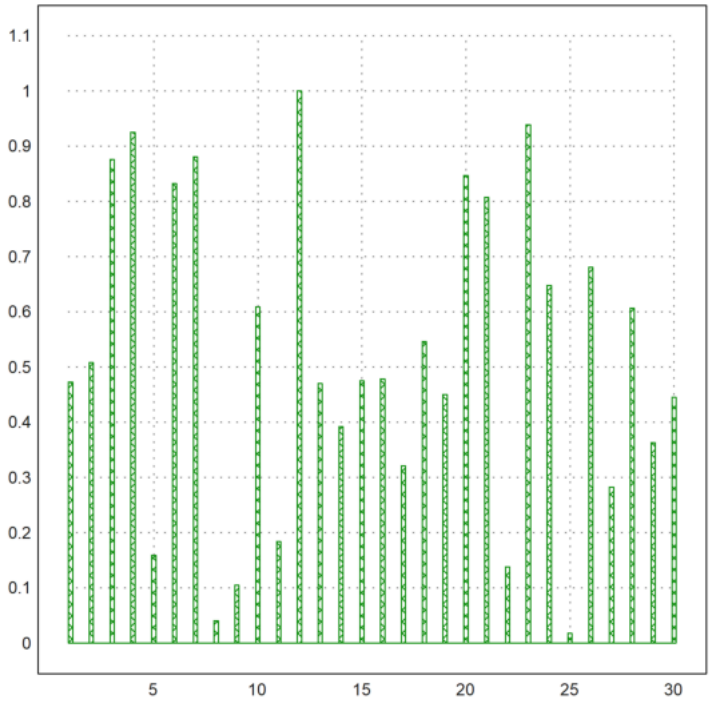
Contoh penggunaan diagram impuls di EMT

Berikut adalah plot impuls dari sebuah data acak yang terdistribusi secara merata didalam interval [0,1].

```
>plot2d(makeimpulse(1:10,random(1,10)),>bar):
```




```
>plot2d(makeimpulse(1:30,random(1,30)),>bar, style="\/"):
```



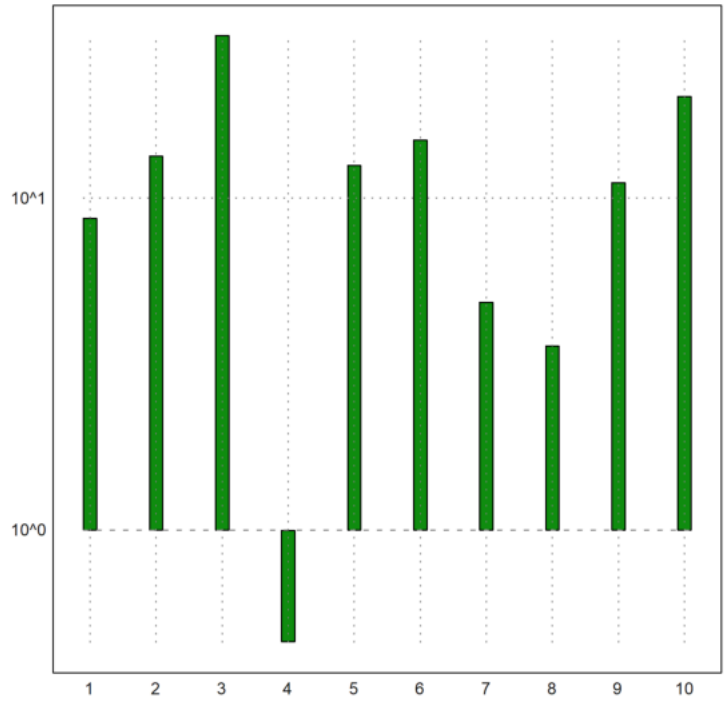
Jika data tersebut merupakan data yang terdistribusi secara

eksponensial, maka diperlukan plot logaritmik untuk membuat diagram impuls nya. Distribusi eksponensial adalah jenis distribusi di mana banyak data memiliki nilai yang kecil, tetapi ada beberapa data yang memiliki nilai yang sangat besar.

Pada sumbu logaritmik, perbedaan skala nilai yang besar dapat

diurutkan. Sehingga akan ada perbedaan kecil yang dapat dilihat lebih jelas. Dikarenakan dapat terlihat perbedaan antara nilai-nilai yang mendekati nol dan nilai yang lebih besar pada grafik logaritmik. Berikut adalah penggunaan diagram impuls dengan fungsi logaritma.

```
>logimpulseplot(1:10,-log(random(1,10))*10):
```



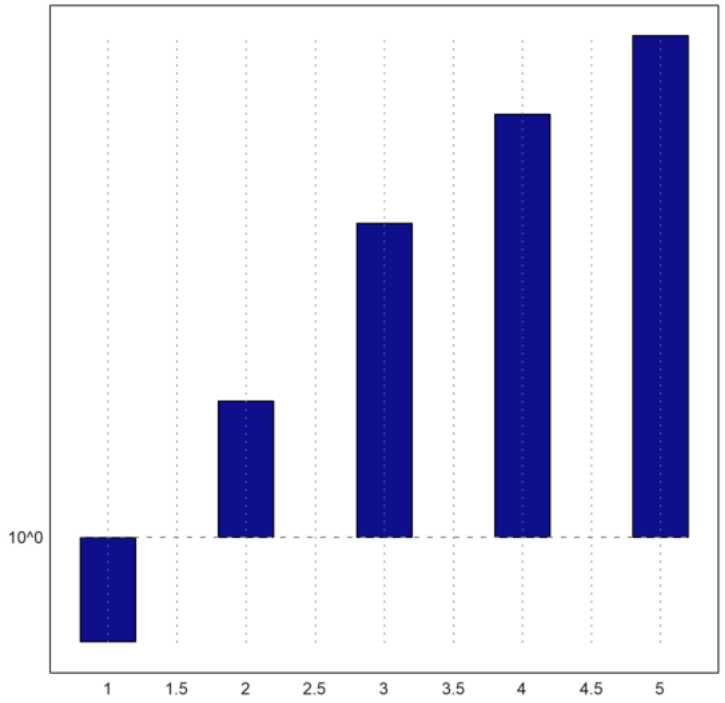

```
#Contoh soal lainnya

Sebuah koloni bakteri tumbuh secara eksponensial. Pada suatu
```

eksperimen, populasi bakteri diukur setiap jam selama 5 jam. Data pertumbuhan populasi bakteri (dalam ribuan) adalah sebagai berikut:

```
>Jam=[1:5];
```

Populasi ini diatur dalam ribuan

```
>Populasi=[2,5,20,80,320];
>logimpulseplot (Jam,log(Populasi), style="O#", color=blue, d=0.2):
```



Penjelasan:
- "Jam" disini digunakan sebagai data pada x bar
- "log(populasi)" disini digunakan untuk menglogaritmakan data yang ada di populasi sebagai daya y bar dalam diagram
- "style" disini digunakan untuk mengatur style atau gaya yang digunakan dalam kotak/batang.
- "color" disini digunakan untuk mengatur warna pada kotak/batang diagram.
- "d' disini digunakan untuk mengatur ketebalan kotak/batang diagram.

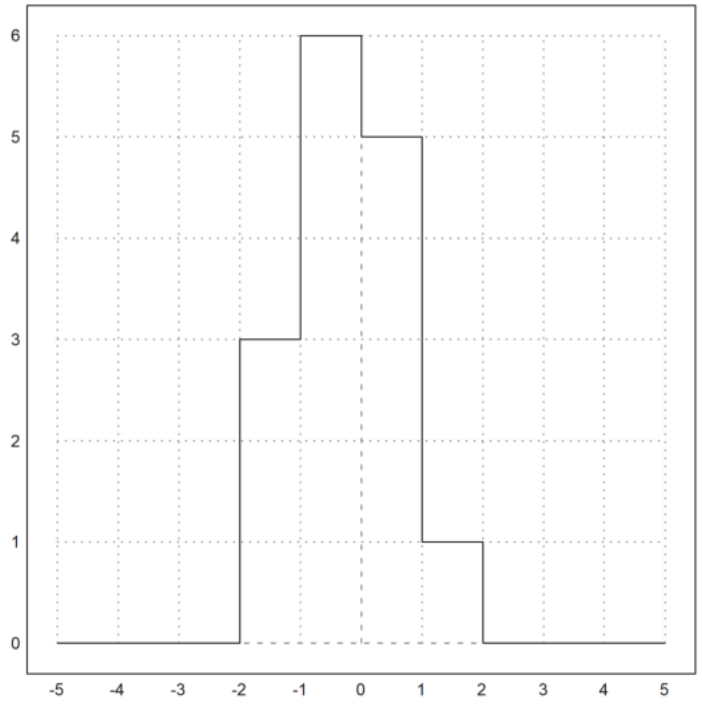
Histogram

Histogram adalah representasi grafis untuk distribusi warna dari citra digital. Sumbu ordinat vertikal merupakan representasi piksel dengan nilai tonal dari tiap-tiap deret bin pada sumbu axis horizontalnya. Sumbu axis terdiri dari deret logaritmik bin densitometry yang membentuk rentang luminasi atau exposure range yang mendekati respon spectral sensitivity visual mata manusia. Deret bin pada density yang terpadat mempunyai interval yang relatif sangat linear dengan variabel mid-tone terletak tepat di tengahnya.

Contoh penggunaan Histogram di EMT

Berikut merupakan contoh histogram pada sebuah data yang terdistribusi normal.

```
>plot2d(histo(normal(1,15),v=-5:5,<bar)):
```




```
#Contoh soal lainnya
```

```
Sebuah universitas melakukan penelitian untuk mengidentifikasi
```

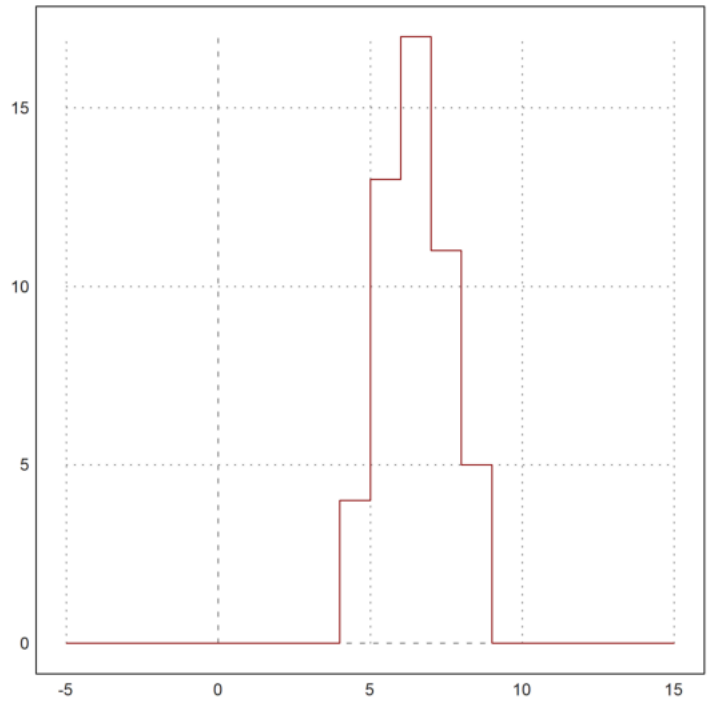
distribusi waktu studi mahasiswa di suatu jurusan. Data waktu studi (dalam tahun) dari 100 mahasiswa lulusan jurusan tersebut telah dikumpulkan. Berikut adalah data tersebut:

```
Data M:
```

```
>M=[5,4,6,7,5,6,7,8,4,6, ...
 5,6,7,5,6,5,7,6,8,6, ...
 4,5,6,7,5,6,5,6,7,8, ...
 6,7,5,6,7,5,6,7,8,6, ...
 7,8,6,5,4,5,6,7,5,6];
```

Akan diplotkan histogram

```
>plot2d(histo(M, v=-5:15, <bar,),color=red):
```



Penjelasan:
- "histo" disini digunakan unutk memanggil fungsi histogram yang ada pada plot 2d di EMT.
- "M" disini merupakan data yang akan diplotkan.
- "v" disini disini digunakan untuk mengatur interval pada grafik histogram yang diperlihatkan.
- "<bar" digunakan untuk membuat bar pada grafik.
- "color" disini digunakan untuk mengatur warna grafik

```
#Contoh soal lainnya
```

Seorang pemilik toko buku ingin menganalisis pola penjualan harian untuk memahami frekuensi pembelian buku oleh pelanggan. Data penjualan harian selama 30 hari terakhir telah dikumpulkan, dan berikut adalah jumlah buku yang terjual setiap hari:

```
>T=[35, 28, 42, 18, 24, 35, 40, 22, 29, 33, ...
 26, 38, 32, 19, 31, 25, 37, 29, 24, 27, ...
 36, 20, 28, 33, 31, 26, 39, 30, 23, 34];
>plot2d(histo(T, v=10:50, <bar),color=blue):
```

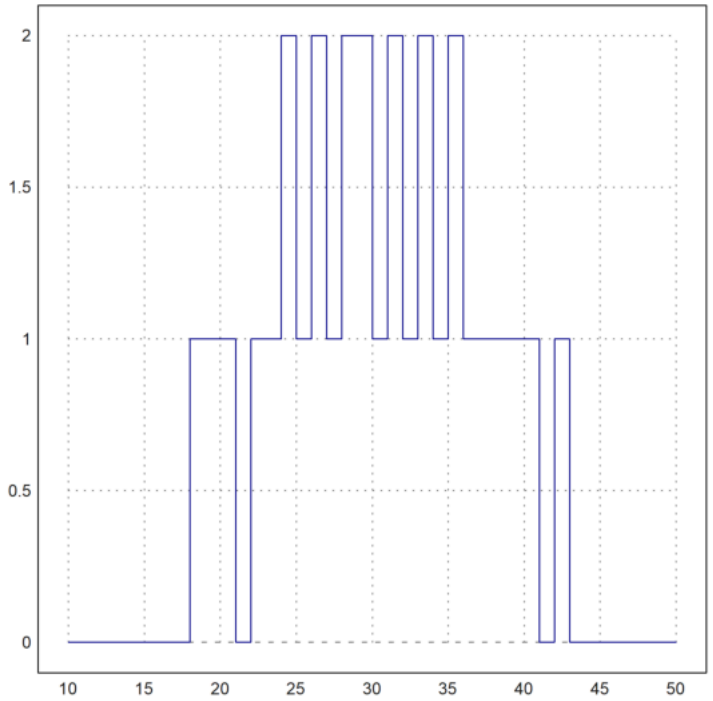


Diagram titik

Diagram titik atau disebut juga sebagai scatter plot, adalah jenis diagram statistik yang menggunakan titik-titik untuk merepresentasikan nilai dari dua variabel yang berbeda. Setiap titik dalam diagram pencar mewakili satu pengamatan atau data dengan nilai-nilai yang sesuai untuk kedua variabel tersebut.

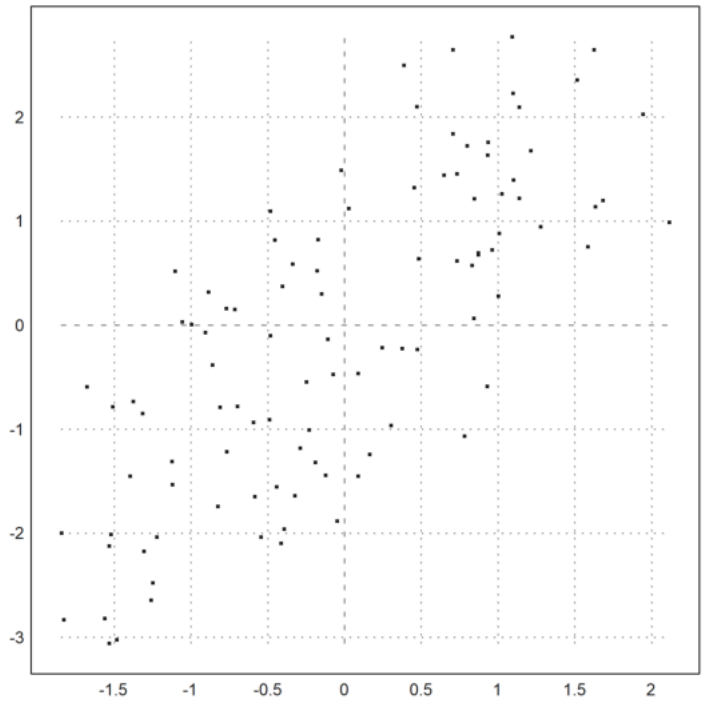
Scatter plot sangat berguna untuk menemukan pola atau hubungan antara dua variabel, serta untuk mengevaluasi distribusi data.

Dalam scatter plot, sumbu horizontal umumnya digunakan untuk variabel independen (bebas), sedangkan sumbu vertikal digunakan untuk variabel dependen (bergantung). Dengan mengamati pola penyebaran titik-titik, kita mendapatkan wawasan tentang apakah ada korelasi antara dua variabel dan jenis korelasi apa yang mungkin ada (positif, negatif, atau tidak ada korelasi).

Berikut contoh menggambar diagram titik di EMT

Contoh 1
Pada contoh ini, kita akan menggambar diagram titik dengan menggunakan fungsi plot2d().

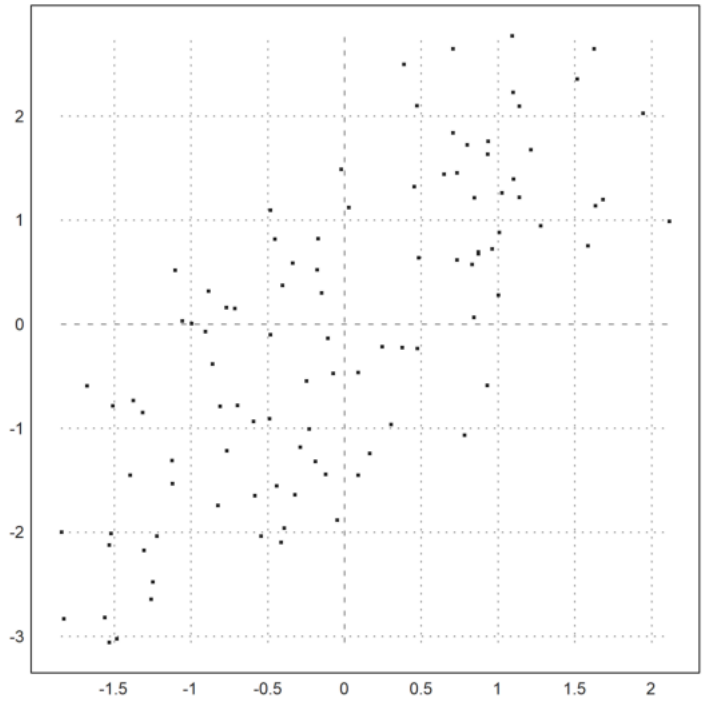
```
>x=normal(1,100);
>plot2d(x,x+rotright(x),>points,style="..");
```



Terdapat banyak style titik yang dapat kita dunakan, yaitu:
"[]", "<>", ".", ",", "..", "...","*", "+", "|", "-", "o",
"[]#", "<>#", "o#" (bentuk terisi)
"[]w", "<>w", "ow" (tidak transparan)

Selanjutnya, akan kita coba gambarkan diagram tersebut menggunakan fungsi statplot(), pilih plottype="p" (karena kita akan menggambar plot titik).

```
>statplot(x,x+rotright(x),plottype="p",pstyle=".."):
```



Contoh 2
Pada contoh ini, akan kita gambarkan diagram titik menggunakan fungsi scatterplot().

```
>{MS,hd}:=readtable("table1.dat",tok2:=["m","f"]);
>writetable(MS,labc=hd,tok2:=["m","f"]);
```

Person	Sex	Age	Mother	Father	Siblings
1	m	29	58	61	1
2	f	26	53	54	2
3	m	24	49	55	1
4	f	25	56	63	3
5	f	25	49	53	0
6	f	23	55	55	2
7	m	23	48	54	2
8	m	27	56	58	1
9	m	25	57	59	1
10	m	24	50	54	1
11	f	26	61	65	1
12	m	24	50	52	1
13	m	29	54	56	1
14	m	28	48	51	2
15	f	23	52	52	1
16	m	24	45	57	1
17	f	24	59	63	0
18	f	23	52	55	1
19	m	24	54	61	2
20	f	23	54	55	1

```
>scatterplots(tablecol(MS,3:5),hd[3:5]):
```

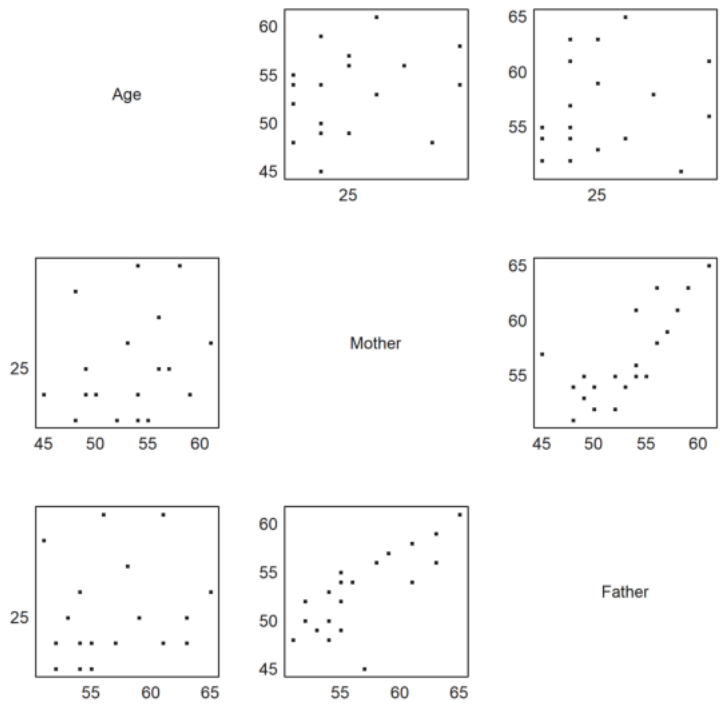


Diagram Garis

Diagram garis adalah penyajian data yang digunakan untuk menggambarkan suatu keadaan berupa data berkala atau berkelanjutan.

Selain itu, diagram ini juga bisa dikatakan berhubungan dengan kurun waktu dan untuk menunjukkan perkembangan suatu keadaan.

Diagram ini sangat tepat untuk menyajikan data untuk mengetahui kecenderungan kelakuan atau tren, seperti produksi minyak tiap tahun, jumlah kelahiran tiap tahun, jumlah produksi tiap jam, dan lain-lain.

Dalam diagram garis, terdapat sumbu vertikal (sumbu y) untuk menunjukkan frekuensi dan sumbu horizontal (sumbu x) untuk menunjukkan variabel tertentu.

Berikut contoh menggambar diagram garis di EMT

Contoh 1
Akan digambarkan diagram garis data banyaknya pelanggan di toko A tahun 2015-2023.

Kita deskripsikan terlebih dahulu matriks x dan y,kemudian akan kita buat tabel datanya.

```
>x=[2015,2016,2017,2018,2019,2020,2021,2022,2023]; y=[600,500,900,1000,800,850,900,1000,1200];
>writetable(x'|y',labc=["Tahun","Banyak Pelanggan"])
```

Tahun	Banyak Pelanggan
2015	600
2016	500
2017	900
2018	1000
2019	800
2020	850
2021	900

```
2022            1000
2023            1200
```

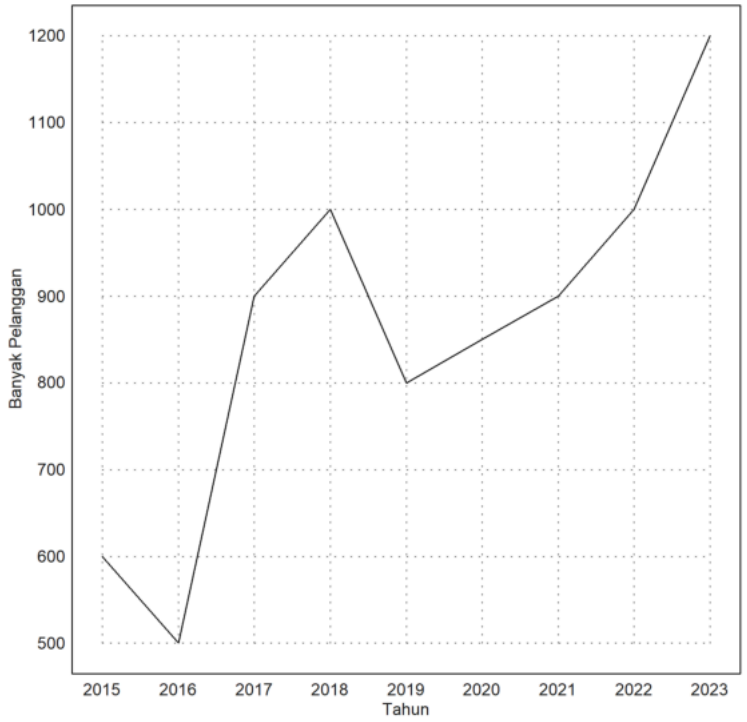
Selanjutnya, akan digambarkan diagram garis dengan menggunakan fungsi statplot, dengan format:

statplot (x, y, plottype="l", lstyle="-", xl="", yl="", color=none, vertical=0)

x : data untuk sumbu x
y : data untuk sumbu y
plotstyle : "l" (kita pilih style "l" karena berupa plot garis)
lstyle : style garis
xl : label sumbu x
yl : label sumbu y
color : warna garis
vertikal : vertikal

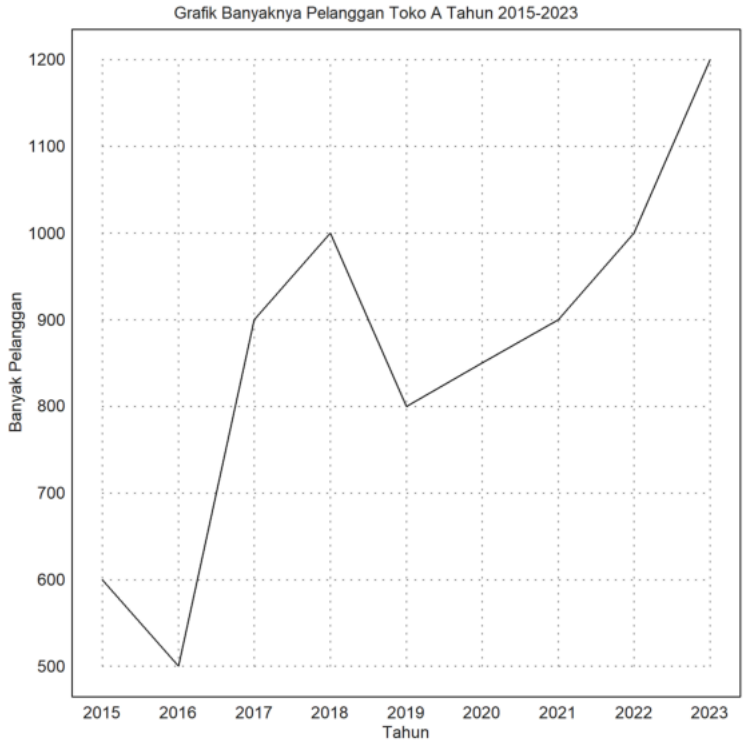
Style garis:
"-", "--", "-.", ".", ".-.", "-.-", "->"

```
>statplot(x,y,plottype="l",lstyle="-",xl="Tahun",yl="Banyak Pelanggan",vertical=50):
```



Kita juga bisa menambahkan judul pada grafik dengan menggunakan fungsi title().

```
>title("Grafik Banyaknya Pelanggan Toko A Tahun 2015-2023"):
```

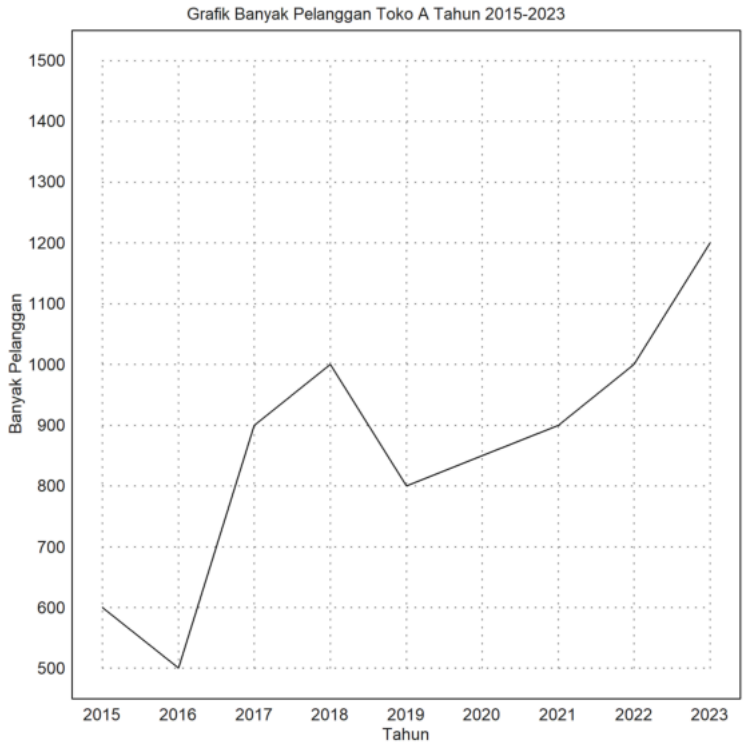


Menyajikan data dalam bentuk diagram dapat memudahkan pembaca untuk memahami data yang disajikan.

Dari diagram garis diatas, dapat kita peroleh informasi bahwa:
Banyaknya pelanggan di toko A tidak tetap (naik-turun)untuk setiap tahunnya.
Banyaknya pelanggan paling sedikit pada tahun 2016 yaitu sebanyak 500 pelanggan, sedangkan banyak pelanggan paling banyak pada tahun 2023 yaitu sebanyak 1200 pelanggan.

Selain menggunakan fungsi statplot, kita juga dapat menggambar diagram garis menggunakan fungsi plot2d() seperti yang sudah pernah kita pelajari sebelumnya, yaitu sebagai berikut.

```
>plot2d(x,y,a=2015,b=2023,c=500,d=1500,style="_",xl="Tahun",yl="Banyak Pelanggan",vertical=50); title("Grafik Banyak Pelanggan Toko A T
```



a dan b : batas untuk sumbu x
c dan d : batas untuk sumbu y
style : gaya garis
xl : label untuk sumbu x
yl : label untuk sumbu y

Selain kita dapat menggambarkan diagram garis saja atau diagram titik saja, kita juga dapat menggambarkan diagram keduanya.

Contoh 2
Akan kita gambar diagram titik dan garis data hasil pemilu Jerman dari tahun 1990 sampai 2013, diukur dalam kursi.

```
>BW := [ ...
 1990,662,319,239,79,8,17; ...
 1994,672,294,252,47,49,30; ...
 1998,669,245,298,43,47,36; ...
 2002,603,248,251,47,55,2; ...
 2005,614,226,222,61,51,54; ...
 2009,622,239,146,93,68,76; ...
 2013,631,311,193,0,63,64];
>P:=["CDU/CSU","SPD","FDP","Gr","Li"];
>BT:=BW[,3:7]; BT:=BT/sum(BT); YT:=BW[,1]';
>writetable(BT*100,wc=6,dc=0,>fixed,labc=P,labr=YT)
```

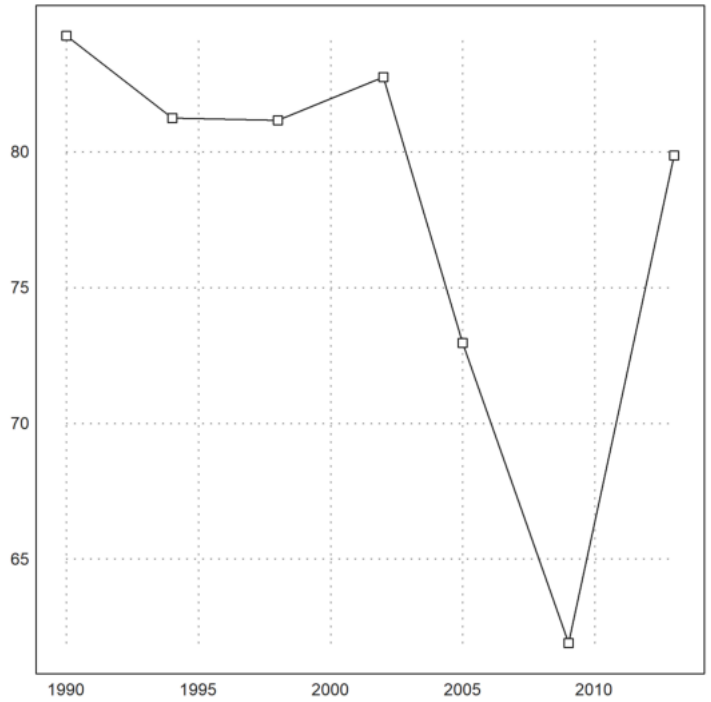
	CDU/CSU	SPD	FDP	Gr	Li
1990	48	36	12	1	3
1994	44	38	7	7	4
1998	37	45	6	7	5
2002	41	42	8	9	0
2005	37	36	10	8	9
2009	38	23	15	11	12
2013	49	31	0	10	10

```
>BT1:=(BT.[1;1;0;0;0])'*100
```

```
     [84.29,  81.25,  81.1659,  82.7529,  72.9642,  61.8971,  79.8732]
```

Akan kita gambarkan plot statistik sederhana, yaitu plot titik dan garis secara bersamaan dengan menggunakan fungsi statplot dan pilih plottype="b".

```
>statplot(YT,BT1,"b"):
```




```
>CP:=[rgb(0.5,0.5,0.5),red,yellow,green,rgb(0.8,0,0)];
```

Untuk menggabungkan deretan data statistik dalam satu plot, dapat kita digunakan fungsi dataplot().

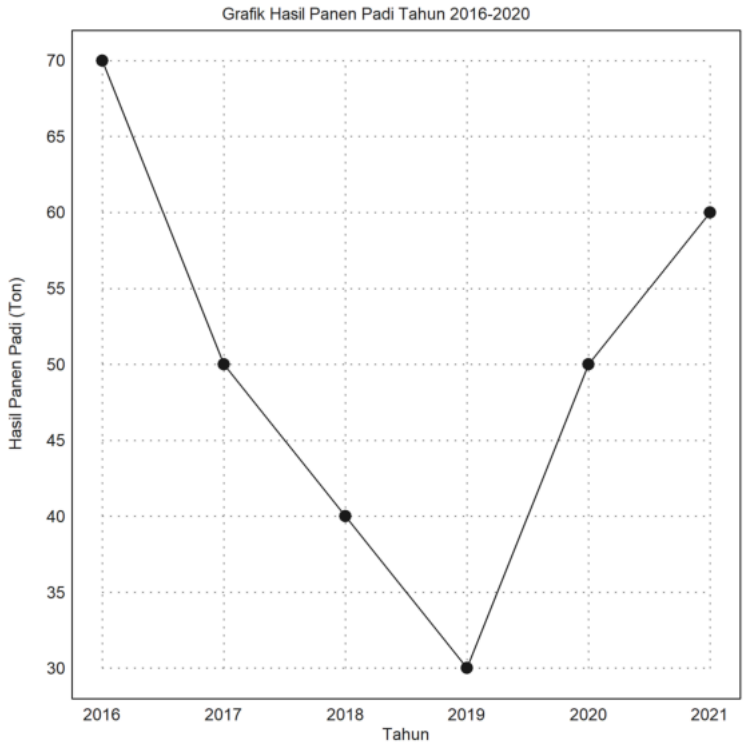
```
>J:=BW[,1]'; DP:=BW[,3:7]'; ...
 dataplot(YT,BT',color=CP);  ...
 labelbox(P,colors=CP,styles="[]",>points,w=0.2,x=0.3,y=0.4):
```

Contoh 3
Akan digambarkan diagram titik dan garis dari data hasil panen padi pada tahun 2016 sampai 2020.

```
>T=[2016,2017,2018,2019,2020,2021]; P=[70,50,40,30,50,60];
>writetable(T'|P',labc=["Tahun","Hasil Panen Padi (Ton)"])
```

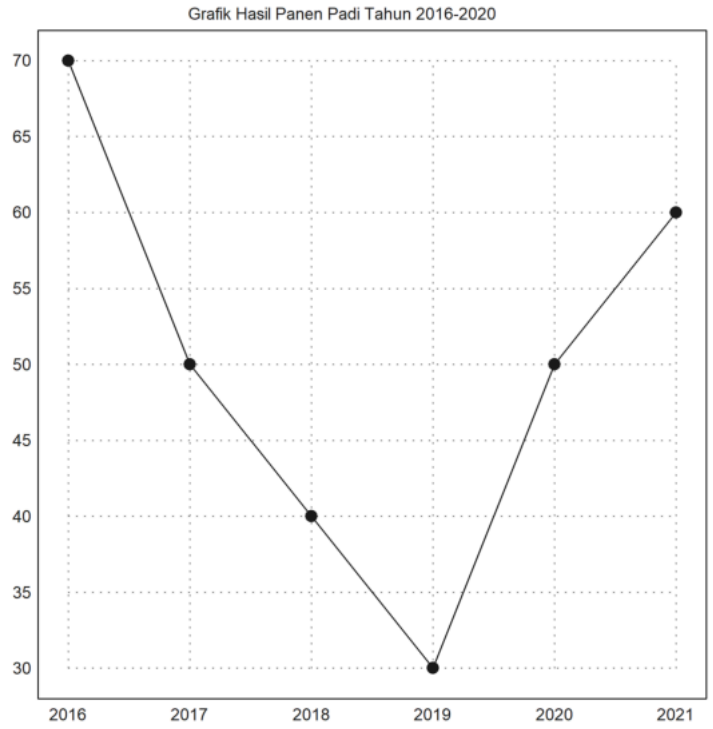
Tahun	Hasil Panen Padi (Ton)
2016	70
2017	50
2018	40
2019	30
2020	50
2021	60

```
>statplot(T,P,"b",pstyle="o#",lstyle="-",xl="Tahun",yl="Hasil Panen Padi (Ton)",vertical=50);
>title("Grafik Hasil Panen Padi Tahun 2016-2020"):
```



Kita juga menggambar diagram titik dan garis secara bersama dengan menggunakan fungsi plot2d(), yaitu sebagai berikut.

```
>plot2d(T,P,2016,2020); plot2d(T,P,>points,style="o#",>add); title("Grafik Hasil Panen Padi Tahun 2016-2020"):
```



Dari grafik di atas, dapat dengan mudah kita ketahui bahwa panen padi paling banyak yaitu pada tahun 2016 (70 ton) dan paling sedikit yaitu pada tahun 2019 (30 ton).

Kurva Regresi

Kurva regresi adalah representasi grafis dari model regresi yang digunakan untuk memodelkan hubungan antara satu atau lebih variabel independen (biasanya dilambangkan sebagai (X) dan variabel dependen (Y). Kurva regresi ini mencoba untuk menunjukkan pola atau tren dalam data dan memungkinkan kita untuk membuat prediksi atau estimasi berdasarkan model tersebut.
Secara umum, terdapat dua jenis kurva regresi yang umum digunakan yaitu regresi linier dan regresi non-linier.

Regresi Linier adalah garis lurus yang digunakan untuk memodelkan hubungan antara variabel independen (X) dan variabel dependen (Y).

Persamaan regresi linier umumnya ditulis sebagai

di mana m adalah kemiringan (slope) dan b adalah perpotongan sumbu-y (intercept).

Regresi linier dapat dilakukan dengan fungsi polyfit () atau berbagai fungsi fit.

Sebagai permulaan kita menemukan garis regresi untuk data univariat dengan polyfit(x, y, 1).

Berikut contoh menggambar kurva regresi di EMT

Contoh 1

```
>x=1:10; y=[2,3,1,5,6,3,7,8,9,8]; writetable(x'|y',labc=["x","y"])
```

```
              x          y
              1          2
```

```
 2          3
 3          1
 4          5
 5          6
 6          3
 7          7
 8          8
 9          9
10          8
```

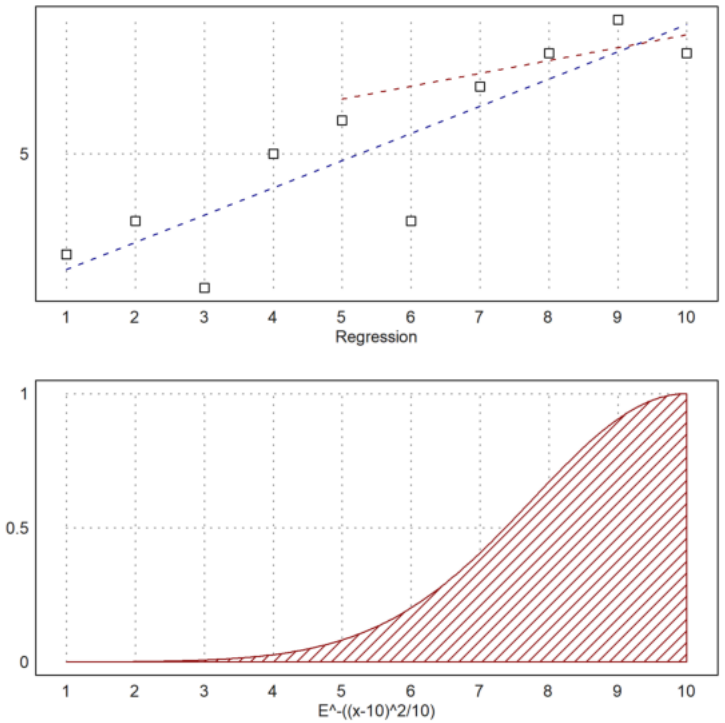
```
>p=polyfit(x,y,1)
```

```
    [0.733333,  0.812121]
```

```
>w &= "exp(-(x-10)^2/10)"; pw=polyfit(x,y,1,w=w(x))
```

```
    [4.71566,  0.38319]
```

```
>figure(2,1); ...
 figure(1); statplot(x,y,"p",xl="Regression"); ...
   plot2d("evalpoly(x,p)",>add,color=blue,style="--"); ...
   plot2d("evalpoly(x,pw)",5,10,>add,color=red,style="--"); ...
 figure(2); plot2d(w,1,10,>filled,style="/",fillcolor=red,xl=w); ...
 figure(0):
```



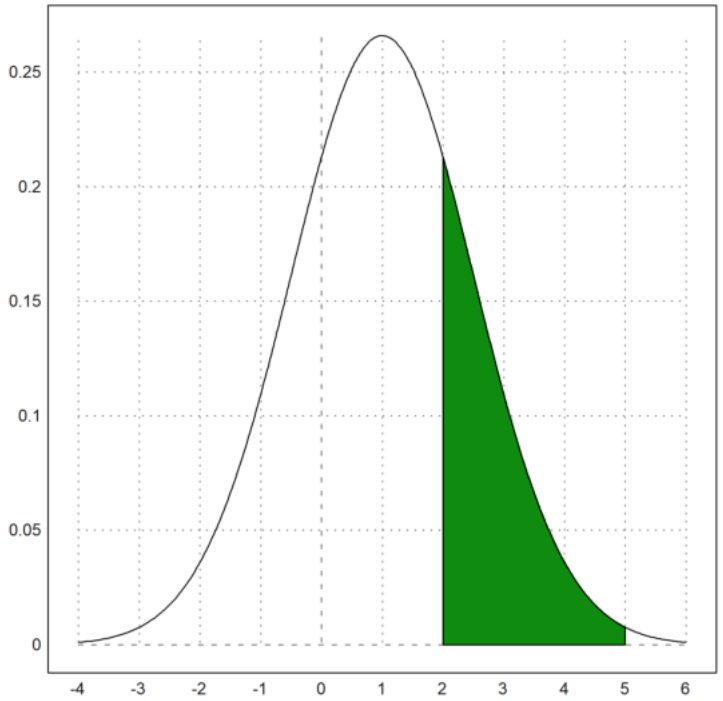
Kurva Fungsi Kerapatan Probabilitas

Fungsi kerapatan/kepadatan probabilitas adalah fungsi yang memberikan kemungkinan bahwa nilai suatu variabel acak akan berada di antara rentang nilai tertentu.

Grafik fungsi kepadatan probabilitas berbentuk kurva lonceng. Area yang terletak di antara dua nilai tertentu memberikan probabilitas hasil observasi yang ditentukan.

Berikut contoh kurva fungsi kepadatan probabolitas.

```
>plot2d("qnormal(x,1,1.5)",-4,6);
>plot2d("qnormal(x,1,1.5)",a=2,b=5,>add,>filled):
```



Kurva Distribusi Kumulatif

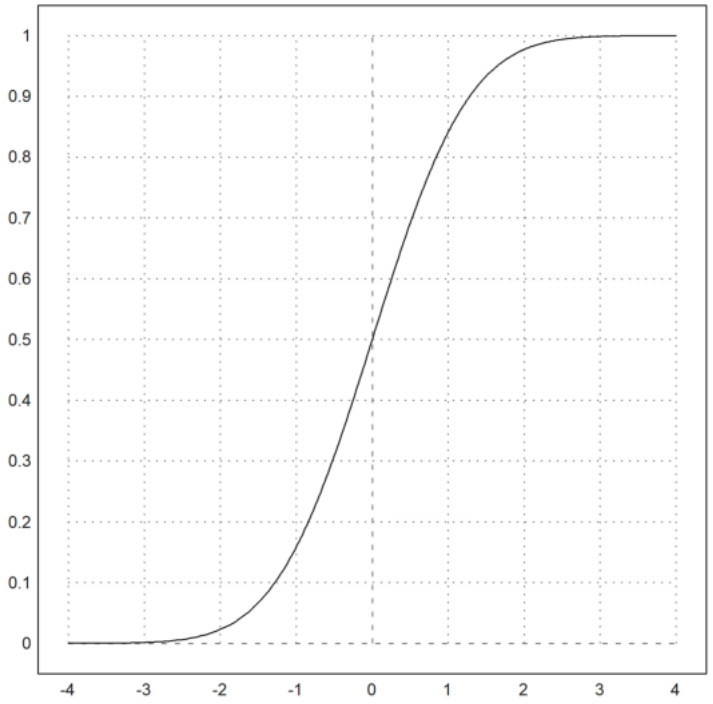
Kurva distribusi kumulatif (CDF) adalah representasi kumulatif dari fungsi distribusi probabilitas suatu variabel acak. CDF memberikan probabilitas bahwa variabel acak tersebut kurang dari atau sama dengan suatu nilai tertentu.

Seringkali, grafik CDF disajikan dalam bentuk kurva monotonik yang terus meningkat, dan ini memberikan gambaran visual yang baik tentang distribusi probabilitas variabel acak.

Berikut menggambar kurva distribusi kumulatif di EMT

Contoh 1

```
>plot2d("normaldis",-4,4):
```



Dapat kita lihat dalam kurva CDF kontinu di atas dibagi menjadi 3 bagian, yaitu:
1. Bernilai 0 untuk x kurang dari batas bawah daerah rentang.
2. Merupakan fungsi monoton naik pada daerah rentang.
3. Bernilai konstan 1 untuk x lebih dari batas atas daerah rentangnya.

Contoh 2
Diberikan variabel acak dengan PDF sebagai berikut:

$$f(x) = \begin{cases} \frac{6}{5}(x^2 + x) & 0 \leq x \leq 1 \\ 0 & \text{x yang lain.} \end{cases}$$

Untuk menggambar grafik CDF-nya, pertama kita cari terlebih dahulu CDF dari fungsi tersebut.

Untuk x pada interval

$$(-\infty, x)$$

$$F(x) = \int_{-\infty}^{x} f(x)\ dx$$

$$F(x) = \int_{-\infty}^{x} 0\ dx$$

$$F(x) = 0$$

Untuk x pada interval [0,1]

$$F(x) = \int_{-\infty}^{x} f(x)\ dx$$

$$F(x) = \int_{-\infty}^{0} f(x)\ dx + \int_{0}^{x} f(x)\ dx$$

$$F(x) = \int_{-\infty}^{0} 0\ dx + \int_{0}^{x} \frac{6}{5}(x^2 + x)\ dx$$

$$F(x) = 0 + \frac{6}{5}(\frac{x^3}{3} + \frac{x^2}{2})$$

$$F(x) = \frac{6}{5}(\frac{x^3}{3} + \frac{x^2}{2})$$

Untuk x pada interval

$$(1, \infty)$$

$$F(x) = \int_{-\infty}^{x} f(x)\ dx$$

$$F(x) = \int_{-\infty}^{0} f(x)\ dx + \int_{0}^{1} f(x)\ dx + \int_{1}^{x} f(x)\ dx$$

$$F(x) = \int_{-\infty}^{0} 0\ dx + \int_{0}^{1} \frac{6}{5}(x^2 + x)\ dx + \int_{1}^{\infty} 0\ dx$$

$$F(x) = 0 + \frac{6}{5}(\frac{1^3}{3} + \frac{1^2}{2}) + 0$$

$$F(x) = \frac{6}{5}\frac{5}{6} = 1$$

Sehingga, diperoleh CDF :

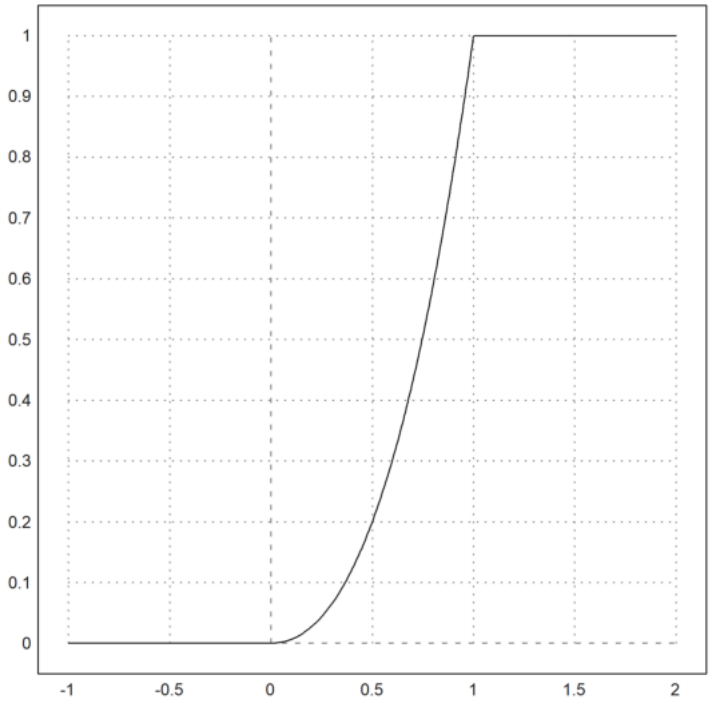
$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{6}{5}(\frac{x^3}{3} + \frac{x^2}{2}) & 0 \le x \le 1 \\ 1 & x > 1. \end{cases}$$

Selanjutnya, akan kita gambarkan grafik CDF tersebut di EMT.
Pertama, kita definisikan terlebih dahulu fungsi f(x) dengan menggunakan fungsi function map.

```
>function map f(x) ...
   if x<0 then return 0
   else if x>=0 and x<=1 then return (6/5)*((x^3/3)+(x^2/2))
   else return 1
   endif;
 endfunction
```

Kemudian, kita akan menggambar grafik fungsi di atas dengan menggunakan fungsi plot2d() pada interval x dari -1 sampai 2.

```
>plot2d ("f(x)",-1,2):
```



Menampilkan Tabel Frame Data

Sebelum menampilkan tabel frame data, terlebih dahulu kita mendownload dan menyimpan file yang telah disediakan di folder Euler. Data yang akan ditampilkan adalah hasil dari sebuah survey. Data tersebut berbentuk notebook, atau yang terdapat di besmart dengan format .dat dan .tab.

Perintah yang digunakan adalah
printfile("...",...);

perintah print digunakan untuk memformat angka dengan cara tertentu. Fungsi ini dapat menambahkan satuan, memisahkan kelompok angka dengan spasi, dan menggunakan koma desimal. kemudian dipisahkan dengan koma lalu diisi dengan angka. Angka disini adalah jumlah baris yang ingin ditampilkan, lalu diakhiri dengan titik koma.

Mari kita coba.

```
>printfile("table.dat",2);
```

```
  Person Sex Age Titanic Evaluation Tip Problem
  1 m 30 n . 1.80 n
```

```
>printfile("table.dat",4);
```

```
  Person Sex Age Titanic Evaluation Tip Problem
  1 m 30 n . 1.80 n
  2 f 23 y g 1.80 n
  3 f 26 y g 1.80 y
```

```
>printfile("test.tab",4);
```

```
    one,two,three
          0.59,      0.15,      0.81
          0.36,       0.8,       0.1
          0.48,      0.39,      0.78
```

```
>printfile("test.tab",7);
```

```
    one,two,three
          0.59,      0.15,      0.81
          0.36,       0.8,       0.1
          0.48,      0.39,      0.78
```

Pada file table.dat yang ditampilkan, data gersebut memuat 7 kolom angka atau token (strings). Kita akan membaca tabel tersebut dari file dengan menggunakan terjemahan untuk token.

Untuk itu, kita definisikan set tokennya terlebih dahulu dengan menggunakan fungsi strtokens() untuk mendapatkan vektor string token dari string yang diberikan.

```
>mf:=["m","f"]; yn:=["y","n"]; ev:=strtokens("g vg m b vb");
```

Jika dijabarkan arti tiap komponen perintah tersebut adalah :
- mf := ["m", "f"];
ni mendefinisikan variabel mf sebagai sebuah array atau daftar yang berisi dua string, yaitu "m" dan "f".
- yn := ["y", "n"];
Ini mendefinisikan variabel yn sebagai array atau daftar yang berisi dua string, yaitu "y" dan "n".
- ev := strtokens("g vg m b vb");
erintah strtokens("g vg m b vb") memisahkan string "g vg m b vb" menjadi token atau elemen yang dipisahkan oleh spasi, dan memasukkannya ke dalam array atau daftar.

Sekarang kita membaca tabel dengan terjemahan ini dengan menggunakan perintah
readtable("...",argumen tok2, tok4, dll. argumen tok tersebut adalah terjemahan dari kolom-kolom tabel.
Argumen-argumen ini tidak ada di dalam readtable(), jadi kita harus menyediakannya dengan ":=".

```
>{MT,hd}=readtable("table.dat",tok2:=mf,tok4:=yn,tok5:=ev,tok7:=yn);
>load over statistics;
```

Perintah load over statistics; pada Euler Math Toolbox berfungsi untuk memuat paket atau modul statistik ke dalam sesi kerja. Dalam Euler Math Toolbox, perintah load digunakan untuk memanggil modul atau library tertentu yang sudah ada dalam perangkat lunak tersebut, sehingga kita dapat menggunakan fungsi atau perintah-perintah terkait statistik tanpa harus mendefinisikan atau menuliskannya ulang.

Untuk mencetak, kita perlu menentukan set token yang sama. Kami mencetak empat baris pertama saja.

Perintah yang digunakan adalah
writetable(...);

```
>writetable(MT[1:4],labc=hd,wc=5,tok2:=mf,tok4:=yn,tok5:=ev,tok7:=yn);
```

```
  Person  Sex  Age Titanic Evaluation  Tip Problem
       1    m   30       n          .  1.8       n
       2    f   23       y          g  1.8       n
       3    f   26       y          g  1.8       y
       4    m   33       n          .  2.8       n
```

Arti dari komponen perintah tersebut adalah :
- writetable: Ini adalah perintah untuk menulis atau menampilkan data dari tabel atau matriks.
- MT[1:4]: Ini menyatakan bahwa kita hanya ingin mengambil elemen pertama hingga elemen keempat dari matriks atau tabel MT.
- labc=hd: Parameter ini sering digunakan untuk menambahkan label atau header pada kolom tabel. hd mungkin merujuk pada penggunaan header (label kolom) yang sudah diset pada MT.
- wc=5: Mengatur lebar kolom menjadi 5 karakter. Nilai 5 menunjukkan bahwa kolom memiliki lebar maksimal lima karakter untuk setiap elemen yang ditampilkan, sehingga nilai-nilai akan disesuaikan agar muat dalam lebar ini.
- tok2:=mf, tok4:=yn, tok5:=ev, tok7:=yn: Parameter ini biasanya merupakan tokens atau pengaturan khusus yang dapat diaktifkan atau dinonaktifkan sesuai kebutuhan:
a. tok2:=mf: Menetapkan token ke-2 dengan kode mf yang bisa mengindikasikan jenis format angka atau

teks tertentu, misalnya mean-fitted atau format lainnya.
b. tok4:=yn: Mengaktifkan token ke-4 dengan pengaturan yn, yang bisa diartikan sebagai yes-no untuk menampilkan atau menyembunyikan fitur tertentu, seperti baris atau kolom spesifik.
c. tok5:=ev: Mengaktifkan token ke-5 dengan ev, mungkin untuk mengevaluasi atau melakukan pemeriksaan ekspresi pada elemen-elemen dalam tabel.
d. tok7:=yn: Mengaktifkan token ke-7 dengan pengaturan yn, yang bisa berarti keputusan untuk menampilkan atau menyembunyikan data tertentu.
Setelah perintah ini dieksekusi, hasilnya akan berupa tampilan tabel yang memuat baris pertama hingga keempat dari matriks MT, disesuaikan dengan header kolom yang telah ditentukan.

Tanda titik "." mewakili nilai yang tidak tersedia.
Jika kita tidak ingin menentukan token untuk terjemahan sebelumnya, kita hanya perlu menentukan kolom mana yang
yang berisi token dan bukan angka.

```
>ctok=[2,4,5,7]; {MT,hd,tok}=readtable("table.dat",ctok=ctok);
```

ctok=[2,4,5,7]; – Variabel ctok didefinisikan sebagai vektor yang berisi elemen [2,4,5,7]. Angka-angka dalam ctok ini biasanya mewakili indeks kolom yang akan dipilih dari file tabel.

Fungsi readtable() di atas mengembalikan satu set token. Cek token tersebut dengan perintah "tok" dan akan menampilkan satu set token yang terdiri dari m, n, f, y, g, dan vg.

```
>tok
```

```
m
n
f
y
g
vg
```

Tabel yang berisi entri dari file dengan token yang diterjemahkan ke angka terdapat string khusus NA="." yang ditafsirkan sebagai "Tidak Tersedia", dan mendapatkan NAN (bukan angka) dalam tabel. Terjemahan ini dapat diubah dengan parameter NA, dan NAval.

```
>MT[1]
```

```
[1,  1,  30,  2,  NAN,  1.8,  2]
```

```
>MT[4]
```

```
[4,  1,  33,  2,  NAN,  2.8,  2]
```

Di atas adalah isi tabel dengan angka yang tidak diterjemahkan.

```
> writetable(MT,wc=5)
```

```
   1    1   30    2    .  1.8    2
   2    3   23    4    5  1.8    2
   3    3   26    4    5  1.8    4
   4    1   33    2    .  2.8    2
   5    1   37    2    .  1.8    2
   6    1   28    4    5  2.8    4
   7    3   31    4    6  2.8    2
   8    1   23    2    .  0.8    2
   9    3   24    4    6  1.8    4
  10    1   26    2    .  1.8    2
  11    3   23    4    6  1.8    4
  12    1   32    4    5  1.8    2
  13    1   29    4    6  1.8    4
  14    3   25    4    5  1.8    4
  15    3   31    4    5  0.8    2
  16    1   26    4    5  2.8    2
```

```
  17    1   37    2    .  3.8    2
  18    1   38    4    5    .    2
  19    3   29    2    .  3.8    2
  20    3   28    4    6  1.8    2
  21    3   28    4    1  2.8    4
  22    3   28    4    6  1.8    4
  23    3   38    4    5  2.8    2
  24    3   27    4    1  1.8    4
  25    1   27    2    .  2.8    4
```

Hasil dari perintah ini adalah tampilan tabel MT di mana setiap kolom memiliki lebar yang seragam, yaitu 5 karakter.

Kita dapat menaruh output dari readtable() ke dalam sebuah daftar.

```
>Table={{readtable("table.dat",ctok=ctok)}};
```

Dengan menggunakan kolom token yang sama dan token yang dibaca dari file, kita dapat mencetak tabel. Kita dapat menentukan
tok, tok, dll. atau menggunakan daftar Tabel.

```
>writetable(Table,ctok=ctok,wc=5);
```

```
Person  Sex  Age Titanic Evaluation  Tip Problem
     1    m   30       n          .  1.8       n
     2    f   23       y          g  1.8       n
     3    f   26       y          g  1.8       y
     4    m   33       n          .  2.8       n
     5    m   37       n          .  1.8       n
     6    m   28       y          g  2.8       y
     7    f   31       y         vg  2.8       n
     8    m   23       n          .  0.8       n
     9    f   24       y         vg  1.8       y
    10    m   26       n          .  1.8       n
    11    f   23       y         vg  1.8       y
    12    m   32       y          g  1.8       n
    13    m   29       y         vg  1.8       y
    14    f   25       y          g  1.8       y
    15    f   31       y          g  0.8       n
    16    m   26       y          g  2.8       n
    17    m   37       n          .  3.8       n
    18    m   38       y          g    .       n
    19    f   29       n          .  3.8       n
    20    f   28       y         vg  1.8       n
    21    f   28       y          m  2.8       y
    22    f   28       y         vg  1.8       y
    23    f   38       y          g  2.8       n
    24    f   27       y          m  1.8       y
    25    m   27       n          .  2.8       y
```

Perintah ini akan menampilkan Table dengan setiap kolom dipisahkan oleh pemisah yang ditetapkan oleh ctok, dan lebar setiap kolom akan diatur menjadi 5 karakter. Hasilnya adalah tabel yang terlihat rapi dengan format kolom yang konsisten, sesuai dengan aturan pemisah dan lebar kolom yang ditentukan.

selanjutnya kita akan menghilangkan baris yang mengandung angka yang tidak dapat diterjemahkan dengan menggunakan Fungsi tablecol() mengembalikan nilai kolom dari tabel, melewatkan setiap baris dengan nilai NAN(“.” dalam file), dan indeks kolom, yang berisi nilai-nilai ini.

```
>{c,i}=tablecol(MT,[5,6]);
```

Fungsi tablecol digunakan untuk mengekstrak kolom-kolom tertentu dari sebuah matriks atau tabel dalam Euler Math Toolbox. Perintah ini mengambil kolom sesuai indeks yang diberikan dan mengembalikan hasilnya.
- Parameter [5,6] menunjukkan bahwa kita ingin mengekstrak kolom ke-5 dan ke-6 dari matriks atau tabel MT.
- c berisi nilai-nilai dalam kolom yang diambil dari MT. dan i menyimpan indeks atau informasi lain yang berkaitan dengan kolom-kolom yang diekstrak.

Kita dapat menggunakan ini untuk mengekstrak kolom dari tabel untuk tabel baru.

```
>j=[1,5,6]; writetable(MT[i,j],labc=hd[j],ctok=[2],tok=tok)
```

Person	Evaluation	Tip
2	g	1.8
3	g	1.8
6	g	2.8
7	vg	2.8
9	vg	1.8
11	vg	1.8
12	g	1.8
13	vg	1.8
14	g	1.8
15	g	0.8
16	g	2.8
20	vg	1.8
21	m	2.8
22	vg	1.8
23	g	2.8
24	m	1.8

Tampilan hasilnya adalah tabel dengan kolom pertama, kelima, dan keenam dari matriks MT, hanya label yang sesuai dengan kolom yang dipilih, yaitu hd[1], hd[5], dan hd[6], kolom ke-2 pada tampilan (yang berasal dari kolom ke-5 di MT) akan diformat sesuai dengan pengaturan dalam ctok=[2], dan tok mengatur tampilan tabel, seperti pemisah antar-kolom atau format keseluruhan, sesuai pengaturan dalam tok.

Kita perlu mengekstrak tabel itu sendiri dari daftar Tabel dalam kasus ini, menggunakan perintah : MT=Table[]

```
>MT=Table[1];
```

kita juga dapat menggunakannya untuk menentukan nilai rata-rata kolom atau nilai statistik lainnya.

```
>mean(tablecol(MT,6))
```

```
    2.175
```

```
>tablecol(MT,6)
```

```
    [1.8,  1.8,  1.8,  2.8,  1.8,  2.8,  2.8,  0.8,  1.8,  1.8,  1.8,  1.8,
    1.8,  1.8,  0.8,  2.8,  3.8,  3.8,  1.8,  2.8,  1.8,  2.8,  1.8,  2.8]
```

Kita dapat menghitung secara manual dengan cara menjumlahkan semua nilai yang ada di MT 6 kemudian membagi dengan banyak data yang ada di MT 6.

$$\bar{X} = \frac{\sum x_i}{n}$$

$$\sum x_i =$$

```
{1.8+1.8+1.8+2.8+1.8+2.8+2.8+0.8+1.8+1.8+
 1.8+1.8+1.8+1.8+0.8+2.8+3.8+3.8+1.8+2.8+
 1.8+2.8+1.8+2.8}=52.2
       n = 24
```

$$\bar{X} = \frac{\sum x_i}{n}$$

$$\bar{X} = \frac{52.2}{24}$$

$$\bar{X} = 2.175$$

Fungsi getstatistics() mengembalikan elemen-elemen dalam sebuah vektor, dan jumlahnya. Kita menerapkannya pada nilai “m” dan “f”
i kolom kedua dari tabel kita.

```
>{xu,count}=getstatistics(tablecol(MT,2)); xu, count,
```

```
  [1,  3]
  [12,  13]
```

```
>tablecol(MT,2)
```

```
  [1,  3,  3,  1,  1,  1,  3,  1,  3,  1,  3,  1,  1,  3,  3,  1,  1,  1,
  3,  3,  3,  3,  3,  3,  1]
```

getstatistics biasanya menghasilkan dua output, yaitu:
- xu: Berisi nilai-nilai unik (distinct) yang ada di kolom kedua.
- count: Berisi jumlah kemunculan (frekuensi) dari setiap nilai unik tersebut.

Kita dapat mencetak hasilnya dalam bentuk tabel baru dengan menggunakan perintah writetable(...,...)

```
>writetable(count',labr=tok[xu])
```

```
            m        12
            f        13
```

labr=tok[xu] menentukan bahwa label-label yang digunakan untuk baris atau kolom pada tabel akan diambil dari elemen-elemen array tok pada indeks xu.

hasilnya adalah "m" berjumlah 12 dan "f" berjumlah 13. Selanjutnya kita dapat mengembalikan tabel baru dengan nilai dalam satu kolom yang dipilih dari vektor indeks dengan fungsi selecttable (). Namun sebelum itu, terlebih dahulu kita mencari indeks dari dua nilai di tabel token dengan fungsi indexof().

```
>v:=indexof(tok,["g","vg"])
```

```
  [5,  6]
```

Fungsi indexof akan memeriksa tok untuk mengetahui di mana posisi pertama dari masing-masing elemen dalam daftar ["g", "vg"]. Pada tok, token v dan vg berada di posisi 5 dan 6 sehingga hasil tampilannya itu 5 dan 6.

Sekarang kita dapat memilih baris tabel, yang memiliki salah satu nilai dalam v di baris ke-5.

```
>MT1:=MT[selectrows(MT,5,v)]; i:=sortedrows(MT1,5);
```

- MT1:= ...: Hasil dari pemilihan baris ini kemudian disimpan dalam variabel MT1, yang menjadi matriks baru berisi baris-baris dari MT yang memenuhi kriteria pada kolom ke-5.
- selectrows(MT,5,v): Perintah ini memilih baris dari matriks MT di mana elemen pada kolom ke-5 memenuhi kondisi tertentu yang diberikan oleh nilai v. Misalnya, jika v adalah sebuah nilai tertentu, maka baris-baris dengan nilai kolom ke-5 yang sesuai akan diseleksi.
- i:= ...: Hasil dari pengurutan ini disimpan dalam variabel i. Variabel i berisi indeks-indeks dari baris MT1 yang menunjukkan urutan berdasarkan kolom ke-5.
- sortedrows(MT1,5): Perintah ini mengurutkan baris-baris dalam matriks MT1 berdasarkan token yang dipilih pada kolom ke-5.

Sekarang kita dapat mencetak tabel, dengan nilai yang diekstraksi dan diurutkan di kolom ke-5.

```
>writetable(MT1[i],labc=hd,ctok=ctok,tok=tok,wc=7);
```

```
     Person    Sex    Age Titanic Evaluation    Tip Problem
          2      f     23       y          g    1.8       n
          3      f     26       y          g    1.8       y
```

```
 6      m     28       y          g    2.8       y
18      m     38       y          g      .       n
16      m     26       y          g    2.8       n
15      f     31       y          g    0.8       n
12      m     32       y          g    1.8       n
23      f     38       y          g    2.8       n
14      f     25       y          g    1.8       y
 9      f     24       y         vg    1.8       y
 7      f     31       y         vg    2.8       n
20      f     28       y         vg    1.8       n
22      f     28       y         vg    1.8       y
13      m     29       y         vg    1.8       y
11      f     23       y         vg    1.8       y
```

Untuk statistik berikutnya, kita dapat menghubungkan dua kolom dari tabel yaitu kolom 2 dan 4 lalu kita mengekstrak kolom 2 dan 4 dan mengurutkan tabel.

```
>i=sortedrows(MT,[2,4]); ...
   writetable(tablecol(MT[i],[2,4])',ctok=[1,2],tok=tok)
```

```
m         n
m         n
m         n
m         n
m         n
m         n
m         n
m         y
m         y
m         y
m         y
m         y
f         n
f         y
f         y
f         y
f         y
f         y
f         y
f         y
f         y
f         y
f         y
f         y
```

Hasil tampilannya adalah dua kolom yaitu kolom 2 dan kolom 4 beserta isinya. Dengan getstatistics (), kita juga bisa melakukan hal yang smaa dengan sebelumnya yaitu menghubungkan hitungan dalam dua kolom tabel satu sama lain.

```
>MT24=tablecol(MT,[2,4]); ...
 {xu1,xu2,count}=getstatistics(MT24[1],MT24[2]); ...
 writetable(count,labr=tok[xu1],labc=tok[xu2])
```

```
             n         y
m            7         5
f            1        12
```

Tabel dapat ditulis ke file. Pertama kita namakan dulu untuk file nya.

```
>filename="test.dat"; ...
 writetable(count,labr=tok[xu1],labc=tok[xu2],file=filename);
```

Kemudian kita dapat membaca tabel dari file tersebut.

```
>{MT2,hd,tok2,hdr}=readtable(filename,>clabs,>rlabs); ...
```

```
writetable(MT2,labr=hdr,labc=hd)
```

```
          n          y
m         7          5
f         1         12
```

Tampilannya adalah file yang isinya sama dengan file yang merupakan hasil hitung m, f, n, dan y. Lalu kita dapat menghapus file tersebut.

```
>fileremove(filename);
```

File pun terhapus.

Simulasi Monte Carlo

Pengertian Simulasi Monte Carlo

```
Simulasi Monte Carlo adalah jenis algoritma komputasi yang
```

menggunakan pengambilan sampel acak berulang untuk memperoleh kemungkinan terjadinya serangkaian hasil. Dikenal juga sebagai Metode Monte Carlo atau simulasi probabilitas ganda, Simulasi Monte Carlo adalah teknik matematika yang digunakan untuk memperkirakan kemungkinan hasil dari suatu peristiwa yang tidak pasti. Metode Monte Carlo diciptakan oleh John von Neumann dan Stanislaw Ulam selama Perang Dunia II untuk meningkatkan pengambilan keputusan dalam kondisi yang tidak pasti. Metode ini dinamai berdasarkan kota kasino terkenal, yang disebut Monaco, karena unsur peluang merupakan inti dari pendekatan pemodelan, mirip dengan permainan rolet.

Manfaat Simulasi Monte Carlo

Simulasi Monte Carlo memberikan beberapa kemungkinan hasil dan probabilitas dari masing-masing dari kumpulan besar sampel data acak. Monte Carlo memberikan gambaran yang lebih jelas daripada prakiraan deterministik.

Cara Kerja Simulasi Monte Carlo

Tidak seperti model peramalan normal, Simulasi Monte Carlo memprediksi serangkaian hasil berdasarkan kisaran nilai yang diestimasikan versus serangkaian nilai masukan yang tetap. Dengan kata lain, Simulasi Monte Carlo membangun model hasil yang mungkin dengan memanfaatkan distribusi probabilitas, seperti distribusi seragam atau normal, untuk setiap variabel yang memiliki ketidakpastian inheren. Kemudian, model tersebut menghitung ulang hasilnya berulang kali, setiap kali menggunakan serangkaian angka acak yang berbeda antara nilai minimum dan maksimum. Dalam eksperimen Monte Carlo yang umum, latihan ini dapat diulang ribuan kali untuk menghasilkan sejumlah besar kemungkinan hasil.

Simulasi Monte Carlo juga digunakan untuk prediksi jangka panjang karena keakuratannya. Seiring bertambahnya jumlah input, jumlah prakiraan juga bertambah, yang memungkinkan Anda memproyeksikan hasil di masa mendatang dengan akurasi yang lebih tinggi. Saat Simulasi Monte Carlo selesai, simulasi ini menghasilkan serangkaian kemungkinan hasil dengan probabilitas terjadinya masing-masing hasil.

Contohnya adalah berapa distribusi jumlah dari 1000 kali lemparan 3 dadu?

```
>ds:=sum(intrandom(1000,3,6))'; fs=getmultiplicities(3:18,ds)
```

```
 [4,  11,  22,  44,  82,  104,  123,  129,  116,  113,  86,  63,  58,
 23,  12,  10]
```

Kita bisa merencanakannya sekarang.

```
>columnsplot(fs,lab=3:18):
```

Fungsi berikut menghitung banyaknya cara bilangan k dapat direpresentasikan sebagai jumlah dari n bilangan dalam rentang 1 hingga m. Ini bekerja secara rekursif dengan cara yang jelas.

```
>function map countways (k; n, m) ...
   if n==1 then return k>=1 && k<=m
   else
     sum=0;
     loop 1 to m;sum=sum+countways(k-#,n-1,m); end;
     return sum;
   end;
 endfunction
```

Ini adalah hasil dari tiga lemparan dadu.

```
>cw=countways(3:18,3,6)
```

```
 [1,  3,  6,  10,  15,  21,  25,  27,  27,  25,  21,  15,  10,  6,  3,
 1]
```

Atau kita bisa menemukan secara manual yaitu :
Jumlah minimal dari 3 dadu tersebut adalah 3 dan jumlah maksimal dari 3 dadu tersebut adalah 18, diperoleh nilai terkecil dan nilai terbesar dari dadu. Sedangkan jumlah seluruh sampelnya yaitu 6^3 = 216, sehingga diperoleh frekuensi dari jumlah 3 hingga jumlah 18 yaitu :

Jumlah 3 = 1
Jumlah 4 = 3
Jumlah 5 = 6
Jumlah 6 = 10
Jumlah 7 = 15
Jumlah 8 = 21
Jumlah 9 = 25
Jumlah 10 = 27
Jumlah 11 = 27
Jumlah 12 = 25
Jumlah 13 = 21
Jumlah 14 = 15
Jumlah 15 = 10
Jumlah 16 = 6
Jumlah 17 = 3
Jumlah 18 = 1

sehingga peluang dari 3 dadu yang dilempar sebanyak 1000 kali adalah :

$$\text{Peluang muncul dadu dengan jumlah } 3$$

$$\text{Peluang muncul dadu dengan jumlah } 3 = \frac{1}{216}1000$$

$$\text{Peluang muncul dadu dengan jumlah } 3 = \frac{1000}{216}$$

$$\text{Peluang muncul dadu dengan jumlah } 3 = 4.629269$$

$$\text{Peluang muncul dadu dengan jumlah } 4$$

$$\text{Peluang muncul dadu dengan jumlah } 4 = \frac{3}{216}1000$$

$$\text{Peluang muncul dadu dengan jumlah } 4 = \frac{3000}{216}$$

$$\text{Peluang muncul dadu dengan jumlah } 4 = 13.8888$$

$$\text{Peluang muncul dadu dengan jumlah } 5$$

$$\text{Peluang muncul dadu dengan jumlah } 5 = \frac{6}{216}1000$$

$$\text{Peluang muncul dadu dengan jumlah } 5 = \frac{6000}{216}$$

$$\text{Peluang muncul dadu dengan jumlah } 5 = 27.7777$$

$$\text{Peluang muncul dadu dengan jumlah } 6$$

$$\text{Peluang muncul dadu dengan jumlah } 6 = \frac{10}{216}1000$$

$$\text{Peluang muncul dadu dengan jumlah } 6 = \frac{10000}{216}$$

$$\text{Peluang muncul dadu dengan jumlah } 6 = 46.296296$$

$$\text{Peluang muncul dadu dengan jumlah } 7$$

$$\text{Peluang muncul dadu dengan jumlah } 7 = \frac{15}{216}1000$$

$$\text{Peluang muncul dadu dengan jumlah } 7 = \frac{15000}{216}$$

$$\text{Peluang muncul dadu dengan jumlah } 7 = 69.44444$$

$$\text{Peluang muncul dadu dengan jumlah } 8$$

$$\text{Peluang muncul dadu dengan jumlah } 8 = \frac{21}{216}1000$$

$$\text{Peluang muncul dadu dengan jumlah } 8 = \frac{21000}{216}$$

$$\text{Peluang muncul dadu dengan jumlah } 8 = 97.22222$$

$$\text{Peluang muncul dadu dengan jumlah } 9$$

$$\text{Peluang muncul dadu dengan jumlah } 9 = \frac{25}{216}1000$$

$$\text{Peluang muncul dadu dengan jumlah } 9 = \frac{25000}{216}$$

$$\text{Peluang muncul dadu dengan jumlah } 9 = 115.740740$$

$$\text{Peluang muncul dadu dengan jumlah } 10$$

$$\text{Peluang muncul dadu dengan jumlah } 10 = \frac{27}{216}1000$$

$$\text{Peluang muncul dadu dengan jumlah } 10 = \frac{27000}{216}$$

$$\text{Peluang muncul dadu dengan jumlah } 10 = 125$$

$$\text{Peluang muncul dadu dengan jumlah } 11$$

$$\text{Peluang muncul dadu dengan jumlah } 11 = \frac{27}{216}1000$$

$$\text{Peluang muncul dadu dengan jumlah } 11 = \frac{27000}{216}$$

$$\text{Peluang muncul dadu dengan jumlah } 11 = 125$$

$$\text{Peluang muncul dadu dengan jumlah } 12$$

$$\text{Peluang muncul dadu dengan jumlah } 12 = \frac{25}{216}1000$$

$$\text{Peluang muncul dadu dengan jumlah } 12 = \frac{25000}{216}$$

$$\text{Peluang muncul dadu dengan jumlah } 12 = 115.740740$$

$$\text{Peluang muncul dadu dengan jumlah } 13$$

$$\text{Peluang muncul dadu dengan jumlah } 13 = \frac{21}{216}1000$$

$$\text{Peluang muncul dadu dengan jumlah } 13 = \frac{21000}{216}$$

$$\text{Peluang muncul dadu dengan jumlah } 13 = 97.22222$$

$$\text{Peluang muncul dadu dengan jumlah } 14$$

$$\text{Peluang muncul dadu dengan jumlah } 14 = \frac{15}{216}1000$$

$$\text{Peluang muncul dadu dengan jumlah } 14 = \frac{15000}{216}$$

$$\text{Peluang muncul dadu dengan jumlah } 14 = 69.44444$$

$$\text{Peluang muncul dadu dengan jumlah } 15$$

$$\text{Peluang muncul dadu dengan jumlah } 15 = \frac{10}{216}1000$$

$$\text{Peluang muncul dadu dengan jumlah } 15 = \frac{10000}{216}$$

$$\text{Peluang muncul dadu dengan jumlah } 15 = 46.296296$$

$$\text{Peluang muncul dadu dengan jumlah } 16$$

$$\text{Peluang muncul dadu dengan jumlah } 16 = \frac{6}{216}1000$$

$$\text{Peluang muncul dadu dengan jumlah } 16 = \frac{6000}{216}$$

$$\text{Peluang muncul dadu dengan jumlah } 16 = 27.7777$$

$$\text{Peluang muncul dadu dengan jumlah } 17$$

$$\text{Peluang muncul dadu dengan jumlah } 17 = \frac{3}{216}1000$$

$$\text{Peluang muncul dadu dengan jumlah } 17 = \frac{3000}{216}$$

$$\text{Peluang muncul dadu dengan jumlah } 17 = 13.8888$$

$$\text{Peluang muncul dadu dengan jumlah } 18$$

$$\text{Peluang muncul dadu dengan jumlah } 18 = \frac{1}{216}1000$$

$$\text{Peluang muncul dadu dengan jumlah } 18 = \frac{1000}{216}$$

$$\text{Peluang muncul dadu dengan jumlah } 18 = 4.6296296$$

```
>plot2d(cw/6^3*1000,>add); plot2d(cw/6^3*1000,>points,>add):
```

Untuk simulasi lain, deviasi nilai rata-rata n 0-1-variabel acak terdistribusi normal adalah 1 / sqrt (n).

```
>longformat; 1/sqrt(10)
```

```
    0.316227766017
```

Pada grafik tersebut, dipilih n=10 karena Jika dihasilkan 10 variabel acak berdistribusi normal, maka rata-rata dari 10 nilai acak tersebut akan memiliki deviasi yang lebih kecil dibandingkan deviasi dari setiap variabel acak secara individu. Sehingga

$$\text{Deviasi rata-rata} = \frac{1}{\sqrt{(n)}}$$

$$\text{Deviasi rata-rata} = \frac{1}{\sqrt{(10)}}$$

$$\text{Deviasi rata-rata} = 0.3162277660$$

Mari kita periksa dengan simulasi jika kita menghasilkan 10.000 kali 10 vektor acak.

```
>M=normal(10000,10); dev(mean(M)')
```

```
    0.316328304681
```

```
>M=normal(10000,10)
```

```
    Real 10000 x 10 matrix
      0.151685     -1.00226      1.94616     0.743747    ...
      0.979657    -0.122011     0.818575     -1.89213    ...
      -1.11448     -0.21814    -0.413708     -1.00679    ...
      0.438735     -1.86911    -0.990688    -0.516679    ...
      -1.34763     -0.62153     -1.74141      1.18461    ...
       1.16604      1.32031    0.0676411     -0.79376    ...
      -1.88976     0.513862    -0.809437     -0.24734    ...
      0.880383     0.113518    0.0139723    0.0294643    ...
       1.50031     0.474103     0.803946     -2.61492    ...
       1.98931     0.046397     -1.50437    -0.818097    ...
      0.940107     0.913105     0.913803      1.27034    ...
      0.157797     0.781666      1.75805      1.81021    ...
     -0.153745     0.744802     -1.24776     0.329338    ...
     -0.582068     0.361094     0.372039    -0.124113    ...
      0.316495     -0.62396    -0.293248      1.03646    ...
       1.33143     0.254778    -0.675555      1.43669    ...
      0.662019     -1.90994      1.84987      1.17101    ...
    -0.0103056     0.469626    -0.169751     0.158544    ...
     -0.805576     -3.36537     0.540737     0.132018    ...
```

```
         0.300512      -1.64979     -0.747102          1.39     ...
        ...
```

Fungsi normal di sini kemungkinan besar menghasilkan sampel acak dari distribusi normal (atau distribusi Gauss) dengan dua parameter: rata-rata (mean) dan simpangan baku (standard deviation).
0000 adalah jumlah sampel yang dihasilkan.
10 adalah simpangan baku distribusi normal yang digunakan (mungkin nilai default untuk rata-rata adalah 0, kecuali jika ada parameter lain yang diberikan). M berisi 10000 nilai acak dari distribusi normal dengan rata-rata 0 dan simpangan baku 10.

```
>mean(M)'
```

```
    [0.054066,  0.137852,  -0.269318,  -0.303338,  -0.215712,  0.179388,
    -0.577525,  0.455971,  0.0964222,  0.0744279,  0.464923,  0.109895,
    0.0359783,  0.0253716,  0.286258,  0.615962,  -0.0123694,  -0.13424,
    -0.524677,  -0.062896,  0.301061,  0.0839325,  0.563085,  -0.153123,
    -0.456268,  0.417449,  0.111746,  -0.127539,  0.0627154,  -0.271042,
    -0.160956,  0.0535369,  0.330689,  0.0260881,  0.0340671,  0.0546837,
    0.139457,  0.103406,  -0.0986348,  -0.372611,  -0.156452,  0.0965679,
    0.180855,  -0.204245,  0.159236,  0.208756,  -0.686867,  0.162544,
    -0.279051,  -0.0641498,  0.283022,  -0.171005,  -0.133081,  -0.305445,
    0.252124,  0.38621,  0.0566469,  0.378854,  0.0752077,  -0.197157,
    0.510207,  0.789611,  -0.201046,  -0.0691992,  -0.701826,  0.150543,
    -0.316121,  -0.0130725,  0.00545695,  -0.270085,  -0.00573537,
    0.381877,  -0.349566,  0.0530836,  0.150669,  0.500059,  0.240934,
    0.560466,  0.213385,  -0.0700648,  -0.126668,  -0.0979828,  -0.579372,
    0.0955774,  0.0861937,  -0.0981197,  -0.248876,  0.326066,  -0.314562,
    -0.0113767,  -0.102068,  0.017878,  0.0990435,  0.524752,  -0.592298,
    0.18421,  0.233783,  -0.257091,  -0.259344,  0.109608,  0.323671,
    0.0476524,  0.195784,  -0.457423,  0.113264,  -0.587725,  0.286818,
    -0.311792,  -0.0271427,  -0.261843,  0.0335133,  -0.064879,
    -0.316554,  -0.250199,  -0.328343,  0.219324,  -0.579584,  -0.0653047,
     ... ]
```

Fungsi mean(M) menghitung rata-rata (mean) dari data yang ada dalam matriks M.
anda ' setelah mean(M) menandakan operasi transpose (mengubah vektor baris menjadi vektor kolom atau sebaliknya). Jadi, ini menghasilkan rata-rata dari data dalam M dalam bentuk vektor kolom.

```
>dev(mean(M)')
```

```
    0.31305194497
```

dev(mean(M)') akan menghasilkan deviasi standar dari rata-rata, yang biasanya akan sangat kecil karena kita hanya menghitung deviasi standar dari satu nilai (rata-rata tersebut

```
>plot2d(mean(M)',>distribution):
```

Median dari 10 bilangan acak terdistribusi normal 0-1 memiliki deviasi yang lebih besar.

```
>dev(median(M)')
```

```
    0.368089256438
```

Fungsi dev() di EMT digunakan untuk menghitung deviasi standar (standard deviation) dari argumen yang diberikan. Dalam hal ini, dev(median(M)') akan menghitung deviasi standar dari vektor hasil transpos median(M).

```
>median(M)'
```

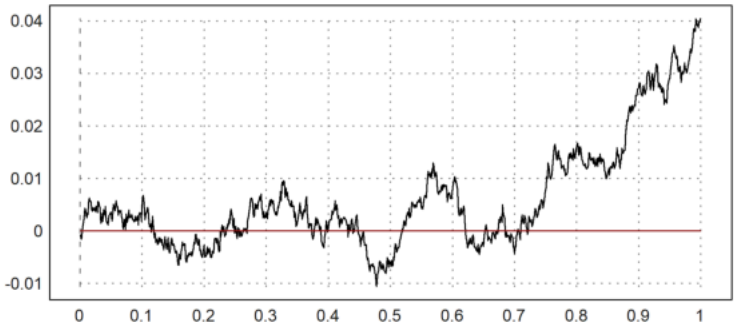
```
    [0.0684484,  0.325667,  -0.386868,  -0.364706,  -0.144964,  -0.199952,
    -0.792367,  0.103994,  0.488095,  0.229225,  0.6333,  0.0431889,
    0.0877964,  0.0127804,  0.344804,  0.394353,  0.355689,  0.0741192,
    -0.304557,  -0.045104,  0.278832,  -0.115618,  0.619516,  0.0112977,
```

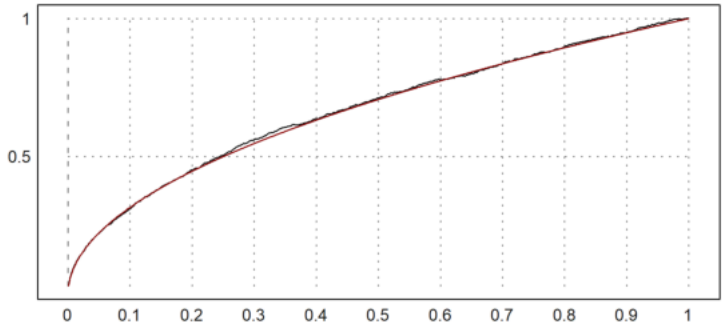
```
  -0.629649,  0.35773,  0.14809,  0.0834768,  0.300811,  -0.312329,
  0.126922,  0.0743472,  0.287485,  -0.128953,  0.0713497,  -0.181692,
  -0.0469569,  -0.0168101,  0.034153,  -0.270542,  -0.404438,  0.265634,
  0.239947,  -0.252063,  0.0664508,  0.633112,  -0.818544,  0.0675659,
  -0.232479,  -0.274845,  -0.142955,  -0.343121,  0.239811,  -0.151973,
  0.0403504,  0.409607,  0.216766,  0.347813,  0.153663,  0.15999,
  0.201931,  0.838007,  0.0456063,  -0.0197394,  -0.880415,  0.0784366,
  -0.165468,  -0.413477,  -0.174709,  -0.0995962,  -0.332236,  0.492711,
  -0.25966,  0.0379527,  0.0446682,  0.263386,  0.260861,  0.514473,
  0.131535,  -0.106826,  -0.30017,  -0.0865173,  -0.677099,  0.0235322,
  0.113935,  0.018397,  -0.231763,  0.162483,  -0.117253,  -0.196423,
  -0.177889,  0.353854,  -0.0769231,  0.293384,  -0.641366,  0.189722,
  0.132624,  -0.306178,  -0.626473,  0.107711,  0.426145,  0.0261286,
  0.101676,  -0.2583,  -0.203698,  -0.50372,  0.251377,  -0.0401274,
  -0.0380675,  -0.396162,  0.286002,  0.0415127,  -0.0195718,
  -0.266532,  -0.26727,  0.186955,  -0.620237,  -0.116954,  0.511446,
   ... ]
```

median(M) akan mengembalikan vektor baris atau kolom yang berisi median dari setiap kolom (untuk matriks). Artinya, setiap elemen di vektor hasil adalah median dari nilai dalam satu kolom dari M. Tanda (') berarti transpos matriks atau vektor. Jadi, median(M)' akan melakukan transpos terhadap hasil dari median(M). Jika hasilnya adalah vektor kolom, maka transpos akan mengubahnya menjadi vektor baris, dan sebaliknya.

Karena kita dapat membuat jalan acak, kita dapat mensimulasikan proses Wiener, yaitu Proses Wiener adalah istilah dalam teori probabilitas yang merujuk pada model matematis yang digunakan untuk menggambarkan gerak acak. Proses ini juga dikenal sebagai Brownian motion dalam konteks fisika dan merupakan contoh dari proses stokastik kontinu. Kita mengambil 1000 langkah dari 1000 proses. kemudian memplot deviasi standar dan mean dari langkah ke-n dari proses ini bersama dengan nilai yang diharapkan berwarna merah.

```
>n=1000; m=1000; M=cumsum(normal(n,m)/sqrt(m)); ...
 t=(1:n)/n; figure(2,1); ...
 figure(1); plot2d(t,mean(M')'); plot2d(t,0,color=red,>add); ...
 figure(2); plot2d(t,dev(M')'); plot2d(t,sqrt(t),color=red,>add); ...
 figure(0):
```





Uji Statistik

Uji statistik adalah alat atau prosedur yang digunakan dalam analisis data untuk menentukan kemungkinan mengamati pola, hubungan, atau perbedaan tertentu dalam kumpulan data secara kebetulan saja. Uji statistik membantu peneliti menarik kesimpulan tentang populasi berdasarkan data sampel. Uji statistik melibatkan perhitungan matematika dan pengujian hipotesis untuk menilai signifikansi hasil dan membuat kesimpulan tentang populasi yang mendasarinya.

Jenis-jenis Uji Statistik

a. Uji Statistik Parametrik

- Uji Regresi : Uji regresi menentukan hubungan sebab-akibat. Uji ini dapat digunakan untuk memperkirakan pengaruh satu atau lebih variabel kontinu terhadap variabel lain. Jenis-jenis uji regresi adalah Uji regresi linier sederhana, Uji regresi linier berganda, dan uji regresi logistik.
- Uji Perbandingan : Uji perbandingan menentukan perbedaan di antara rata-rata kelompok. Uji ini dapat digunakan untuk menguji pengaruh variabel kategoris terhadap nilai rata-rata karakteristik lainnya. Jenis-jenis uji perbandingan adalah Uji T, Uji T berpasangan, Uji T Independen, Uji T satu sampel, Analisis Variaans (ANOVA), Multivariate Analysis of Variance (MANOVA), dan Uji z.
- Uji Korelasi : Uji korelasi memeriksa apakah variabel-variabel tersebut saling terkait tanpa mengasumsikan adanya hubungan sebab- akibat. Uji ini dapat digunakan untuk memeriksa apakah dua variabel yang ingin digunakan dalam uji regresi berganda saling berkorelasi. Jenis-jenis uji korelasi adalah koefisien korelasi pearson.

b. Uji Statistik nonParametrik
Uji nonparametrik tidak membuat banyak asumsi tentang data dibandingkan dengan uji parametrik. Uji ini berguna ketika satu atau beberapa asumsi statistik umum dilanggar. Namun, kesimpulan ini tidak seakurat uji parametrik. Salah satu uji yang termasuk uji statistik nonparametrik adalah Uji Chi Kuadrat atau Uji ChiSquare.

Uji Chi-Square

Uji Chi-square atau uji Khi-Kuadrat bertujuan untuk menghitung nilai harapan dan menggunakan uji goodness-of-fit dengan uji khi kuadrat pada tiga atau lebih proporsi populasi.

Jenis data yang dapat diuji dengan Chi-Square

a. Data kategorik univariat, goodness of fit tests
- Data kategorik seringkali disajikan dalam suatu distribusi frekuensi.
- Fokusannya hanya pada banyaknya observasi atau pengamatan dalam masing-masing kategori dan masing-masing kategori muncul.
- prosedur uji hipotesis yang disajikan pada bagian kategorik univariat dirancang untuk membandingkan sekumpulan proporsi yang dihipotesiskan dengan sekumpulan proporsi sebenarnya, untuk memeriksa goodness of fit (Gof).

Hipotesis

$$H_0 : p_1 = p_{10}, p_2 = p_{20}, ..., p_k = p_{k0}$$

$$\text{(Setiap proporsi kategori benar sama}$$

$$\text{dengan nilai hipotesis yang ditentukan)}$$

$$H_1 : p_i \text{tidak sama dengan } p_{i0}$$

$$\text{bagi minimal satu i}$$

$$\text{(Setidaknya ada satu proporsi kategori benar}$$

$$\text{yang tidak sama dengan nilai hipotesis yang}$$

$$\text{ditentukan yang sesuai)}$$

statistik uji

$$X^2 = \sum_{i=1}^{k} \frac{(n_i - e_i)^2}{e_i}$$

dengan

$$e_i = np_{i0}$$

Kriteria keputusan

$$H_0 \text{ditolak jika } X^2 > X^2_{\alpha,k-1}$$

Tes ini sesuai jika semua jumlah sel yang diharapkan setidaknya 5 bagi semua i.

Contoh uji chi-Square untuk data kategorik univariat

Sebagai contoh, kami menguji lemparan dadu untuk distribusi seragam. Pada 600 lemparan, kami mendapatkan nilai berikut, yang kami masukkan ke dalam uji chi-square.

```
>chitest([90,103,114,101,103,89],dup(100,6)')
```

```
  0.498830517952
```

Data pengamatan atau observasi yang diperoleh adalah 90,103,114,101,103,89. Jika tipe pelemparan dadu tersebut sama, maka proporsi pelemparan berada dalam masing-masing katgori, yaitu

$$\frac{1}{k} = \frac{1}{6} = 0.16$$

Sehingga diperoleh frekuensi harapan (e_i) yaitu

$$e_i = \frac{600}{6} = 100$$

Hipotesis

$$H_0 : p_1 = 0.16, p_2 = 0.16, ..., p_6 = 0.16$$

$$H_1 : \text{terdapat } p_i \text{tidak sama dengan}$$

$$p_{i0}, i = 1, 2, 3, 4, 5, 6$$

Taraf signifikansi:

$$\alpha = 0.05$$

Statistik uji

$$X^2 = \sum_{i=1}^{k} \frac{(n_i - e_i)^2}{e_i}$$

kriteria keputusan:

$$k = 4, X^2_{0.05(5)} = 11.070$$

$$H_0 \text{ditolak jika} X^2 > 11.070$$

Sekarang kita akan melakukan statistik uji :

$$X^2 = \sum_{i=1}^{6} \frac{(n_i - e_i)^2}{e_i}$$

$$\frac{(90-100)^2}{100} + \frac{(103-100)^2}{100} + \frac{(114-100)^2}{100}$$

$$+ \frac{(101-100)^2}{100} + \frac{(103-100)^2}{100} + \frac{(89-100)^2}{100}$$

$$X^2 = 1 + 0.09 + 1.96 + 0.01 + 0.09 + 1,21$$

$$X^2 = 4.36$$

oleh karena

$$X^2 = 4.36 < 11.070$$

dan

$$\text{p-value} 0.499 > 0.05$$

maka H_0 tidak ditolak. pada taraf signifikansi alpha = 0.05, tidak ada bukti untuk menyimpulkan bahwa ada proporsi sebenarnya yang berbeda dari 0.16.

Uji chi-square juga memiliki mode, yang menggunakan simulasi Monte Carlo untuk menguji statistik. Hasilnya harusnya hampir sama. Parameter> p mengartikan vektor y sebagai vektor probabilitas.

```
>chitest([90,103,114,101,103,89],dup(1/6,6)',>p,>montecarlo)
```

```
  0.487
```

Selanjutnya kita menghasilkan 1000 lemparan dadu menggunakan generator nomor acak, dan melakukan tes yang sama.

```
>n=1000; t=random([1,n*6]); chitest(count(t*6,6),dup(n,6)')
```

```
0.0456038958572
```

- Fungsi n=1000; Mendeklarasikan variabel n yang menyimpan jumlah total lemparan dadu yang akan dilakukan. Dalam hal ini, n diatur menjadi 1000. -t = random([1, n*6]); Fungsi random digunakan untuk menghasilkan angka acak dalam rentang tertentu. Di sini, random([1, n*6]) mungkin mencoba menghasilkan array angka acak yang memiliki ukuran tertentu. Namun, sintaks ini tampaknya sedikit tidak tepat jika tujuan Anda adalah menghasilkan angka acak antara 1 hingga 6 untuk 1000 lemparan dadu.
- chitest(count(t*6, 6), dup(n, 6)') untuk menghitung frekuensi observasi dari hasil lemparan dan melakukan uji chi-square dengan membandingkan frekuensi yang diamati dengan distribusi yang diharapkan.

Mari kita uji nilai rata-rata 100 dengan uji-t.

```
>s=200+normal([1,100])*10; ...
 ttest(mean(s),dev(s),100,200)
```

```
0.34142223121
```

Fungsi ttest () membutuhkan nilai mean, deviasi, jumlah data, dan nilai mean untuk diuji.

Sekarang mari kita periksa dua pengukuran untuk mean yang sama. Kami menolak hipotesis bahwa mereka memiliki mean yang sama, jika hasilnya <0,05.

```
>s=200+normal([1,100])*10
```

```
[189.006,  203.518,  205.781,  211.546,  216.448,  211.663,  205.449,
186.222,  201.539,  201.946,  199.065,  188.347,  202.124,  193.428,
215.124,  200.759,  205.076,  195.09,  190.896,  198.217,  189.65,
186.439,  193.552,  197.141,  194.763,  181.19,  200.599,  195.3,
205.465,  192.096,  217.975,  202.911,  177.938,  188.34,  203.369,
208.461,  200.263,  206.772,  177.055,  172.294,  204.237,  193.373,
197.297,  210.604,  186.875,  198.243,  191.572,  195.299,  193.279,
202.902,  179.697,  202.407,  212.861,  193.375,  197.515,  202.237,
193.609,  197.319,  195.396,  202.553,  208.082,  201.648,  199.913,
198.654,  182.717,  195.548,  195.831,  195.217,  205.61,  172.636,
221.952,  216.335,  182.094,  193.421,  189.289,  213.19,  219.407,
195.728,  202.254,  213.923,  209.843,  201.368,  187.211,  199.067,
200.916,  196.585,  198.542,  225.116,  209.583,  181.694,  198.768,
193.169,  193.974,  207.596,  226.426,  200.514,  208.877,  205.728,
210.514,  200.136]
```

```
>mean(s)
```

```
199.225442249
```

```
>dev(s)
```

```
10.7573505018
```

Uji t adalah Uji t adalah uji statistik yang digunakan untuk membandingkan rata-rata sampel dengan nilai yang diharapkan (misalnya rata-rata populasi) atau untuk membandingkan rata-rata dua sampel.

Adapun cara menghitung uji-t adalah :
1. menentukan hipotesis

$$H_0 : \text{rata-rata sampel} = 200$$

$$H_1 : \text{rata-rata sampel tidak sama dengan} 200$$

2. menentukan signifikansi

$$\alpha = 0.05$$

3. menghitung statistik uji :

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

dengan

$$\bar{x} = \text{rata-rata sampel}$$

$$\mu = \text{rata-rata populasi atau yang diharapkan}$$

$$s = \text{deviasi standar sampel}$$

$$n = \text{jumlah sampel}$$

sehingga

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

$$t = \frac{199.225 - 200}{\frac{10.757}{\sqrt{100}}}$$

$$t = \frac{-0.775}{1.0757}$$

$$t = -0.720$$

Untuk tingkat signifikansi alpha=0.05 dan uji dua sisi, kita menggunakan tabel distribusi t untuk menemukan nilai kritis t. Berdasarkan tabel distribusi t, untuk df=99 dan alpha=0.05, nilai kritis t adalah sekitar ±1.984. Berdasarkan distribusi t, untuk t=-0.720 dan df=99, p-value yang didapat adalah sekitar 0.3414. Dengan t = -0.720 dan p-value > 0.05, kita gagal menolak hipotesis nol, yang berarti rata-rata sampel tidak berbeda signifikan dari nilai yang diuji, yaitu 200.

```
>tcomparedata(normal(1,10),normal(1,10))
```

```
  0.381011874607
```

Fungsi tcomparedata() digunakan untuk membandingkan dua set data yang diberikan. Fungsi ini mengasumsikan bahwa kedua set data yang dibandingkan berasal dari distribusi yang sama dan melakukan uji perbandingan, seperti uji t, untuk melihat apakah terdapat perbedaan yang signifikan antara kedua set data tersebut. tcomparedata(normal(1,10), normal(1,10)) akan membandingkan dua sampel acak yang dihasilkan dari distribusi normal dengan parameter rata-rata 1 dan simpangan baku 10.

Jika kita menambahkan bias ke satu distribusi, kita mendapatkan lebih banyak penolakan. Ulangi simulasi ini beberapa kali untuk melihat efeknya.

```
>normal(1,10)
```

```
  [-1.17136,  -1.82033,  1.18563,  0.242431,  0.0464708,  -1.64705,
  -0.605288,  0.37594,  -0.102427,  -2.20438]
```

```
>normal(1,10)
```

```
  [0.525392,  -0.404127,  -0.136861,  -1.16369,  -1.2074,  0.852233,
  1.50818,  0.592556,  0.866333,  -0.804296]
```

```
>R=random(100,20); R=sum(R*6<=1)'; mean(R)
```

```
  3.25
```

Sekarang kami membandingkan jumlah satuan dengan distribusi binomial. Pertama kami memplot distribusi satu.

```
>R=random(100,20)
```

```
  Real 100 x 20 matrix
       0.924035      0.841409     0.0129681      0.349666     ...
       0.748544      0.668704      0.405077      0.141343     ...
       0.437177      0.638373      0.849585       0.92516     ...
       0.372093       0.63486      0.698864      0.976892     ...
       0.944048      0.653211      0.973297      0.612211     ...
       0.864933      0.169111      0.548412    0.00913312     ...
        0.13826      0.985917      0.319179      0.144756     ...
      0.0992828      0.574177     0.0241684      0.702153     ...
       0.881473       0.70026      0.547032       0.78667     ...
       0.188297      0.957198      0.460827      0.251874     ...
       0.948249      0.549084       0.27058       0.93615     ...
       0.434448      0.766206      0.602589      0.512131     ...
       0.402645    0.00341134      0.545179      0.077726     ...
       0.500287      0.342416      0.187569      0.323336     ...
       0.896779      0.354839      0.439077      0.977004     ...
        0.18772     0.0534695      0.543357     0.0914732     ...
     0.00084127      0.278015      0.937448     0.0290271     ...
       0.414482      0.856088      0.592102      0.106549     ...
        0.21762        0.1263      0.665896      0.244132     ...
       0.348405      0.168132      0.473588      0.870012     ...
      ...
```

Fungsi random(100, 20): Fungsi ini menghasilkan sebuah matriks acak dengan ukuran 100 x 20, yang berarti 100 baris dan 20 kolom. Setiap elemen dalam matriks R adalah angka acak dalam interval (0, 1).

```
>R=sum(R*6<=1)'
```

```
  [2,  4,  4,  5,  2,  8,  5,  4,  3,  0,  4,  0,  7,  2,  3,  5,  2,  3,
  5,  1,  2,  1,  2,  4,  3,  3,  3,  5,  4,  2,  4,  3,  3,  4,  9,  4,
  4,  2,  5,  2,  2,  3,  4,  0,  6,  4,  2,  3,  4,  6,  2,  2,  7,  4,
  2,  1,  3,  2,  4,  6,  1,  3,  3,  7,  3,  2,  2,  3,  2,  4,  1,  6,
  6,  3,  5,  2,  3,  3,  4,  1,  2,  1,  3,  1,  3,  4,  3,  3,  3,  3,
  3,  1,  3,  2,  2,  2,  4,  4,  2,  3]
```

sum(R*6<=1): Fungsi sum menghitung jumlah nilai True (atau 1) di setiap kolom dari matriks yang memenuhi kondisi R×6<=1. Dengan kata lain, fungsi ini menghitung berapa banyak elemen dalam setiap kolom yang memenuhi kondisi tersebut. Tanda ' (transpose): Setelah itu, hasil penjumlahan ini ditranspose (diubah dari vektor baris menjadi vektor kolom). Hasilnya adalah vektor kolom yang berisi jumlah elemen yang memenuhi kondisi R×6<=1 untuk setiap kolom matriks R.

```
>mean(R)
```

```
  3.21
```

mean(R): Fungsi mean menghitung nilai rata-rata dari elemen-elemen dalam vektor kolom R.

```
>plot2d(R,distribution=max(R)+1,even=1,style="\/"):
>t=count(R,21);
```

count(R, 21): Fungsi count menghitung jumlah kemunculan nilai 21 dalam array R. Jadi, jika ada elemen-elemen di dalam array R yang bernilai 21, perintah ini akan mengembalikan jumlah berapa kali angka 21 muncul dalam R.

Kemudian kami menghitung nilai yang diharapkan.

```
>n=0:20; b=bin(20,n)*(1/6)^n*(5/6)^(20-n)*100;
```

Perintah ini menghasilkan sebuah array b yang berisi probabilitas dalam persen untuk setiap jumlah kemunculan angka "1" dari 0 hingga 20 kali dalam 20 lemparan dadu.

Kami harus mengumpulkan beberapa nomor untuk mendapatkan kategori yang cukup besar.

```
>t1=sum(t[1:2])|t[3:7]|sum(t[8:21]); ...
 b1=sum(b[1:2])|b[3:7]|sum(b[8:21])
```

```
[13.042,  19.8239,  23.7887,  20.2204,  12.941,  6.47051,  3.71353]
```

```
>chitest(t1,b1)
```

```
0.751629631763
```

b. Data dengan kategorik bivariat
merupakan kumpulan data yang diperoleh dari dua pengamatan kategorik yang dilakukan pada individu atau objek yang sama.

Dapat dilakukan beberapa uji dari data kategorik bivariat, salah satunya yaitu uji independensi untuk dua variabel kategorik, yaitu
- uji yang dilakukan dalam satu sampel acak n individu.
- Dalam tabel frekuensi dua arah I x J yang dihasilkan, misalkan n_ij menyatakan cacah yang diamati dalam sel (ij) dan e_ij menyatakan cacah harapan dalam sel (ij).
- Uji hipotesis untuk independensi dari dua variabel kategori dengan taraf signifikansi alpha memiliki bentuk berikut.

Hipotesis:

$$H_0 : \text{Dua variabel independen}$$

$$H_1 : \text{Dua variabel dependen}$$

Statistik uji:

$$X^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(n_{ij} - e_{ij})^2}{e_i j}$$

dimana

$$e_{ij} = \frac{(\text{ith baris total})(\text{jth kolom total})}{\text{total ij}}$$

$$e_{ij} = \frac{n_i \times n_j}{n}$$

Kriteria keputusan:

$$H_0 \text{ditolak jika } X^2 > X^2_{\alpha,(I-1)(J-1)}$$

Uji ini baik digunakan jika semua cacah harapan minimal 5.

Contoh Uji Chi-Square data kategorik bivariat

Contoh berikut berisi hasil dari dua kelompok orang (misalnya laki-laki dan perempuan) yang memberikan suara untuk satu dari enam partai.

```
>A=[23,37,43,52,64,74;27,39,41,49,63,76]; ...
  writetable(A,wc=6,labr=["m","f"],labc=1:6)
```

	1	2	3	4	5	6
m	23	37	43	52	64	74
f	27	39	41	49	63	76

Kami ingin menguji independensi suara dari jenis kelamin. Tes tabel chi ^ 2 melakukan ini. Hasilnya adalah cara yang besar untuk menolak kemerdekaan. Jadi kami tidak bisa mengatakan, apakah voting tergantung jenis kelamin dari data ini.

dari data tersebut, nilai n total i dan j adalah

$$n_{11} = 293$$

$$n_{21} = 295$$

$$n_{i1} = 50$$

$$n_{i2} = 76$$

$$n_{i3} = 84$$

$$n_{i4} = 101$$

$$n_{i5} = 127$$

$$n_{i6} = 150$$

n=11176

perhitungan :

$$e_{11} = \frac{50 \times 293}{1176}$$

$$e_{11} = 12.457$$

$$e_{21} = \frac{50 \times 295}{1176}$$

$$e_{21} = 12.543$$

$$e_{21} = \frac{76 \times 293}{1176}$$

$$e_{21} = 18.935$$

$$e_{22} = \frac{76 \times 295}{1176}$$

$$e_{22} = 19.065$$

$$e_{31} = \frac{84 \times 293}{1176}$$

$$e_{31} = 20.929$$

$$e_{32} = \frac{84 \times 295}{1176}$$

$$e_{32} = 21.071$$

$$e_{41} = \frac{101 \times 293}{1176}$$

$$e_{41} = 25.164$$

$$e_{42} = \frac{101 \times 295}{1176}$$

$$e_{42} = 25.336$$

$$e_{51} = \frac{127 \times 293}{1176}$$

$$e_{51} = 31.642$$

$$e_{52} = \frac{127 \times 295}{1176}$$

$$e_{52} = 31.858$$

$$e_{61} = \frac{150 \times 293}{1176}$$

$$e_{62} = 37.372$$

$$e_{62} = \frac{150 \times 295}{1176}$$

$$e_{62} = 37.628$$

Taraf signifikansi

$$\alpha = 0.01$$

Kriteria keputusan :

$$(I-1)(J-1) = (2-1)(6-1) = 5, X^2_{0.01,5} = 15.086$$

Statistik Uji :

$$X^2 = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{(n_{ij} - e_{ij})^2}{e_ij}$$

$$X^2 = \frac{(23-12.457)^2}{12.457} + \frac{(27-12.543)^2}{12.543} + \frac{(37-18.935)^2}{18.935}$$
$$+ \frac{(39-19.065)^2}{19.065} + \frac{(43-20.929)^2}{20.929} + \frac{(41-21.071)^2}{21.071}$$
$$+ \frac{(52-25.164)^2}{25.164} + \frac{(49-25.336)^2}{25.336} + \frac{(64-31.642)^2}{31.642}$$
$$+ \frac{(63-31.858)^2}{31.858} + \frac{(74-37.372)^2}{37.372} + \frac{(76-37.628)^2}{37.628}$$

$$X^2 = 8.923 + 16.663 + 17.235 + 20.845 + 23.275 + 18.849$$
$$+ 28.619 + 22.102 + 33.090 + 30.442 + 35.899 + 39.130$$

$$X^2 = 295.072$$

karena nilai statistik uji X^2=295.072 > 15.086, maka H_0 ditolak, sehingga pada taraf signifikansi 0.05 dapat disimpulkan bahwa ada bukti bahwa suara laki-laki dan perempuan adalah dependen.

```
>tabletest(A)
```

```
0.990701632326
```

Berikut adalah tabel yang diharapkan, jika kita mengasumsikan frekuensi pemungutan suara yang diamati.

```
>writetable(expectedtable(A),wc=6,dc=1,labr=["m","f"],labc=1:6)
```

```
        1     2     3     4     5     6
m  24.9  37.9  41.9  50.3  63.3  74.7
f  25.1  38.1  42.1  50.7  63.7  75.3
```

Kita dapat menghitung koefisien kontingensi yang dikoreksi. Karena sangat mendekati 0, kami menyimpulkan bahwa pemungutan suara tidak bergantung pada jenis kelamin.

```
>contingency(A)
```

```
0.0427225484717
```

Uji Variansi

ANOVA melibatkan analisis data sample dari dua atau lebih
populasi. Satu-satunya perbedaan di antara populasi adalah faktor tunggal.

Asumsi-asumsi dalam prosedur pengujian ANOVA

a. k distribusi populasi adalah normal
b. k variansi populasi adalah sama

$$\sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2$$

c. sampel-sampel dipilih secara acak dan independen dari populasi masing-masing.

Simbol

a. Pengamatan :

$$y_{ij} = \text{Pengukuran ke-j yang diperoleh dari populasi ke-i}$$

b. Rata-rata pengamatan dalam sampel ke-i

$$\bar{y_i} = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij} = \frac{1}{n_i}(y_{i1} + y_{i2} + ... + y_{in_i})$$

c. Rata-rata dari semua pengamatan (grand mean)

$$\bar{y_{...}} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}$$

d. Total Pengamatan dalam sampel ke-i :

$$y_{i.} = \sum_{j=1}^{n_i} y_{ij}$$

e. Total semua pengamatan :

$$y_{...} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}$$

Dekomposisi keragaman total

a. Keragaman total dalam data (jumlah kuadrat total/ total sum of squares) didekomposisi menjadi jumlah keragaman antar sampel (jumlah kuadrat karena faktor/ sum of squares due to factor) dan keragaman dalam sampel (jumlah kuadrat galat/ sum of squares due to error).

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y_{...}})^2=\sum_{i=1}^{k}n_i(\bar{y_{i.}}-\bar{y_{..}})^2+\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y_{i.}})^2$$

$$\mathrm{JKT}=\mathrm{JKA}+\mathrm{JKG}$$

b. statistik uji ANOVA satu arah yang didasarkan pada dua estimator bagi variansi umum, sigma kuadrat, yang dihitung dari JKA dan JKG

$$F=\frac{KTA}{KTG}=\frac{\frac{JKA}{(k-1)}}{\frac{JKG}{(n-k)}}$$

Selanjutnya kami menggunakan analisis varians (uji-F) untuk menguji tiga sampel data terdistribusi normal untuk nilai rata-rata yang sama. Metode tersebut dinamakan ANOVA (analysis of variance). Di Euler, fungsi varanalysis () digunakan.

```
>x1=[109,111,98,119,91,118,109,99,115,109,94]; mean(x1),
```

```
    106.545454545
```

```
>x2=[120,124,115,139,114,110,113,120,117]; mean(x2),
```

```
    119.111111111
```

```
>x3=[120,112,115,110,105,134,105,130,121,111]; mean(x3),
```

```
    116.3
```

```
>varanalysis(x1,x2,x3)
```

```
    0.0138048221371
```

Ada k=3 sampel atau grup dan n=11+9+10=30 total pengamatan.
Kita dapat mencari

a. total sampel, yaitu

$$y_{1.}=1172$$
$$y_{2.}=1072$$
$$y_{3.}=1163$$

b. rata-rata sampe, yaitu

$$\bar{y_{1.}}=106.5454$$
$$\bar{y_{2.}}=119.1111$$
$$\bar{y_{3.}}=116.3$$

c. Variansi sampel, yaitu

$$s_1^2 = 92.4724$$

$$s_2^2 = 73.6108$$

$$s_3^2 = 97.7888$$

Jumlah semua pengamatan (grand total) adalah

$$y_{...} = 341.9565$$

Artinya, kami menolak hipotesis dengan nilai mean yang sama. Kami melakukan ini dengan probabilitas kesalahan 1,3%.

Hipotesis

$$H_0 : \mu_1 = \mu_2 = \mu_3 \text{semua rata-rata populasi sama}$$

$$H_1 : \text{terdapat} \mu_i \text{tidak sama dengan} \mu_j, i \text{tidak sama dengan} j; i, j = 1, 2, 3 \text{minimal ada dua rata-rata yang berbeda}$$

taraf signifikansi

$$\alpha = 0.013$$

statistik uji :

$$F = \frac{KTA}{KTG}$$

kriteria keputusan :

$$k = 3, n = 30, F_{\alpha,(k-1,n_1)} = F_{0.013(2,27)} = 2.51061$$

$$H_0 \text{ditolak jika} F > 2.51061$$

Sekarang kita melakukan perhitungannya :

$$F = \frac{KTA}{KTG}$$

$$F = \frac{KTA}{KTG} = \frac{\frac{JKA}{(k-1)}}{\frac{JKG}{(n-k)}}$$

.....

.....

Contoh Uji non-prametrik

Selanjutnya ada juga median test atau uji non-parametrik yang digunakan untuk membandingkan median dari dua atau lebih kelompok untuk melihat apakah mereka memiliki median yang sama. Uji ini sering digunakan saat data tidak berdistribusi normal, sehingga tidak cocok untuk uji parametris seperti uji-t atau ANOVA.

```
>a=[56,66,68,49,61,53,45,58,54];
>b=[72,81,51,73,69,78,59,67,65,71,68,71];
>mediantest(a,b)
```

```
  0.0241724220052
```

Langkah untuk menentukan median test tersebut adalah :
1. Menggabungkan kedua data tersebut menjadi satu data

$$56, 66, 68, 49, 61, 53, 45, 58, 54]72, 81, 51, 73, 69, 78, 59, 67, 65, 71, 68, 71$$

2. Mengurutkan data tersebut dari yang terkecil hingga terbesar

$$45, 49, 51, 53, 54, 56, 58, 59, 61, 65, 66, 67, 68, 68, 69, 71, 71, 72, 73, 78, 81$$

3. Menentukan median dari data tersebut

$$66.0$$

4. Menentukan median atas dan median bawah dari masing-masing kelompok

$$\text{Median atas kelompok a} = 68$$

$$\text{median bawah atau sama dengan kelompok a} = 45, 49, 53, 54, 56, 61, 66, 68$$

$$\text{Median atas kelompok b} = 68, 69, 71, 71, 72, 73, 78, 81$$

$$\text{median bawah atau sama dengan kelompok b} = 51, 59, 65, 67, 68$$

5. Melakukan uji chi-square untuk mendapatkan nilai p-value lalu membandingkan dengan tingkat signifikansi yang diinginkan.

....
....

Selanjutnya, ada rank test atau sekelompok uji statistik non-parametrik yang digunakan untuk membandingkan median atau distribusi dari dua atau lebih kelompok, terutama ketika data tidak memenuhi asumsi distribusi normal yang diperlukan oleh uji parametris. Rank test mengurutkan data dalam urutan nilai dan menggunakan peringkat tersebut untuk analisis, bukan nilai data mentah.

Langkah menentukan rank test:
1. Mengurutkan dua data tersebut

$$45, 49, 51, 53, 54, 56, 58, 59, 61, 65, 66, 67, 68, 68, 69, 71, 71, 72, 73, 78, 81$$

2. Menentukan peringkat dari setiap nilai seperti :

nilai 68 muncul 2 kali di urutan 13 dan 14
peringkat nilai 68 = (13+14)/2 = 27/2 = 13.5

3. Menghitung statistik U

```
>ranktest(a,b)
```

```
  0.00199969612469
```

Dalam contoh berikut, kedua distribusi memiliki mean yang sama.

```
>ranktest(random(1,100),random(1,50)*3-1)
```

```
  0.0351496951249
```

Sekarang mari kita coba meniru dua perlakuan a dan b yang diterapkan pada orang yang berbeda.

```
>a=[8.0,7.4,5.9,9.4,8.6,8.2,7.6,8.1,6.2,8.9];
>b=[6.8,7.1,6.8,8.3,7.9,7.2,7.4,6.8,6.8,8.1];
```

Tes signum memutuskan, jika a lebih baik dari b. Tes Signum atau Sign Test adalah uji statistik non-parametrik yang digunakan untuk membandingkan dua sampel berpasangan atau untuk menguji median suatu sampel. Uji ini berguna terutama jika kita tidak membuat asumsi tentang distribusi data (seperti normalitas) atau jika data yang dianalisis tidak berskala interval.

```
>signtest(a,b)
```

```
  0.0546875
```

Kita tidak dapat menolak bahwa a sama baiknya dengan b.

Tes Wilcoxon lebih tajam dari tes ini, tetapi bergantung pada nilai kuantitatif perbedaannya. Tes Wilcoxon adalah uji statistik non-parametrik yang digunakan untuk menguji perbedaan antara dua sampel berpasangan atau dua kondisi yang terkait. Tes ini digunakan ketika data tidak memenuhi asumsi normalitas yang diperlukan oleh uji t berpasangan (paired t-test). Tes Wilcoxon dapat digunakan untuk data ordinal atau interval dan lebih fleksibel dibandingkan dengan uji t berpasangan dalam hal distribusi data.

```
>wilcoxon(a,b)
```

```
  0.0296680599405
```

Mari kita coba dua tes lagi menggunakan seri yang dihasilkan.

```
>wilcoxon(normal(1,20),normal(1,20)-1)
```

```
  0.00401721442778
```

```
>wilcoxon(normal(1,20),normal(1,20))
```

```
  0.686276983339
```

Latihan Soal

1. Bagi peternak lebah yang ingin memanen lebih banyak madu, ada empat kemungkinan untuk mendapatkan lebih banyak lebah: package bees, nucs, colonies, swarms. Departemen ilmu pertanian dari universitas tertentu memperoleh sampel acak dari pembelian lebah, dan masing-masing diklasifikasikan ke dalam salah satu dari empat kategori tersebut. Gunakan tabel frekuensi satu arah berikut untuk menguji hipotesis bahwa empat kemungkinan pembelian lebah terjadi dengan frekuensi yang sama. Gunakan taraf signifikansi Alpha = 0.05

```
>chitest([31,36,26,20],dup(28.25,4)')
```

```
  0.173097757603
```

Diperoleh

$$\text{p-value} = P(X^2 > 4.9823) = 0.1731$$

Oleh karena

$$X^2 = 4.9823 < 7.8147 ataup - value = 0.1731 > 0.05$$

maka H_0 tidak ditolak. Pada taraf signifikansi alpha = 0.05, tidak ada bukti untuk menyimpulkan bahwa ada proporsi sebenarnya yang berbeda dari 0.25. Proporsi tipe pembelian lebah itu sama.

2. National Advisory Committte bagian imunisasi di Canada menyediakan rekomendasi kesehatan tertentu dan laporan ringkasan. Suatu sampel acak pasien-pasien yang dites positif influenza selama musim flu tahun 2012-2013 dicatat. Masing-masing orang diklasifikan dalam grup umur dan jenis flu. Ringkasan data disajikan dalam tabel berikut:

```
>B=[207,849,1592,617;66,630,754,895;319,1207,1617,512;300,1196,1946,510;120,3599,5860,597]; ...
  writetable(B,wc=6,labr=["KDL","LEP","DPEE","ELEE","ELP"],labc=1:4)
```

```
              1     2     3     4
      KDL   207   849  1592   617
      LEP    66   630   754   895
     DPEE   319  1207  1617   512
     ELEE   300  1196  1946   510
      ELP   120  3599  5860   597
```

```
>contingency(B)

     0.359159327147

>writetable(expectedtable(B),wc=6,dc=1,labr=["KDL","LEP","DPEE","ELEE","ELP"],labc=1:4)

               1     2     3     4
       KDL 141.21044.11642.6   437
       LEP 101.4 749.91179.8 313.9
      DPEE 158.11168.91838.8 489.2
      ELEE   1711263.81988.2 528.9
       ELP 440.23254.25119.5  1362

>contingency(expectedtable(B))

     0
```

Oleh karena (p-value < 0.01) maka H_0 ditolak, sehingga pada taraf signifikansi 0.01 dapat disimpulkan bahwa ada bukti bahwa grup umur dan jenis influenza adalah dependen.

3. Suatu studi penelitian menunjukkan bahwa diet tinggi garam pada wanita tua meningkatkan risiko patah tulang. Meskipun mekanisme biologisnya masih tidak jelas, namun terlihat ada hubungan antara asupan natrium yang berlebihan dan kerapuhan tulang. Salah satu ukuran kesehatan tulang adalah kadar vitamin D dalam darah. Seperti yang ditentukan oleh kuesioner makanan, sampel acak independen dari wanita tua dalam empat kategori asupan garam diperoleh. Kadar vitamin D dalam darah (dalam nmol / L) diukur di masing-masing. Data diberikan dalam tabel berikut. Adakah bukti yang menunjukkan bahwa setidaknya dua dari rata-rata populasi kadar vitamin D dalam darah berbeda? Gunakan alpha = 0.05.

```
>a1=[91.5,77.5,94.5,77.5,92.0]; mean(a1),

     86.6

>a2=[89.0,92.0,98.2,80.0,86.7]; mean(a2),

     89.18

>a3=[92.5,100.7,94.0,93.3,106.3]; mean(a3),

     97.36

>a4=[100.1,98.0,99.1,103.9,97.6]; mean(a4),

     99.74

>varanalysis(a1,a2,a3,a4)

     0.0116566288294
```

Oleh karena F = 5.08 > 3.24 dan p-value <0.05 maka H_0 ditolak. Jadi pada taraf signifikansi 0.05 dapat disimpulkan ada bukti bahwa minimal ada dua rata-rata populasi kadar vitamin D dalam darah yang berbeda.

4. Dengan taraf nyata 5% ujilah apakah (peringkat) pendapatan di departemen Q lebih kecildibandingkan departemen Z?

```
>p1=[6,10,15,32];
>p2=[12,13,15,15,20,31,38,40];
>ranktest(p1,p2)
```

```
0.117241303547
```

Karena didapat z hitung = -1.19 ada di daerah penerimaan H_0 maka H_0 diterima, sehingga peringkat Pendapatan di kedua departemen sama.

5. Untuk melihat apakah ada perbedaan produksi per hektar tanaman jagung karena pengaruh duametode penanaman yang digunakan, pertumbuhan tanaman jagung dipilih dari sejumlah plottanah yang berbeda secara random. Kemudian produksi per hektar dari masing-masing plot dihitung dan hasilnya adalah sebagai berikut: (? = 5%)

Metode 1 : 83 91 94 89 96 91 92 90 92 85
Metode 2 : 91 90 81 83 84 83 88 91 90 84 80 85

Apakah dua metode tersebut memiliki nilai median yang sama?

```
>b1=[83,91,94,89,96,91,92,90,92,85];
>b2=[91,90,81,83,84,83,88,91,90,84,80,85];
>mediantest(b1,b2)
```

```
0.985013438574
```

Harga Chi kuadrat tabel dk =1 dan alpha 5%= 3.841 karena Chi hitung < chi kuadrat tabel maka H_0 tidak ditolak. Sehingga disimpulkan dua metode mempunyai nilai median yang sama untuk produksi per hektar.

Subtopik 9

menyimpan data dan hasil analisis statistika ke dalam

file/berkas dengan berbagai format

Penyimpanan data dalam EMT adalah proses menyimpan informasi numerik, hasil analisis, dan visualisasi ke dalam bentuk file/berkas yang dapat diakses kembali. Format penyimpanan menentukan bagaimana data akan terstruktur dan dapat dibaca oleh program.

1. Format teks (.txt)

File teks adalah format paling sederhana yang menyimpan data dalam bentuk teks biasa dan mudah dibaca manusia. Format ini cocok untuk data sederhana atau hasil analisis yang berupa teks, karena sifatnya yang tidak memerlukan format khusus.

```
>file = "data.txt";  // Mendefinisikan nama file sebagai "data.txt"
>M = random(5, 5);  // Membuat matriks acak berukuran 5x5
>writematrix(M, file, separator = " "); // Menyimpan matriks ke dalam file dengan elemen dipisahkan oleh spasi
>printfile(file)   // Menampilkan isi file "data.txt"
```

```
0.6554163483485531 0.2009951854518572 0.8936223876466522 0.2818865431288053 0.5250003829714993
0.3141267749950177 0.4446156782993733 0.2994744556282315 0.2826898577756425 0.8832274852666707
0.2709059757306277 0.7044190158488305 0.217693385437102 0.4453627564273003 0.3084110353211584
0.9145409086421026 0.1935854779811416 0.4633868194128757 0.0951529788263157 0.5950170162024192
0.4311837813121366 0.7286804774486648 0.4651641815616542 0.3230318624794988 0.5251835885646776
```

```
>nilai = [75,82,90,68,95,73,88,78,85,92];
>file = "hasil_analisis.txt";
>writematrix(nilai, file)
>printfile(file)
```

```
75,82,90,68,95,73,88,78,85,92
```

2. Format data (.dat)

Format .dat digunakan secara khusus oleh EMT untuk menyimpan data numerik dengan menjaga presisi angka. Format ini efisien untuk menyimpan data dalam bentuk matriks atau array, yang penting untuk keakuratan dalam perhitungan numerik.

```
>a=random(1,100); mean(a), dev(a),
```

```
0.474656416958
0.273269044817
```

Untuk menulis data ke file, kita menggunakan fungsi writematrix().

Karena pengenalan ini kemungkinan besar ada di direktori, di mana pengguna tidak memiliki akses tulis, kami menulis data ke direktori home pengguna. Untuk buku catatan sendiri, hal ini tidak diperlukan, karena file data akan ditulis ke dalam direktori yang sama.

```
>filename="test.dat";
```

Sekarang kita menulis vektor kolom a' ke file. Ini menghasilkan satu nomor di setiap baris file.

```
>writematrix(a',filename);
```

Untuk membaca data, kami menggunakan readmatrix().

```
>a=readmatrix(filename)'
```

```
[0.655416,  0.200995,  0.893622,  0.281887,  0.525,  0.314127,
0.444616,  0.299474,  0.28269,  0.883227,  0.270906,  0.704419,
0.217693,  0.445363,  0.308411,  0.914541,  0.193585,  0.463387,
0.095153,  0.595017,  0.431184,  0.72868,  0.465164,  0.323032,
0.525184,  0.502255,  0.168603,  0.262253,  0.866587,  0.536137,
0.493453,  0.601344,  0.659461,  0.967468,  0.193151,  0.935921,
0.0728753,  0.988966,  0.0104376,  0.356626,  0.52143,  0.428893,
0.168134,  0.182742,  0.288048,  0.750042,  0.472935,  0.324407,
0.340388,  0.195494,  0.44607,  0.216545,  0.598477,  0.729271,
0.950019,  0.791698,  0.418637,  0.469159,  0.898763,  0.388543,
0.674114,  0.0752676,  0.854693,  0.123222,  0.205006,  0.465049,
0.0281269,  0.60809,  0.10999,  0.615836,  0.217391,  0.962872,
0.842267,  0.328758,  0.561146,  0.918802,  0.997492,  0.473212,
0.185709,  0.421585,  0.720843,  0.247804,  0.834083,  0.44372,
0.371866,  0.775718,  0.37468,  0.224336,  0.0102885,  0.164816,
0.356866,  0.408866,  0.350091,  0.347412,  0.161257,  0.588325,
0.83884,  0.906718,  0.916402,  0.0960726]
```

Dan hapus file tersebut.

```
>fileremove(filename);
>mean(a), dev(a),
```

```
0.474656416958
0.273269044817
```

```
>file = "data.dat";
>V = [10, 20, 30, 40, 50];  // Vektor contoh
>writematrix(V', file);  // Menyimpan vektor sebagai kolom di file
>printfile(file);
```

```
10
20
30
40
50
```

contoh lainnya

```
>file="test.dat"; open(file,"w"); ...
 writeln("A,B,C"); writematrix(random(3,3)); ...
 close();
```

Filenya terlihat seperti ini.

```
>printfile(file)
```

```
A,B,C
0.4101190650184269,0.5586998134336826,0.7113081484386663
0.08434731707420291,0.3202409265658254,0.3533112015721631
0.4332551573527457,0.4248850131943944,0.02207403531125421
```

Fungsi readtable() dalam bentuknya yang paling sederhana dapat membaca ini dan mengembalikan kumpulan nilai dan baris judul.

```
>L=readtable(file,>list);
```

Koleksi ini dapat dicetak dengan writetable() ke buku catatan, atau ke file.

```
>writetable(L,wc=10,dc=5)
```

```
         A         B         C
   0.41012    0.5587   0.71131
   0.08435   0.32024   0.35331
   0.43326   0.42489   0.02207
```

Matriks nilai adalah elemen pertama dari L. Perhatikan bahwa mean() di EMT menghitung nilai rata-rata baris matriks.

```
>mean(L[1])
```

```
  0.560042
  0.252633
  0.293405
```

3. Format CSV (.csv)

CSV, atau Comma Separated Values, adalah format yang ideal untuk data berbentuk tabel. File CSV dapat dibuka di perangkat lunak spreadsheet seperti Microsoft Excel, sehingga memudahkan analisis dan pengelolaan data dalam format tabular.

Fungsi writematrix() atau writetable() dapat dikonfigurasi untuk bahasa lain.

Misalnya, jika Anda memiliki sistem Indonesia (titik desimal dengan koma), Excel Anda memerlukan nilai dengan koma desimal yang dipisahkan dengan titik koma dalam file csv (defaultnya adalah nilai yang dipisahkan koma). File berikut "test.csv" akan muncul di folder saat ini Anda.

```
>filename="test.csv"; ...
 writematrix(random(5,3),file=filename,separator=",");
```

nahh sekarang kita dapat membuka file ini dengan Excel bahasa Indonesia secara langsung.

```
>fileremove(filename);
```

Terkadang kita memiliki string dengan token seperti berikut.

```
>s1:="f m m f m m m f f f m m f";  ...
 s2:="f f f m m f f";
```

Untuk melakukan tokenisasi ini, kami mendefinisikan vektor token.

```
>tok:=["f","m"]
```

```
 f
 m
```

Kemudian kita dapat menghitung berapa kali setiap token muncul dalam string, dan memasukkan hasilnya ke dalam tabel.

```
>M:=getmultiplicities(tok,strtokens(s1))_ ...
  getmultiplicities(tok,strtokens(s2));
```

Tulis tabel dengan header token.

```
>writetable(M,labc=tok,labr=1:2,wc=8)
```

```
                f       m
        1       6       7
        2       5       2
```

contoh lain kita menulis matrik ke dalam file

```
>file="test.csv";  ...
 M=random(3,3); writematrix(M,file);
```

Berikut isi file ini

```
>printfile(file)
```

```
 0.5909668510929599,0.0638118965663068,0.1650645844780326
 0.7963239116070477,0.718715244015118,0.4867280858437937
 0.2701174495555382,0.4301235887969402,0.172917093009666
```

4. Format Tabel .tab

Format .tab adalah format file yang menggunakan karakter tab sebagai pemisah (separator) antar data dalam tabel. Mari saya jelaskan dengan contoh penggunaan format .tab di EMT untuk data statistika.

Tabel dapat digunakan untuk membaca atau menulis data numerik. Misalnya, kita menulis tabel dengan header baris dan kolom ke sebuah file.

```
>file="test.tab"; M=random(3,3);  ...
 open(file,"w");  ...
 writetable(M,separator=",",labc=["one","two","three"]);  ...
 close(); ...
 printfile(file)
```

```
one,two,three
      0.28,      0.28,      0.07
      0.22,      0.04,      0.45
      0.49,      0.52,       0.7
```

Ini dapat diimpor ke Excel.

Untuk membaca file di EMT, kami menggunakan readtable().

```
>{M,headings}=readtable(file,>clabs); ...
 writetable(M,labc=headings)
```

```
          one       two     three
         0.28      0.28      0.07
         0.22      0.04      0.45
         0.49      0.52       0.7
```

5. Format Euler .e

Format .e dalam EMT digunakan untuk menyimpan data atau informasi terkait sesi pembelajaran, seperti variabel atau hasil analisis, yang diperlukan untuk melanjutkan pekerjaan tanpa kehilangan data.

kita dapat menulis variabel dalam bentuk definisi Euler ke file atau ke baris perintah.

```
>writevar(pi,"mypi");
```

```
    mypi = 3.141592653589793;
```

Untuk pengujian, kami membuat file Euler di direktori kerja EMT.

```
>file="test.e"; ...
 writevar(random(2,2),"M",file); ...
 printfile(file,3)
```

```
    M = [ ..
    0.3846545367015691, 0.6790765424824814;
    0.1244805858381324, 0.4892517195951175];
```

```
>load(file); show M,
```

```
    M =
         0.384655      0.679077
         0.124481      0.489252
```

jika writevar() digunakan pada suatu variabel, definisi variabel dengan nama variabel tersebut akan dicetak.

```
>writevar(M); writevar(inch$)
```

```
    M = [ ..
    0.3846545367015691, 0.6790765424824814;
    0.1244805858381324, 0.4892517195951175];
    inch$ = 0.0254;
```

Kita juga bisa membuka file baru atau menambahkan file yang sudah ada. Dalam contoh kita menambahkan file yang dibuat sebelumnya.

```
>open(file,"a"); ...
 writevar(random(2,2),"M1"); ...
 writevar(random(3,1),"M2"); ...
```

```
 close();
>load(file); show M1; show M2;
```

```
    M1 =
         0.361611     0.0303537
         0.572118      0.245745
    M2 =
         0.197327
         0.265461
         0.791105
```

Untuk menghapus file apa pun, gunakan fileremove().

```
>fileremove(file);
>open(file,"w"); writeln("M = ["); ...
 for i=1 to 5; writeln(""+random()); end; ...
 writeln("];"); close(); ...
 printfile(file)
```

```
    M = [
    0.276178975345
    0.908965623594
    0.641265982998
    0.602454940886
    0.704626651431
    ];
```

```
>load(file); M
```

```
    [0.276179,  0.908966,  0.641266,  0.602455,  0.704627]
```

6. Format gambar (.png, .jpg)

Format gambar seperti .png dan .jpg berguna untuk menyimpan visualisasi dan grafik. Kualitas gambar dapat diatur sesuai kebutuhan, dan format ini fleksibel karena dapat digunakan di berbagai platform untuk berbagai keperluan visual.

contoh, kita asusmikan jumlah rata-rata suatu data dalam rentang ukuran tertentu.

```
>r=155.5:4:187.5; v=[22,71,136,169,139,71,32,8];
```

Berikut adalah alur pendistribusiannya.

```
>plot2d(r,v,a=150,b=200,c=0,d=190,bar=1,style="\/"):
>M=fold(r,[0.5,0.5])
```

```
    [157.5,  161.5,  165.5,  169.5,  173.5,  177.5,  181.5,  185.5]
```

```
>{m,d}=meandev(M,v); m, d,
```

```
    169.901234568
    5.98912964449
```

```
>plot2d("qnormal(x,m,d)*sum(v)*4", ...
   xmin=min(r),xmax=max(r),thickness=3,add=1):
```

contoh lain

Berikut ini adalah salah satu cara untuk memplot kuantil.

```
>plot2d("qnormal(x,1,1.5)",-4,6);  ...
 plot2d("qnormal(x,1,1.5)",a=2,b=5,>add,>filled):
```

Ini adalah jenis plot lainnya.

```
>starplot(normal(1,10)+4,lab=1:10,>rays):
```

Beberapa plot di plot2d bagus untuk statika. Berikut adalah plot impuls dari data acak, terdistribusi secara seragam di [0,1].

```
>plot2d(makeimpulse(1:10,random(1,10)),>bar):
```

contoh plot logaritmik

```
>logimpulseplot(1:10,-log(random(1,10))*10):
```

7. Format HTML (.html)

Format file yang digunakan untuk membuat halaman web. HTML (HyperText Markup Language) menyusun struktur konten seperti teks, gambar, dan link agar dapat ditampilkan di browser web.

File format HTML yang satunya dihasilkan dengan menu Extras-> Export Current Notebook to HTML

8. Format PDF (.pdf)

Format file yang digunakan untuk menyimpan dokumen yang dapat dibuka dengan cara yang konsisten di berbagai perangkat. PDF (Portable Document Format) mempertahankan tampilan asli dokumen, termasuk teks, gambar, dan format tata letak.

File PDF yang satunya dihasilkan dengan menu Extras-> Generate PDF with Latex.

~~TERIMA KASIH~~