

Wprowadzenie do Mathcad 15 – ćwiczenie 4

Statystyka i analiza danych

Mathcad pozwala na wykonywanie zarówno podstawowej analizy statystyk opisowych, badania rozkładów oraz tzw. data fitting. Podstawowe funkcje statystyczne zawarte w programie zostały zestawione w poniższej tabeli.

Komenda	opis	Komenda	opis
mean(A)	Oblicza średnią arytmetyczną z elementów wektora A	histogram(int,A)	Zwraca macierz złożoną z częstości (jak w hist(A)) oraz punktów granicznych przedziałów
gmean(A)	Oblicza średnią geometryczną z elementów wektora A	stdev(A)	Zwraca odchylenie standardowe elementów wektora A
hmean(A)	Średnia harmoniczna elementów wektora A	var(A)	Zwraca wariancję elementów wektora A
kurt(A)	Kurtoza elementów wektora A	corr(A,B)	Zwraca współczynnik korelacji Pearsona (R) dla dwóch wektorów
median(A)	Mediana elementów wektora A		
mode(A)	Moda elementów wektora A		
hist(int,A)	Zwraca wektor częstości z jakimi elementy a wpadają do odpowiednich przedziałów (int=ilość przedziałów)		

Kolejną grupą funkcji, które należy wyróżnić, są funkcje związane z wykonywaniem operacji na gęstościach prawdopodobieństwa. Rozróżniamy cztery grupy tego typu funkcji, które łatwo rozpoznać, dzięki dwuczłonowym nazwom: pierwszej litery definiującej zastosowanie, oraz

nazwy stosowanego rozkładu np.: $dnorm(x,a,b)$ oznacza obliczenie dystrybuanty rozkładu normalnego o zadanych parametrach (średnia i odchylenie standardowe) w punkcie x .

Opis tych czterech grup funkcji przedstawiono poniżej.

- Gęstości prawdopodobieństwa – (litera d) zwraca prawdopodobieństwo tego, że zmienna przyjmie konkretną wartość.
- Dystrybuanta – (litera p) zwraca prawdopodobieństwo, że zmienna losowa będzie miała wartość mniejszą lub równą od zadanej. (klasyczne zastosowanie znanej Państwu dystrybuanty).
- Odwrotna dystrybuanta – (litera q) odwraca działanie powyższej funkcji, tj. Dla danego prawdopodobieństwa p zwraca taką wartość, że dowolna zmienna losowa będzie od niej mniejsza lub równa z prawdopodobieństwem p .
- Generator liczb losowych – (litera r) generuje wektor m losowych wartości o zadanych parametrach rozkładu

Ćwiczenie

Wygeneruj wektor długości 100 opierając się na rozkładzie normalnym oraz chi-kwadrat(chisq). Oblicz podstawowe statystyki opisowe dla tego zestawu danych. Przedstaw otrzymane dane na wykresie oraz narysuj ich histogram składający się z 10 przedziałów.

Data fitting

Oprócz operacji typowo statystycznych przy pomocy Mathcada możemy również dokonywać dopasowania krzywych do danych. Mathcad oferuje nam 2 grupy funkcji – interpolujące oraz liczące regresję.

Interpolacja:

- $linterp(vx,vy,x)$ – łączy punkty liniowymi odcinkami. Wektor vx musi być podany w kolejności rosnącej,
- $lspline(vx,vy)$ – interpolacja przy pomocy splinów kubicznych o liniowych końcach,
- $cspline(vx,vy)$ – spliny kubiczne o kubicznych końcach,
- $pspline(vx,vy)$ – spliny kubiczne o parabolicznych końcach,
- $interp(vs,vx,vy,x)$ – zwraca z interpolowane wartości y dla poszczególnych x . Wektor vs jest wynikiem działania jednej z powyższych funkcji.

Regresja:

Otrzymanie zwykłej liniowej regresji w programie Mathcad, wymaga “zbudowania” równania funkcji liniowej przy pomocy dwóch następujących funkcji:

- **slope(vx,vy)** – zwraca nachylenie prostej(współczynnik a)
- **intercept(vx,vy)** – zwraca punkt przecięcia z osią OY (współczynnik b)

Mając te dwa parametry należy zdefiniować równanie linii regresji.

Użycie komendy **line(vx,vy)** zwraca obydwie parametry za jednym zamachem (uwaga na kolejność!).

Oczywiście możemy również korzystać z dużo bardziej dokładnych modeli regresji:

- **regress(vx,vy,n)** – zwraca współczynniki wielomianu n -tego stopnia najlepiej dopasowującego zadane wartości
- **loess(vx,vy,span)** – lokalna regresja w otoczeniu kolejnych punktów (wielkość otoczenia definiujemy parametrem *span*. Zwraca wektor vs , którego można użyć w przedstawionej wcześniej funkcji *interp*

Zastosowanie innych, niewielomianowych modeli regresji wymaga wcześniejszego zdefiniowania wektora (vg) wartości startowych specyficznych dla wybranego modelu. Wynikiem zastosowania poszczególnych modeli jest trzelementowy wektor zawierający odpowiednie wartości a, b oraz c .

- **expfit(vx,vy,vg)** – krzywa eksponentialna dana wzorem: $a*exp(b*x)+c$
- **lgsfit(vx,vy,vg)** – krzywa hiperboliczna: $a/(1+b*exp(-cx))$
- **logfit(vx,vy,vg)** – krzywa logarytmiczna $a*ln(x+b)+c$
- **pwrfit(vx,vy,vg)** – krzywa potęgowa $a*x^b+c$
- **sinfit(vx,vy,vg)** – sinusoida $a*sin(x+b)+c$

Mathcad pozwala również na zdefiniowanie własnych funkcji regresji. W tym celu należy skorzystać z jednej z poniższych funkcji:

- **linfit(vx,vy,F)** – dla zdefiniowanych, w postaci wektora F , funkcji polecenie generuje współczynniki kombinacji liniowej zadanych funkcji aby otrzymać jak najlepsze dopasowania,
- **genfit(vx,vy,vg,F)** – uogólnienie powyższej procedury o przypadki nieliniowe

Wyglądanie danych

Jeżeli podejrzewamy, lub wiemy, że posiadane przez nas dane mogą być zaszumione (szczególnie te pochodzące z cyfrowych analizatorów, przydatne może okazać się ich wcześniejsze wygładzenie przed przystąpieniem do dalszej analizy.

W Mathcadzie służą do tego trzy następujące funkcje:

- **medsmooth(vy,n)** –running median methods (n must be an odd number lower than number of elements in vy)
- **ksmooth(vx,vy,b)** – uses Gaussian kernel method to calculate the weighted averages for vy elements (use b which is few times bigger than the spacing between vx elements)
- **supsmooth(vx,vy)** – uses linear fitting for k-nearest neighbors (k is automatically taken) and returns the vector of fitted elements.

Polecenie Minerr

Działa bardzo podobnie do polecenia *Find* poznanego już wcześniej. Różnica polega na tym, że polecenie jeśli polecenie *Find* nie znajdzie dokładnego rozwiązania, wyświetli komunikat błędu, natomiast funkcja *Minerr* znajdzie rozwiązanie będące najbliższym poszukiwanego rozwiązania/zadanego zbioru wartości (minimalizacja błędu). Składnia polecenia jest identyczna jak polecenia *Find*, czyli dla przypomnienia wygląda następująco

Given

Określone warunki

$Var := Minerr(x,y,z,....)$

x, y, z... oznacza parametry który należy wyznaczyć w celu minimalizacji błędu (przykład w zadaniu 3)

ĆWICZENIE

1. W tabeli zestawiono pomiar pewnej wielkości w zależności od parametru początkowe „c” dla pięciu różnych temperatur. Dla każdej temperatury, wyznaczyć prosta regresji liniowej $y=1/q_{voc}$ $x=1/c$ Zdecyduj czy dopasowanie jest dobre (na podstawie współczynnika korelacji R). Dla wybranej pary wartości c i q narysować wykres przedstawiający punkty pomiarowe (oznaczone + o wielkości „3”) oraz przebieg prostej regresji (linia ciągła, grubość 2).

C[ppm]	Temp [°C]	q_{voc}				
		33.6	41.5	57.4	76.4	99
200		0,080	0,069	0,052	0,042	0,027
500		0,093	0,083	0,072	0,056	0,042
1000		0,101	0,088	0,076	0,063	0,045
2000		0,105	0,092	0,083	0,068	0,052
3000		0,112	0,102	0,087	0,072	0,058

2. W pliku “Exercie_03.xlsm” zawarty jest zestaw danych dotyczących procesu zgazowania. Przenieś dane do Mathcada i na ich podstaw przedstaw kinetykę procesu zgazowania (krzywe kinetyczne dla poszczególnych produktów). W tym celu:
 - a. Dokonaj wygładzenia otrzymanych danych,
 - b. Przedstawić wykres dV/dt dla każdego składnika – w tym celu należy dokonać operacji różniczkowania poszczególnych wektorów wygładzonych danych.Wzory na różniczkowanie przedstawiono poniżej:

$$dX = \frac{X_n + X_{n+1}}{2} \quad dY = \frac{Y_{n+1} - Y_n}{X_{n+1} - X_n}$$

3. Dla podanych w tabeli wartości zmiennych x i y, korzystając z funkcji Minerr znaleźć parametry α oraz β , dające najlepsze dopasowanie rozkładu Weibulla danego funkcją:

$$F(x, \alpha, \beta) = \alpha * \beta * x^{\beta-1} \exp(-\alpha * x^{\beta})$$

x	,132	,322	,511	,701	,891	1,081	1,27	1,46	1,65	1,839	2,029	2,219
y	,1	,258	,543	,506	,606	,622	,569	,453	,438	,316	,29	,195