



LUND UNIVERSITY

DATPAC programpaket för mätdataanalys

Jensen, Lars

1976

Document Version:
Förlagets slutgiltiga version

[Link to publication](#)

Citation for published version (APA):
Jensen, L. (1976). *DATPAC programpaket för mätdataanalys*. (Research Reports TFRT-3141a). Department of Automatic Control, Lund Institute of Technology (LTH).

Total number of authors:
1

General rights

Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

DATPAC PROGRAMPAKET FÖR MÄTDATAANALYS

Lars Jensen

Denna rapport avser anslag 750212-0 från Statens råd för byggnadsforskning till institutionen för Byggnadskonstruktionslära vid Lunds Tekniska Högskola, Lund.

INNEHALLSFÖRTECKNING

1	Inledning	1
2	Skillnader DATPAC-IDPAC	2
3.1	Matematiska funktioner	3
3.2	Polynomfunktion	4
3.3	Linjär transformation	4
3.4	Linjär transformation	4
3.5	Begränsningsfunktioner	4
4.1	Sortering	9
4.2	Frekvensfunktion - normaliserad summa	9
4.3	Fördelningsfunktion - kumulativ summa	9
4.4	Normalisering - normalfördelningskurva	9
5.1	Statistik	11
5.2	Delstatistik	11
5.3	Periodisk statistik	12

1 Inledning

DATPAC är ett interaktivt programpaket för analys av stora datamängder. Programpaketet innehåller alla de enklare kommandona i IDPAC. Utökningen består av att följande typer av beräkningar kan utföras

- matematiska funktioner
- polynom- och segmentfunktioner
- begränsningsfunktioner
- periodisk och delstatistik
- sortering
- frekvens och fördelningsfunktioner

I de följande avsnitten kommer de olika kommandona att behandlas, men först redogörs för skillnaderna gentemot IDPAC.

Det förutsätts att IDPAC är bekant för läsaren. Rapporten beskriver endast de kommandon som är unika för DATPAC. Programpaketet IDPAC finns dokumenterat i: Wieslander, J, IDPAC Users guide revision I. Report 7605, 1976, Department of Automatic Control, Lund Institute of Technology.

2 Skillnader DATPAC - IDPAC

För alla kommandon som beskrivs här så gäller det att om inga kolonner anges så avses samtliga. Detta görs för att de flesta beräkningarna som sker ofta avser samtliga kolonner, vilket underlättar arbetet för användaren.

För att kunna klara av stora datamängder med ett stort antal kolonner så har antalet utökats till 25. Detta gäller också ett flertal IDPAC kommandon. Vidare har antalet kolonner i konverteringskommandot CONV utökats till 100 vid inläsning. En annan skillnad är att samtliga datafiler som står som vänsterargument ej får anges med kolonner. Denna inskränkning har gjorts för att förenkla programmeringen och för att förhindra att för många buffertar utnyttjas än vad som är tillgängligt.

De IDPAC kommandon som ej finns med i DATPAC är:

PLLOG
ACOF
ASPEC
CCOF
CSPEC
ML
SPTRF
RANPA
RESID

Dessa resterande kommandon kan införas i DATPAC, varvid IDPAC skulle bli ett subset av DATPAC. Skälet till att alla IDPAC kommandon ej är med, är den långa körtiden med CHAIN och den låga utnyttjningsgraden för dessa kommandon.

Hakparenteser [] i en kommandorad anger att det som står inuti kan utelämnas. [()] anger att kolonner ej behöver anges. Observera att då antas att samtliga kolonner skall ingå i beräkningen.

3.1 Matematiska funktioner

Ett flertal matematiska funktioner kan anropas. Kommandot är följande

```
FUNC SUB FNEW ← FOLD [( )]
```

Här betecknar SUB en av följande funktioner:

kommando	funktion
ABS	absolutbelopp
CUBE	kub
EXP	exponent
INV	invers
NLOG	naturliga logaritmen
QUAD	kvadrat
SQRT	kvadratroten
TLOG	tiologaritmen

För att undvika onödiga avbrott så tas absolutbelopp före funktionsberäkning med NLOG, SQRT och TLOG.

Ändringar i en datafil kan erhållas med kommandot

```
FUNC DIFF FNEW ← FOLD [( )]
```

Beräkningen sker på följande sätt:

$$\begin{aligned} &(\text{FOLD}(0, C(J)) = 0.0, J=1, NC) \\ &((\text{FNEW}(I, J) = \text{FOLD}(I, C(J)) - \text{FOLD}(I-1, C(J)), J=1, NC) \quad I=1, NS) \end{aligned}$$

Första tidpunkten är oförändrad.

Potensfunktioner

Kommandot är

```
POWER FNEW ← FOLD [( )]  FDAT [( )]
```

Beräkningen kan beskrivas med följande:

$$FNEW(I,J) = FOLD(I,K_J) * * FDAT(1,L_J)$$

Första raden anger de individuella potenserna.

Exponentialfunktioner

Generella exponentialfunktioner kan beräknas med följande kommando

$$GEXP \quad FNEW \leftarrow FOLD \left[\left(\right) \right] \quad FDAT \left[\left(\right) \right]$$

Beräkningen kan beskrivas med följande:

$$FNEW(I,J) = FDAT(1,L_J) * * FOLD(I,K_J)$$

3.2 Polynomfunktioner

Kommandot är

$$LIN \quad POL \quad FNEW \leftarrow FOLD \left[\left(\right) \right] \quad FDAT \left[\left(\right) \right]$$

Polynomfunktionen kan beskrivas på följande sätt

$$\begin{aligned} y &= FNEW(I,J) \\ u &= FOLD(I,K_J) \\ a_{I-1} &= FDAT(I,L_J) \end{aligned}$$

$$y = a_0 + u \cdot (a_1 + u \cdot (a_2 + \dots))$$

Första raden i filen FDAT anger faktorn för den nollte potensen. Den andra raden för den linjära termen. Gradtalet anges av antalet rader minus ett i filen FDAT. Antalet rader i FDAT är begränsat till 100.

3.3 Segmentfunktioner

Ofta kan en funktion ej anges analytiskt utan anges istället som en tabell eller som segment. Kommandot är

$$LIN \quad SEG \quad FNEW \leftarrow FOLD \left[\left(\right) \right] \quad FDAT \left[\left(\right) \right]$$

Om antalet segment är n st så är antalet rader i FDAT $2n+2$. De $n+1$ första raderna anger segmentsfunktionens x - eller ingångskordinater och de $n+1$ sista motsvarande y - eller utgångskordinater. Beräkningen sker med linjär interpolation. Ingångsdata utanför intervallet begränsas till intervallet. Maximala antalet segment är 49 st.

3.4 Linjär transformation

Detta kan ske med följande kommandoord:

```
TRANS FNEW ← FOLD [( )] FDAT [( )]
```

Beräkning sker på följande sätt för kolonn J och rad I

$$FNEW(I,J) = (FDAT(J,K_N) * FOLD(I,L_N), N=1,NC)$$

3.5 Begränsningsfunktioner

För att kunna rätta till enstaka felaktiga mätpunkter automatiskt och begränsa data på olika sätt så finns kommandot

```
LIMIT SUB FNEW ← FOLD [( )] FDAT [( )]
```

Här är SUB något av subkommandona DIFF, NORM eller SPEC. Datafilen FDAT innehåller data som anger hur beräkningarna skall ske.

Subkommando NORM innebär att varje datavektor begränsas till för varje datavektor individuellt intervall. Den första raden i FDAT anger undre gränsen och den andra anger övre gränsen. Observera att antalet kolonner som anges i FOLD och FDAT måste vara lika, men samma kolonn kan anges ett flertal gånger för att operera på olika kolonner i FOLD.

Exempel

Begränsa datafilen TEST med tre kolonner till intervallet

```
(0.0,100.0)
```

```
EDIT FDAT
```


INPUT

0 100

↓

>E

CONV FDAT FDAT(1) 1

LIMIT NORM TEST ←TEST FDAT(1 1 1)

När en datapunkt ligger utanför intervallet så kan datapunkter tilldelas ett annat värde än begränsningsvärdet. Detta kan ske med subkommandot SPEC. Den tredje raden anger då det värde som skall ersätta värden mindre än den undre gränsen och den fjärde raden används för värden större än den övre gränsen.

För att kunna åtgärda enstaka fel så kan subkommandot DIFF användas. Linjär interpolation används för att rätta felaktiga data. Testet tillgår så att mätpunkten före och efter den mätpunkt som skall testas, bildar ett intervall som utökas med det värde, som anges i första raden i FDAT och för motsvarande kolonn. Observera att FDAT kan mycket väl innehålla endast en kolonn som används för test av samtliga kolonner i FOLD.

Exempel

Antag att mätdatafilen TEST med tre kolonner skall begränsas till intervallet (0.0,20.0), därefter linjäriseras med ett tredjegradspolynom $y = 3+x - x^2 - 2x^3$ och slutligen skall avvikelser större än 0.5 rättas till.

```

EDIT LDAT
INPUT
0 20
↓
>E
CONV LDAT ← LDAT (1) 1
LIMIT NORM TEST ← TEST LDAT(1 1 1)
EDIT PDAT
INPUT
3 1 -1 -2
↓
>E
CONV PDAT ← PDAT(1) 1
LIN POL TEST ← TEST PDAT(1 1 1)
EDIT PDAT
INPUT
0.5
↓
>E
CONV DDAT ← DDAT(1) 1
LIMIT DIFF TEST ← TEST DDAT(1 1 1)

```

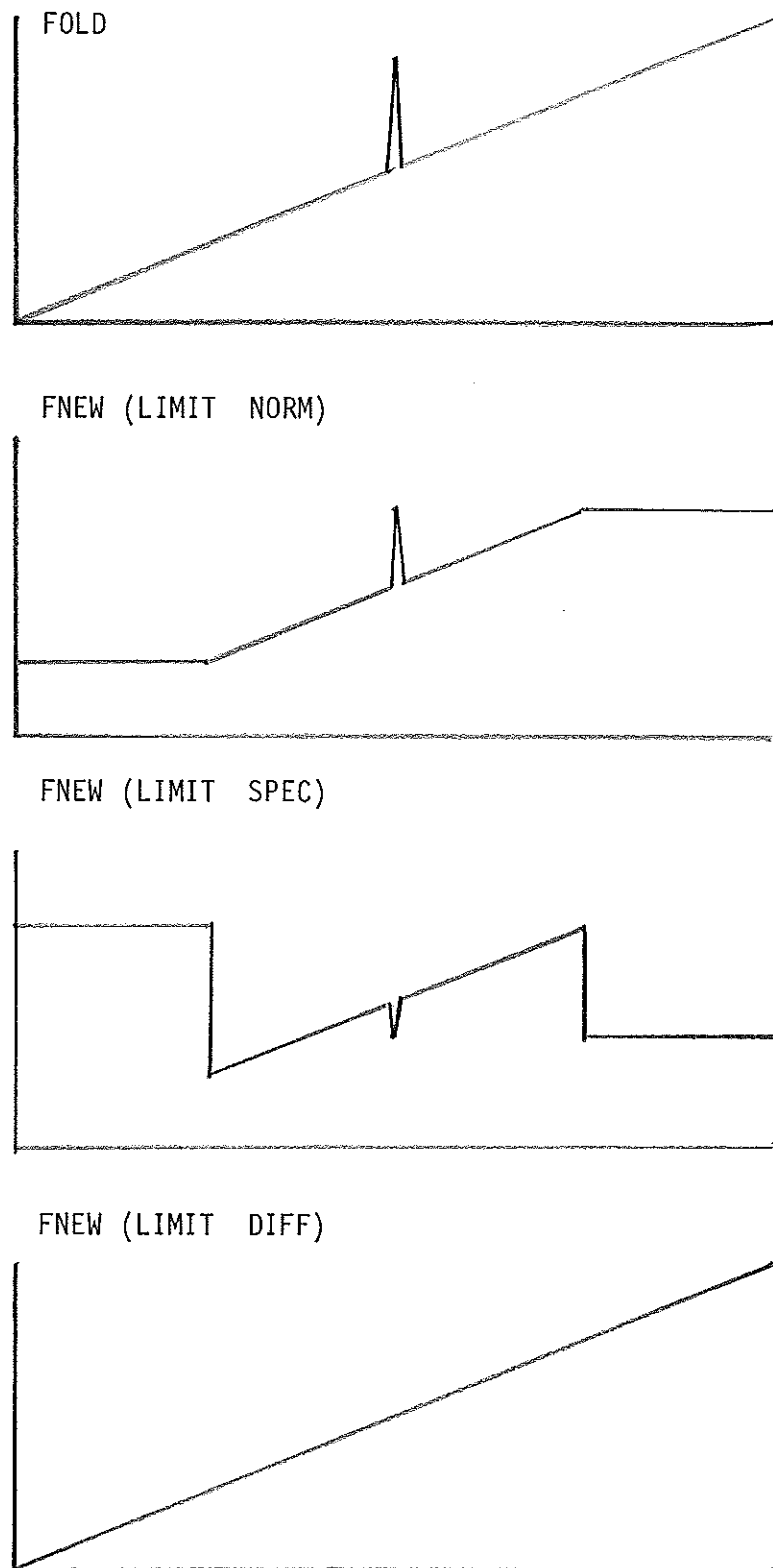


FIG.3.1 Exempel på olika LIMIT subkommandon, Överst ingångsdata, därefter resultatet efter subkommandona NORM SPEC och DIFF

4.1 Sortering i klasser

För att undersöka hur en variabel har varierat eller för att beräkna dess frekvens- eller fördelningsfunktion så kan en datafil sorteras i klasser. Detta sker med kommandot

```
SORT FNEW ← FOLD [( )] NX XMIN DX
```

Här är

```
NX    antalet klasser
XMIN  undre klassgränsen
DX    klassbredden
```

Observera att klassindelningen gäller för samtliga kolonner. Alla data som ligger utanför intervallet (XMIN,XMIN+NX·DX) ingår ej.

4.2 Frekvensfunktion - normaliserad summa

När en datafil har sorterats i klasser så kan frekvensfunktionen erhållas genom att normalisera summan till ett. Detta sker med kommandot

```
FUNC NSUM FNEW ← FOLD [( )]
```

4.3 Fördelningsfunktion - kumulativ summa

När en frekvensfunktion har erhållits så kan motsvarande fördelningsfunktion beräknas med kommandot

```
FUNC CSUM FNEW ← FOLD [( )]
```

4.4 Normalisering - normalfördelningskurva

En datafil kan normaliseras så att dess medelvärde och standardavvikelse blir 0.0 resp 1.0. Detta för att en jämförelse med normalfördelningsfunktionen skall kunna ske. Kommandot är

```
FUNC NORM FNEW ← FOLD [ ( ) ]
```

Normalfördelningskurvan kan beräknas med följande kommando om medelvärdet och standardavvikelsen är kända.

```
NORM FX ← MEAN STD NX XMIN XMAX
```

Här är

```
MEAN  medelvärde
STD   standardavvikelse
NX    antal punkter
XMIN  undre gräns
XMAX  övre gräns
```

Den erhållna frekvensfunktionen har summan 1.0 dvs svansarna har försumrats.

Exempel

Normalisera datafilen TEST, beräkna dess frekvensfunktion och jämför den med beräknade normalfördelningskurvan. Datapunkterna i TEST ligger i intervallet (0.0,10.0).

```
FUNC NORM NTEST ← TEST
SORT FTEST ← NTEST 200 -5.0 0.05
FUNC NSUM FTEST ← FTEST
NORM NF ← 0.0 1.0 200 -5.0 5.0
PLOT FTEST NF
```

5.1 Statistik

För en stor datamängd, som kan vara uppdelad i ett flertal datafiler, kan det vara av intresse att beräkna statistiska data för ett flertal datafiler, som om de vore en enda datafil. Kommandot är

```
STAT SNEW ← FOLD [( )] [SOLD [( )]]
```

Här anger SNEW och SOLD datafiler som innehåller statistiska data på följande sätt:

rad nr	innehåll
1	summa
2	medelvärde
3	kvadratsumma
4	varians
5	minimum
6	maximum
7	standardavvikelse
8	antal data
9	sampel för rad 5
10	sampel för rad 6
11	största ändring
12	minsta ändring
13	sampel för rad 11
14	sampel för rad 12

Statistiska data från tidigare beräkningar överförs med datafilen SOLD. Vid beräkning på den första datafilen så ingår SOLD ej i kommandoraden.

5.2 Delstatistik

För datamängder som har periodiska variationer, som är i fas med samplingar, så kan det vara av intresse att jämföra de olika perioderna genom att beräkna största, minsta, medelvärde eller summan för varje period. Kommandot är följande

```
PART SUB FNEW ← FOLD [ ( ) ] NP
```

Här betecknar SUB någon av följande subkommandon.

```
MAX    ger största värde
MEAN   ger medelvärde
MIN    ger minsta värde
SUM    ger summa
```

Periodens längd som skall analyseras anges med NP. Antalet tidpunkter i den nya filen och den gamla är NNEW resp NOLD då gäller att

$$NNEW = NOLD/NP$$

5.3 Periodisk statistik

I samband med undersökningar av periodiska förlopp kan det vara av intresse att känna till medelsvängningen eller det största eller minsta värdet för en viss tidpunkt i perioden. Observera att det förutsätts att svängningen är i fas med samplingen. Kommandot är:

```
PER SUB FNEW ← FOLD [ ( ) ] NP
```

Här betecknar SUB samma subkommandon som för kommandot PART. Parametern NP betecknar periodens längd och antalet tidpunkter i den nya filen.

Exempel

Antag att man för datafilen TEST önskar plotta medelsvängningen och den största avvikelser från denna (både över och under). Periodens längd är 25 sampel.

```
PER MAX M1 ← TEST 25
PER MEAN M2 ← TEST 25
PER MIN M3 ← TEST 25
PLOT 25 M1 M2 M3
```