"DATALOOK", A SERIES OF MACROS FOR DESCRIPTIVE STATISTICS

William R. Gjertsen
Lipid Research Clinics Program, Department of Biostatistics
University of North Carolina, Chapel Hill

INTRODUCTION

DATALOOK is a multipurpose series of SAS macros which produces useful descriptive information on samples from univariate continuous variables or discrete variables with a large number of values.

The design for DATALOOK has as its criteria:

1) a listing of both moment and percentile information by a by-variable list (BYVAR)
2) an indication of possible farout values

To accomplish i) - iii) in a typical run we would perform the following operations. Create a data set which outputs selected moments by a by-variable list (BYVAR) and another which outputs selected percentiles by BYVAR. Merge by BYVAR and print the resulting data set with PROC PRINT. Next, put ID information and/or counts of farout values with PUT STATEMENTS. Finally produce overlay plots of mean, median and interquartile range with PROC SCATTER. DATALOOK has built-in flexibility by having the user specify the options he wants for each PROC called within DATALOOK.

Following Tukey [1], we will define farout hi (1o) values as those values which are greater than (less than) two interquartile ranges above (below) the 75th (25th) percentile. The definition of what a user wants to specify as farout may change with different applications and is thus left to be defined within DATALOOK.

PROC PCTL

PROC PCTL is a user written SAS procedure which outputs whole number percentiles with default percentiles corresponding to p=.01, .05, .10, .25, .50, .75, .90, .95, and .99. The percentiles are those one would obtain directly from an empirical distribution of the sample without any interpolation between ordered sample values.

EXAMPLES

Two examples of DATALOOK follow. Both are based on data for response variables Y1 and Y2 broken down by agegroup A (A=25,30,35,40), clinic SITE (SITE=1,2,3) and treatment T (T=1,2). Y1 is continuous and somewhat normal looking, Y2 has integer values and is highly skewed. Although the data is based on real data from several clinics within the Lipid Research Clinics Program, Y1 and Y2 have been randomly altered and disguised and do not represent real data as such.

In the first example we give a DATALOOK description of Y1 on just one plot using the technique of horizontal expansion (see MACRO XRBFNBR and the resulting plot).

In the second example we give a goodness-of-fit description for a linear model by plotting residuals vs. group ranked predicted values. Only output for farout counts and the residual plot for Y2 are shown. The plot (conical L-U band, means always higher than medians) and the farout counts (38 hi, 0 lo) indicate that the

residuals from the Y2 model are highly skewed and suggest that a log transformation on Y2 may help. Similar residual plots of residuals vs. grouped ranked regressors can be obtained by forcing all the non-class regressors to be on the output data set with an ID statement within PROC GLM.

OTHER APPLICATIONS

In a quality control situation such as monitoring normal ranges for blood chemistry, it may be more appropriate to classify values as farout if they have some clinical meaning. For instance, if a batch of values was constantly being updated, a new value might be considered physiologically farout if it was below the pth or above the qth percentile based on the previous sample values. In this case, farout values would be redefined appropriately and in addition to the L-U interval which contains half the data two or more plotting symbols say, P and Q, could appear in the overlay plots. Furthermore, these plots over time on split samples taken at different sites can indicate whether a particular site is going out of control.

For other time series applications DATALOOK might prove useful. For instance, a stock investor might decide to sell (buy) a stock if its price exceeds an upper farout limit (drops below a lower farout limit). His BYVAR list might include TIME in months and the STOCKS in his portfolio.

ACKNOWLEDGEMENTS

Special thanks go to Frank Harrell for writing PROC PCTL and imparting some clever ideas and to Jane Waddell and Ann Rutledge for their suggestions and improvements to DATALOOK. This work was supported by contract NIH-NHLI-71-2243.

REFERENCES

APPENDIX

EXAMPLE 1: DATALOOK DESCRIPTION OF Y1

DATA ONE; SET ON2.STOC1265;

IF A=201|A=451|SITE=0 THEN DELETE;

MACRO DAT ONE %
MACRO YYVAR Y1 %
MACRO ID ID %
MACRO BYVAR SITE T A %
MACRO B YVAR BY %
MACRO PIRSTBY %
MACRO BYF %
MACRO BYPRINT SITE= T = A= %
MACRO SORTIT
PROC SORT DATA=DAT;
   BY BYVAR %
   * SORTS DATA SET DAT BY BYVAR LIST
   
PROC NMOOPTS
   MEAN STD STDEEP MIN MAX %
   * SPECIFY INPUT MEANS STATISTICS
   
PROC NMOOPTS OUTPUT OUT=MNS MEAN=MEAN STD=STD STDEEP=STDEEP MIN=MIN MAX=MAX %
   VAR YYVAR; BY BYVAR;
   
PROC NMOOPTS
   MNLOPT %
   * PRODUCE STATS FROM PROC MEANS & OUTPUT ACCORDING TO SPSS IN MNLOPT
   
PROC BYVAR
   Y101=PCT01 Y105=PCT05 Y110=PCT10 Y112=PCT25 Y150=PCT50
   Y175=PCT75 Y190=PCT90 Y195=PCT95 Y199=PCT99 %
   
PROC PCTOPTS DATA=DAT OUT=PCTS (RENAME=(RENAME))
   MIN=4 MAX=7 DIV=.01 N=N %
   * SPECIFY PARAMETERS FOR PROC PCTL
   
PROC PCTL
   PROC PCTL PCTOPTS; VAR YYVAR; BY BYVAR %
   * OUTPUTS PERCENTILES TO DS PCTS
   
PROC PFORMAT VAR YYVAR MEAN STD MIN MAX PCT01 PCT05 PCT10 PCT25 PCT50
   PCT75 PCT90 PCT95 PCT99 %
   * SPECIFY PRINT FORMAT
   
PROC PRINTIT
   DATA PRINTS; MERGE MNS PCTS; BY BYVAR;
   PROC DELETE DATA= MNS PCTS;
   
PROC PRINT DATA=PRINTS; ID N; PFORMAT;
   FORMAT PCT01 PCT05 PCT10 PCT25 PCT50 PCT75 PCT90 PCT95 PCT99 4.2;
   
TITLE SIMPLE STATISTICS FOR VARIABLE YYVAR;
   * MERGE MNS PCTS BY YYVAR & PRINT RESULTING DS
   
MACRO PARGRTS
   DATA _NULL_; MERGE DAT PRINTS; BY YYVAR; FILE PRINT;
   TITLE PARGRTS VALUES;
   IF YYVAR>PCT75-2*(PCT75=PCT25) THEN PUT / *PARGRTS HI VALUE* ID= YYVAR= BYPRINT;
   IF YYVAR<PCT25-2*(PCT75=PCT25) THEN PUT / *PARGRTS LO VALUE* ID= YYVAR= BYPRINT;
   
MACRO XDEFINE
   XVAR=**(SITE=1)*T=1 (A+20)*SITE=1*THE=1) + ((A+100)*SITE=2*THE=2) + (A+100)*THE=3*SITE=3 + (A+100)*THE=3*THE=2 %
   * DEFINE THE X VARIABLE FOR THE PLOTS
   
MACRO XVAR AGEGROUP %
   * NAME THE X VARIABLE
   
MACRO SCATOPT %
   * SPECIFY ANY PROC SCATTER PARAMETERS
   
PROC SCATOPT DATA=PCTS; ID A; IDDEFINE; KEEP XVAR FIRSTBY MEAN PCT25 PCT50 PCT75;
   PROC DELETE DATA=PRENTS;
   PROC SCATTER SCATOPT; BY FIRSTBY;
   PLOT XVAR=MEAN * XVAR=PCT75 * U; XVAR=PCT25-XVAR=PCT25 * L; / OVERLAY;
   TITLE DATA LOOK PLOTS OF XVAR VS XVAR BY FIRSTBY;
   % Produces OVERLAY 'DATA LOOK' PLOTS BY FIRSTBY;
   
MACRO DATALOG SORTIT; MEANIT; PCTIT; PRINTIT; PARGRTS; PLOTIT; % * DO IT ALL
   PAGE;
   DATALOG;

122
### SIMPLE STATISTICS FOR VARIABLE Y5

<table>
<thead>
<tr>
<th>N</th>
<th>SITE</th>
<th>F</th>
<th>MEAN</th>
<th>STD</th>
<th>MIN</th>
<th>MAX</th>
<th>PCT05</th>
<th>PCT10</th>
<th>PCT25</th>
<th>PCT50</th>
<th>PCT75</th>
<th>PCT90</th>
<th>PCT95</th>
<th>PCT99</th>
</tr>
</thead>
<tbody>
<tr>
<td>145</td>
<td>1</td>
<td>1</td>
<td>25</td>
<td>5.0804</td>
<td>0.170730</td>
<td>0.01414784</td>
<td>4.73230</td>
<td>4.66</td>
<td>4.68</td>
<td>5.07</td>
<td>5.24</td>
<td>5.35</td>
<td>5.43</td>
<td>5.59</td>
</tr>
<tr>
<td>109</td>
<td>1</td>
<td>1</td>
<td>30</td>
<td>5.10972</td>
<td>0.200730</td>
<td>0.0192238</td>
<td>4.65949</td>
<td>5.63335</td>
<td>4.07</td>
<td>4.09</td>
<td>4.96</td>
<td>5.32</td>
<td>5.47</td>
<td>5.67</td>
</tr>
<tr>
<td>78</td>
<td>1</td>
<td>2</td>
<td>35</td>
<td>5.149275</td>
<td>0.176825</td>
<td>0.0197288</td>
<td>4.67175</td>
<td>5.82289</td>
<td>4.88</td>
<td>4.94</td>
<td>4.99</td>
<td>5.36</td>
<td>5.47</td>
<td>5.66</td>
</tr>
<tr>
<td>94</td>
<td>1</td>
<td>3</td>
<td>40</td>
<td>5.21530</td>
<td>0.152131</td>
<td>0.15707</td>
<td>5.97520</td>
<td>4.07</td>
<td>4.06</td>
<td>5.05</td>
<td>5.33</td>
<td>5.42</td>
<td>5.47</td>
<td>5.50</td>
</tr>
<tr>
<td>196</td>
<td>1</td>
<td>2</td>
<td>25</td>
<td>5.14799</td>
<td>0.130934</td>
<td>0.0129238</td>
<td>4.57739</td>
<td>5.66016</td>
<td>4.60</td>
<td>4.68</td>
<td>5.09</td>
<td>5.63</td>
<td>5.72</td>
<td>5.80</td>
</tr>
<tr>
<td>85</td>
<td>1</td>
<td>2</td>
<td>30</td>
<td>5.16848</td>
<td>0.175099</td>
<td>0.0139922</td>
<td>4.73022</td>
<td>5.95954</td>
<td>4.78</td>
<td>4.90</td>
<td>4.96</td>
<td>5.18</td>
<td>5.29</td>
<td>5.41</td>
</tr>
<tr>
<td>47</td>
<td>1</td>
<td>1</td>
<td>35</td>
<td>5.16665</td>
<td>0.171655</td>
<td>0.0121049</td>
<td>4.86328</td>
<td>5.29</td>
<td>5.38</td>
<td>5.60</td>
<td>5.80</td>
<td>5.86</td>
<td>5.90</td>
<td>5.94</td>
</tr>
<tr>
<td>44</td>
<td>1</td>
<td>2</td>
<td>40</td>
<td>5.273117</td>
<td>0.159946</td>
<td>0.0240449</td>
<td>4.97998</td>
<td>5.64087</td>
<td>4.98</td>
<td>5.01</td>
<td>5.07</td>
<td>5.12</td>
<td>5.23</td>
<td>5.38</td>
</tr>
<tr>
<td>154</td>
<td>1</td>
<td>1</td>
<td>25</td>
<td>5.12924</td>
<td>0.172968</td>
<td>0.039982</td>
<td>4.70959</td>
<td>5.96614</td>
<td>4.75</td>
<td>4.86</td>
<td>4.96</td>
<td>5.08</td>
<td>5.20</td>
<td>5.38</td>
</tr>
<tr>
<td>94</td>
<td>1</td>
<td>1</td>
<td>30</td>
<td>5.16853</td>
<td>0.168253</td>
<td>0.0239324</td>
<td>4.71293</td>
<td>5.75244</td>
<td>4.75</td>
<td>4.93</td>
<td>5.08</td>
<td>5.33</td>
<td>5.68</td>
<td>5.40</td>
</tr>
<tr>
<td>79</td>
<td>1</td>
<td>1</td>
<td>35</td>
<td>5.14263</td>
<td>0.189563</td>
<td>0.0219993</td>
<td>4.96143</td>
<td>5.64764</td>
<td>4.93</td>
<td>4.95</td>
<td>5.07</td>
<td>5.28</td>
<td>5.31</td>
<td>5.42</td>
</tr>
<tr>
<td>75</td>
<td>2</td>
<td>1</td>
<td>40</td>
<td>5.20748</td>
<td>0.159105</td>
<td>0.0383728</td>
<td>4.73315</td>
<td>5.99885</td>
<td>4.73</td>
<td>4.96</td>
<td>5.03</td>
<td>5.09</td>
<td>5.20</td>
<td>5.33</td>
</tr>
<tr>
<td>263</td>
<td>1</td>
<td>1</td>
<td>25</td>
<td>5.13574</td>
<td>0.139780</td>
<td>0.0102499</td>
<td>4.77685</td>
<td>5.65799</td>
<td>4.79</td>
<td>4.93</td>
<td>4.98</td>
<td>5.10</td>
<td>5.29</td>
<td>5.39</td>
</tr>
<tr>
<td>87</td>
<td>2</td>
<td>3</td>
<td>30</td>
<td>5.21042</td>
<td>0.172966</td>
<td>0.0385422</td>
<td>4.59921</td>
<td>5.67510</td>
<td>4.60</td>
<td>4.95</td>
<td>5.03</td>
<td>5.19</td>
<td>5.29</td>
<td>5.44</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
<td>2</td>
<td>25</td>
<td>5.19239</td>
<td>0.17649</td>
<td>0.0200012</td>
<td>4.89846</td>
<td>5.00832</td>
<td>4.90</td>
<td>4.92</td>
<td>4.97</td>
<td>5.00</td>
<td>5.23</td>
<td>5.33</td>
</tr>
<tr>
<td>72</td>
<td>2</td>
<td>4</td>
<td>40</td>
<td>5.26698</td>
<td>0.164450</td>
<td>0.0193806</td>
<td>4.97095</td>
<td>5.77229</td>
<td>4.97</td>
<td>5.04</td>
<td>5.05</td>
<td>5.13</td>
<td>5.22</td>
<td>5.34</td>
</tr>
<tr>
<td>79</td>
<td>3</td>
<td>1</td>
<td>25</td>
<td>5.10074</td>
<td>0.264677</td>
<td>0.0185466</td>
<td>4.71925</td>
<td>5.93760</td>
<td>4.72</td>
<td>4.88</td>
<td>4.95</td>
<td>5.00</td>
<td>5.20</td>
<td>5.33</td>
</tr>
<tr>
<td>85</td>
<td>2</td>
<td>1</td>
<td>30</td>
<td>5.11361</td>
<td>0.158032</td>
<td>0.0286667</td>
<td>4.70239</td>
<td>5.07532</td>
<td>4.70</td>
<td>4.93</td>
<td>4.98</td>
<td>5.09</td>
<td>5.24</td>
<td>5.38</td>
</tr>
<tr>
<td>122</td>
<td>3</td>
<td>1</td>
<td>40</td>
<td>5.22744</td>
<td>0.169987</td>
<td>0.0150621</td>
<td>4.71526</td>
<td>5.73315</td>
<td>4.79</td>
<td>4.95</td>
<td>5.03</td>
<td>5.13</td>
<td>5.23</td>
<td>5.33</td>
</tr>
<tr>
<td>128</td>
<td>2</td>
<td>2</td>
<td>35</td>
<td>5.17936</td>
<td>0.245798</td>
<td>0.0140749</td>
<td>4.74565</td>
<td>5.29728</td>
<td>4.80</td>
<td>4.98</td>
<td>5.10</td>
<td>5.19</td>
<td>5.29</td>
<td>5.39</td>
</tr>
<tr>
<td>51</td>
<td>3</td>
<td>2</td>
<td>40</td>
<td>5.28443</td>
<td>0.175022</td>
<td>0.0245080</td>
<td>4.84912</td>
<td>5.68115</td>
<td>4.65</td>
<td>4.90</td>
<td>5.05</td>
<td>5.24</td>
<td>5.26</td>
<td>5.36</td>
</tr>
</tbody>
</table>

---

### FAROUT VALUES

- **FAROUT HI VALUE ID-2132 Y1=5.746582 SITE=1 T1=1 A=25**
- **FAROUT LO VALUE ID-3202 Y1=4.577883 SITE=1 T2=2 A=25**
- **FAROUT HI VALUE ID=2813 Y1=5.766113 SITE=2 T1=1 A=25**
- **FAROUT HI VALUE ID=2374 Y1=5.752248 SITE=2 T2=3 A=30**
- **FAROUT HI VALUE ID=2778 Y1=5.682129 SITE=2 T2=3 A=30**
- **FAROUT LO VALUE ID=2633 Y1=5.959225 SITE=2 T2=3 A=30**
- **FAROUT LO VALUE ID=3026 Y1=4.610532 SITE=3 T2=3 A=35**
- **FAROUT LO VALUE ID=2477 Y1=4.543709 SITE=3 T2=3 A=25**
DATA ONE; SET OM2.STO;12S5;
PROC GLM; CLASS SITE T;
MODEL Y1 Y2 = T SITE T*SITE A*SITF A*SITE;
OUTPUT OUT=NEW
PREDICTED=Y1HAT Y2HAT
RESIDUAL=Y1RES Y2RES;
DATA NEW1(KEEP=ID Y1HAT Y1RES) NEW2 (KEEP=ID Y2HAT Y2RES); SET NEW;
PROC RANK DATA=NEW1 OUT=NEW1 GROUPS=10; VAR Y1HAT; RANKS R1;
PROC RANK DATA=NEW2 OUT=NEW2 GROUPS=10; VAR Y2HAT; RANKS R2;
PROC DELETE DATA= NEW1 NEW2;
MACRO DAT NW1 %
MACRO YVAR Y1ER %
MACRO XVAR RHT %
MACRO YYVAR RHT % * THE GROUP RANKED PREDICTED VALUE IS BOTH
BY VARIABLE YYVAR & THE ABSCISSA XVAR;
MACRO SORTIT
PROC SORT DATA=DAT;
BY BYVAR %
MACRO MNINOPTS
MEAN STD STDERR MIN MAX %
MACRO MEANOPT OUT PUT
OUT=MNS MEAN=MEAN STD=STD
STDERR=STD MIN=MIN MAX=MAX %
MACRO MN XVARIT PROC MEANS NOPRINT MNINOPTS;
VAR YYVAR; BY BYVAR; MEANOPT %
Y1ER75=PCT75 Y1ER90=PCT90 Y1ER95=PCT95 Y1ER99=PCT99 %
MACRO PCTOPTS DATA=DAT OUT=PCTS (RENAME=(PCTOPTS))
MIN=-3 MAX=3 DIV=.01 N=K %
MACRO PCGIT
PROC PCTL PCTOPTS; VAR YYVAR; BY BYVAR; %
MACRO PFORMAT YYVAR N MEAN STD STR MIN MAX PCT01 PCT05 PCT10 PCT25 PCT50 PCT75 PCT90 PCT95 PCT99 %
MACRO PRINTIT
DATA PRINTS; MEAN N MISS PCTS; BY BYVAR; %
PROC DELETE DATA= MNS PCTS;
PROC PRINT DATA=PRINTS; ID BYVAR; PFORMAT; %
Y1ER75=PCT75 Y1ER90=PCT90 Y1ER95=PCT95 Y1ER99=PCT99 %
MACRO PCTOPTS DATA=DAT OUT=PCTS (RENAME=(PCTOPTS))
MIN=-3 MAX=3 DIV=.01 N=K %
MACRO PCGIT
PROC PCTL PCTOPTS; VAR YYVAR; BY BYVAR; %
MACRO PFORMAT YYVAR N MEAN STD STR MIN MAX PCT01 PCT05 PCT10 PCT25 PCT50 PCT75 PCT90 PCT95 PCT99 %
MACRO PRINTIT
DATA PRINTS; MEAN N MISS PCTS; BY BYVAR; %
PROC DELETE DATA= MNS PCTS;
PROC PRINT DATA=PRINTS; ID BYVAR; PFORMAT; %
Y1ER75=PCT75 Y1ER90=PCT90 Y1ER95=PCT95 Y1ER99=PCT99 %
MACRO PCTOPTS DATA=DAT OUT=PCTS (RENAME=(PCTOPTS))
MIN=-3 MAX=3 DIV=.01 N=K %
MACRO PCGIT
PROC PCTL PCTOPTS; VAR YYVAR; BY BYVAR; %
MACRO PFORMAT YYVAR N MEAN STD STR MIN MAX PCT01 PCT05 PCT10 PCT25 PCT50 PCT75 PCT90 PCT95 PCT99 %
MACRO PRINTIT
DATA PRINTS; MEAN N MISS PCTS; BY BYVAR; %
PROC DELETE DATA= MNS PCTS;
PROC PRINT DATA=PRINTS; ID BYVAR; PFORMAT; %
Y1ER75=PCT75 Y1ER90=PCT90 Y1ER95=PCT95 Y1ER99=PCT99 %
MACRO PCTOPTS DATA=DAT OUT=PCTS (RENAME=(PCTOPTS))
MIN=-3 MAX=3 DIV=.01 N=K %
DATALOOK;
PROC DELETE DATA=NW1;
MACRO DAT NW2 %
MACRO YYVAR Y2PR %
MACRO BBNAME Y2B01=PCT01 Y2B05=PCT05 Y2B10=PCT10 Y2B25=PCT25 Y2B50=PCT50 Y2B75=PCT75 Y2B90=PCT90 Y2B95=PCT95 Y2B99=PCT99 %
MACRO PCTOPTS DATA=DAT OUT=PCTS (RENAME=(PCTOPTS))
MIN=-300 MAX=300 N=K %
DATALOOK;
<table>
<thead>
<tr>
<th># of Farout Values Nt=56 Rule</th>
<th>DATA CODE PLOTS OF Y2LR VS HHT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td></td>
</tr>
<tr>
<td>30+</td>
<td>U</td>
</tr>
<tr>
<td>50+</td>
<td>U</td>
</tr>
<tr>
<td>10+</td>
<td>+</td>
</tr>
<tr>
<td>0 ++</td>
<td>+</td>
</tr>
<tr>
<td>-10 ++</td>
<td>+</td>
</tr>
<tr>
<td>-20</td>
<td>L</td>
</tr>
<tr>
<td>-50</td>
<td>L</td>
</tr>
<tr>
<td>-40</td>
<td>L</td>
</tr>
</tbody>
</table>
FUNCTION ROUTINE FOR COMPUTING DEFAULT OR SELECTED PERCENTILES FOR ONE VARIABLE AND OUTPUTTING THEM IN A SAS DATA SET. NO PRINTING IS DONE.

PROC PRINT MAY BE CALLED AFTER THIS FOR A NICE LISTING OF THE PERCENTILES, IF ANY VARIABLES ARE PRESENT, CALCULATIONS ARE DONE SEPARATELY FOR EACH BY GROUP AND ONE RECORD IS WRITTEN FOR EACH BY GROUP WITH THE VALUES OF THE BY VARIABLES INCLUDED. DATA ARE ASSUMED TO BE DISCRETE OR ELSE THE RESULTS WILL BE ACCURATE TO ONLY A SPECIFIED PRECISION (PARAMETER DIV).

IF 'X' IS THE INPUT VARIABLE NAME, NEW VARIABLES WILL BE CREATED FOR THE DIFFERENT PERCENTILES WITH NAMES OF THE FORM 'INH' WHERE NN IS THE PERCENTILE (E.G. 95) OR 'NX' (MAXIMUM) FOR THE 100TH PERCENTILE. IF 'X' IS OVER 6 BYTES LONG, THE LAST 2 BYTES WILL BE REPLACED WITH 'NN'. VARIABLE LABELS WILL BE CREATED FOR EACH VARIABLE OF THE FORM 'NN' PERCENTILE, IF THE INPUT VARIABLE HAD A LABEL, THE NEW LABELS WILL CONSIST OF THE ORIGINAL LABEL WITH 'NN' PRECEDED APPENDED TO IT. NEW VARIABLE NAMES CAN BE OVERWRITTEN THROUGH A PARM CLAUSE ON THE OUT=DATASET_NAME PARAMETER.

IF OPTION 'ALL' IS SPECIFIED, ALL PERCENTILES 0-100 WILL BE COMPUTED, IF A PARMED STATEMENT IS NOT GIVEN AND 'ALL' IS NOT SPECIFIED, THE PERCENTILES COMPUTED WILL BE 1, 5, 10, 25, 50, 75, 90, 95, 99. Optionally, a NUMBER OF OBSERVATIONS VARIABLE WILL BE INCLUDED IN THE NEW DATA SET. IT WILL CONTAIN THE NUMBER OF OBSERVATIONS IN EACH BY GROUP. MISSING DATA ARE IGNORED. CELLS ARE ALLOCATED FOR COUNTING FREQUENCIES. THE NUMBER OF CELLS REQUIRED IS (MAX-MIN)/DIV+1 AND EACH CELL REQUIRES 4 BYTES. VALUE X IS COUNTED IN CELL I WHERE I IS THE GREATEST INTEGER < (X-MIN)/DIV+1. ANY VALUES OUTSIDE THE INTERVAL (MIN,MAX) ARE SET TO THE NEAREST BOUNDARY AND A WARNING IS PRINTED.

ENTRY PROC PCTL OUT=NEW_DATASET_NAME (ALL) MIN=A MAX=B DIV=C N=VAR_NAME;

VARIABLES X BY ...;

PARMCARDS; (OPTIONAL)

5 0 5 50 75 100 (FOR EXAMPLE)

PARMS OUT NAME OF OUTPUT DATA SET.

IF OMITTED, NORMAL SAS DEFAULT DATASET NAME WILL BE USED.

ALL OPTION TO GET ALL 101 PERCENTILES

MIN, MAX MINIMUM AND MAXIMUM VALUES FOR VARIABLE X. THESE CAN BE LIBRARY ESTIMATES SUBJECT TO THE CONSTRAINT THAT THE NUMBER OF CELLS REQUIRED SHOULD FIT IN THE CORE REGION.

DIV MAXIMUM DISTANCE BETWEEN ANY TWO NON-EQUAL DATA VALUES OF A VALUE GREAT ENOUGH SO THAT THE NUMBER OF CELLS IS NOT TOO LARGE. DEFAULTS TO 1 IF NOT GIVEN. 'DIV' MAY BE ABBREVIATED 'D'.

N NAME OF VARIABLE TO HOLD NUMBER OF OBSERVATIONS FOR DATA SET OR BY GROUP IF NOT GIVEN, NO SUCH VARIABLE WILL BE ADDED TO THE OUTPUT DATA SET.

X NAME OF INPUT VARIABLE (EXACTLY ONE VARIABLE).

'VAR' MAY BE SUBSTITUTED FOR 'VARIABLES'.

THE DATA AFTER PARMCARDS, IF PRESENT, IS IN (2613,2X) FORMAT WITH THE NUMBER OF PERCENTILES TO CALCULATE BEING THE VERY FIRST NUMBER. IT MUST BE BETWEEN 0 AND 100 INCLUSIVE. THE PARMCARDS DATA MAY TAKE UP TO 9 CARDS.

JCL // EXEC SAS

//STEPLIB DD

//DD

//DD DSM=UNC.B.SASANAL,DISP=SHR

//TT157001 DD UNIT=SYSUD,SPACE=(60,(50,50)) (OPTIONAL FOR PARMCARDS)

EXIT STOP

LANGUAGE PORTAN

AUTHOR FRANK HARIEL

DATE 10MAR76

***