AN EASILY UNDERSTOOD STATISTICAL SYSTEM

GREG LARSEN
WESTMAR COLLEGE

THE COMPUTER IS NOT ONLY THE INSTRUMENT OF THE MATHEMATICIANS OR THE PHYSICISTS. THE COMPUTER IS THE INSTRUMENT OF THE DATA PROCESSOR, HOW EVER PROCESSING IS DEFINED. THE GENERAL PROBLEM HAS BEEN LACK OF MEANINGFUL USE BY STUDENTS AND FACULTY MEMBERS OF VARIOUS DEPARTMENTS. STAT IS AN ATTEMPT TO OVERCOME THIS PROBLEM. IT IS DESIGNED FOR SIMPLICITY IN USE, YET IS VERSATILE IN SCOPE. STAT IS ONE ATTEMPT TO FILL THE VOID EXISTING BETWEEN A COMPLEX PROGRAMMING LANGUAGE AND ITS GREATER FLEXIBILITY AND THE SPECIALIZED, HOWEVER STILL COMPLEX PROGRAMMING SYSTEMS IN EXISTANCE.

IN PAST YEARS, GREAT EMPHASIS HAS BEEN PLACED ON INDIVIDUALS TO KNOW HOW TO PROGRAM A COMPUTER. TO USE THE COMPUTER IT WAS AND TO A LARGE EXTENT STILL IS NECESSARY TO KNOW SOMETHING MORE ABOUT COMPUTER LANGUAGER THAN IS DESIRABLE. IN MOST UNDERGRADUATE COURSES STUDENTS ARE LED TO BELIEVE THAT TO MAKE USE OF THE COMPUTER, THEY NEED TO KNOW A MYSTICAL LANGUAGE CALLED FORTRAN OR COBOL OR BASIC, OR THE UNFORTUNATE AND MISTAKEN CONCLUSION IS DRAWN THAT COMPUTERS ARE ONLY FOR MATHEMATICIANS, AND PHYSICISTS, AND PEOPLE WHO ARE SUPPOSED TO KNOW ABOUT THEM. THE IDEA IS SOON IMPLANTED NOT TO DELVE FURTHER INTO THE SUBJECT.

MOVEMENT HAS BEEN TOWARDS INVOLVING UNDERGRADUATE PEOPLE IN THE VARIOUS DISCIPLINES WITH THE COMPUTER, IN AN ATTEMPT TO FAMILIARIZE THEM WITH ITS CAPABILITIES AND BENEFITS AS AN AID IN THEIR ENDEAVORS. FOR INSTANCE THE PSYCHOLOGISTS, SOCIOLOGISTS, BUSINESS ADMINISTRATORS SOMEWHERE IN THEIR CAREERS WILL RUN INTO LARGE SETS OF INFORMATION TO BE PROCESSED. KNOWING THE COMPUTER COULD PROCESS THIS INFORMATION FOR THEM AND RELIEVE THE EXPECTED DRUDGERY, MOST INDIVIDUALS ARE INCLINED TO TURN TO THE COMPUTER CENTER SEEKING HELP. WHAT THEY FIND IS USUALLY DISHEARTENING OR TIME CONSUMING ON SOMEONE'S PART. THEY FIND THEY MUST LEARN A RELATIVELY COMPLICATED PROGRAMMING LANGUAGE IN ORDER TO USE THE FACILITY OR THEY FIND SOMEONE WILLING TO DO IT FOR THEM, BUT WHO FOR VARIOUS REASONS MAY NOT QUITE UNDERSTAND WHAT IS WANTED, OR JUST PLAIN DOESN'T HAVE THE TIME TO HELP THEM. ANOTHER ROADBLOCK IS THE MUCH-TOUTED PACKAGE SYSTEM THATS USE IS CONTINGENT UPON KNOWING THAT COMPLICATED PROGRAMMING LANGUAGE, WITH ITS MYRIAD RULES OF OPERATION.

ATTEMPTS TO ALLEVIATE THIS SITUATION HAVE RESULTED IN ''ELEMENTARY COURSES'' IN COMPUTER PROGRAMMING, WHICH EVEN TO THE MOTIVATED INDIVIDUAL OFTEN BECOME AN ARDUOUS LEARNING TASK. ANOTHER ATTEMPT HAS BEEN TO EDUCATE

THE INSTRUCTORS IN PROGRAMMING LANGUAGES IN THE HOPES THAT HE WILL UNDERSTAND HOW TO USE SPECIAL SYSTEMS DEVELOPED FOR HIS EXISTING FACILITIES. THEN PASS THIS UNDERSTANDING ON TO THE STUDENT. WHAT HAPPENS? THE STUDENT UNLESS HIGHLY MOTIVATED IS GENERALLY TURNED OFF TO THE IDEA OF MAKING USE OF THE COMPUTER. AND ONLY A FEW STUDENTS EVER GAIN THE UNDERSTANDING THE INSTRUCTOR INTENDED AND USE THE COMPUTER TO THEIR BENEFIT.

JUST AS KNOWING THE INTRICATE WORKINGS OF MEDICINE ARE NOT A PREREQUISITE TO SEEING A DOCTOR. KNOWING COMPLEX COMPUTER PROGRAMMING SHOULD NOT BE A REQUISITE TO USING THE FACILITIES.

USING THE STAT SYSTEM

THE STATISTICAL SYSTEM IS USED TO GENERATE AND PRINT STATISTICS, FROM INFORMATION SUPPLIED TO IT, NORMALLY IN THE FORM OF PUNCHED CARDS.

CONTROL OF THE COMPUTER IS BY MEANS OF CONTROL STATEMENTS...PRESENTLY THERE ARE 15 TYPES OF CONTROL STATEMENTS.

AN EXPLANATION OF EACH OF THE CONTROL STATEMENTS IS AS FOLLOWS. EACH LINE OF A CONTROL STATEMENT REPRESENTS ONE PUNCHED CARD.

```
// JOB T
// XEQ STAT
```

THESE TWO CARDS ARE IBM CONTROL CARDS AND INITIATE THE START OF YOUR PROGRAM

```
*BEGIN          .....NAME.....DATE.....COMMENTS.....
```

THIS CONTROL RECORD IS APPROPRIATELY CALLED A BEGIN RECORD. IT INITIALIZES EACH JOB IN YOUR PROGRAM AND SHOULD BE USED PRIOR TO THE BEGINNING OF EACH PROGRAM. IF IT IS NOT USED UNPREDICTABLE RESULTS MAY OCCUR IN SOME INSTANCES. ALL INFORMATION IN CARD COLUMS 7-80 IS PRINTED AT THE TOP OF EACH PAGE OF OUTPUT AS A HEADER.

```
*VARIABLES (NAME,LRN,IF,IL,NDEC),(NAM2,LRN,IF,IL,NDEC).........
```

THIS RECORD DEFINES A VARIABLE TO THE COMPUTER...ALL INFORMATION PER VARIABLE MUST BE ENCLOSED BY PARENTHESIS. THE FOLLOWING INFORMATION MUST BE SUPPLIED.

1. NAME---THE NAME OF THE VARIABLE MUST BE PRESENT. IT MAY BE ANY FIVE CHARACTERS SUCH AS A,B,1,2,$/*,APPLE. NO NAME LONGER THAN FIVE

CHARACTERS IS PERMITTED.

2. LRN----THE LOGICAL RECORD NUMBER MAY BE SUPPLIED. FOR INSTANCE SUPPOSE ALL THE DATA INFORMATION COULD NOT FIT IN 80-COLUMNS OF ONE CARD AND TWO CARDS WERE NEEDED. THEN THAT LOGICAL SET IS MADE UP OF TWO LOGICAL RECORDS. A LOGICAL SET MAY HAVE AS MANY AS 99 LOGICAL RECORDS.

3. IF,IL--THE CARD COLUMN WHERE THE INFORMATION BEGINS (IF) AND THE INFORMATION ENDS (IL) MAY BE SPECIFIED ALSO.

4. NDEC---A FINAL SPECIFICATION POSSIBLE IS CALLED THE NUMBER OF IMPLIED DECIMAL PLACES. SUPPOSE YOU DID NOT TYPE A DECIMAL POINT IN YOUR DATA CARDS. FOR EXAMPLE YOU TYPE ALL YOUR DATA IN THE FORM 12345, BUT YOU REALLY HAD DATA OF THE FORM 123.45, THEN NDEC WOULD BE THE DIGIT VALUE 2.

THERE ARE THREE ALTERNATE METHODS OF DEFINING A VARIABLE. THEY ALL INVOLVE ELIMINATION OF INFORMATION.

1. IF NDEC IS NOT SPECIFIED ITS VALUE IS ASSUMED ZERO AND DECIMAL POINT IN THE DATA FIELD WILL BE ASSUMED CORRECT. THAT MEANS ANY DECIMAL WHERE PUNCHED, WILL OVERRIDE THE ZERO SPECIFICATION AND SUPPLY THE DECIMAL WHERE PUNCHED.

2. IF NO BEGINNING AND END INFORMATION IS INCLUDED, A FREE SEARCH WILL BE MADE FOR THE VARIABLE ON THE LOGICAL RECORD SPECIFIED. THAT VARIABLE WILL ASSUME THE VALUE OF THE FIRST SET OF INFORMATION ENCLOSED BY BLANK CHARACTERS.

3. AND IF ONLY A VARIABLE NAME IS SPECIFIED, THE LOGICAL RECORD NUMBER WILL BE ASSUMED ONE AND A FREE SEARCH MADE FOR THE VARIABLE. NDEC IN CASE 2, AND 3 IS ASSUMED ZERO.

MORE THAN ONE VARIABLES CONTROL RECORD IS PERMITTED, BUT NO MORE THAN 30 VARIABLES MAY BE USED IN ANY ONE JOB.

---

*GROUPS 1(NAM1,NAM2,....NAMN).....30(NAM1,NAM2,....NAMN)

CERTAIN STATISTICS MAKE SENSE ONLY WHEN ONE SPEAKS OF GROUPS. THE GROUP CONTROL RECORD DEFINES A GROUP NUMBER AND A LIST OF VARIABLES ASSOCIATED WITH THE GROUP NUMBER. NO MORE THAN 30 GROUPS CAN BE HANDLED IN A SINGLE JOB. GROUP NUMBERS MUST RANGE IN VALUE FROM 1-30, AND PRECEDE THE LIST OF VARIABLES

ASSOCIATD WITH IT. SINCE VARIABLE NAMES ARE NEEDED IN DEFINING A GROUP, A VARIABLES CONTROL RECORD MUST PRECEDE THE GROUPS RECORD.

THE GROUPS CONTROL RECORD IS USED FRO STATISTICAL COMPUTAITONS THAT CONNOT BE PFRFORMED ON A SINGLE VARIABLE. EACH GROUPS RECORD MUST HAVE A GROUP NUMBER 1-30 AND TWO OR MORE VARIABLE NAMES ASSOCIATED WITH IT.

---

*STAT ST1(NAM1,NAM2,NAM3,...NAMN),ST2(G1,G2,G3,....GN)

THE STAT CONTROL RECORD REQUESTS STATISTICS DESIRED FOR EACH VARIABLE OR GROUP. SINGULAR STATISTICS THOSE STATISTICS THAN CAN BE COMPUTED FOR A SINGLE VARIABLE SUCH AS MINIMUMS AND MAXIMUMS, ARE SPECIFIED BY A THREE-LETTER CODE FOLLOWED BY A LIST IN PARENTHESIS OF THOSE VARIABLES FOR WHICH THAT STATISTIC IS WANTED. A STATISTIC FOR A GROUP IS SPECIFIED SIMILARLY, ONLY A GROUP NUMBER IS SPECIFIFD INSTEAD OF A VARIABLE NAME.

SFE APPENDIX 1 FOR A LISTING OF CURRENTLY AVAILABLE STATISTICS.

---

*DATA

THIS CONTROL RECORD SIGNALS TO THE COMPUTER THAT DATA INFORMATION IS TO FOLLOW. THE PROGRAM WILL READ DATA RECORDS UNTIL ANOTHER CONTROL RECORD IS ENCOUNTFRED. LOGICAL RECORD SETS MUST ALWAYS BE COMPLETE.

---

*SELECT (LRN,IF,IL) OP CRITERIA

THE SELECTION RECORD PERMITS THE SELECTION OF DATA FROM A SET OF DATA. THE RECORD MUST CONTAIN THREE DIGITS ENCLOSED BY PARENTHESIS, AN OPERATION CODE AND CRITERIA OF SELECTION.

1. (LRN,IF,IL)---LRN SPECIFIES THE LOGICAL RECORD ON WHICH SELECTION IS BEING MADE. IF IS THE NUMBER OF THE FIRST COLUMN WHERE

THE SELECTION INFORMATION IS CONTAINED.

2. OP----------A SELECTION OPERATOR IS SPECIFIED. VALID OPERATORS ARE

A. LT WHICH MEANS LESS THAN

B. EQ WHICH MEANS EQUAL TO

C. GT WHICH MEANS GREATER THAN

D. LE WHICH MEANS LESS THAN OR EQUAL TO

E. GE WHICH MEANS GREATER THAN OR EQUAL TO

F. NE WHICH MEANS NOT EQUAL TO

3. CRITERIA-------AND FINALLY A CRITERIA OF SELECTIN MUST BE PRESENT, THAT IS WHAT IS THE CRITERIA OF ACCEPTANCE OR REJECTION OF THE DATA AFTER FUNCTIONAL APPLICATION OF THE SELECTION OPERATOR.

TEN SELECTION SPECIFICATIONS MAY BE MADE PER JOB. IF MORE THAN ONE SELECTION IS REQUESTED, IT IS ASSUMED THAT ALL SELECTION REQUIREMENTS MUST BE MET BEFORE A LOGICAL SET IS ACCEPTED AS CONTAINING VALID DATA RECORDS. IN OTHER WORDS,REJECTION OF A LOGICAL RECORD SET IS MADE IF ONLY ONE SELECTION REQUIREMENT IS NOT MET, AND ACCEPTANCE OF A LOGICAL RECORD SET IS MADE IF AND ONLY IF ALL SELECTION REQUIREMENTS ARE MET, AS DEFINED.

---

*INPUT II

THE INPUT CONTROL RECORD SHOULD BE ONLY USED BY INDIVIDUALS FAMILIAR WITH COMPUTER PROGRAMMING METHODS, THOUGH IT MAY BE USED BY THE UNINITIATED WITH EXTREME CAUTION. THE INPUT CONTROL RECORD DEFINES WHERE ALL FOLLOWING INFORMATION SUPPLIED TO THE COMPUTER IS TO COME. IF NO INPUT ROCORD IS ENCOUNTERED, THE CARD READER IS ASSUMED TO BE THE INPUT DEVICE.

II----IS THE INPUT DEVICE NUMBER AND MAY TAKE ON TWO VALUES, THEY ARE

2 MEANS INPUT FROM THE CARD READER.

6 MEANS INPUT FROM THE CONSOLE KEYBOARD.

---

*OUTPUT IO

THE OUTPUT CONTROL RECORD IS SIMILAR TO THE INPUT RECORD. IO IS THE OUTPUT DEVICE NUMBER, WITH TWO POSSIBLE VALUES. IF NO OUTPUT RECORD IS ENCOUNTERED THE 1132 PRINTER IS THE ASSUMED OUTPUT DEVICE.

3 MEANS OUTPUT ON THE 1132 PRINTER (ASSUMED IF NOT PRESENT)

1 MEANS OUTPUT ON THE CONSOLE TYPEWRITER

---

*RECORDS NLR

THE RECORDS CONTROL SPECIFIES THE NUMBER OF LOGICAL RECORDS PER LOGICAL SET. FOR INSTANCE FIVE CARDS PER SET WOULD MAKE NLR EQUAL TO FIVE.

---

*FREQ (NAME,UP,LOW,INC),(NAME,INC)

THIS CONTROL RECORD IS A SPECIAL CONTROL RECORD NEEDED WHEN A FREQUENCY DISTRIBUTION OR HISTOGRAM IS REQUESTED BY A STAT COMMAND. THERE ARE TWO METHODS OF DEFINING VALUES TO CONTROL A FREQUENCY DISTRIBUTION OR HISTOGRAM. INFORMATION WITHIN PARENTHESIS CAN CONTAIN.

2. THE VARIABLE NAME, AND AN INCREMENT VALUE.

IF OPTION 2 IS CHOSEN THE UPPER AND LOWER BOUND OF THE FREQUENCY DISTRIBUTION IS ASSUMED AS MAXIMUM AND MINIMUM VALUES OF THE DATA IN QUESTION. AN INCREMENT VALUE MUST NOT RESULT IN MORE THAN 50 INTERVALS.

---

*PROCESS

NO COMPUTATION OR PRINTING OF STATISTICS IS INITIATED UNTIL THIS CONTROL RECORD IS ENCOUNTERED. IT SERVES AS A SIGNAL TO THE COMPUTER TO BEGIN PROCESSING ALL MATERIAL FROM THE LAST BEGIN RECORD. IMMEDIETELY AFTER PROCESSING HAS TAKEN PLACE, THE PROGRAM SYSTEM PREPARES TO ACCEPT ANOTHER CONTROL RECORD.

*END

THIS CONTROL RECORD MUST ALWAYS BE THE LAST CARD OF YOUR ENTIRE DECK. IT FUNCTIONS TO TERMINATE CONTROL OF THE COMPUTER BY THE STAT SYSTEM AND ALLOWS A RETURN TO NORMAL IBM SYSTEM CONTROL. ONCE AN END CONTROL HAS BEEN ENCOUNTERED, THE SYSTEM CAN ONLY BE REUSED BY INSERTING IN FRONT OF YOUR DECK. // JOB T AND // XEQ STAT CONTROL RECORDS.

*CALL NAME

THE LINKING CONTROL RECORD IS TO BE USED BY COMPETENT PROGRAMMERS OR THOSE INDIVIDUALS WITH SPECIAL INSTRUCTION IN ITS USE. SEE THE SYSTEMS MANUAL.

A FEW GENERAL INSTRUCTIONS CONCERNING CONTROL RECORDS ARE THESE

1. ALL CONTROL RECORDS MUST BE PRECEDED BY AN ASTERIK, FOLLOWED BY THE CONTROL FUNCTION.
2. A BLANK MUST SEPERATE THE CONTROL FUNCTION AND THE INFORMATION ON THE CONTROL RECORD.
3. NO DATA CARDS WHICH BEGIN WITH AN ASTERIK SHOULD BE USED.

THE STAT MONITOR SYSTEM DOES DETECT AND DISPLAY ERRORS WHEN POSSIBLE. IF AN ERROR IS DETECTED AN ERROR MESSAGE IS PRINTED TO THE USER IMMEDIATELY FOLLOWING THE CONTROL RECORD WHICH CONTAINS THE ERROR(S). THOUGH ERRORS ARE DETECTED THERE EXIST POSSIBILITIES OF BAD DATA OR MISINFORMATION SPOILING STATISTIC VALUES. A DETECTION OF ZERO ERRORS IN YOUR PROGRM
BUT NOT A SUFFICIENT CONDITION FOR ERRORLESS OUTPUT.

APPENDIX 1

CURRENTLY AVAILABLE STATISTICS AND OPERATION CODES

STATISTIC	CODE
MINIMUM	MIN
MAXIMUM	MAX
VARIANCE	VAR
MEAN AVERAGE	MEA
STANDARD DEVIATION	STD
STANDARD ERROR OF THE MEAN	STE
INDEPENDENT T-TEST	ITT
FREQUENCY DISTRIBUTION	FRE
PRINT DATA CARDS	PRC
PRINT VARIABLE OBSERVATION VALUES	PRO

SAMPLE USE

```
// JOB T
// XEQ STAT
*BEGIN      NAME...DATE....COMMENTS...
*RECORDS 2
*VARIABLES  (A,1,1,5,0)      (B,1,6,10,0)  ,  (C,2,1,5,2)
*DATA
      *
      *      YOUR DATA DECK GOES HERE
      *
*STAT MIN(A,B,C),MAX(B,A),STE(C,A),MEA(A)     STD(B)
*PROCESS
*END
```

```
// JOB T
// XEQ STAT
*BEGIN      NAME
*VARIABLES  (A)     (D,2,5,10)
*VARIABLES   (B),(C,1,5,10,0)
*RECORDS     2
*DATA
      *
      *     YOUR DATA DECK GOES HERE
      *
*GROUPS   1  (B,A), 2(C,D)
*STAT   STD(A,B,C,D)
*STAT ITT(1,2)
*PROCESS
*END
```

```
// JOB T
// XEQ STAT
*BEGIN
*VARIABLE   (3,1)  ,  (APPLE,1,50,55)
*STAT  MIN(3,APPLE),  MEA(3,APPLE),STD(APPLE,3)
```

```
*GROUPS  12(3,APPLE)
*STAT ITT(12)
*DATA
      *
      *   YOUR DATA DECK HERE
      *
*PROCESS
*BEGIN        NAME AND DATE.....COMMENTS....
*VARIABLE  (A,2)
*RECORDS   2
*DATA
      *
      *        YOUR DATA HERE
      *
*STAT      MIN(A),  MAX(A)
*PROCESS
*END
```

```
PAGE   1
// JOB    OECA
LOG DRIVE   CART SPEC   CART AVAIL  PHY DRIVE
  0000        OECA        OECA        0000
V2 M09   ACTUAL  8K  CONFIG  8K
// *FOLLOWING IS A SAMPLE SAMPLE RUN OF THE STAT SYSTEM
// XEQ STAT
*BEGIN      TEST GREG LARSEN
*VARIABLES (A)
*VARIABLES (B)
*GROUPS 1 (A,B)
*DATA
*STAT MIN(A,B),MAX(A,B),MEA(A,B),VAR(A,B),STD(A,B),STE(A,B)
*PROCESS
```

TEST GREG LARSEN

SINGULAR STATISTICS

VARIABLE NAME	STATISTIC		VALUE
A	MINIMUM	=	10.00000
A	MAXIMUM	=	26.00000
A	MEAN AVERAGE	=	18.30000
A	VARIANCE	=	19.69474
A	STANDARD DEVIATION	=	4.43787
A	STAND. ERR. OF MEAN	=	0.99233

VARIABLE NAME	STATISTIC		VALUE
B	MINIMUM	=	10.00000
B	MAXIMUM	=	26.00000
B	MEAN AVERAGE	=	18.30000
B	VARIANCE	=	19.69474
B	STANDARD DEVIATION	=	4.43787
B	STAND. ERR. OF MEAN	=	0.99233

*END

```
// JOB T

LOG DRIVE   CART SPEC   CART AVAIL  PHY DRIVE
  0000        0ECA        0ECA        0000

V2 M09   ACTUAL  8K  CONFIG  8K

// *THIS SAMPLE RUN CONTAINS ERRORS IN THE CONTROL CARDS

// XEQ STAT

*BEGIN      TEST GREG LARSEN
*VARIABLES (A)
*VARIABLES (B),(APPLES)
CONTROL HAS DETECTED A VARIABLE NAME LONGER THAN FIVE CHARACTERS.


*GROUPS 1 (A,B)
*DATA
*STAT MAN(A,B),MAX(A,B),MEA(A,B),VAR(A,B),STD(A,B),STE(A,B)
A STATISTIC COMMAND HAS MADE AN INVALID REQUEST. CHECK SPELLING AND THE LIST OF
VALID STATISTICS.

*PROCESS
EXECUTION HAS BEEN SUPPRESSED FOR THIS JOB AND ALL JOBS FOLLOWING DUE TO
THE ABOVE ERRORS.
```