Abstract

In many statistical investigations accepting or rejecting a statistical hypothesis is dependent upon the p-value attained by a calculated test statistic. Many tables exist for specification of a critical value of a test statistic that is to be compared to the calculated test statistic; if the calculated test statistic exceeds the critical value chosen (often based on a p-value desired and possibly a number of other parameters) the null hypothesis is rejected. These tables can also be used to find an upper bound and lower bound for p-values once a test statistic has been calculated. An exact p-value cannot be determined. Many SAS procedures report p-values to the fourth or fifth decimal place and although four or five decimals are adequate for judging "borderline significance" exact p-values for test statistics with extreme values cannot be determined. Often a quick and simple manner would be helpful in determining exact p-values. This paper outlines a CMS EXEC written in Rexx that uses four SAS functions to calculate exact p-values for a Z-test, a T-test, an F-test, or a Chi-square test.

Keywords:
P-value, PROBNORM, PROBT, PROBCHI, PROBF, Rexx, CMS, EXEC

I. Introduction

The acceptance or rejection of a statistical hypothesis often depends on the p-value attained by a test statistic. Tables used to obtain reliability coefficients from desired p-values exist in most statistics books (e.g. Kirk, 1984; Hollander and Wolfe, 1973; and Hogg and Craig, 1978) and those tables can be used to find an upper bound and a lower bound which correspond to the test statistic calculated. However, often one would prefer to report an exact p-value rather than a range of p-values. Therefore this problem has been answered by SAS by reporting p-values to the fourth decimal place; however, in extreme cases SAS often reports "p < 0.0001". Although this method of reporting p-values solves the problem of having a range of p-values to report and makes conclusions concerning borderline significance easier, often exact p-values cannot be reported.

II. Problem

Even if one writes a SAS program that will calculate p-values and it remains resident on one's minidisk it takes time (albeit not a tremendous amount of time) to call up the program, customize the program for the specific situation, access SAS, execute the program, and obtain the results. It would be very beneficial if a manner existed in which exact p-values could be calculated with a minimal amount of effort required.

III. Solution

The solution presented here is a CMS EXEC (named PVALUE EXEC) that can calculate p-values for four of the more popular statistical tests using four SAS functions: PROBNORM, PROBT, PROBCHI, and PROBF. PVALUE EXEC, written in Rexx, uses the above four functions, noninteractive SAS, and the PTYPE option for noninteractive SAS to compute p-values and return them to the screen for a Z-Test, a T-Test, a Chi-square test, or an F-Test.

The EXEC has been well documented and comes with a "safety valve", so that the first time the EXEC is invoked one is aware that the SAS disks for noninteractive SAS are released and detached. If a test other than a Z-test, a T-test, a Chi-square test, or an F-test is requested the EXEC will abort calculations. Calculations will also be aborted if negative degrees of freedom are entered. Negative test statistics are acceptable values for a Z-test or a T-test. In those cases the test statistic is multiplied by negative one. If a negative test statistic is entered as a Chi-square test statistic or an F-test statistic, p-value calculations are aborted.

The EXEC has been fashioned for both the beginning user of PVALUE EXEC and the experienced user of PVALUE EXEC. After the PVALUE EXEC has been transferred to one's minidisk and before running the EXEC, one should make sure to delete the safety valve lines if one has no reason to think the EXEC will interfere with one's minidisk or other parts of a site's mainframe computer operations.

Novice Approach:

Type: "PVALUE" and one will be prompted for one of four statistical tests: a Z-test, a T-test, a Chi-square test, or an F-test. Enter the test desired. (If anything other than the keywords (ZTEST, TTEST, CHISQUARE, or FTEST are entered the p-value calculations will be aborted.) After entering one of the keywords, one will be prompted to enter the test statistic. Do so. If the keyword ZTEST was entered p-value calculations will be done. Otherwise one is prompted for the degrees of freedom. If negative degrees of freedom are entered p-value calculations are aborted. After the degrees of freedom have been entered (unless p-value calculations for an F-test are desired, in which case one is prompted for numerator degrees of freedom and denominator degrees of freedom) the EXEC is submitted and the SAS program that was constructed returns the results to one's screen.

Approach for the Experienced User:

Once one becomes familiar with the PVALUE EXEC one does not have to be a slave to the prompts provided by the EXEC. One can invoke the EXEC and at the same time specify the statistical test desired, the test statistic, and degrees of freedom, if needed. An example follows:

PVALUE FTTEST 6.045.3 8

Also the full names of the tests are not required for p-value calculations to be done. Acceptable abbreviations for the four statistical tests are the first letter of each test. (i.e. Z, T, C, or F)

IV. Conclusions

P-values attained by test statistics can determine the success or failure of one's investigation. When computations are carried out by hand only a range of p-values can be determined for test statistics employing the use of tables used to establish reliability coefficients. Exact p-values cannot be reported using these tables! SAS and other statistical software have helped in determining exact p-values; however, often p-values are reported only to four or five decimal places. In extreme cases of a p-value of "< 0.0001" is reported. This paper has presented a relatively simple manner in which one can determine a wider range of exact p-values for four of the more popular statistical tests: the Z-test, the T-test, the Chi-square test, and the F-test.
V. Obtaining the PVALUE EXEC

The PVALUE EXEC may be obtained in one of the three following manners.

1. Requests may be sent to the author at his electronic mail address:
   C09195@UNIVSVCVM (Bitnet).

2. Either a formatted Macintosh diskette, a formatted IBM PC diskette (high or low density), or a formatted IBM PS/2 diskette can be sent to the author at the address below. [Because of the time it takes to format diskettes the author reserves the right to return unformatted diskettes to the individuals requesting the EXEC without writing the EXEC to the disk (though this procedure will only be followed in extreme circumstances)]. It would be much appreciated if phone requests were not made.

Mr. Gregory E. Gilbert
Academic Services
1244 Blossom Street
Columbia, SC 29208 USA

The PVALUE EXEC will be downloaded using either TINCan software (Macintosh) or YTERM software (IBM) and written to diskettes as an 80 byte ASCII file.

For those parties who send Macintosh diskettes documentation will accompany the EXEC saved as a Microsoft Word 4.0 text file that should be able to be incorporated into any of the more popular Macintosh word processing software and for those who furnish an IBM PC diskette or an IBM PS/2 diskette documentation will be in a similar form should be able to be incorporated into a word-processing package of one's choice. Also included on IBM PC and IBM PS/2 diskettes will be a Microsoft Word 5.0 file (PVWORD.OOC) and a Wordperfect 5.0 file (PYALUEWP.DOC).

3. The program can be entered by hand or scanned from the copy of code included on this diskette.

VI. Technical Note

This EXEC has been programmed to link to a SAS diskpack(s) upon execution. In almost all cases the linking procedures from site to site will differ. Before attempting to invoke the PVALUE EXEC one should investigate how linking to the SAS disk is done at one's site. Please do not contact the author if problems and idiosyncrasies were reported to the author.

VII. The PVALUE EXEC

/* P-value EXEC */
/* The author is Greg Gilbert and by no means whatsoever */
/* does he claim that this EXEC is bug free! He cannot be */
/* held accountable for the nasty things it might do to one's */
/* CMS minidisk. */

/* Update Record: */
/* DATE DESCRIPTION OF WORK DONE */
/* */
/* 062489 Modified the linking procedure to reflect changes in CMS SAS 5.18 linkage procedures. */
/* 060189 Changed names of variables assigned to represent the numerical values of the statistics. */
/* They used to be the same as the name of the statistical test in each step. */
/* 061489 Modified so that only first three letters of test are needed to invoke the test. Also */
/* abbreviated (with the help of Karen Maloney) so */
/* that when invoking the EXEC the test, the test statistic, and degrees of freedom (if needed) */
/* can be specified. EXAMPLE: */
/* */
/* PVALUE CHI 3.841 1 */
/* */
/* If further information is needed for the calculations prompts are issued. */
/* */
/* 062489 Safety value lines that need to be deleted were added to the safety value message and the line */
/* suggesting the renaming of the EXEC was deleted because it was thought to be useless. */
/* 062599 A message was inserted in the safety valve to advise persons to make the step that links the SAS minidisk to one's account "site specific". */
/* 121290 Shortened abbreviations needed to run the various statistical tests to Z, T, C, and F. Previously */
/* abbreviations were the first three letters of each statistical test keyword. */
/* */
/* Define On-Highlighting and Off-Highlighting */
/* hi = '1 DE8'; lo = '1 D60' */
/* */
/* The Proverbial Safety Valve */
/* say */
/* say 'This EXEC accesses the SAS disks as */
/* say 'and disk'hi'P'lo'and then'hi'releases'lo'those disks.' */
/* say */
/* say 'Make sure to customize the step that links to the */
/* say 'SAS minidisks. (It needs to be made 'site specific').' */
/* say */
/* say 'Delete safety valve lines before executing the program.' */
/* say */
/* End the proverbial safety valve */

ARG test stat nu1 nu2

"OP SET EMSG OFF"
"MAKEBUF"

if test = 'Z' then do
/* Query what test wanted */
say "What kind of p-value would you like to calculate?";
say "hi" ZTEST;
say "hi" CHISQUARE;
say "hi" FTTEST;
pull test stat nu1 nu2
upper test
/* End Query */
end

/* Abbreviate test to first three letters of test desired */
/* test = substr(test, 1 , 3) */
/* End Abbreviation Step */

/* Select Subroutine for P-value Calculations */
select
  when test = 'Z' then call ZTEST
  when test = 'T' then call TTEST
  when test = 'C' then call CHISQUARE
  when test = 'F' then call FTTEST
otherwise do
  say "P-value Calculation Aborted! Test not found." lo
  exit
end
/* End Selection Process */
end

/* Write main body of SAS Program */
line1 = "OPTION NODATE NONUMBER;"
line2 = "DATA;"

1117
line4 = "FILE PRINT;"
line5 = "PUT 'The P-value is 'SM'.';
line6 = "RUN;"
  " End Create
/
/* Link to SAS Disk Pack — must be customized to each site
' */
'SET CMSTYPE HT' /* Masks output to screen */
'CP LINK PROGPROD 20a 20a RR' 'ACC 20a N'
'CP LINK PROGPROD 200 200 RR' 'ACC 200 O'
'CP LINK PROGPROD 240 240 RR' 'ACC 240 P'
'SET CMSTYPE RT' /* Turns off masking */
' End Link */
' /* Create SAS Source Program with Fixed Record Format and LRECL = 80 */
EXECIO 1 DISKW "PVA$$LUE SAS A 1 F 80 (STRING"line1
EXECIO 1 DISKW "PVA$$LUE SAS A 2 F 80 (STRING"line2
EXECIO 1 DISKW "PVA$$LUE SAS A 3 F 80 (STRING"line3
EXECIO 1 DISKW "PVA$$LUE SAS A 4 F 80 (STRING"line4
EXECIO 1 DISKW "PVA$$LUE SAS A 5 F 80 (STRING"line5
EXECIO 1 DISKW "PVA$$LUE SAS A 6 F 80 (STRING"line6
' /* End Write */
'/* Submit SAS Program */
'EXEC SAS PVA$$LUE (PTYPE'
'/* Erase SAS Program and SASLOG file on A Disk */
'ERASE PVA$$LUE SAS A'
'ERASE PVA$$LUE SASLOG A'
'/* Release and Detach SAS Disks */
'RELEASE N (DET'
'RELEASE O (DET'
'RELEASE P (DET'
'DROPBUF'
'CP SET EMSG ON'
exit
'/* Statistical Test Subroutines */
'ZTEST: /* Subroutine for Z-Test calculations */
  if stat = " then do
    say "Z-Test Statistic:"
    pull stat
  end
  if stat < 0 then do
    stat = 1 - stat
  end
  say "This might take a couple of seconds so please be patient."
  line3 = "SLM = 1-PROBNORM(stat);"
' /* End Z-Test Subroutine */
return
'TTEST: /* Subroutine for T-Test Calculations */
  if stat = " then do
    say "T-Test Statistic:"
    pull stat
  end
  if stat < 0 then do
    stat = 1 - stat
  end
  if nu1 = " then do
    say "How many degrees of freedom do you wish?"
    pull nu1
  end
  if nu1 <= 0 then do
    say "Invalid Degrees of Freedom!!" lo "Bye-Bye..."
  end
  exit
' /* End of Program */
VIII. Acknowledgments

The author would like to gratefully acknowledge the assistance of the Academic Services group of the University of South Carolina Computer Services Division, Phil Moore, Barbara Koski, and Karen
Fleming. He benefitted much from their advice, suggestions, and help. He would also like to thank Sharon Gilbert for her assistance in reviewing this paper. The author would especially like to thank Karen Maloney for her help in revising the EXEC for more expedient usage.

Microsoft Word 3.01 word processing software. Copyright 1987 Microsoft Corporation.

Microsoft Word 4.00 word processing software. Copyright 1989 Microsoft Corporation.

Aldus Pagemaker 3.01 desktop publishing software. Copyright 1988 Aldus Corporation.

Microsoft Word 5.00 word processing software. Copyright 1989 Microsoft Corporation.

CMS (Conversational Monitoring System [version 5.0]), IBM PC, IBM PS/2, and Rexx are registered trademarks of the International Business Machines Corporation.

Macintosh is a registered trademark of Apple Computer, Inc.

SAS is a registered trademark of Statistical Analysis Systems; Cary, NC.

TinCan 3.0 communications software copyright 1987, Yale University.

WordPerfect 5.00 word processing software. Copyright 1988 WordPerfect Corporation.

Yterm 1.4 communications software copyright 1987, Yale University.

IX. References

