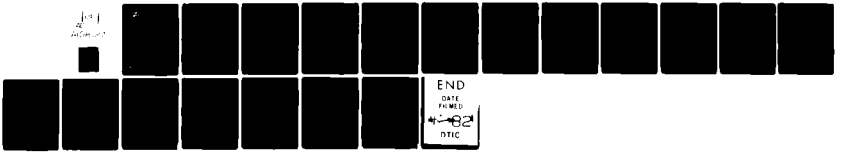
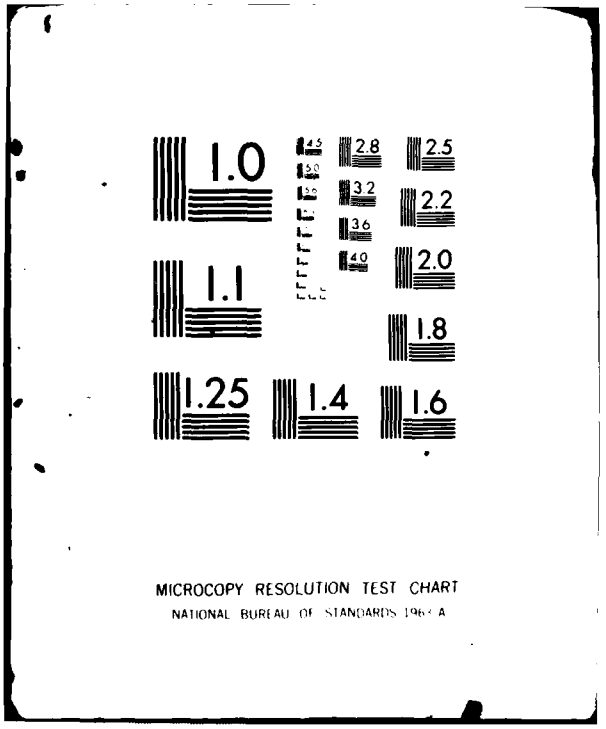
AD-A108 127 MISSOURI UNIV-COLUMBIA DEPT OF STATISTICS F/G 12/1
P-VALUES FOR MULTI-STAGE AND SEQUENTIAL TESTS.(U)
OCT 81 R W MADSEN, K B FAIRBANKS N00014-78-C-0655
UNCLASSIFIED TR-112 NL



END
DATE FILMED
4-82
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 196

University of Missouri-Columbia

P-Values for Multi-Stage and Sequential Tests

Richard Madsen and Kenneth Fairbanks
University of Missouri-Columbia

Technical Report No. 112
Department of Statistics
October 1981

Mathematical Sciences

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 112	2. GOVT ACCESSION NO. AD-A108 127	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) P-values for Multi-stage and Sequential Tests		5. TYPE OF REPORT & PERIOD COVERED Technical 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Richard Madsen Kenneth Fairbanks		8. CONTRACT OR GRANT NUMBER(s) ONR Grant N00014-78-C-0655
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics University of Missouri, Columbia Columbia, MO. 65201		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Department of the Navy Arlington, VA.		12. REPORT DATE October 1981 13. NUMBER OF PAGES 17
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

Accession For
NTIS GRA&I
DTIC TAB
Unannounced
Justification
Distribution/
Availability Codes
Dist | Avail and/or Special
A

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

P-values, multi-stage tests, sequential tests, exponential distribution

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

P-values are commonly given for ordinary single stage statistical tests. In this note we give a general method for calculating p-values for a large class of multi-stage and sequential tests. We also give some tables of p-values for multi-stage tests about the parameter of an exponential distribution when test plans from MIL-STD-781C are used.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601 |

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

P-VALUES FOR MULTI-STAGE AND SEQUENTIAL TESTS

by

Richard W. Madsen and Kenneth B. Fairbanks
University of Missouri-Columbia and Murray State University

Summary

P-values are commonly given for ordinary single stage statistical tests. In this note we give a general method for calculating p-values for a large class of multi-stage and sequential tests. We also give some tables of p-values for multi-stage tests about the parameter of an exponential distribution when test plans from MIL-STD-781C are used.

Key words: P-values, multi-stage tests, sequential tests, exponential distribution.

1. INTRODUCTION

It is quite common for investigators to report the results of a statistical test by giving a p-value rather than simply stating that the test was (or was not) significant using an α-level test. However when the statistical test used is a multi-stage test or a sequential test rather than a single stage test, p-values are generally not given. It is the purpose of this note to give a general method for calculating p-values for a large class of multi-stage and se-

quential tests. We also give some tables of p-values for multi-stage tests about an exponential parameter using test plans from MIL-STD-781C.

2. DEFINITION OF P-VALUES

Say that X is a random variable having distribution function $F(x;\theta)$. Let H_o denote some statistical hypothesis about F, perhaps a hypothesis about θ such as: $H_o: \theta \in \Theta_o$. In the single sample case a random sample $X_1, X_2, \ldots, X_n$ is typically chosen from the distribution of X and a test statistic T is calculated. A critical region C_α is chosen so that

$$\sup_{\theta \in \Theta_o} P[T \in C_\alpha] = \alpha.$$

Generally C_α will consist of the extreme values of T, perhaps

$$C_\alpha = \{t: t \geq t_\alpha\}.$$

In this case we would have $C_{\alpha_1} \subset C_{\alpha_2}$ if $\alpha_1 < \alpha_2$.
It is at this point that the concept of a p-value may be introduced. (Note that some authors use the term prob-value while others use the term significance probability instead of p-value.) Dudewicz (1976, p.313) defines it as..."the smallest α for which we would surely reject if we observed" $T = t$. Bickel and Doksum (1977, p.170) and Bhattacharyya and Johnson (1977, p.175), to name just two

others, give similar definitions. However if we try to use this same definition for sequential tests we run into a problem, as we shall see.

If we consider a sequential probability ratio test (SPRT) of a simple null against a simple alternative hypothesis (Wald (1947) or Ghosh (1970)) and if Z_n denotes the value of the test statistic at stage n, then the decision boundaries, a and b, are determined by the desired values of α and β. The general procedure is to observe the values of Z_n sequentially and to

accept H_o if $Z_n < b$

reject H_o if $Z_n \geq a$

continue by observing the next value Z_{n+1} otherwise.

(See Figure 1.) The values of a and b can be found approximately by taking

$$(1) \qquad a = \ell n((1-\beta)/\alpha) \quad , \quad b = \ell n(\beta/(1-\alpha)).$$

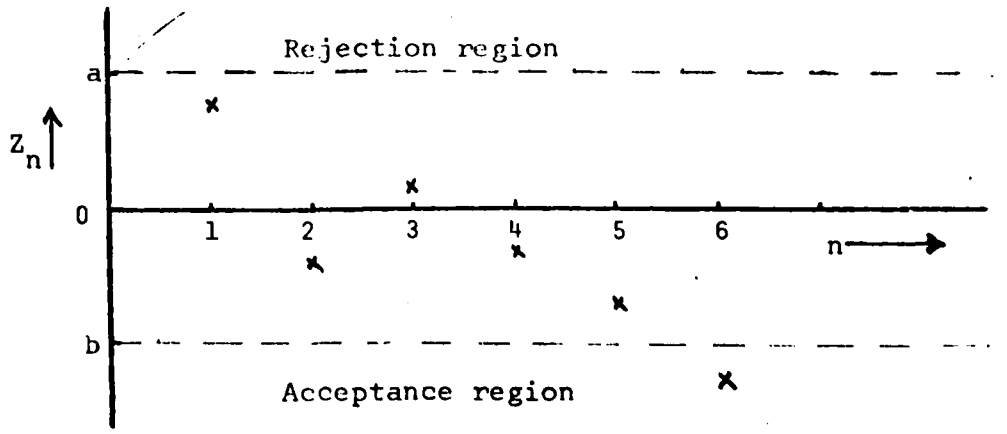


Figure 1. Graphical Representation of a Sequential Test.

Given a sample path as shown in Figure 1 it becomes obvious that there are difficulties in trying to extend the definition of p-values to a sequential test. For one thing, in order to find "the smallest α for which we would surely reject" the null hypotheses, we would have to use Equations (1) changing both a and b. We would then also have to know the entire sample path and not just the value of the test statistic at the time at which a decision is made.

If we now return to the simpler single sample test we can find an alternative characterization of p-values which lends itself more easily to generalization. Specifically, "using the distribution of T under H_o, calculate the probability P* [the significance probability or p-value] of the occurrence of the observed value or more extreme values" (Bhattacharyya and Johnson, 1977, p.180). In order to do this we must determine which values should be considered more extreme than the observed value. This determination is generally not difficult in single sample tests but is more difficult for sequential tests.

Assume that test boundaries a_n, b_n have been given such that for test statistic Z_n, we

accept H_o if $Z_n < b_n$

reject H_o if $Z_n \geq a_n$

continue by observing the next value Z_{n+1} otherwise.

Note that by setting $b_N = a_N$ we can obtain a truncated sequential test or, equivalently, an N-stage (multi-stage) test. If $b_1 = a_1$ we obtain an ordinary single stage test. (In some cases the directions of the inequalities for acceptance and rejection will have to be reversed. This causes no real problem, however.) Our convention for determining which values should be considered more extreme than the observed values will be as follows:

(1) A decision to reject at stage n is more extreme than one to reject at stage n + 1.

(2) A reject decision at stage n with observed value z_n is more extreme than a reject decision at stage n with observed value z_n' if $z_n > z_n'$.

(3) A decision to accept at stage n is more extreme than one to accept at stage n - 1.

(4) An accept decision at stage n with observed value z_n is more extreme than an accept decision at stage n with observed value z_n' if $z_n < z_n'$.

In Figure 2 we show "decision points" $d_1, d_2, \ldots, d_6$ which are possible final observed values of a test statistic. These points are "ordered" in the sense that d_1 is "more extreme" than d_2 which is "more extreme" than d_3, and so

on. With this convention of determining extremeness of the final values of the test statistic we can find p-values.

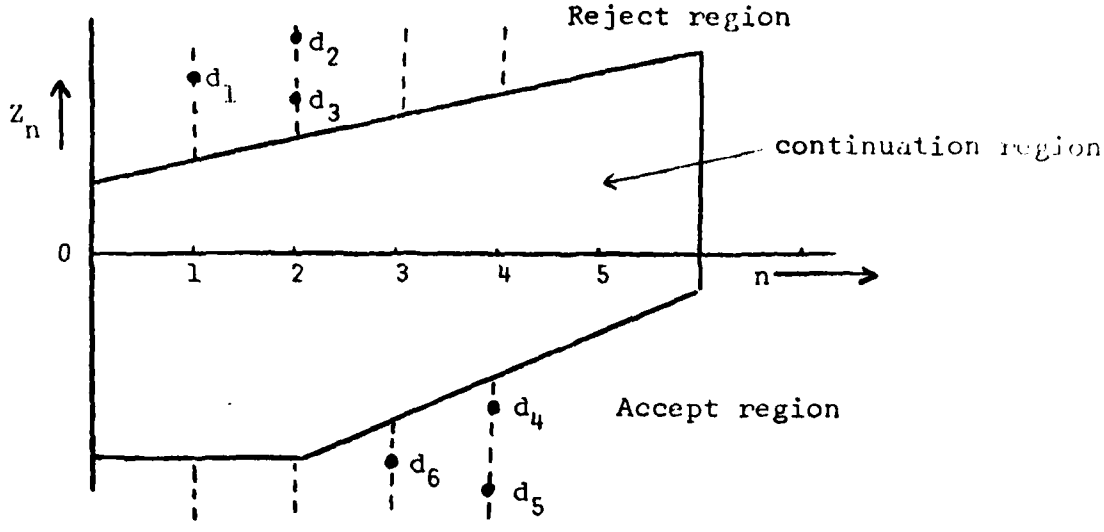


Figure 2. Ordered Decision Points

Definition. For a sequential test, truncated or not, with given test boundaries a_n, b_n, if the test terminates at stage k with observed test statistic z_k, then the p-value is defined by

p-value = P[a test statistic as or more extreme than z_k will be observed when H_o is true].

If H_o is composite and not simple, then the maximum probability found when H_o is true will be the p-value.

For simplicity we will assume that H_o is a simple hypothesis so that we will not have to be concerned with finding the maximum probability under H_o. Notationally we

can define

$$\alpha_i = P[\text{reject } H_o \text{ at stage } i | H_o \text{ true}]$$
$$= P[\text{Continuation at stages } 1,2,\ldots i-1 \text{ and } Z_1 \geq a_i | H_o].$$
$$(2) \quad = P[(b_1 \leq Z_1 < a_1), \ldots (b_{i-1} \leq Z_{i-1} < a_{i-1}), (Z_i \geq a_i) | H_o].$$

and for $z_{io} \geq a_i$ define

$$(3) \quad p_i^* = P[(b_1 \leq Z_1 < a_1), \ldots (b_{i-1} \leq Z_{i-1} < a_{i-1}), (Z_i \geq z_{io}) | H_o].$$

The p-value for a reject decision at stage i with final observed value z_{io} can then be found from

$$\text{p-value} = \sum_{j=1}^{i-1} \alpha_j + p_i^*$$

(Note that the overall level of significance of the test will be given by $\alpha = \Sigma\alpha_j$ with the sum taken over all possible test stages. Also if a test is curtailed with rejection at the i^{th} stage, the p-value will be bounded since

$$\sum_{1}^{i-1} \alpha_j < \text{p-value} \leq \sum_{1}^{i} \alpha_j.)$$

In a similar way we define

$$(4) \quad \gamma_i = P[(b_1 \leq Z_1 < a_1), \ldots (b_{i-1} \leq z_{i-1} < a_{i-1}), (Z_i < b_i) | H_o]$$

and for $z_{io} < b_i$

$$(5) \quad q_i^* = P[(b_1 \leq Z_1 < a_1), \ldots, (b_{i-1} \leq Z_{i-1} < a_{i-1}), (Z_i < z_{io}) | H_o]$$

then the p-value for an accept decision at stage i with final observed value z_{io} is

$$\text{p-value} = 1 - \left(\sum_{j=1}^{i-1} \gamma_j + q_i^* \right).$$

Here too, if the test is curtailed at an acceptance boundary bounds for the p-value can be given.

If the p-value is defined in this way, then when H_o is true the distribution of the p-value will be uniform over the interval [0, 1]. This same property, of course, holds in the single sample testing situation.

3. APPLICATION TO MIL-STD-781C TEST DESIGNS

While conceptually it is quite straightforward to use Equations (2) - (5) to find p-values, the actual calculations will typically involve numerical integration to find the α_i, γ_i, etc. We will illustrate the method of finding p-values by considering just three of the test plans given in MIL-STD-781C (1977) where the underlying random variable of interest has an exponential distribution. Bryant and Schmee (1979) considered the problem of finding confidence intervals for the parameter θ of the exponential distribution when using these test plans. Although the method is applicable to all test plans, we will only consider test plans IVC, VIC, and VIIC. We will begin with Plan VIC. Here the discrimination ratio is 3, so we may consider the

test of

$$H_o: \theta = \theta_o = 3 \quad \text{vs} \quad H_1: \theta = \theta_1 = 1$$

with $\alpha = \beta = .20$ (nominal values). If $X_1, X_2, \ldots$ are independent exponential random variables with parameter θ, then we will use as test statistic $Z_n = X_1 + \cdots + X_n$. The decision boundaries a_n and b_n are shown in Table 1. Note that because of the relative magnitudes of θ_o and θ_1 the accept-reject regions will be interchanged, i.e. here H_o will be accepted if $Z_n > b_n$ and rejected if $Z_n \leq a_n$. The necessary modifications in Equations (2) to (4) are easy to make.

	(Reject boundary)	(accept boundary)
i	a_i	b_i
1	0	2.67
2	0.36	4.32
3	4.50	4.50

Table 1. Decision Boundaries for Test Plan VIC.

By using numerical integration we were able to find the values of α_i and γ_i as well as p-values for various terminal values in the rejection region. Since the test plans call for curtailment of the tests when an acceptance boundary is reached it is only possible to give bounds for the p-value at acceptance but not the actual p-value. The results for Test Plan VIC are shown in Table 2.

Test Plan IVC has nominal values of $\alpha = \beta = .2$ with a discrimination ratio of 2 while Test Plan VIIC has nominal values of $\alpha = \beta = .30$ with discrimination ratio 1.5. The decision boundaries for these test plans are shown in Table 3. The p-values for these test plans are given in Tables 4 and 5.

4. AN EXAMPLE

The examples we give here follow the examples given by Bryant and Schmee (1979). Specifically, Neathammer, Pabst, and Wigginton (1965) describe a production reliability acceptance test of a black box term for an aircraft. In this problem the risks are to be $\alpha = \beta = .2$ and the discrimination ratio $d = 2$. Consequently test plan IV C would be appropriate.

Now assume that in an actual test the failures occurred at (scaled) accumulated test times of 1.0, 1.8, 2.4, 5.0, and 7.8 hours. By looking at the test boundaries shown in Table 3 we see that the test should be continued at each of the first five stages. Assume that the sixth failure does not occur prior to the accumulated time of 9.74 hours. Then the test will be curtailed with acceptance at this time. Without knowing the actual time of the sixth failure it is not possible to give the p-value exactly, but lower and upper bounds on the p-value can be found from Table 4b. Since the decision occurs at the

sixth stage we find the bounds to be:

lower bound = .299 $\leq$ p-value $\leq$ .334 = upper bound

Next we will consider a test which ends in rejection rather than acceptance. If in an actual test the failures occured at accumulated test times of 1.0, 1.8, 2.4, and 3.0 hours, then the test would result in rejection at the fourth stage. The p-value can be found from Table 4a. At stage 4 with a final observed value of 3.00, the p-value can be seen to be .127.

5. ACKNOWLEDGMENTS

This research was supported in part by Office of Naval Research Grant ONR N00014-78-C-0655.

(a) P-Values at Rejection

Stage	1	2			
Observed z	0	.10	.20	.30	.36
P-value	0	.001	.002	.005	.007

Stage	3							
Observed z	.40	.60	.80	1.00	1.20	1.40	1.60	1.80
P-value	.007	.007	.008	.010	.013	.017	.021	.027

Stage	3							
Observed z	2.00	2.20	2.40	2.60	2.80	3.00	3.20	3.40
P-value	.034	.042	.051	.061	.071	.083	.095	.108

Stage	3					
Observed z	3.60	3.80	4.00	4.20	4.40	4.50
P-value	.121	.135	.148	.162	.176	.182

(b) P-values at Acceptance

Stage	1	2	3
Lower bound	.589	.378	.182
Upper bound	1.000	.589	.378

Table 2. P-values for Test Plan VI C

Stage	Plan IV C		Plan VII C	
	Reject Boundary a_i	Accept Boundary b_i	Reject Boundary a_i	Accept Boundary b_i
1	0	2.80	0	3.15
2	.70	4.18	0	4.37
3	2.08	5.58	1.22	5.58
4	3.46	6.96	2.43	6.80
5	4.86	8.34	3.65	6.80
6	6.24	9.74	6.80	6.80
7	7.62	9.74		
8	9.74	9.74		

Table 3. Decision Boundaries for Test Plans IV C and VII C .

(a) P-Values at Rejection

Stage	1	2			
Observed z	0	.20	.40	.60	.70
P-value	0	.005	.018	.037	.049

Stage	3							
Observed z	.80	1.00	1.20	1.40	1.60	1.80	2.00	2.08
P-value	.049	.052	.057	.065	.075	.088	.103	.109

Stage	4							
Observed z	2.20	2.40	2.60	2.80	3.00	3.20	3.40	3.46
P-value	.110	.112	.115	.120	.127	.136	.146	.149

Stage	5							
Observed z	3.60	3.80	4.00	4.20	4.40	4.60	4.80	4.86
P-value	.149	.151	.153	.157	.162	.167	.174	.176

Stage	6							
Observed z	5.00	5.20	5.40	5.60	5.80	6.00	6.20	6.24
P-value	.176	.177	.179	.181	.185	.188	.193	.194

Stage	7							
Observed z	6.40	6.60	6.80	7.00	7.20	7.40	7.60	7.62
P-value	.194	.195	.196	.198	.200	.202	.205	.206

Stage	8							
Observed z	8.00	8.40	8.80	9.00	9.20	9.40	9.60	9.74
P-value	.206	.208	.211	.213	.216	.218	.221	.223

Table 4(a) P-values at Rejection for Test Plan IV C .

(b) P-values at Acceptance

Stage	1	2	3	4	5	6	7	8
Lower bound	.753	.580	.464	.386	.334	.299	.253	.223
Upper bound	1.000	.753	.580	.464	.386	.334	.299	.253

Table 4(b) P-values at Acceptance for Test Plan IV C .

(a) P-Values at Rejection

Stage	1	2	3					
Observed z	0	0	.20	.40	.60	.80	1.00	1.22
P-value	0	0	.000+	.003	.008	.017	.030	.049

Stage	4						
Observed z	1.40	1.60	1.80	2.00	2.20	2.40	2.43
P-value	.050	.054	.061	.070	.082	.097	.099

Stage	5						
Observed z	2.60	2.80	3.00	3.20	3.40	3.60	3.65
P-value	.100	.103	.108	.116	.125	.136	.140

Stage	6						
Observed z	4.00	4.20	4.40	4.60	4.80	5.00	5.20
P-value	.142	.146	.152	.159	.168	.178	.190

Stage	7							
Observed z	5.40	5.60	5.80	6.00	6.20	6.40	6.60	6.80
P-value	.203	.217	.233	.249	.266	.283	.301	.319

Table 5(a) P-values at Rejection for Test Plan VII C .

(b) P-values at Acceptance

Stage	1	2	3	4	5	6
Lower bound	.878	.764	.669	.593	.455	.319
Upper bound	1.000	.878	.764	.669	.593	.455

Table 5(b) P-values at Acceptance for Test Plan VII C .

REFERENCES

Bhattacharyya, G.K. and Johnson, R.A. (1977), Statistical Concepts and Methods, New York: John Wiley.

Bickel, P.J. and Doksum, K.A. (1977), Mathematical Statistics, San Francisco: Holden-Day.

Bryant, C.M. and Schmee, J. (1979), "Confidence Limits on MTBF for Sequential Test Plans of MIL-STD 781", Technometrics, 21, 33-42.

Dudewicz, E.J. (1976), Introduction to Statistics and Probability, New York: Holt, Rinehart and Winston.

Ghosh, B.K. (1970), Sequential Tests of Statistical Hypotheses, Reading, Massachusetts: Addison-Wesley.

MIL-STD 781C. (1977), Military Standard Reliability Design Qualification and Production Acceptance Tests: Exponential Distribution. Washington, D.C., Department of Defense.

Neathammer, R.C., Pabst, W.R. and Wigginton, C.G. (1969), Mil-Std 781B, Reliability Tests: Exponential Distribution. J. Quality Technology, 1, 58-67.

Wald, A. (1947), Sequential Analysis, New York: John Wiley.


FILMED
8