Two Sample Tests in the Weibull Distribution

Won Joon Park*

ABSTRACT

In Thoman and Bain [4] and Schafer and Sheffield [3], procedures for testing the equality of the scale parameters of two Weibull populations with a common shape parameter and procedures for selecting the Weibull population with the largest scale parameter are given. We give, in this paper, a modified procedure for the above testing and selection problems, which is more powerful than those previously given.

1. Introduction

Suppose that $X_{11}, X_{12}, \ldots, X_{1n}$ and $X_{21}, X_{22}, \ldots, X_{2n}$ are independent random samples from two Weibull distrisbutions

$$F_{X1}(x_1) = 1 - \exp\left[-(x_1/b_1)^{c_1}\right] \text{ and}$$

$$F_{X2}(x_2) = 1 - \exp\left[-(x_2/b_2)^{c_2}\right],$$

respectively.

Thoman and Bain [4] and Schafer and Sheffield [3] have considered the following problems of inference:

(A) Testing the hypothesis $H_0 : b_1 = b_2,\ c_1 = c_2$
against the alternative hypothesis $H_1 : b_1 = kb_2,\ c_1 = c_2$.

(B) A procedure for selecting the Weibull population with the largest scale parameter from two populations.

The problem (A) is a test for the equality of two scale parameters when the two Weibull populations have the same but unknown shape parameter $c_1 = c_2 = c$.

*Department of Mathematics, Wright State University, Dayton, Ohio 45431, and Visiting Professor, Department of Computer Seience and Statistics, Seoul National University.

This is a frequently occurring problem and the assumption of a common c is not as restrictive as it might seem since the common value of c need not be known.

It is noted that a test for the equality of shape parameters c_1 and c_2 was considered in [4]. The problem (B) above was also considered by Qureishi [2]. The initial procedure in [4] for the problems (A) and (B) was modified by Schafer and Sheffield [3], whose procedure gave improved power of the test.

The improvement of power in [3] is basically due to an improved method of estimation for the common shape parameter c. In this paper we use another method of estimating the common shape parameter c, which is better than those given in [3] and [4], and give a modified procedure for the problems (A) and (B). At first we give various methods of estimation for the common shape parameter c.

2. Pooled Estimations of the Parameter

Although various methods of pooled estimation for the common shape parameter c were given in Park [1] for the case of m-samples($m \geq 2$), we present them($m=2$) here for the completeness.

(a) Averaging M.L.E. c:

Let $\hat{c}_i$ be the maximum likelihood estimate (M.L.E.) of c_i based on the observations $x_{i1}, x_{i2}, \ldots, x_{in}$ for $i=1$ or 2, that is, $\hat{c}_i$ is the non-negative solution of the following equation:

$$\frac{\sum_{j=1}^{n} x_{ij}^{\hat{c}_i} \ln x_{ij}}{\sum_{j=1}^{n} x_{ij}^{\hat{c}_i}} - \frac{1}{\hat{c}_i} - \frac{\sum_{j=1}^{n} \ln x_{ij}}{n} = 0 \qquad (1)$$

$$i=1, 2.,$$

then $\bar{c}=(\hat{c}_1+\hat{c}_2)/2$ is the averaging M.L.E.of c.

(b) Joint M.L.E. c^*:

The maximum likelihood equation for the pooled data, x_{ij}, for $i=1, 2$ and $j=1, 2 \cdots, n$, is given by

$$\sum_{i=1}^{2}\left(\frac{\sum_{j=1}^{n} x_{ij}^{c^*} \ln x_{ij}}{\sum_{j=1}^{n} x_{ij}^{c^*}}\right) - \frac{2}{c^*} - \frac{\sum_{i=1}^{2}\sum_{j=1}^{n} \ln x_{ij}}{n} = 0 \qquad (2)$$

and a solution c^* of the equation (2) is called the joint M.L.E. of c.

(c) M.L.E. by Normalization $\tilde{c}$:

Since the shape parameter $\tilde{c}$ is free from scale changes (normalization), pooled estimation of c can be obtained by normalizing the data, that is, by letting

$y_{ij}=x_{ij}/b_i$, for $i=1,2$ and $j=1,2,\cdots,n$.

The M.L.E. by normalization c is the non-negative solution of the following maximum likelihood equation for the normalized data:

$$\frac{\sum_{i=1}^{2}\sum_{j=1}^{n} y_{ij}^{\tilde{c}} \ln y_{ij}}{\sum_{j=1}^{2}\sum_{j=1}^{n} y_{ij}^{\tilde{c}}} - \frac{1}{\tilde{c}} - \frac{\sum_{i=1}^{2}\sum_{j=1}^{n} \ln x_{ij}}{2n} = 0 \qquad (3)$$

When normalizing the data, the scale parameters b_i are usually unknown and they may be replaced by the M.L.E. $\hat{b}_i$, where

$$\hat{b}_i=\left(\sum_{j=1}^{n} x_{ij}^{\hat{c}_i}/n\right)1/\hat{c}_i, \quad i=1,2, \qquad (4)$$

where $\hat{c}_i$ are given in (1).

Let us denote

$$b_i^*=\left(\sum_{j=1}^{n} x_{ij}^{c^*}/n\right)\ 1/c^* \qquad (5)$$

$$\tilde{b}_i=\left(\sum_{j=1}^{n} x_{ij}^{\tilde{c}}\ /n\right)\ 1/\tilde{c} \qquad (6)$$

for $i=1,2$.

We note that estimates $\hat{b}_i$, b_i^* and $\tilde{b}_i$ are M.L.E.'s of b_i correspondcng the various pooled estimates of c.

In [1], the distributions of the above estimators $\bar{c}$, c^* and $\tilde{c}$, and $\bar{c}\ \ln\hat{b}_i$, $c^*\ \ln\hat{b}_i^*$ and $\tilde{c}\ \ln\tilde{b}_i$ were obtained by Monte Carlo methods. It is also noted that

$\tilde{c}$ is a better estimate than $\bar{c}$ or c^* in [1].

3. Testing the Equality of the Scale Parameters.

To test $H_0 : b_1=b_2,\ \ c_1=c_2$ *v.s.* $H_1 : b_1=kb_2,\ \ c_1=c_2,\ \ (k>1)$, the following statistics.

$$S(\hat{c}_1,\ \hat{c}_2,\ \hat{b}_1,\ \hat{b}_2)=\frac{\hat{c}_1+\hat{c}_2}{2}[\ln\hat{b}_1-\ln\hat{b}_2],$$
$$=\hat{c}[\ln\hat{b}_1-\ln\hat{b}_2], \tag{7}$$

and $T\ (c^*,\ b_1{}^*,\ b_2{}^*)=c^*\ [\ln b_1{}^*-\ln b_2{}^*]$ (8)

were used in [4] and [3], respectively. The distributions of both statistics S and T, which are independent of the unknown parameters c, b_1 and b_2 under the null hypothesis, were also obtained by Monte Carlo methods in [4] and [3], respectively.

We propose a modified statistic,

$$R(\tilde{c},\ \tilde{b}_1,\tilde{b}_2)=\tilde{c}[\ln\tilde{b}_1-\ln\tilde{b}_2]. \tag{9}$$

Under the null hypothesis, the distribution of $R=R(\tilde{c}_1,\ \tilde{b}_1,\ \tilde{b}_2)$ are independent of the parameters c, b_1 and b_2 and has the same distibution as R^*, where $R^*=R(\tilde{c},^*\ \tilde{b}_1,^*\ \tilde{b}_2{}^*)=\tilde{c}^*\ [\ln\tilde{b}_1{}^*-\ln\tilde{b}_2{}^*]$ and $\tilde{c}^*$ and $\tilde{b}_1{}^*$ are satisfying the equations (3) and (6) respectively when the data y_{ij} (and x_{ij}) are from the standard exponential distribution.

The distributions of R^* was obtained by Monte Carlo method s and percentage points r_{1-r} such that

$Pr[R^*\leq r_{1-r}]=1-r$ are given in Table 1 as a function of r and the common sample size n.

Now the testing hypothesis problem above with the level of significance r can be solved by using the fact that

$$Pr[\tilde{c}[\ln\tilde{b}_1-\ln\tilde{b}_2]<r_{1-r}\,|\,H_0]=1-r. \tag{10}$$

Thus the critical region (with r) for the test is

$\{R \geq r_{1-r}\}$ where r_{1-r}

satisfies (10).

As an aid in evaluating the accuracy of the results, the standard error of the entries in Table 1 is computed and it is approximately 0.005 at $n=5$ decreasing to 0.001 at $n=100$ for $(1-r) \leq 80$. and 0.021 at $n=5$ decreasing to 0.002 at $n=100$ for $(1-r) > .80$.

n \ $1-r$	.60	.70	.75	.80	.85	.90	.95	.98
5	.187	.392	.498	.618	.749	.900	1.142	1.360
6	.170	.342	.429	.531	.656	.822	1.011	1.230
7	.139	.305	.388	.488	.596	.746	.932	1.130
8	.138	.289	.371	.456	.570	.693	.872	1.048
9	.129	.267	.343	.425	.526	.644	.815	1.011
10	.125	.258	.332	.413	.494	.605	.769	.935
11	.116	.231	.301	.378	.467	.574	.726	.887
12	.108	.229	.292	.366	.451	.552	.699	.864
13	.104	.213	.271	.342	.420	.517	.671	.826
14	.100	.206	.263	.326	.395	.499	.645	.783
15	.096	.204	.260	.324	.399	.491	.621	.751
16	.091	.197	.253	.311	.381	.470	.600	.735
17	.090	.186	.235	.291	.363	.449	.577	.723
18	.087	.182	.231	.286	.349	.431	.560	.686
19	.085	.180	.231	.284	.346	.424	.549	.684
20	.084	.170	.218	.275	.333	.414	.526	.650
24	.076	.157	.202	.251	.307	.379	.481	.598
28	.068	.140	.181	.228	.282	.353	.456	.564
32	.065	.135	.172	.216	.266	.328	.419	.526
36	.062	.125	.163	.206	,249	.309	.400	.495
40	.059	.124	.157	.194	.238	.290	.371	.461
50	.056	.107	.136	.171	.210	.263	.336	.420
60	.047	.097	.124	.155	.189	.234	.302	.377
70	.044	.093	.117	.146	.180	.224	.285	.356
80	.038	.083	.108	.134	.166	.205	.264	.322
90	.037	.078	.099	.125	.156	.193	.247	.307
100	.036	.073	.095	.120	.148	.183	.232	.287

Table 1. Values r_{1-r} Such that $Pr(R \leq r_{1-r}) = 1-r$

The power of this test can be expressed by

$$Pr[R>r_{1-r}|H_1]=Pr\{\tilde{c}[\ln\tilde{b}_1-\ln\tilde{b}_2]>r_{1-r}|H_1: b_1=kb_1,\ c_1=c_2=c,k>1\}$$

From the definition of R in (9), the above probability is clearly

$$Pr\{\tilde{c}[\ln(\tilde{b}_1/b_1)-\ln(\tilde{b}_2/b_2)+\ln k]>r_{1-1}\},\ \text{that is,}$$

$$Pr\tilde{c}\{[\ln(\tilde{b}_1/b_1)-\ln(\tilde{b}_2/b_2)+\frac{1}{c}\ln(k^c)]>r_{1-r}\}. \tag{11}$$

The distribution of this random variable does not depend on b_1 and b_2 and depends only c through k^c. Monte Carlo methods can be applied a gain for each fixed k^c value.

The comparison of the powers of S, T and R, based on 10,000 Monte Carlo runs for $n=5, 10, 20, 40$ and $k^c=1.3,\ 1.5,\ 2.5,$ is given in Table 2,

It should be noted that the test of this section on b_1 and b_2 under the assumption of equal shape parameter is equivalent to a test on the means of the two Weibull distribution since the mean $\mu=b\Gamma(1+\frac{1}{c})$.

k^c	1.3			1.5			2.0			2.5		
n	S	T	R	S	T	R	S	T	R	S	T	R
5	.17	.18	.21	.22	.24	.29	.36	.38	.45	.48	.51	.58
10	.23	.24	.26	.33	.34	.37	.56	.58	.62	.74	.75	.78
20	.30	.32	.34	.47	.48	.51	.78	.80	.82	.93	.94	.94
40	.42	.46	.46	.67	.69	.70	.95	.96	.96	.99	1.00	1.00

Table 2. Comparison of Power for S,T and R(for Test Size $r=.10$)

4. A Selection Procedure for Scale Parameters

Consider two Weibull populations with the same shape parameter c but different and unknown scale parameter b_1 and b_2. Suppose that it is desired to choose the population with the largest scale parameter, or equivalently, the population with the largest mean. When the procedure of the previous section is applied to this problem, it leads to the following procedure (corresponding the procedure R_1 discussed in [2]) Compute $\tilde{b}_1/\tilde{b}_2$ when $\tilde{b}_i$ are M.L.E.'s of b_i given in (6) and choose population 1 if $\tilde{b}_1/\tilde{b}_2>1$ and population 2 if $\tilde{b}_1/\tilde{b}_2<1$. The probability of correct selection for this procedure can be determined by probability given in (11). If $b_1/b_2=k$, then the probability of correct selection is

$$\begin{aligned} Pr(CS) &= Pr\{\tilde{b}_1/\tilde{b}_2>1 \mid b_1/b_2=k>1\} \\ &= Pr\{R>0 \mid b_1=kb_2\} \\ &= Pr\{\tilde{c}[\ln(\tilde{b}_1/\tilde{b}_1)-\ln(\tilde{b}_2/b_2)+\frac{1}{c}\ln(k^c)]>0\}. \end{aligned}$$

The probability of correct selection, corresponding the statistics S and T, was considered in [4] and [3], respectively. The comparison of $Pr(CS)$ for S, T and R is given in Table 3. Again, the entries are based on 10,000 Monte Carlo runs.

k^c	1.10			1.30			1.50		
n	S	T	R	S	T	R	S	T	R
10	.581	.588	.589	.706	.724	.727	.806	.817	.821
20	.613	.616	.621	.785	.794	.796	.889	.896	.898
30	.636	.646	.650	.835	.844	.848	.928	.942	.944
40	.656	.667	.669	.856	.880	.882	.953	.962	.962

Table 3. Comparison of $Pr(CS)$ for S,T and R

REFERENCES

[1] Park, W.J., "Pooled Estimations of the Parameters on Weibull Distributions," Technical Report, Air Force Materials Laboratory, AFML-TR-79-4112, 1979.

[2] Qureishi, A.S., "The Discrimination between Two Weibull Processes," *Technometrics*, 6, 1964, 57-75.

[3] Schafer, R.E. and Sheffield, T.S., "On Procedures for Comparing Two Weibull Populations," *Technometrics*, 18. 1976, 231-235.

[4] Thoman, D.R. and Bain, L.J., "Two Sample Tests in the Weibull Distribution," *Technometrics*, 4, 1969, 805-815.

[5] Thoman, D.R., Bain, L.J. and Antle, C.E., "Inferences on the Parameters of Weibull Distribution," *Technometrics*, 11, 1969, 445-460