Thailand Statistician
January 2019; 17(1): 84-92
http://statassoc.or.th
Contributed paper

Interval Estimation of the Overlapping Coefficient of Two Normal Distributions: One Way ANOVA with Random Effects

Sibil Jose*[a] and Seemon Thomas [b]
[a] Department of Statistics, St. George's College Aruvithura, Kottayam, Kerala, India.
[b] Department of Statistics, St. Thomas College Pala, Kottayam, Kerala, India.
*Corresponding author; e-mail: sibiljose60@yahoo.com

Received: 3 May 2018
Revised: 4 August 2018
Accepted: 18 August 2018

Abstract

We consider the problem of deriving confidence interval for the Overlapping Coefficient (OVL) between one way random model with balanced and unbalanced data. The confidence intervals are derived using the concept of a Generalized Pivotal Quantity (GPQ). The accuracy of the proposed solutions are assessed using estimated coverage probabilities, and are also compared with other approximate solutions. The numerical results show that the GPQ method performs well in the estimation of OVL of two normal distributions under one way random model set-up. The results are illustrated with simulated examples.

Keywords: Overlapping coefficient, generalized pivotal quantity, one way random model, coverage probability.

1. Introduction

A similarity measure explains the amount of overlap between two statistical populations. Similarity measure has a variety of applications in different fields. Overlapping Coefficient (OVL) is a measure of similarity between two probability distributions. It is the common area under two probability density functions. The value of OVL lies in between 0 and 1. If the two distributions are identical OVL becomes 1 and 0 if the two distributions are entirely different.

Models with random effects have a wide variety of applications. For example, Kromhout et al. (1993) considered the one-way random model to study the within and between worker component of occupational exposure to chemical agents. Bhaumik and Kulkarni (1996) gave an exact method of constructing tolerance interval for the one-way ANOVA with random effects. The relevance of one-way random model for analyzing occupational exposure data was pointed out in Lyles et al. (1997).

The concept of generalized pivotal quantity for the construction of confidence interval was used by many authors. Krishnamoorthy and Mathew (2002) used generalized confidence interval for assessing occupational exposure via the one-way random effects model with balanced data. Krishnamoorthy and Mathew (2003) constructed generalized confidence interval and p-values for the means of lognormal distributions. Roy and Mathew (2005) constructed generalized confidence limits for the reliability function of two parameter exponential distribution. Krishnamoorthy and Mathew (2004) constructed one sided tolerance limits in balanced and unbalanced one way random models based on generalized confidence intervals. This paper introduces the construction of confidence intervals using generalized pivotal quantities of the parameters involved in the random effects models.

2. One Way Random Model with Balanced Data

Let X_{ij} and Y_{ij} be two independent observations following the one-way random models given by

$$X_{ij} = \mu_1 + \alpha_i + \varepsilon_{ij};\quad i=1,2,\ldots,k_1,\ j=1,2,\ldots,n_1 \tag{1}$$

$$Y_{ij} = \mu_2 + \beta_i + \varepsilon_{ij};\quad i=1,2,\ldots,k_2,\ j=1,2,\ldots,n_2, \tag{2}$$

where μ_1 and μ_2 are fixed unknown parameters, α_i, β_i, ε_{ij} and e_{ij} are independent random variables such that

$$\alpha_i \sim N(0,\sigma_\alpha^2),\quad \varepsilon_{ij} \sim N(0,\sigma_\varepsilon^2),$$

$$\beta_i \sim N(0,\sigma_\beta^2),\quad e_{ij} \sim N(0,\sigma_e^2).$$

Therefore, $X_{ij} \sim N(\mu_1, \sigma_\alpha^2 + \sigma_\varepsilon^2)$ and $Y_{ij} \sim N(\mu_2, \sigma_\beta^2 + \sigma_e^2)$.

Here we consider Matusita's measure of OVL of two normal distributions (see Minami and Shimizu 1999) with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 given by

$$\rho = \sqrt{\frac{2\sigma_1\sigma_2}{(\sigma_1{}^2 + \sigma_2{}^2)}}\exp\left\{-\frac{1}{4}\frac{(\mu_1-\mu_2)^2}{(\sigma_1{}^2+\sigma_2{}^2)}\right\}. \tag{3}$$

In our application dealing with one-way random model, the two normal populations are $N(\mu_1, \sigma_\alpha^2 + \sigma_\varepsilon^2)$ and $N(\mu_2, \sigma_\beta^2 + \sigma_e^2)$. In (3), now replace σ_1^2 and σ_2^2 with $\sigma_\alpha^2 + \sigma_\varepsilon^2$ and $\sigma_\beta^2 + \sigma_e^2$. Our problem is the derivation of a confidence interval for ρ.

2.1. Generalized confidence interval

In order to derive a confidence interval for ρ, we shall use the generalized confidence interval idea due to Weerahandi (1993, 1994, 2004). For this, it is necessary to exhibit a generalized pivotal quantity (GPQ) for ρ. By definition, a GPQ is a function of underlying basic random variables, and the corresponding observed values, along with the parameters. A GPQ is required to satisfy two conditions: (i) given the observed data, the distribution of the GPQ is free of any unknown parameters, and (ii) if the random variables in the definition of the GPQ are replaced by the corresponding observed values, the GPQ simplifies to a quantity that is free of the nuisance parameters. In order to exhibit a GPQ for ρ, let us define,

$$\bar{X}_{i.} = \frac{1}{n_1}\sum_{j=1}^{n_1} X_{ij}, \qquad \bar{Y}_{i.} = \frac{1}{n_2}\sum_{j=1}^{n_2} Y_{ij},$$

$$\bar{X}.. = \frac{1}{k_1 n_1}\sum_{i=1}^{k_1}\sum_{j=1}^{n_1} X_{ij}, \qquad \bar{Y}.. = \frac{1}{k_2 n_2}\sum_{i=1}^{k_2}\sum_{j=1}^{n_2} Y_{ij},$$

$$SS_\alpha = n_1\sum_{i=1}^{k_1}(\bar{X}_{i.} - \bar{X}..)^2, \qquad SS_\beta = n_2\sum_{i=1}^{k_2}(\bar{Y}_{i.} - \bar{Y}..)^2,$$

$$SS_{\varepsilon}=\sum_{i=1}^{k_1}\sum_{j=1}^{n_1}(X_{ij}-\bar{X}_{i.})^2,\qquad SS_{e}=\sum_{i=1}^{k_2}\sum_{j=1}^{n_2}(Y_{ij}-\bar{Y}_{i.})^2.$$

Then

$$\bar{X}_{i.}\sim N\left(\mu_1,\sigma_\alpha^2+\frac{\sigma_\varepsilon^2}{n_1}\right),\qquad \bar{Y}_{i.}\sim N\left(\mu_2,\sigma_\beta^2+\frac{\sigma_e^2}{n_2}\right),$$

$$\bar{X}..\sim N\left(\mu_1,\frac{n_1\sigma_\alpha^2+\sigma_\varepsilon^2}{k_1n_1}\right),\qquad \bar{Y}..\sim N\left(\mu_2,\frac{n_2\sigma_\beta^2+\sigma_e^2}{k_2n_2}\right),$$

and $\bar{X}..$, SS_α, SS_ε, $\bar{Y}..$, SS_β and SS_e are independently distributed with

$$Z_1=\sqrt{k_1n_1}\,\frac{\left(\bar{X}..-\mu_1\right)}{\sqrt{n_1\sigma_\alpha^2+\sigma_\varepsilon^2}}\sim N(0,1),\quad Z_2=\sqrt{k_2n_2}\,\frac{\left(\bar{Y}..-\mu_2\right)}{\sqrt{n_2\sigma_\beta^2+\sigma_e^2}}\sim N(0,1),$$

$$U_\alpha=\frac{SS_\alpha}{n_1\sigma_\alpha^2+\sigma_\varepsilon^2}\sim\chi^2_{(k_1-1)},\qquad U_\beta=\frac{SS_\beta}{n_2\sigma_\beta^2+\sigma_e^2}\sim\chi^2_{(k_2-1)},$$

$$V_\varepsilon=\frac{SS_\varepsilon}{\sigma_\varepsilon^2}\sim\chi^2_{(k_1(n_1-1))},\qquad V_e=\frac{SS_e}{\sigma_e^2}\sim\chi^2_{(k_2(n_2-1))}.$$

Let $\bar{x}..$, $\bar{y}..$, ss_α, ss_β, ss_ε, and ss_e be the observed values of $\bar{X}..$, $\bar{Y}..$, SS_α, SS_β, SS_ε, and SS_e. The GPQs of σ_ε^2, σ_e^2, σ_α^2, σ_β^2, μ_1 and μ_2 are

$$T_{\sigma_\varepsilon^2}=\frac{ss_\varepsilon}{V_\varepsilon},\tag{4}$$

$$T_{\sigma_e^2}=\frac{ss_e}{V_e},\tag{5}$$

$$T_{\sigma_\alpha^2}=\frac{1}{n_1}\left(\frac{ss_\alpha}{U_\alpha}-\frac{ss_\varepsilon}{V_\varepsilon}\right),\tag{6}$$

$$T_{\sigma_\beta^2}=\frac{1}{n_2}\left(\frac{ss_\beta}{U_\beta}-\frac{ss_e}{V_e}\right),\tag{7}$$

$$T_{\mu_1}=\bar{x}..-Z_1\sqrt{\frac{n_1T_{\sigma_\alpha^2}+T_{\sigma_\varepsilon^2}}{k_1n_1}},\tag{8}$$

$$T_{\mu_2}=\bar{y}..-Z_2\sqrt{\frac{n_2T_{\sigma_\beta^2}+T_{\sigma_e^2}}{k_2n_2}}.\tag{9}$$

It can be verified that the distributions of the quantities in (4) to (9) are free of unknown parameters and their observed values are the respective parameters. On substituting (4) to (9) in (3) we get T_ρ as the generalized pivotal quantity for ρ. The $(\alpha/2)^{\text{th}}$ and $(1-\alpha/2)^{\text{th}}$ percentiles of T_ρ give a $100(1-\alpha)\%$ generalized confidence interval for ρ.

2.2. Simulation study

A simulation study was conducted to assess the performance of generalized confidence interval of ρ. A comparison is done with bootstrap percentile method. Tables 1-3 give coverage probabilities of the confidence intervals for the following parameter values

(i) $\mu_1 = -3, \mu_2 = 4, \sigma_\alpha^2 = 2, \sigma_\beta^2 = 1, \sigma_\varepsilon^2 = 1, \sigma_e^2 = 2$ give the value of ρ =0.1298,

(ii) $\mu_1 = 1, \mu_2 = 5, \sigma_\alpha^2 = 1, \sigma_\beta^2 = 2, \sigma_\varepsilon^2 = 1, \sigma_e^2 = 2$ give the value of ρ =0.4985,

(iii) $\mu_1 = 2, \mu_2 = 0, \sigma_\alpha^2 = 3, \sigma_\beta^2 = 2, \sigma_\varepsilon^2 = 2, \sigma_e^2 = 1$ give the value of ρ =0.8684.

Tables 1-3 give the estimated coverage probabilities for several treatment-replication combinations in the balanced case, and for the three parameter combinations (i), (ii) and (iii) given above. In the table titles, these three cases are identified using the corresponding value of ρ. A nominal 95% confidence level is assumed throughout. All coverage probabilities are estimated using 5,000 simulated samples. In addition, each generalized confidence interval was calculated using 5,000 generated values of the GPQ, and the percentile bootstrap was implemented using 5,000 parametric bootstrap samples. The tables give the coverage probabilities (labelled "Central") and also the values of the right and left tails. From the numerical results, it should be very clear that the GPQ methodology is very satisfactory, and is to be preferred over the percentile bootstrap method.

Table 1 Estimated coverage probabilities when ρ = 0.1298

(k_1, k_2)	(n_1, n_2)	GPQ Method			Bootstrap Perc. Method		
		Coverage Probabilities			Coverage Probabilities		
		Central	Left	Right	Central	Left	Right
(15,15)	(5, 5)	0.9641	0.0257	0.0102	0.9364	0.0581	0.0055
	(10, 10)	0.9626	0.0210	0.0102	0.9685	0.0150	0.0165
	(50,50)	0.9419	0.0046	0.0535	0.9198	0.0051	0.0751
	(100,100)	0.9555	0.0198	0.0247	0.9199	0.0075	0.0726
	(100,200)	0.9640	0.0189	0.0171	0.9099	0.0102	0.0799
	(200,200)	0.9625	0.0201	0.0174	0.9189	0.0064	0.0747
	(20,20)	0.9621	0.0210	0.0169	0.9588	0.0145	0.0267
	(50,50)	0.9586	0.0215	0.0199	0.9575	0.0105	0.0320
(20,20)	(100,100)	0.9543	0.0243	0.0212	0.9432	0.0101	0.0457
	(100,200)	0.9514	0.0213	0.0223	0.9446	0.0112	0.0442
	(200,200)	0.9568	0.0211	0.0221	0.9492	0.0098	0.0410
	(15,8)	0.9554	0.0205	0.0241	0.9138	0.0714	0.0148
	(10,20)	0.9552	0.0198	0.0250	0.9270	0.0654	0.0076
(50,50)	(25,15)	0.9560	0.0205	0.0230	0.9561	0.0413	0.0026
	(100,100)	0.9520	0.0245	0.0235	0.9534	0.0318	0.0148
	(100,200)	0.9554	0.0213	0.0233	0.9762	0.0402	0.0126
	(200,200)	0.9570	0.0214	0.0216	0.9494	0.0171	0.0235

3. Unbalanced Data

Now consider the one-way random model with unbalanced data, given by

$$X_{ij} = \mu_1 + \alpha_i + \varepsilon_{ij}; i = 1, 2, \ldots, k_1; j = 1, 2, \ldots, n_{1i},$$

$$Y_{ij} = \mu_2 + \beta_i + e_{ij}; i = 1, 2, \ldots, k_2; j = 1, 2, \ldots, n_{2i}.$$

Let us define

Table 2 Estimated coverage probabilities when ρ = 0.4985

(k_1,k_2)	(n_1,n_2)	GPQ Method			Bootstrap Perc. Method		
		Coverage Probabilities			Coverage Probabilities		
		Central	Left	Right	Central	Left	Right
(15,15)	(5, 5)	0.9602	0.0188	0.0210	0.9471	0.0079	0.0450
	(10, 10)	0.9572	0.0204	0.0224	0.9284	0.0101	0.0615
	(50,50)	0.9591	0.0192	0.0217	0.9198	0.0051	0.0751
	(100,100)	0.9555	0.0225	0.0220	0.9199	0.0075	0.0726
	(100,200)	0.9580	0.0209	0.0211	0.9099	0.0102	0.0799
	(200,200)	0.9551	0.0217	0.0232	0.9189	0.0064	0.0747
	(20,20)	0.9552	0.0210	0.0238	0.9226	0.0055	0.0719
(20,20)	(50,50)	0.9525	0.0265	0.0210	0.9198	0.0060	0.0742
	(100,100)	0.9575	0.0178	0.0247	0.9277	0.0094	0.0629
	(100,200)	0.9562	0.0182	0.0256	0.8994	0.0121	0.0885
	(200,200)	0.9625	0.0185	0.0190	0.9264	0.0064	0.0672
	(15,8)	0.9554	0.0194	0.0252	0.9655	0.0040	0.0305
	(10,20)	0.9531	0.0249	0.0220	0.9265	0.0086	0.0649
(50,50)	(25,15)	0.9468	0.0292	0.0240	0.9311	0.0102	0.0587
	(100,100)	0.9505	0.0231	0.0264	0.9334	0.0074	0.0592
	(200,200)	0.9543	0.0257	0.0200	0.9213	0.0076	0.0711
	(100,200)	0.9508	0.0226	0.0266	0.9330	0.0069	0.0601

Table 3 Estimated coverage probabilities when ρ = 0.8684

(k_1,k_2)	(n_1,n_2)	GPQ Method			Bootstrap Perc. Method		
		Coverage Probabilities			Coverage Probabilities		
		Central	Left	Right	Central	Left	Right
(15,15)	(5, 5)	0.9415	0.0045	0.0540	0.9396	0.0404	0.0700
	(10, 10)	0.9451	0.0043	0.0506	0.9391	0.0291	0.0318
	(50,50)	0.9419	0.0046	0.0535	0.9421	0.0158	0.0421
	(100,100)	0.9499	0.0051	0.0450	0.9372	0.0136	0.0492
	(200,200)	0.9520	0.0038	0.0442	0.9210	0.0174	0.0616
	(100,200)	0.9516	0.0035	0.0449	0.9422	0.0155	0.0432
	(20,20)	0.9418	0.0054	0.0528	0.9125	0.0262	0.0613
(20,20)	(50,50)	0.9426	0.0045	0.0529	0.9002	0.0172	0.0826
	(100,100)	0.9475	0.0034	0.0491	0.9461	0.0141	0.0398
	(200,200)	0.9430	0.0065	0.0505	0.9385	0.0261	0.0354
	(100,200)	0.9481	0.0042	0.0477	0.9492	0.0196	0.0312
	(10,20)	0.9245	0.0070	0.0685	0.9710	0.0211	0.0079
	(25,15)	0.9246	0.0078	0.0676	0.9069	0.0705	0.0226
(50,50)	(15,8)	0.9246	0.0071	0.0683	0.9711	0.0152	0.0137
	(100,100)	0.9296	0.0071	0.0633	0.9486	0.0185	0.0329
	(200,200)	0.9322	0.0064	0.0614	0.9209	0.0314	0.0477
	(100,200)	0.9380	0.0063	0.0587	0.9491	0.0198	0.0311

$$\bar{X}_{i.} = \frac{\sum_{j=1}^{n_{1i}} X_{ij}}{n_{1i}}, \qquad \bar{Y}_{i.} = \frac{\sum_{j=1}^{n_{2i}} Y_{ij}}{n_{2i}},$$

$$SS_\varepsilon = \sum_{i=1}^{k_1}\sum_{j=1}^{n_{1i}} (X_{ij} - \bar{X}_{i.})^2, \qquad SS_e = \sum_{i=1}^{k_2}\sum_{j=1}^{n_{2i}} (Y_{ij} - \bar{Y}_{i.})^2,$$

$$N_1 = \sum_{i=1}^{k_1} n_{1i}, \qquad N_2 = \sum_{i=1}^{k_2} n_{2i},$$

$$\tilde{n}_1 = \frac{1}{k_1}\sum_{i=1}^{k_1} n_{1i}, \qquad \tilde{n}_2 = \frac{1}{k_2}\sum_{i=1}^{k_2} n_{2i},$$

$$\bar{\bar{X}} = \frac{1}{k_1}\sum_{i=1}^{k_1} \bar{X}_{i.}, \qquad \bar{\bar{Y}} = \frac{1}{k_2}\sum_{i=1}^{k_2} \bar{Y}_{i.},$$

$$SS_{\bar{x}} = \sum_{i=1}^{k_1} (\bar{X}_{i.} - \bar{\bar{X}})^2, \qquad SS_{\bar{y}} = \sum_{i=1}^{k_2} (\bar{Y}_{i.} - \bar{\bar{Y}})^2.$$

The sampling distributions of the above quantities are given by; see Krishnamoorthy and Mathew (2004),

$$\bar{\bar{X}} \sim N\left(\mu_1, \frac{\sigma_\alpha^2 + \tilde{n}_1\sigma_\varepsilon^2}{k_1}\right),$$

$$\bar{\bar{Y}} \sim N\left(\mu_2, \frac{\sigma_\beta^2 + \tilde{n}_2\sigma_e^2}{k_2}\right),$$

$$U_{\bar{X}} = \frac{SS_{\bar{x}}}{\sigma_\alpha^2 + \tilde{n}_1\sigma_\varepsilon^2} \sim \chi^2_{(k_1-1)},$$

$$U_{\bar{Y}} = \frac{SS_{\bar{y}}}{\sigma_\beta^2 + \tilde{n}_2\sigma_e^2} \sim \chi^2_{(k_2-1)},$$

$$U_\varepsilon = \frac{SS_\varepsilon}{\sigma_\varepsilon^2} \sim \chi^2_{(N_1-k_1)},$$

$$U_e = \frac{SS_e}{\sigma_e^2} \sim \chi^2_{(N_2-k_2)}.$$

3.1. Generalized confidence intervals

Using lower case letters to denote the observed values of the corresponding random variables, the GPQs of the respective parameters are given by

$$T_{\sigma_\varepsilon^2} = \frac{SS_\varepsilon}{U_\varepsilon}, \tag{10}$$

$$T_{\sigma_e^2} = \frac{SS_e}{U_e}, \tag{11}$$

$$T_{\sigma_\alpha^2} = \frac{ss_{\bar{x}}}{U_{\bar{x}}} - \tilde{n}_1 T_{\sigma_\varepsilon^2}, \tag{12}$$

$$T_{\sigma_\beta^2} = \frac{ss_{\bar{y}}}{U_{\bar{y}}} - \tilde{n}_2 T_{\sigma_e^2}, \tag{13}$$

$$T_{\mu_1} = \bar{\bar{x}} - Z_1 \sqrt{\frac{T_{\sigma_\alpha^2} + \tilde{n}_1 T_{\sigma_\varepsilon^2}}{k_1}}, \tag{14}$$

$$T_{\mu_2} = \bar{\bar{y}} - Z_2 \sqrt{\frac{T_{\sigma_\beta^2} + \tilde{n}_1 T_{\sigma_e^2}}{k_2}}. \tag{15}$$

On substituting the GPQs of the parameters given in (10) to (15) in (3) we get the GPQ of the OVL ρ.

3.2. Simulation study

Simulation studies are conducted for three different values of ρ for $k = 12$ and three different choices of **n** using 5,000×5,000 simulations as in the balanced case. Tables 4-6 give the estimated coverage probabilities for several replication combinations. The combination of following replication sizes are used for the simulation.

(1) $\mathbf{n}_1$ = (3, 15, 30, 14, 2, 3, 13, 22, 8, 6, 9, 11),

(2) $\mathbf{n}_2$ = (3, 4, 3, 4, 2, 3, 3, 2, 2, 2, 2, 2),

(3) $\mathbf{n}_3$ = (13, 40, 7, 14, 22, 30, 3, 2, 12, 2, 21, 4).

We notice once again that the GPQ approach provides satisfactory coverage in all cases, and is to be preferred over the percentile bootstrap method.

Table 4 Estimated coverage probabilities when ρ = 0.1298

(k_1,k_2)	(n_1,n_2)	GPQ Method			Bootstrap Perc. Method		
		Coverage Probabilities			Coverage Probabilities		
		Central	Left	Right	Central	Left	Right
(12,12)	$(\mathbf{n}_1,\mathbf{n}_1)$	0.9404	0.0101	0.0495	0.9854	0.0128	0.0018
	$(\mathbf{n}_2,\mathbf{n}_2)$	0.9508	0.0044	0.0448	0.9298	0.0670	0.0032
	$(\mathbf{n}_3,\mathbf{n}_3)$	0.9498	0.0029	0.0473	0.9838	0.0143	0.0019

Table 5 Estimated coverage probabilities when ρ = 0.4985

(k_1,k_2)	(n_1,n_2)	GPQ Method			Bootstrap Perc. Method		
		Coverage Probabilities			Coverage Probabilities		
		Central	Left	Right	Central	Left	Right
(12,12)	$(\mathbf{n}_1,\mathbf{n}_1)$	0.9522	0.0024	0.0434	0.9442	0.0214	0.0344
	$(\mathbf{n}_2,\mathbf{n}_2)$	0.9548	0.0032	0.0420	0.9452	0.0478	0.0070
	$(\mathbf{n}_3,\mathbf{n}_3)$	0.9560	0.0021	0.0419	0.9692	0.0264	0.0044

Table 6 Estimated coverage probabilities when ρ = 0.8684

(k_1,k_2)	(n_1,n_2)	GPQ Method			Bootstrap Perc. Method		
		Coverage Probabilities			Coverage Probabilities		
		Central	Left	Right	Central	Left	Right
(12,12)	$(\mathbf{n}_1,\mathbf{n}_1)$	0.9582	0.0012	0.0406	0.8892	0.0024	0.1084
	$(\mathbf{n}_2,\mathbf{n}_2)$	0.9606	0.0011	0.0383	0.9240	0.0054	0.0706
	$(\mathbf{n}_3,\mathbf{n}_3)$	0.9560	0.0015	0.0335	0.9050	0.0008	0.0942

4. Examples

We shall now illustrate our methodology by applying it to some simulated data sets; balanced and unbalanced.

4.1. Balanced case

We shall illustrate our methodology in the case of balanced data using two simulated data sets. In the first set, 15 treatments are considered and each treatment is repeated 10 times. That is, $k_1 = k_2 = 15$, $n_1 = n_2 = 10$. The simulated data give the following observed values: $\bar{x}_{..}$ = 1.0674, $\bar{y}_{..}$ = 4.4929, ss_α = 150.4879, ss_β = 267.221, ss_ε = 134.8539 and ss_e = 277.6038. The estimated value of ρ is 0.5844. The 95% generalized confidence interval is (0.4712, 0.7678) and the bootstrap confidence interval is (0.4249, 0.7656).

In the second dataset 50 treatments from the first population and 30 from the second population are considered. Each of the 50 treatments are repeated 5 times and 30 treatments are repeated 3 times. The simulated data gave the following observed values: $\bar{x}_{..}$ = 1.9517, $\bar{y}_{..}$ = 0.1325, ss_α = 267.777, ss_β = 104.4906, ss_ε = 133.1577 and ss_e = 31.858. The estimated value of ρ is 0.7464. The 95% confidence interval based on GPQ method is (0.6472, 0.8809) and that based on bootstrap method is (0.6581, 0.9080). We note the GPQ based confidence interval is similar to, or shorter than the corresponding percentile bootstrap confidence interval.

4.2. Unbalanced case

In the first example of an unbalanced dataset, 12 treatments are considered from both the populations. The replications of each treatment are given as **n**= (3, 4, 3, 4, 2, 3, 3, 2, 2, 2, 2, 2). The simulated data gave the following observed values: $\bar{\bar{x}}$ = 1.0722, $\bar{\bar{y}}$ = 5.3776, $ss_{\bar{x}}$ = 7.1086, $ss_{\bar{y}}$ = 32.3491, ss_ε = 0.7592, ss_e = 2.5051. The estimated value of ρ is 0.2484. The 95% generalized confidence interval is (0.1058, 0.5514) and bootstrap confidence interval is (0.2850, 0.6909).

In the second datasets, 10 treatments are considered and the replications are n = (15, 14, 30, 3, 13, 22, 9, 8, 6, 11) and the observed values are $\bar{\bar{x}}$ = 2.4580, $\bar{\bar{y}}$ = 4.3890, $ss_{\bar{x}}$ = 15.9006, $ss_{\bar{y}}$ = 49.9430, ss_ε = 2.2068 and ss_e = 0.8482. Here the estimate of OVL is 0.8157. The 95% confidence interval based on GPQ is (0.6173, 0.9670) and bootstrap confidence interval is (0.6908, 0.9602). Once again, the examples have brought out the differences between the two approaches for computing confidence intervals.

5. Conclusions

The GPQ approach has found numerous applications in the literature for several interval estimation problems. Furthermore, numerical results have demonstrated the accuracy of the resulting solutions. This article addresses yet another application: the interval estimation of the overlap coefficient under one way random models with balanced or unbalanced data. We have derived the GPQ based confidence interval, and have assessed its performance using estimated coverage probabilities. The only other approach that naturally comes to mind is the bootstrap, implemented parametrically. Our numerical results show that the GPQ based solution is to be preferred over the bootstrap solution.

References

Bhaumik DK, Kulkarni PM. A Simple and exact method of constructing tolerance intervals for the one way ANOVA with random effects. Am Stat. 1996; 50: 319-323.

Krishnamoorthy K., Mathew T. Assessing Occupational exposure via the one way random effects model with balanced data. J Agric Biol Environ Statist. 2002; 4: 440-457.

Krishnamoorthy K, Mathew T. Inferences on the means of log normal distributions using generalized p-values and generalized confidence Intervals. J Stat Plann Infer. 2003; 115: 103-121.

Krishnamoorthy K, Mathew T. One sided tolerance limits in balanced and unbalanced one way random models based on generalized confidence intervals. Technometrics. 2004; 46: 44-52.

Kromhout H, Symanski E, Rappaport SM. A comprehensive evaluation of within and between worker components of occupational exposure to chemical agents. Ann Occup Hyg. 1993; 37: 253-270.

Lyles RH, Kupper LL, Rappaport SM. Assessing regulatory compliance of occupational exposures via the balanced one way random effects ANOVA model. J Agric Biol Environ Statist. 1997; 2: 64-86.

Minami M, Shimizu K. Estimation of similarity measure for multivariate normal distributions. Environ Ecol Stat. 1999; 6: 229-248.

Roy A, Mathew T. A generalized confidence limit for the reliability function of a two parameter exponential distribution. J Stat Plann Infer. 2005; 128: 1-7.

Weerahandi S. Generalized confidence intervals. J Am Stat Assoc. 1993; 88: 899-905.

Weerahandi S. Exact statistical methods for data analysis. New York: Springer; 1994.

Weerahandi S. Generalized inference in repeated measures: Exact methods in MANOVA and mixed models. New Jersey: Wiley; 2004.